Title
Diversity and dynamics of synechococcus populations in the southern California bight

Author
Tai, Vera

Publication Date
2009

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Diversity and dynamics of *Synechococcus* populations in the Southern California Bight

A dissertation submitted in partial satisfaction of the
requirements for the degree of Doctor of Philosophy

in

Marine Biology

by

Vera Tai

Committee in charge:

- Brian Palenik, Chair
- Farooq Azam
- Katherine Barbeau
- Bianca Brahamsha
- Ronald Burton
- Lin Chao

2009

The dissertation of Vera Tai is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

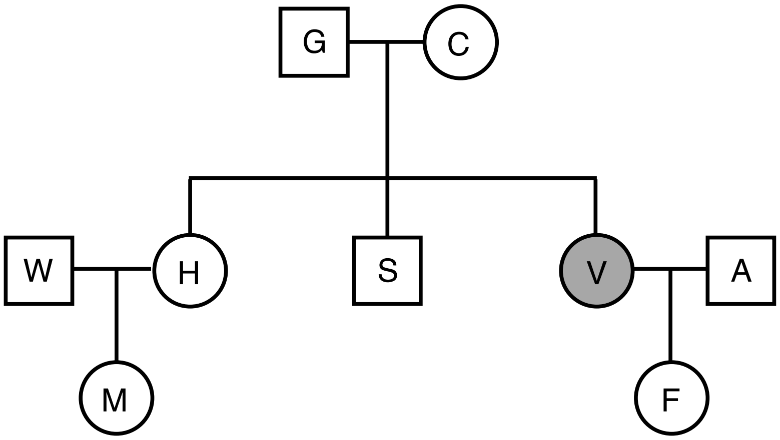
Chair

University of California, San Diego

2009

DEDICATION

To my family



from whose roots I sprang forth.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

bp – base pair

CalCOFI – California Cooperative Oceanic Fisheries Investigations

°C – degrees Celsius

DNA – deoxyribonucleic acid

EDTA – ethylenediaminetetraacetic acid

h – hour

m – meter

mg – microgram

min – minute

μg – microgram

μl – microliter

ml – milliliter

μm – micrometer, micron

μM – micromolar

mM – millimolar

ng – nanogram

nmoles – nanomoles

PCR – polymerase chain reaction

pg – picogram

pmoles – picomoles

qPCR – quantitative polymerase chain reaction

rpoC1 – RNA polymerase C1 (gamma) gene

RNA – ribonucleic acid

rRNA – ribosomal ribonucleic acid

s – second

SDS – sodium dodecyl sulfate

SIO – Scripps Institution of Oceanography

strep-PE – streptavidin-phycoerythrin

TMAC – tetramethyl ammonium chloride

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Brian Palenik, for guiding me through my doctoral degree. I thoroughly enjoyed my time in his lab and am grateful for his support and insight into my doctoral research, scientific career, and life beyond science.

Additionally I would like to give special thanks to two committee members, Bianca Brahamsha and Ron Burton, who were ever present during my degree for advice, encouragement, and a shared laugh.

I am honored to have had Farooq Azam, Kathy Barbeau, and Lin Chao as members of my doctoral committee. Thank you for having open ears and always willing to offer perspectives and ideas.

Without my husband, Art Poon, this dissertation would not have been possible. Love, support, but also intellect, inspiration, and even lessons in evolutionary and computational biology; words cannot begin to express my gratitude.

Many thanks to Rhona Stuart. Whether it was scientific or otherwise, I couldn’t have done without our countless discussions. I feel fortunate to have been paired with the ideal lab and office mate. And to Benjamin Pister and Melissa Roth for providing distractions from the rigors of research while simultaneously empathizing with the demands of progressing through our doctoral degrees. Thank you all for being wonderful colleagues and even better friends.

I am privileged to have collaborated with many excellent scientists during my doctoral degree. Specifically I would like to thank Ian Paulsen for including me in many collaborations involving microarray, genome sequencing, and metagenomic research. Also thanks to Andy Allen and Qinghu Ren for guidance through the world of metagenomics and to Chris Dupont and Drew Lucas for our collaborations at sea.

I could not have done without Eddie Kisfaludy for assisting me with field sampling; Mary Hilbern, Melissa Carter, and John McGowan for providing access to the Scripps Pier Chlorophyll Program dataset and assistance with field measurements; and Alan Saluk, Cody Fine, and Alex Ilic at the TSRI Flow Cytometry Core.

Thanks as well to members of the Palenik/Brahamsha lab from those who welcomed me to those who I will be leaving for creating a friendly and co-operative research environment. Specifically I would like to acknowledge Emy Daniels, Jillian Lai, Elizabeth Latham, Kimberly Tsai, and Rudy Urbano for assisting me in the lab or field and often generating critical data.

I would also like to thank the many denizens of Hubbs Hall for indulging my requests for equipment, reagents, and advice whenever asked and for the friendships that I will take with me wherever I go.

Finally, my sincerest thanks to my parents and parents-in-law for supporting my graduate education and my transition into motherhood. An especially big thank you to my mother-in-law, Magdalene Poon, for generously volunteering to Fiona-sit so I could complete this dissertation.

Chapter 1, in full, is a reprint of the material as it appears in Vera Tai and Brian Palenik. 2009. *The ISME Journal*. Volume 3, pages 903-915. The dissertation author was the primary investigator and first author of this publication.

Chapter 2 is currently being prepared for submission for publication by Vera Tai, Ronald Burton, and Brian Palenik. The dissertation author was the primary investigator and will be first author of this paper.

VITA

1998	Bachelor of Science (Honours), University of New Brunswick
2001	Master of Science, University of British Columbia
2001-2003	Research Assistant, The Scripps Research Institute
2003-2004	Research Lab Coordinator, Zoological Society of San Diego
2009	Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Tai, V., Paulsen, I. T., Phillipy, K., Johnson D. A., and Palenik, B. 2009. Whole genome microarray analyses of *Synechococcus-Vibrio* interactions. *Environmental Microbiology.* 11: 2698-2709.

Tai, V. and Palenik, B. 2009. Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site. *The ISME Journal*. 3: 903-915.

Tetu, S. G., Brahamsha, B., Johnson, D. A., **Tai, V.**, Phillippy, K., Palenik, B., and Paulsen, I. T. Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp. WH8102. 2009. *The ISME Journal*. 3: 835-849.

Palenik, B., Ren, Q., **Tai, V.**, and Paulsen, I. T. 2009. Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environmental Microbiology*. 11: 349-359.

Palenik, B., Grimwood, J., Aerts, A., Rouze, P., Salamovd, A., Putnam, N., Dupont, C., Jorgensen, R., Derelle, E., Rombauts, S., Zhou, K., Otillar, R., Merchant, S. S., Podell, S., Gaasterland, T., Napoli, C., Gendler, K., Manuell, A.,**Tai, V.**, Vallon, O., Piganeau, G., Jancek, S., Heijde, M., Jabbari, K., Bowler, C., Lohr, M., Robbens, S., Werner, G., Dubchak, I., Pazour, G. J., Ren, Q., Paulsen, I., Delwiche, C., Schmutz, J., Rokhsar, D., Van de Peer, Y., Moreau, H., and Grigoriev, I. V. 2007. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proceedings of the National Academy of Sciences, U. S. A.* 104: 7705-7710.

Calderwood, D. A., **Tai, V.**, Di Paolo, G., De Camilli, P., and Ginsberg, M. H. 2004. Competition for talin results in trans-dominant inhibition of integrin activation. *Journal of Biological Chemistry*. 279: 28889-28895.

Tadokoro, S., Shattil, S. J., Eto, K., **Tai, V.**, Liddington, R. C., de Pereda, J. M., Ginsberg, M. H., and Calderwood, D. A. 2003. Talin binding to integrin β tails: a final common step in integrin activation. *Science*. 302: 103-106.

Tai, V., Lawrence, J. E., Lang, A. S., Chan, A. C., Culley, A. I., and Suttle, C. A. 2003. Characterization of HaRNAV, a novel single-stranded RNA virus causing lysis of *Heterosigma akashiwo* (Raphidophyceae). *Journal of Phycology*. 39: 343-352.

Tai, V., Lindstrom, S. C., and Saunders, G. W. 2001. Phylogeny of the Dumontiaceae (Gigartinales, Rhodophyta) and associated families based on SSU rDNA and internal transcribed spacer sequence data. *Journal of Phycology*. 37: 184-196.

ABSTRACT OF THE DISSERTATION

Diversity and dynamics of *Synechococcus* populations in the Southern California Bight

by

Vera Tai

Doctor of Philosophy in Marine Biology

University of California, San Diego, 2009

Professor Brian Palenik, Chair

Marine cyanobacteria from the genus *Synechococcus* are found throughout the world's oceans and as primary producers, they have a significant role in the global carbon cycle. They are also genetically diverse

with at least 10 described clades and physiologically diverse capable of motility, using different sources of nitrogen or phosphorus, or modifying the composition of their photosynthetic pigments, for example. As more is learned of their diversity, however, it has become less clear how their genetic diversity reflects their physiological diversity and the ecological niches that they occupy.

By developing novel, culture-independent methods, biogeographic patterns were uncovered to provide a framework to link the genetic, physiological, and ecological diversity of marine *Synechococcus*. Quantitative PCR was used to show that the coastal Southern California Bight is dominated by two *Synechococcus* clades. Relative to each other, the abundance of these clades varied seasonally indicating that they occupied distinct ecological niches.

Synechococcus biogeography was explored further using a novel high-throughput hybridization method (Luminex) to examine the distribution of sub-clades within the two co-existing clades. Even at this scale of diversity, ecological differentiation was apparent. Most interestingly, in mesotrophic environments of the Southern California Bight a sub-clade was abundant in deep water (50 m) while most other sub-clades preferred surface waters.

Environmental metagenomic sequences were used to investigate the role of selection in shaping the diversity of *Synechococcus* populations. The large majority of genes have evolved under purifying selection. Genes that were well represented in environmental populations and may be essential for the population in the coastal Southern California Bight tended to be more evolutionarily conserved. Genes that may have been positively selected were mostly hypothetical genes and were more rare in the environment.

For this dissertation, the diversity, biogeography, and evolution of *Synechococcus* were examined at a level of detail not previously available. The results of this research formed a foundation for further investigations into the genetic and environmental factors regulating the ecological functions and niches of *Synechococcus* populations. More generally, this research addressed the scales at which bacterial evolution and diversification occur, and how bacterial populations are adapted to their environment.

INTRODUCTION

Like Leewenhoek's lenses, molecular biology has provided a window into the microbial world never before seen. With every completed microbial genome sequence and every environmental survey, microbiologists are beginning to understand the extent and scope of microbial diversity. But as we peer into the microbial world with our molecular tools, how do we make sense of the diversity before us?

This is the question driving the research I have conducted for my doctoral dissertation. By developing methods to examine the diversity and dynamics of marine *Synechococcus* (Cyanobacteria), a framework was generated to link the genetic/genomic variation observed in the environment with their biogeography, ecology, and evolution.

Marine *Synechococcus* are bacterial phototrophs whose abundance and global significance were discovered only thirty years ago (Johnson and Sieburth, 1979; Waterbury *et al.*, 1979). They have proven to be important ecologically as primary producers (Glover, 1985; Li, 1994; Agawin *et al.*, 2000), evolutionarily as descendents of the first oxygen-producing organisms (Mulkidjanian *et al.*, 2006), and physiologically by exemplifying the diversity of photosynthetic mechanisms and other biochemical functions of marine bacteria (see for eg. Palenik, 2001; Six *et al.*, 2007; Dupont *et al.*, 2008). My dissertation research has built upon this knowledge using data from phylogenetic markers, genomics, and metagenomics to show how marine *Synechococcus* can be used to understand the scale at which bacterial evolution and diversification occur, and how bacterial communities are adapted to their environment.

Only since the advent of molecular biology and DNA sequencing methodologies has the extent of bacterial diversity been wholly appreciated. Until recently, we have never before had the ability to fully explore microbial diversity. Using conserved gene sequences such as the 16S ribosomal RNA (rRNA) gene, the breadth of bacterial diversity has been revealed and this has fundamentally altered our knowledge of the evolution of life. We now know the tree of life is comprised of three primary branches, or domains (Woese et al., 1990). Two domains consist entirely of prokaryotic microbes, the Eubacteria and the Archaea. The third is the Eukarya, of which plants and animals are positioned at very tips of the eukaryotic branch.

Using similar methods, the evolution of the marine *Synechococcus* has also been investigated. This genus has been divided into several phylogenetic clades (Ferris and Palenik, 1998; Rocap *et al.*, 2002; Fuller *et al.*, 2003; Ahlgren and Rocap, 2006; Penno *et al.*, 2006), but as the classification is based solely on the DNA sequences of a few marker genes, it has been challenging to associate these clades with defined ecological functions and niches. The phylogenetic scales at which diversification and adaptation occur are also not well understood for *Synechococcus* or any microbial taxon. The goals of my dissertation were to develop methods to address these issues by providing basic but essential information regarding the biogeography and evolutionary history of *Synechococcus* in the Southern California Bight.

The Southern California Bight is one of the best-studied coastal systems in the world. Because of this, the circulation patterns and physical characteristics of the water are known and provide a backdrop from which to interpret *Synechococcus* biogeography and evolution. This coastal embayment stretches 1000 km from Point Conception, north of the Santa Barbara Channel to Cape Colnett, Baja California, just south of the U.S.-Mexico border. Its width extends from shore to the Patton Escarpment and reaches a maximum of 300 km (Jones, 1971; Dailey *et al.*, 1993).

Circulation in the Southern California Bight is generally counter-clockwise, bounded offshore by the equatorward flowing Callifornia Current, the eastern boundary current of the North Pacific gyre (Jones, 1971; Hickey, 1979; Lynn and Simpson, 1987; Di Lorenzo, 2003). During the summer, offshore water turns towards the coast then enters the Bight from the south and flows poleward. The water is warmer and slightly saltier than in the winter. Spring and summer winds also favor coastal upwelling, where deep nutrient-rich water is transported to the surface. Temperature and the temperature gradient with depth increase and reach a maximum in the early fall. In the winter, flow from the south is weakened and by springtime, the majority of the flow originates from the north traveling equatorward carrying colder and more oxygenated water.

Physical characteristics of the water in the Southern California Bight also change from off to on-shore and with depth. Light intensity and temperature both decrease with depth, whereas nutrients are low or nil at the surface but increase with depth (Eppley and Holm-Hansen, 1986). Offshore, the nutricline and pycnocline are deep, below 100m and chlorophyll concentrations are at a maximum in a layer just above the nutricline but are very low at the surface (Hayward and Venrick, 1998). Approaching shore, the nutricline and pycnocline shoal becoming very shallow inshore, sometimes out-cropping. Inshore, the maximum chlorophyll concentrations are higher and occur at shallower depths.

These physico-chemical variations undoubtedly have a significant role in shaping the composition of the *Synechococcus* community. By considering these variations alongside the biogeography of *Synechococcus* in the Southern California Bight, this dissertation has helped to understand the genetic and environmental factors regulating their abundances and ecological functions.

In chapter 1, a quantitative PCR method is described for examining the temporal dynamics of *Synechococcus* clades from a 3-year time series at the Scripps Institution of Oceanography (SIO) pier, located near the southern end of the Southern California Bight. *Synechococcus* diversity was further probed in Chapter 2 by developing the Luminex bead array to assess biogeography at a lower phylogenetic scale. In Chapter 3, population level diversity was analyzed using metagenomic data and the role of selection in shaping the diversity observed from coastal *Synechococcus* populations in the Southern California Bight was examined. Finally, in Appendix 1, I present preliminary data describing the genetic diversity of *Synechococcus* populations, specifically addressing the variability and evolution of horizontally transferred genes.

By describing the biogeography and evolutionary history of marine *Synechococcus*, this dissertation has provided crucial knowledge about an abundant, globally distributed, and ecologically important group of bacteria. As this research demonstrates, with continuing analysis of microbial community diversity, further ecological and evolutionary patterns await discovery.

References

Agawin, N.S.R., Duarte, C.M., and Agusti, S. 2000. Nutrient and temperature control of the contribution of picoplankton to phytoplankton biomass and production. *Limnol Oceanogr* 45:591-600.

Ahlgren, N.A. and Rocap, G. 2006. Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies. *Appl Environ Microbiol* 72:7193-7204.

Dailey, M.D., Anderson, J.W., Reish, D.J., and Gorsline, D.S. 1993. The California Bight: Background and setting. *In* Dailey, M. D., Reish, D. J. and Anderson, J. W. (eds). *Ecology of the Southern California Bight: a synthesis and interpretation*. University of California Press, Berkeley. pp 1-18.

Di Lorenzo, E. 2003. Seasonal dynamics of the surface circulation in the Southern California Current System. *Deep Sea Res Part II* 50:2371-2388.

Dupont, C.L., Barbeau, K., and Palenik, B. 2008. Ni uptake and limitation in marine *Synechococcus* strains. *Appl Environ Microbiol* 74:23-31.

Eppley, R.W. and Holm-Hansen, O. 1986. Primary production in the Southern California Bight. *In* Eppley, R. W. (eds). *Plankton Dynamics of the Southern California Bight*. Springer-Verlag, New York. pp 176-215.

Ferris, M.J. and Palenik, B. 1998. Niche adaptation in ocean cyanobacteria. *Nature* 396:226-228.

Fuller, N.J., Marie, D., Partensky, F., Vaulot, D., Post, A.F., and Scanlan, D.J. 2003. Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ Microbiol* 69:2430-2443.

Glover, H.E. 1985. The physiology and ecology of the marine cyanobacterial genus *Synechococcus*. *Adv Aquat Microbiol* 3:49-107.

Hayward, T.L. and Venrick, E.L. 1998. Nearsurface pattern in the California Current: coupling between physical and biological structure. *Deep Sea Res Part II* 45:1617-1638.

Hickey, B.M. 1979. The California Current System - hypotheses and facts. *Progress in Oceanography* 8:191-279.

Johnson, P.W. and Sieburth, J.M. 1979. Chroococcoid cyanobacteria in the sea: a ubiquitous and diverse phototropic biomass. *Limnol Oceanogr* 24:928-935.

Jones, J.H. 1971. *General Circulation and Water Characteristics in the Southern California Bight*. Southern California Coastal Water Research Project, Los Angeles.

Li, W.K.W. 1994. Primary production of prochlorophytes, cyanobacteria, and eukaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnol Oceanogr* 39:169-175.

Lynn, R.J. and Simpson, J.J. 1987. The California Current System - the seasonal variability of its physical characteristics. *J Geophys Res-Oceans* 92:12947-12966.

Mulkidjanian, A.Y., Koonin, E.V., Makarova, K.S., Mekhedov, S.L., Sorokin, A., Wolf, Y., Dufresne, A., Partensky, F., Burd, H., Kaznadzey, D., Haselkorn, R., and Galperin, M.Y. 2006. The cyanobacterial genome core and the origin of photosynthesis. *Proc Natl Acad Sci U S A* 103:13126-13131.

Palenik, B. 2001. Chromatic adaptation in marine *Synechococcus* strains. *Appl Environ Microbiol* 67:991-994.

Penno, S., Lindell, D., and Post, A.F. 2006. Diversity of *Synechococcus* and *Prochlorococcus* populations determined from DNA sequences of the N-regulatory gene *ntcA*. *Environ Microbiol* 8:1200-1211.

Rocap, G., Distel, D.L., Waterbury, J.B., and Chisholm, S.W. 2002. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* 68:1180-1191.

Six, C., Thomas, J.C., Garczarek, L., Ostrowski, M., Dufresne, A., Blot, N., Scanlan, D.J., and Partensky, F. 2007. Diversity and evolution of phycobilisomes in marine Synechococcus spp.: a comparative genomics study. *Genome Biol* 8:R259.

Waterbury, J.B., Watson, S.W., Guillard, R.R.L., and Brand, L.E. 1979. Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* 277:293-294.

Woese, C.R., Kandler, O., and Wheelis, M.L. 1990. Towards a natural system of organisms - proposal for the Domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* 87:4576-4579.

CHAPTER 1

The ISME Journal (2009) 3, 903–915

www.nature.com/ismej

ORIGINAL ARTICLE

Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site

Vera Tai and Brian Palenik
Marine Biology Research Division, Scripps Institution of Oceanography, University of California, La Jolla, San Diego, CA, USA

Marine cyanobacteria from the genus *Synechococcus* are found throughout the world's oceans and are important contributors to global primary productivity and carbon cycling. Cultured isolates and environmental DNA clone libraries of *Synechococcus* have demonstrated the diversity of these microbes. However, the natural distribution of this diversity through space and time and the ecological significance of their distribution are still poorly understood. To understand the seasonal dynamics of *Synechococcus* diversity, we have developed a quantitative PCR strategy using the gene encoding as a subunit of DNA-dependent RNA polymerase (*rpoC1*) and applied it to a 3-year time series of surface samples from the Scripps Institution of Oceanography pier (La Jolla, CA, USA), a coastal site in the northeastern Pacific Ocean. *Synechococcus* from clades I and IV were dominant throughout the time series and correlated with total *Synechococcus* abundance. The relative abundance of these two dominant clades showed evidence of a seasonal cycle. *Synechococcus* from clade IV were typically more abundant, but those from clade I dominated during periods just before the annual spring bloom of *Synechococcus*. *Synechococcus* from clades II and III were absent during spring and early summer, but appeared at low abundances in late summer and winter possibly due to changes in circulation in the Southern California Bight. As the first long-term time series describing *Synechococcus* population diversity, these temporal dynamics were used to interpret the genetic/genomic diversity observed in the environment and the potential factors regulating their distribution.
The ISME Journal (2009) **3,** 903–915; doi:10.1038/ismej.2009.35; published online 9 April 2009
Subject Category: microbial population and community ecology
Keywords: biogeography; cyanobacteria; microbial diversity; population ecology; quantitative PCR

Introduction

Marine cyanobacteria from the genus *Synechococcus* have a fundamental biogeochemical role in the world's oceans. Due to their photosynthetic lifestyle and sheer abundance, they are significant participants in oceanic primary production and the global carbon cycle (Johnson and Sieburth, 1979; Waterbury *et al.*, 1979; Glover, 1985; Li, 1994; Smith and Kemp, 2001; Richardson and Jackson, 2007). As a component of the picophytoplankton, this group accounts for 40% of global ocean primary productivity (Agawin *et al.*, 2000). In some coastal environments, *Synechococcus* accounts for 20% of the primary productivity (Li, 1994). But to properly understand their role in the environment, we need a better understanding of the ecological niches they occupy, how they adapt to changing environments and how these adaptations are linked to their evolution and diversification.

There is a strong foundation of genetic information describing the diversity of marine *Synechococcus*. Most marine *Synechococcus* form a monophyletic group based on phylogenetic trees constructed with 16S rRNA, *rpoC1* and other conserved gene sequences and are closely related with the other major marine cyanobacterial genus, *Prochlorococcus* (Palenik, 1994; Toledo and Palenik, 1997; Rocap *et al.*, 2002; Fuller *et al.*, 2003; Zeidner *et al.*, 2003; Chen *et al.*, 2004; Gibson *et al.*, 2006; Jenkins *et al.*, 2006; Penno *et al.*, 2006).

Within the marine *Synechococcus* group, distinct clades have been recognized based on *rpoC1* gene phylogenies (Palenik, 1994; Ferris and Palenik, 1998). These clades are supported by other gene phylogenies and have been named numerically with approximately 10 clades currently identified (Rocap *et al.*, 2002; Fuller *et al.*, 2003; Ahlgren and Rocap, 2006; Penno *et al.*, 2006). However, as these clades have been delineated based on gene

---

Correspondence: B Palenik, Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Drive, La Jolla, CA, 92093-0202, USA.
E-mail: bpalenik@ucsd.edu
Received 5 January 2009; revised 2 March 2009; accepted 4 March 2009; published online 9 April 2009

sequence phylogenies, phenotypic traits that define each *Synechococcus* clade have not been identified except for motility, which is found only in clade III strains (Toledo *et al*., 1999). In addition, little is known about biogeographical patterns in *Synechococcus* clades and their causes.

The spatial biogeography of *Synechococcus* diversity has been broadly described. *Synechococcus* from clades I, II and IV are the most common throughout the world's oceans (Zwirglmaier *et al*., 2008). Clades I and IV appear to be the dominant type in temperate, coastal environments (Brown and Fuhrman, 2005; Brown *et al*., 2005; Zwirglmaier *et al*., 2007; Zwirglmaier *et al*., 2008) and clade II is distributed in subtropical/tropical environments (Fuller *et al*., 2006; Zwirglmaier *et al*., 2008) and in off-shore, oligotrophic environments (Ferris and Palenik, 1998; Toledo and Palenik, 2003; Ahlgren and Rocap, 2006).

Few studies, however, have examined the temporal changes in *Synechococcus* diversity and abundance. These few studies have all sampled the Gulf of Aqaba in the Red Sea where *Synechococcus* from clade II are dominant (Fuller *et al*., 2005; Muhling *et al*., 2005, 2006; Penno *et al*., 2006). The genetic composition of *Synechococcus* from the Gulf of Aqaba has been shown to change over time, possibly due to changes in nutrient concentrations or viral infections (Muhling *et al*., 2005) but these studies were of short duration, typically 1 year, and used Southern hybridization, clone library or DNA fingerprinting methods which provided relative quantification only.

In this study, we have developed a quantitative PCR (qPCR) method based on *rpoC1* phylogeny to assess the abundance of the three major *Synechococcus* clades, clades I, II and IV and a fourth clade, clade III. The *rpoC1* gene sequence has proven to be a robust genetic marker for describing *Synechococcus* diversity from several marine environments (Palenik, 1994; Toledo and Palenik, 1997; Ma *et al*., 2004; Muhling *et al*., 2006). We have applied this qPCR method to a 3-year time series located at the Scripps Institution of Oceanography (SIO) pier, a coastal Pacific site of the Southern California Bight, and the data were analyzed in relation to physical and chemical parameters from the same site. As the first long-term time series describing *Synechococcus* population diversity, these temporal dynamics provide essential information to understand the interplay between genetic/genomic diversity, physiology and the environment.

Materials and methods

Sample collection

Surface seawater samples were collected from the end of the SIO pier in La Jolla, CA (32°53′N, 117°15′W) weekly in the morning from January 2005 to January 2008. The sampling site is beyond the surf zone and has a bottom depth of 7 m on average with a spring tidal range of 2 m. From the seawater samples, duplicates of 1 ml seawater were preserved with 0.25% gluteraldehyde (Sigma-Aldrich, St Louis, MO, USA), incubated at room temperature for 15 min, and stored at −80 °C until analyzed by flow cytometry. For DNA analysis, either 500 or 1000 ml of the sample was filtered through a 0.2 μm Supor filter disc (Pall Life Sciences, Ann Arbor, MI, USA) to collect microorganisms and the filter disc was stored at −80 °C.

Sample collection was timed to coincide with chlorophyll, temperature, salinity and nutrient (nitrate (NO_3^-), ammonia (NH_4^+), phosphate (PO_4^{3-}), nitrite (NO_2^-) and reactive silicate) measurements collected biweekly through the Southern California Coastal Ocean Observing System (SCCOOS) program.

Flow cytometry

Flow cytometry was used to assess the abundance of total *Synechococcus* cells. When using flow cytometry, *Synechococcus* cells are distinguished from other picoplankton by the fluorescence of their phycoerythrin pigments. The samples were analyzed at the highest flow rate using a FACSort flow cytometer (Becton-Dickinson, Franklin Lakes, NJ, USA) as described by Collier and Palenik (2003) but without applying a compensation matrix. Duplicates were counted for samples also used for DNA extraction and qPCR analysis.

DNA extraction

DNA was extracted from microorganisms collected on 0.2 μm Supor filter discs (Pall) from monthly to bimonthly samples. The filter discs were first incubated in 12.5 mg/ml lysozyme (Roche Applied Science, Indianapolis, IN, USA), 20 mM EDTA (Sigma), and 50 mM Tris (Sigma), pH 8.0 for 30 min at 37 °C. SDS (Bio-Rad, Hercules, CA, USA) and proteinase K (Sigma) were added to final concentrations of 1% and 1 mg ml^{-1}, respectively, and incubated for 2.5 h at 55 °C. RNA was removed by adding 2 mg ml^{-1} RNAse (Sigma) for 5 min at room temperature. Proteins were removed by extracting twice with an equal volume of 25:24:1 phenol/chloroform/isoamyl alcohol (Invitrogen, Carlsbad, CA, USA) and once with 24:1 chloroform/isoamyl alcohol (Fisher Scientific, Pittsburg, PA, USA). Using phenol/chloroform dissolved the Supor filter and increased the DNA extraction efficiency. The DNA in the aqueous phase was purified using a silica gel column (DNeasy Blood and Tissue Kit; Qiagen, Valencia, CA, USA) following the manufacturer's protocol and eluted from the column using 400 μl of the manufacturer's buffer. DNA was quantified using PicoGreen (Quant-iT; Qiagen), a DNA-binding fluorescent dye, following the manufacturer's protocol.

Cultured representatives of *Synechococcus* clades (CC9311 for clade I, CC9605 for clade II, WH8102 for clade III and CC9902 for clade IV) were counted, filtered and their genomic DNA extracted following the same procedure as for environmental samples. On average, 10^{10} cells were filtered for DNA extraction in a final volume of 400 µl. The DNA extractions from cultured strains were used to test the specificity of clade-specific qPCR primers and as concentration standards.

qPCR

For each clade, two oligonucleotide primers were designed to amplify 145–304 bp fragments of the *rpoC1* gene (Table 1). *rpoC1* is a single copy gene and thus the problems when using the 16S rRNA gene as the marker locus of normalizing cell number with multiple 16S rRNA gene copies were avoided.

The primer sequences are unique for a given clade based on an alignment of all available *Synechococcus rpoC1* gene sequences in the nonredundant (nr) GenBank database from both environmental and cultured sources. A partial alignment of primer regions is provided in Supplementary Table S1. Using DNA from cultured representatives (see above), the primers were tested for clade specificity using a range of annealing temperatures in PCR reactions. An appropriate annealing temperature was chosen (Table 1) so that DNA only from the targeted clade amplified.

qPCR reactions consisted of 1X Brilliant SYBR Green QPCR Master Mix (Stratagene, La Jolla, CA, USA), 30 nM Rox reference dye (Stratagene), 0.8 µM of each primer (Integrated DNA Technologies, Coralville, IA, USA), 0.1 mg/ml BSA (New England BioLabs, Ipswich, MA, USA), and template DNA. Ten-fold serial dilutions of environmental DNA (generally ranging from 5 ng to 0.05 pg DNA) were used as templates in the qPCR reactions. The serial dilutions served as replicates and to dilute any PCR inhibitors occurring in the DNA extractions. DNAs from cultured strains of *Synechococcus* were serially diluted over 6 orders of magnitude for use as concentration standards. Using a Stratagene 3000P machine to acquire SYBR Green fluorescence data, the reactions were incubated at 95 °C for 10 min followed by 40 cycles of 95 °C for 30 s, the clade-specific annealing temperature for 30 s and 72 °C for 30 s, and ended with 95 °C for 1 min, a ramp down to 55 °C, and a ramp up to 95 °C at 0.2 °C s^{-1} to assess the dissociation curve of the amplified product.

Template DNAs for the standard curves were obtained from cultured cells that were counted before the DNA extraction, thus the amount of DNA added to each standard qPCR reaction represents the DNA from a known number of cells taking into account any inefficiencies in the DNA extraction. For environmental samples, the threshold cycle (C_t) determined from each qPCR was converted to the number of cells based on the standard curve. The number of cells in each qPCR was converted to cells per ml based on the volume of seawater filtered for each sample and the volume of the DNA extraction used in the qPCRs (no. of cells in the qPCR/(volume of seawater filtered/(volume of DNA extracted × volume of DNA in qPCR)).

rpoC1 *clone library*

Two cyanobacteria-specific primers (SAN-157F = 5′ YTN AAR CCN GAR ATG GAY GG and SAC-1039R = 5′ CYT GYT TNC CYT CDA TDA TRT C, the numbers in the primer names denote the first nucleotide where the primers are located along the *rpoC1* gene) were designed (by J Collier) to amplify an 841 bp fragment of the *rpoC1* gene. These primers generate a longer fragment of the *rpoC1* gene and are a better match for marine cyanobacteria than previously published primer pairs (Palenik and Haselkorn, 1992; Muhling *et al.*, 2006). Using the DNA extracted from a sample collected from the SIO pier on 12 May 2005, PCR was used to amplify the *rpoC1* gene fragment from the cyanobacterial population. Template DNA (20 ng) was combined with 1.33 µM of each primer and 25 µl of Platinum PCR SuperMix (Invitrogen). This mixture was incubated at 94 °C for 2 min, followed by 30 cycles at 94 °C for 30 s, 52 °C for 1 min and 72 °C for 1 min, and ending with an incubation at 72 °C for 10 min.

Table 1 *Synechococcus* clade-specific qPCR primers

Clade	*Model strain*	*Primer*	*Primer sequence (5′–3′)*	*Temp. (°C)*	*Size (bp)*	*Eff. (%)*
I	CC9311	RPOC362F-I	TGA AAG GGA TYC CCA GTT ATG T	59	304	96.3
		RPOC665R-I	CCC TTA CTI CCA GCA ATC TC			
II	CC9605	RPOC378F-II	CTA CGT GGC CAT CCT GCT	60	170	93.6
		RPOC547R-II	TCG GAR TCT TCG GCG TAG ATC			
III	WH8102	RPOC578F-III	GTG CTG AGG CGC TCA AGC AA	65	144	94.4
		RPOC721R-III	TGG CCA CGA AGT TAT CGA TGA C			
IV	CC9902	RPOC306F-IV	CCG TCA TCG GAT GGG CTT	56	159	99.1
		RPOC464R-IV	TCA CCT GGI TCA AGA ACG AC			

The number in the primer name denotes the first nucleotide where the primers are located along the *rpoC1* gene.
Temp. is the annealing temperature used for clade-specific amplification. Size is the nucleotide length of the PCR product in base pairs (bp). Eff. is the amplification efficiency calculated from the slope of the standard curves (Figure 2).

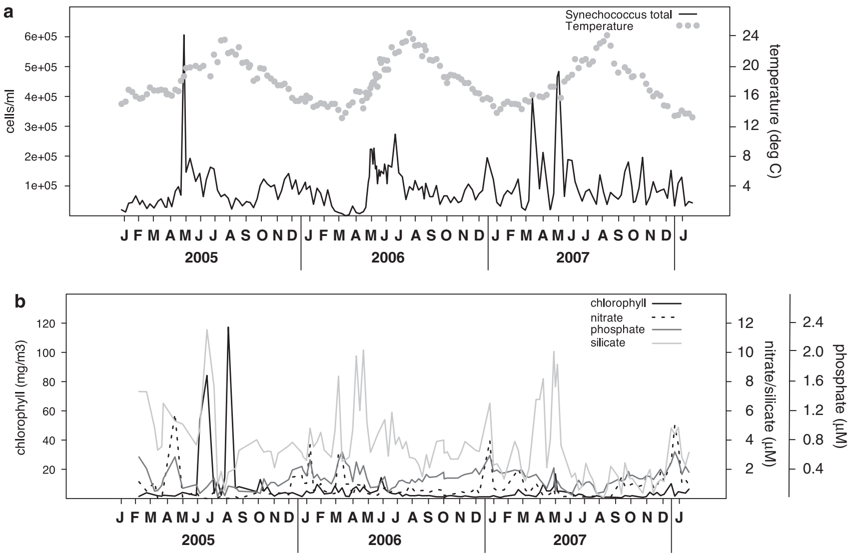


Figure 1 Three-year time series of weekly surface samples from the Scripps Institution of Oceanography (SIO) pier of (**a**) total *Synechococcus* abundances and temperature; (**b**) chlorophyll, nitrate, phosphate and silicate concentrations. Time is presented as the first letter of the month for each year.

The resulting PCR product was ligated into pCR4-TOPO plasmid vectors and cloned via transformation into TOP10 chemically competent *Escherichia coli* cells following the manufacturer's protocol (Invitrogen). Plasmid DNA was prepared from 21 clones using the QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's protocol. The *rpoC1* gene fragment insertion was sequenced (SeqXcel, San Diego, CA, USA) from both ends using the T3 and T7 sequencing primers provided by the manufacturer (Invitrogen).

The *rpoC1* gene fragment sequences were trimmed of plasmid and primer sequences and deposited in GenBank (accession numbers FJ790799–FJ790819). The sequences were aligned to known *rpoC1* gene sequences from cultured representatives of marine cyanobacteria, that is, *Synechococcus* and *Prochlorococcus*, available through GenBank. A phylogenetic tree was calculated using the neighbor-joining algorithm in PAUP (v.4.0b10) with distances calculated using the hky85 model. A 50% majority-rule consensus tree calculated from 1000 bootstrap replicates provided statistical support.

Correlation analyses

Weekly total *Synechococcus* abundances, temperature, chlorophyll, salinity and nutrient measurements were subjected to pairwise correlation analyses. A separate dataset of bimonthly to monthly clade-specific abundances, total *Synechococcus* abundances, temperature, chlorophyll, salinity and nutrient concentrations was also analyzed. Except for temperature, the data were log transformed (log $(x+1)$) before analysis to obtain an approximately normal distribution of data values. The Pearson's product-moment correlation coefficient (r) and the nonparametric Spearman's correlation coefficient (ρ) were calculated using JMP v. 5.0.

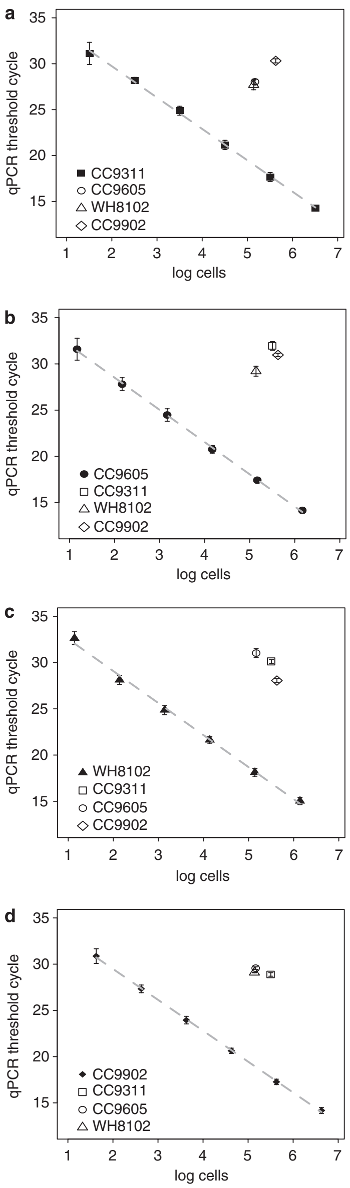
Results

From weekly samples over 3 years, total *Synechococcus* abundances were determined using flow cytometry. By examining the total abundances, a clear annual pattern emerged (Figure 1). Consistently in mid-May, *Synechococcus* abundance increased above basal fluctuations indicating that *Synechococcus* blooms on an annual basis. The timing of the blooms is very consistent, with the annual abundance peaks occurring on 12 May 2005 (6.06×10^5 cells ml^{-1}), 18 May 2006 (2.27×10^5 cells ml^{-1}) and 17 May 2007 (4.84×10^5 cells ml^{-1}). Second abundance peaks occurred in 2006 at the end of June (29 June 2006, 2.73×10^5 cells ml^{-1}) and in 2007 at the end of March (26 March 2007, 3.95×10^5 cells ml^{-1}), but, in general, *Synechococcus* abundances remained below 2×10^5 cells ml^{-1} throughout the rest of the year. Abundance minimums occurred throughout the year; the lowest in the time-series was 319 cells ml^{-1} on 23 March 2006. Total *Synechococcus* abundances weakly correlated with temperature (Pearson's $r = 0.30$, $P = 0.0001$, Spearman's $\rho = 0.28$, $P = 0.0002$) and negatively with phosphate ($r = -0.31$, $P = 0.005$, $\rho = 0.22$, $P = 0.016$), but did not correlate significantly with salinity, chlorophyll or any of the other measured nutrients. The lack of correlation with chlorophyll concentration is not surprising as larger phytoplankton, such as diatoms and dinoflagellates, can dominate more than 70% of the chlorophyll (Melissa Carter, personal communication).

Primers were developed for qPCR to assess the genetic composition of the *Synechococcus* population during this time series. The primers were tested for specificity against genomic DNA from cultured representatives. At the annealing temperature used, the primers were clade specific and did not amplify DNA from any nontarget strain (Figure 2; Supplementary Figure S1). All primers quantified DNA over 6 orders of magnitude equivalent to 10^2 to 10^7 cells. Based on the amount of seawater filtered for the DNA extraction, the detection limit for this method is approximately 10 cells ml^{-1}.

From bimonthly to monthly samples, qPCR was used to evaluate the diversity of the *Synechococcus* population from the SIO pier. *Synechococcus* from clades I and IV consistently dominated this coastal environment throughout the year (Figure 3a and b). Both clade I and clade IV abundances significantly correlated with total *Synechococcus* concentrations (Figure 4a and b; Supplementary Table S2), and the abundance maxima of these clades coincided with the annual blooms in total *Synechococcus* (Figure 3a and b). The clade I and clade IV abundances were also strongly correlated to one another (Figure 4c; Supplementary Table S2).

These two clades accounted for an average of 168 ± 54%, of the total *Synechococcus* abundance (Figure 3c). The values ranged from 83% to 335% with the exception of 12 May 2005 where the two clades accounted for only 34% of the total abundance (see below). The qPCR assay may have overestimated the clade-specific abundances because dividing cells with replicated chromosomes



◄ **Figure 2** Range and specificity of quantitative PCR (qPCR) reactions using dilutions of genomic DNA from cultured *Synechococcus* strains: (**a**) clade I qPCR primers targeting CC9311; (**b**) clade II qPCR primers targeting CC9605; (**c**) clade III qPCR primers targeting WH8102; (**d**) clade IV qPCR primers targeting CC9902. The threshold cycle is the PCR cycle at which the fluorescence signal begins to increase. The number of cells in each qPCR reaction, rather than the amount of DNA, is plotted on the *x* axis. Closed symbols indicate qPCR reactions using DNA from *Synechococcus* strains targeted by the clade-specific qPCR primers. Open symbols indicate qPCR reactions where DNAs from nontarget strains were assayed.

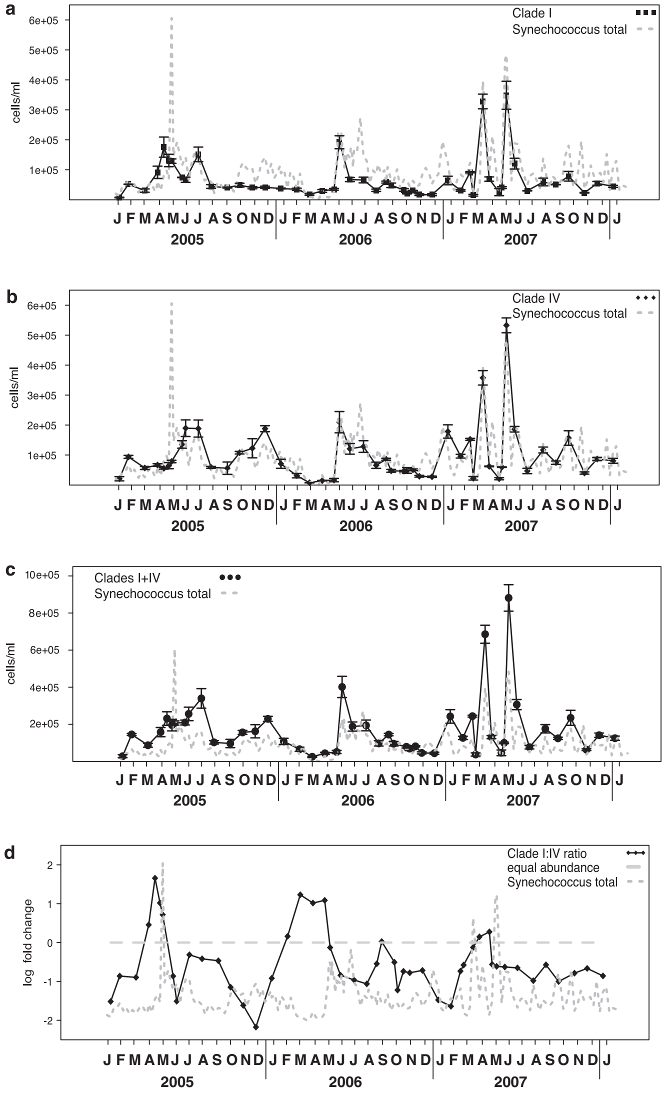
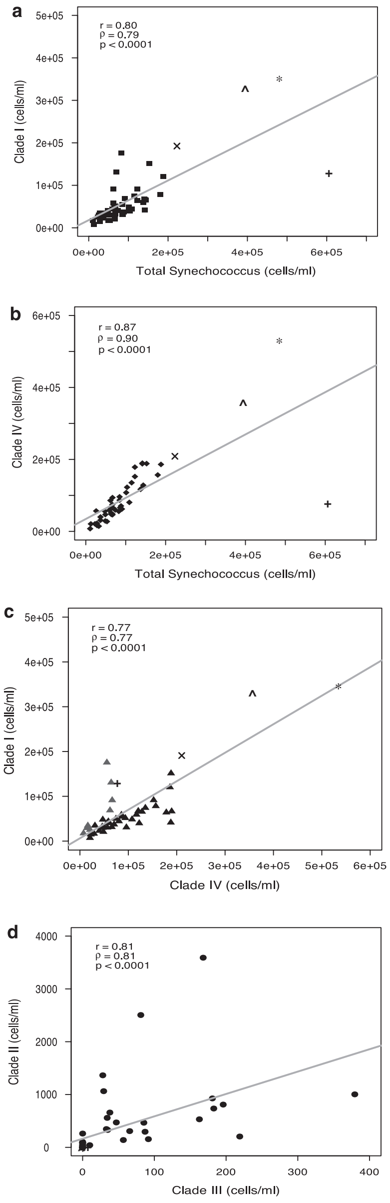


Figure 3 Abundances of the two dominant clades over 3 years from Scripps Institution of Oceanography (SIO) pier surface samples: (**a**) clade I abundances; (**b**) clade IV abundances; (**c**) clade I + clade IV abundances; (**d**) $\log_2$ fold change of clade I to clade IV abundances. Positive values indicate a greater abundance of clade I in the environment. Negative values indicate a greater abundance of clade IV. Zero (0) indicates that each clade was in equal abundance (dashed gray line). In each panel, the dotted gray line indicates total *Synechococcus* abundances, measured weekly. Clade-specific abundances were measured monthly.



would be counted as single cells by the flow cytometer. This suggests that cells in the environment typically contained replicated chromosomes, but not the cultured cells used for qPCR standards.

The ratio of clade I to clade IV abundances was used to assess changes in the abundance of these two dominant clades relative to one another. For most of the year, clade IV *Synechococcus* were at least twice as abundant as clade I ($\log_2$ fold change < -1.0; Figure 3d). However, each year just before the annual peak in total *Synechococcus*, clade I increased to twice the abundance of clade IV ($\log_2$ fold change > 1.0; Figure 3d).

Synechococcus from clades II and III rarely occurred in this environment or occurred in low abundances (Figure 5). Clade II was typically more abundant than clade III, but both clades appeared in the winter when the water is colder and more oligotrophic, and also during the summer when the water is more likely to be stratified. The abundance of these clades was not correlated with total *Synechococcus* (Supplementary Table S2), but clade II abundances were significantly correlated with clade III abundances suggesting their dynamics may be influenced by similar mechanisms (Figure 4d).

An *rpoC1* clone library was generated to determine the composition of the *Synechococcus* population on 12 May 2005. On this date, a bloom of total *Synechococcus* was sampled, but qPCR failed to indicate which were the dominant clade types. This sample is clearly an outlier based on the plots of clade I and clade IV abundances versus total *Synechococcus* (Figures 4a and b). However, phylogenetic analyses of the sequences from the clone library indicated that the majority of the *Synechococcus* from this sample belonged to clades I and IV (Figure 6).

On examining the *rpoC1* sequences from the library, the clade I and clade IV qPCR primers do match the respective *rpoC1* sequences from this clone library, so it is not known why the qPCR assay failed to recognize the dominance of these two clades on this date. Chemical inhibitors coextracted with the environmental DNA may have inhibited the qPCR reaction. However, dilution of the environmental DNA should have avoided this problem.

◀

Figure 4 Scatter plots of clade-specific and total *Synechococcus* abundances from samples assayed by quantitative PCR (qPCR) over the 3-year time series: (**a**) clade I versus total *Synechococcus* abundances; (**b**) clade IV versus total *Synechococcus* abundances; (**c**) clade I versus clade IV abundances; (**d**) clade II versus clade III abundances. In each panel, samples from peaks in *Synechococcus* abundance are represented by the symbols: + (12 May 2005), × (14 May 2006), ^ (26 March 2007) and * (17 May 2007). In (**c**), gray symbols represent samples just before the *Synechococcus* bloom where clade I abundances exceeds clade IV. Results from correlation analyses (Pearson's correlation coefficient (r) and Spearman's correlation coefficient (ρ) and its associated P value) are provided. The best linear fit is shown as a light gray line.

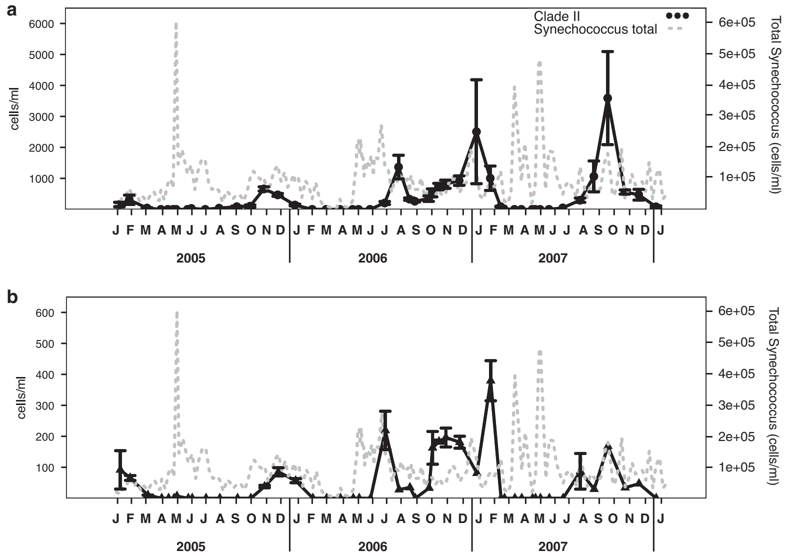


Figure 5 Abundances of the two minor clades over 3 years from Scripps Institution of Oceanography (SIO) pier surface samples: (**a**) clade II abundances; (**b**) clade III abundances. In each panel, the dotted gray line indicates total *Synechococcus* abundances, measured weekly. Clade-specific abundances were measured monthly.

Seventeen and three environmental sequences grouped with strong support to cultured clade I and clade IV strains, respectively. For clade I, the environmental sequences grouped as two separate, but strongly supported, clades distinct from cultured strains. This suggests that the genetic diversity of the *Synechococcus* population in the environment can differ significantly from that of isolated strains as also seen recently in a metagenomic study at our monitoring site (Palenik *et al.*, 2009).

Discussion

We have developed a qPCR strategy to quantify *Synechococcus* clades, as defined by *rpoC1*, 16S rRNA and 16S–23S internal transcribed spacer sequences (Ferris and Palenik, 1998; Rocap *et al.*, 2002; Fuller *et al.*, 2003). This approach enabled us to document the changes in *Synechococcus* diversity from a 3-year time series of a coastal Pacific Ocean site. From this time series, it is clear that the *Synechococcus* population was dominated by clades I and IV throughout the year (Figure 3a and b). Secondly, the abundances of these two dominant clades were correlated with the total *Synechococcus* abundance, which showed annual abundance peaks in the month of May (Figures 4a and b). A third intriguing result was that the abundances of *Synechococcus* from clade I relative to those from clade IV did not remain static, but changed throughout the year with evidence of a seasonal cycle. In particular, although *Synechococcus* from clade IV were generally the dominant type, *Synechococcus* from clade I dominated over clade IV just before the annual abundance peaks of total *Synechococcus* (Figure 3d).

Seasonal pattern of total Synechococcus abundance

The annual pattern of total *Synechococcus* abundance from this coastal Pacific Ocean site differs from that of a temperate coastal Atlantic site (Woods Hole Harbor) and an oligotrophic site (Sargasso Sea). At the SIO pier, abundance maxima above 2×10^5 cells ml^{-1} last approximately 1 week and generally occur as the temperature rises above 17 °C. In contrast, from Woods Hole Harbor, *Synechococcus* blooms when the water reaches 6 °C and the bloom period lasts months, from April until June, with high abundances of *Synechococcus* (10^5 per ml) observed until November (Waterbury *et al.*, 1986). In the Sargasso Sea, *Synechococcus* peaks in March/April with the bloom period lasting several months (Durand *et al.*, 2001).

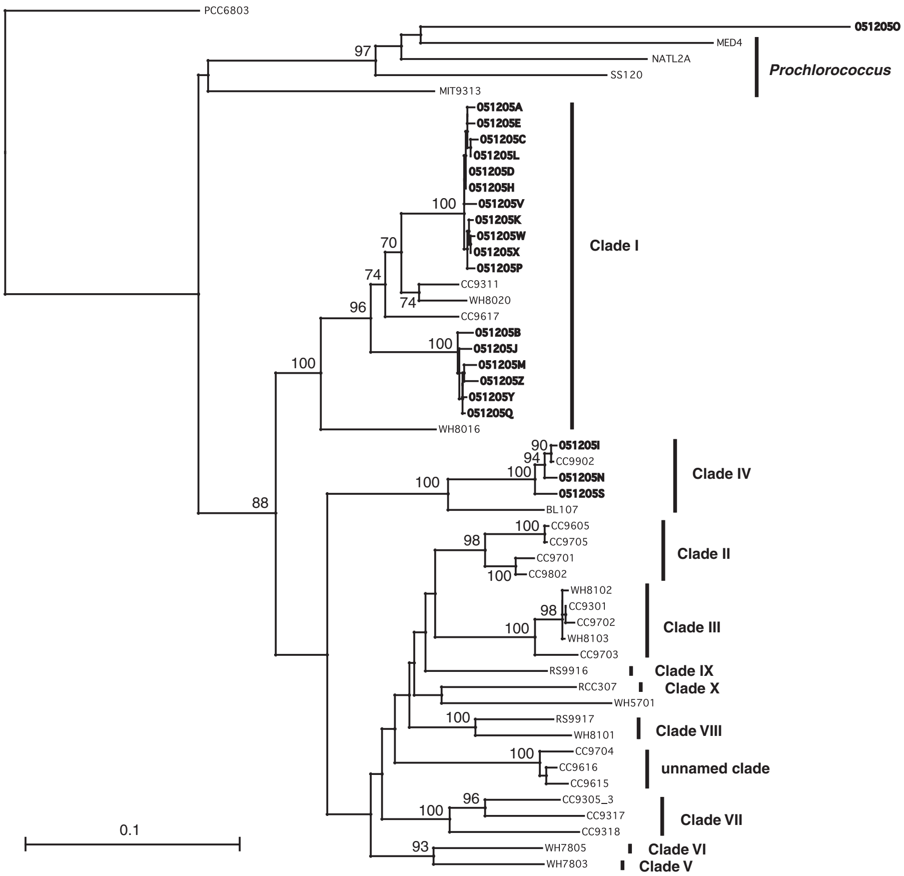


Figure 6 Neighbor-joining tree calculated from an alignment of *rpoC1* sequences from a Scripps Institution of Oceanography (SIO) pier sample on 12 May 2005 and cultured isolates. Sequences from the SIO pier sample are in bold. Values at nodes indicate percentage bootstrap support based on 1000 replicates. Values less than 70% are not shown. The scale bar indicates 10% divergence.

A rise beyond a threshold temperature appears to be a consistent factor that coincides with the timing of bloom initiation in both coastal Pacific and Atlantic oceans. Higher temperature likely causes an increased growth rate of *Synechococcus* resulting in greater abundances (Agawin *et al.*, 1998). The duration of the blooms, however, is probably more dependent on other environmental factors such as grazing pressure or nutrient availability. At the SIO pier, the relatively short bloom period suggests that mortality rates (for example, through grazing or viral lysis) also increased causing a rapid decrease in *Synechococcus* abundance.

Codominance of Synechococcus from clades I and IV

Using qPCR, we have shown that *Synechococcus* from both clades I and IV dominate the SIO pier coastal environment, yet they have different temporal dynamics. The codominance of the *Synechococcus* population by clades I and IV has been documented in other coastal environments of the Atlantic Ocean, Mediterranean Sea, eastern South Pacific Ocean, California coast and Arctic Ocean and their dominance is thought to be restricted to higher latitudes above 30°N and below 30°S (Brown and Fuhrman, 2005; Brown *et al.*, 2005; Zwirglmaier *et al.*, 2007, 2008). In the Eastern South Pacific and the Atlantic, clade IV is more abundant than clade I, but in the Arctic they occur equally (Zwirglmaier *et al.*, 2008). However, these biogeographic surveys may have missed the temporal aspects of *Synechococcus* diversity.

From our 3-year time series, we show that the dominance of one clade over the other is dynamic over time and exhibits seasonal patterns. In particular, *Synechococcus* from clade I exceeds those from clade IV just before the annual abundance peak suggesting that the environment regularly favors clade I (or disfavors clade IV) to the disadvantage of clade IV. However, the relative importance of genetic compared to environmental factors driving these clade-specific changes in *Synechococcus* abundance is poorly understood, but it is likely that both play significant roles.

A matching temporal dataset from the SIO pier of temperature, salinity, and nitrate, phosphate, silicate and ammonia concentrations was analyzed to examine potential bottom-up environmental effects driving *Synechococcus* clade dynamics. Unfortunately none of these significantly correlated with clade-specific abundances (Supplementary Table S2). However, this does not mean that bottom-up effects are insignificant factors in dictating *Synechococcus* abundance or diversity. Other nutrients that have not been measured could have a significant influence on *Synechococcus* dynamics. *Synechococcus* have been shown to use urea as a nitrogen source (Collier *et al.*, 1999) and other organic nutrients are likely important to their survival (Moore *et al.*, 2002, 2005; Zubkov *et al.*, 2003; Bronk *et al.*, 2007). Trace metals may also play a significant role (Hutchins *et al.*, 1999; Leao *et al.*, 2007; Dupont *et al.*, 2008).

The top-down environmental effects on *Synechococcus* populations are also not well understood. A wide diversity of both protozoan grazers (Caron *et al.*, 1991; Strom, 1991; Christaki *et al.*, 1999, 2002; Guillou *et al.*, 2001; Jeong *et al.*, 2005) and viruses (Suttle and Chan, 1993; Waterbury and Valois, 1993; Wilson *et al.*, 1993) have been identified that prey on *Synechococcus*. These mortality agents can be the dominating factor regulating total *Synechococcus* abundances (see for example, Baudoux *et al.*, 2007), but their overall diversity, specificity to *Synechococcus* types and impact as selective agents remain largely unknown.

Genetic information in association with biogeographic data provides another basis for understanding the factors driving the dominance of clades I and IV in coastal environments and their coexistence. Genome sequences from cultured representatives of clades I and IV are available (Palenik *et al.*, 2006; Dufresne *et al.*, 2008) as are sequences from representatives of several other clades (Palenik *et al.*, 2003) providing a strong backbone of genetic information for this microbial taxon.

CC9311, a representative of clade I isolated from the California Current, has a genome with increased sensory and response regulator genes, fewer phosphate scavenging genes, a greater capacity for metal transport and usage and more genes related to organic nutrient use than an open ocean strain (Palenik *et al.*, 2006). These genes may be key to their survival in a coastal location with higher nutrient availability and a more variable, dynamic environment compared to an oligotrophic one.

The genomes of two clade IV isolates, CC9902 and BL107, have also been sequenced. These strains have fewer sensory histidine kinase and response regulator genes than CC9311 and could represent a ‘specialist’ lifestyle (Dufresne *et al.*, 2008). The genomes of CC9902 and BL107 are approximately 0.38 and 0.33 Mb smaller than CC9311, respectively, and have fewer clade-specific genes. With a larger genetic repertoire, clade I types may be more versatile or ‘opportunists’ (Dufresne *et al.*, 2008). For example, CC9311 has genes capable of transporting iron (II) or halogenating compounds via a vanadium-dependent bromoperoxidase (Palenik *et al.*, 2006), but without more specific physiological investigations, the correlations between genetic composition and niche adaptation remain hypothetical.

The correlations between genetic composition and ecological niche are further complicated by the tremendous genetic diversity that exists within *Synechococcus* clades. By examining the diversity captured by our *rpoC1* clone library, *Synechococcus* that were dominant in the environment grouped separately from cultured representatives (Figure 6). In addition, as with other microbial taxa (Medini *et al.*, 2005; Tettelin *et al.*, 2005; Kettler *et al.*, 2007), *Synechococcus* genomes contain mostly conserved core genes, but many highly variable regions exist containing genes with a diversity of functions, likely the result of horizontal gene transfer events, that may be significant to their physiology and ecological niche (Palenik *et al.*, 2009).

Occasional occurrence of Synechococcus from clades II and III

For the two minor clades, clades II and III, their presence in this coastal environment may be more reflective of physical oceanographic patterns. Their occurrence in late summer coincides with periods of strong stratification (Palacios *et al.*, 2004). Stratification has the effect of reducing the input of colder, more saline and nutrient-rich waters from below, therefore surface waters become nutrient depleted. In the fall and winter, stratification is weaker, but there is a consistent northward flow, the Inshore Countercurrent, in the coastal regions of the Southern California Bight in contrast to the southward flow that dominates all other times of the year (Hickey, 1979; Lynn and Simpson, 1987; Di Lorenzo, 2003). The source of the northward flowing waters is the off-shore eastern boundary current of the North Pacific gyre, the California Current, that has turned toward the shore south of the US-Mexico border, then divides and turns north into the Southern California Bight as the Inshore Countercurrent and south alongside Baja California (Hickey, 1979; Lynn and Simpson, 1987). This northward flow brings more oligotrophic, nutrient-poor waters to the coast and may advect *Synechococcus* from clades II and III that are relatively more abundant off-shore toward the coast (Ferris and Palenik, 1998; Toledo and Palenik, 2003). Although our sampling site is on the coastal shelf, the site is influenced by these major circulation trends, as well as by alongshore currents, cross-shelf tidal currents, and currents generated by local wind stress (Winant and Bratkovich, 1981).

Both stratification and the northward Inshore Countercurrent may have the effect of lowering nutrient concentrations of surface waters in the Southern California Bight and these conditions may explain the occurrence of *Synechococcus* from clades II and III at the SIO pier. However, we were unable to find a significant relationship between nutrient concentrations and the abundance of these clades.

An alternative explanation may be the advection of *Synechococcus* from clade II from the south. If clade II is predominantly distributed in subtropical/tropical coastal environments (Zwirglmaier *et al.*, 2008), then northward flow may bring these *Synechococcus* types to Southern California from the south. Observing a transition from clade I and clade IV dominance in temperate environments to clade II closer to the equator would support this hypothesis. A similar shift in *Synechococcus* clade dominance has been observed in the Atlantic Ocean (Zwirglmaier *et al.*, 2007).

The observation of *Synechococcus* from clades II and III in late summer and winter loosely coincides with the presence of *Prochlorococcus* at this site. *Prochlorococcus* is usually absent at the SIO pier during the spring months, although in addition to advection to and away from this location, their absence may also be due to increased grazing pressure (Worden *et al.*, 2004).

Conclusion

The biogeography of environmental microbes is a fundamental piece of information that is critical to understanding their evolution, adaptability and function in their respective ecosystems (Ramette and Tiedje, 2007). In addition, as *Synechococcus* are major contributors to primary production, their biogeography and the genetic and environmental factors influencing the composition of the population has important implications for understanding carbon cycling and other globally important biogeochemical processes. Through the use of qPCR, we have provided an important tool for studying *Synechococcus* biogeography, ecology and evolution. Although we have used qPCR to assess the temporal dynamics of four globally distributed *Synechococcus* clades, this method can be applied to any scale of spatial and temporal sampling. With these tools in hand, the key is to discover the scales at which adaptation occurs and how genetic diversity influences and informs us of their ecological niches.

Acknowledgements

Funding for this project was provided through a National Science Foundation grant (NSF OCE05-26979) to BP and Ron Burton (SIO). We thank Ron Burton for comments on the experimental protocols and data analysis. We also thank J McGowan, M Carter and M Hilbern for collecting and providing access to the Scripps Pier Chlorophyll Program dataset, funded by NOAA and COTS through the Southern California Coastal Ocean Observing System (SCCOOS).

References

Agawin NSR, Duarte CM, Agusti S. (1998). Growth and abundance of *Synechococcus* sp. in a Mediterranean Bay: seasonality and relationship with temperature. *Mar Ecol Prog Ser* **170**: 45–53.

Agawin NSR, Duarte CM, Agusti S. (2000). Nutrient and temperature control of the contribution of picoplankton to phytoplankton biomass and production. *Limnol Oceanogr* **45**: 591–600.

Ahlgren NA, Rocap G. (2006). Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies. *Appl Environ Microbiol* **72**: 7193–7204.

Baudoux AC, Veldhuis MJW, Witte HJ, Brussaard CPD. (2007). Viruses as mortality agents of picophytoplankton in the deep chlorophyll maximum layer during IRONAGES III. *Limnol Oceanogr* **52**: 2519–2529.

Bronk DA, See JH, Bradley P, Killberg L. (2007). DON as a source of bioavailable nitrogen for phytoplankton. *Biogeosciences* **4**: 283–296.

Brown MV, Fuhrman JA. (2005). Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. *Aquat Microb Ecol* **41**: 15–23.

Brown MV, Schwalbach MS, Hewson I, Fuhrman JA. (2005). Coupling 16S-ITS rDNA clone libraries and automated ribosomal intergenic spacer analysis to show marine microbial diversity: development and application to a time series. *Environ Microbiol* **7**: 1466–1479.

Caron DA, Lim EL, Miceli G, Waterbury JB, Valois FW. (1991). Grazing and utilization of chroococcoid cyanobacteria and heterotrophic bacteria by protozoa in laboratory cultures and a coastal plankton community. *Mar Ecol Prog Ser* **76**: 205–217.

Chen F, Wang K, Kan JJ, Bachoon DS, Lu JR, Lau S *et al.* (2004). Phylogenetic diversity of *Synechococcus* in the Chesapeake Bay revealed by ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) large subunit gene (rbcL) sequences. *Aquat Microb Ecol* **36**: 153–164.

Christaki U, Courties C, Karayanni H, Giannakourou A, Maravelias C, Kormas KA *et al.* (2002). Dynamic characteristics of *Prochlorococcus* and *Synechococcus* consumption by bacterivorous nanoflagellates. *Microb Ecol* **43**: 341–352.

Christaki U, Jacquet S, Dolan JR, Vaulot D, Rassoulzadegan F. (1999). Growth and grazing on *Prochlorococcus* and *Synechococcus* by two marine ciliates. *Limnol Oceanogr* **44**: 52–61.

Collier JL, Brahamsha B, Palenik B. (1999). The marine cyanobacterium *Synechococcus* sp. WH7805 requires urease (urea amidohydrolase, EC 3.5.1.5) to utilize urea as a nitrogen source: molecular-genetic and biochemical analysis of the enzyme. *Microbiology* **145**: 447–459.

Collier JL, Palenik B. (2003). Phycoerythrin-containing picoplankton in the Southern California Bight. *Deep Sea Res Part II* **50**: 2405–2422.

Di Lorenzo E. (2003). Seasonal dynamics of the surface circulation in the Southern California Current System. *Deep Sea Res Part II* **50**: 2371–2388.

Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP *et al.* (2008). Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* **9**: R90.

Dupont CL, Barbeau K, Palenik B. (2008). Ni uptake and limitation in marine *Synechococcus* strains. *Appl Environ Microbiol* **74**: 23–31.

Durand MD, Olson RJ, Chisholm SW. (2001). Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea. *Deep Sea Res Part II* **48**: 1983–2003.

Ferris MJ, Palenik B. (1998). Niche adaptation in ocean cyanobacteria. *Nature* **396**: 226–228.

Fuller NJ, Marie D, Partensky F, Vaulot D, Post AF, Scanlan DJ. (2003). Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ Microbiol* **69**: 2430–2443.

Fuller NJ, Tarran GA, Yallop M, Orcutt KM, Scanlan DJ. (2006). Molecular analysis of picocyanobacterial community structure along an Arabian Sea transect reveals distinct spatial separation of lineages. *Limnol Oceanogr* **51**: 2515–2526.

Fuller NJ, West NJ, Marie D, Yallop M, Rivlin T, Post AF *et al.* (2005). Dynamics of community structure and phosphate status of picocyanobacterial populations in the Gulf of Aqaba, Red Sea. *Limnol Oceanogr* **50**: 363–375.

Gibson AH, Jenkins BD, Wilkerson FP, Short SM, Zehr JP. (2006). Characterization of cyanobacterial *glnA* gene diversity and gene expression in marine environments. *FEMS Microbiol Ecol* **55**: 391–402.

Glover HE. (1985). The physiology and ecology of the marine cyanobacterial genus *Synechococcus*. *Adv Aquat Microbiol* **3**: 49–107.

Guillou L, Jacquet S, Chretiennot-Dinet MJ, Vaulot D. (2001). Grazing impact of two small heterotrophic flagellates on *Prochlorococcus* and *Synechococcus*. *Aquat Microb Ecol* **26**: 201–207.

Hickey BM. (1979). The California Current System–hypotheses and facts. *Prog Oceanogr* **8**: 191–279.

Hutchins DA, Witter AE, Butler A, Luther GW. (1999). Competition among marine phytoplankton for different chelated iron species. *Nature* **400**: 858–861.

Jenkins BD, Zehr JP, Gibson A, Campbell L. (2006). Cyanobacterial assimilatory nitrate reductase gene diversity in coastal and oligotrophic marine environments. *Environ Microbiol* **8**: 2083–2095.

Jeong HJ, Park JY, Nho JH, Park MO, Ha JH, Seong KA *et al.* (2005). Feeding by red-tide dinoflagellates on the cyanobacterium *Synechococcus*. *Aquat Microb Ecol* **41**: 131–143.

Johnson PW, Sieburth JM. (1979). Chroococcoid cyanobacteria in the sea: a ubiquitous and diverse phototropic biomass. *Limnol Oceanogr* **24**: 928–935.

Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S *et al.* (2007). Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet* **3**: 2515–2528.

Leao PN, Vasconcelos MTSD, Vasconcelos VM. (2007). Role of marine cyanobacteria in trace metal bioavailability in seawater. *Microb Ecol* **53**: 104–109.

Li WKW. (1994). Primary production of prochlorophytes, cyanobacteria, and eukaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnol Oceanogr* **39**: 169–175.

Lynn RJ, Simpson JJ. (1987). The California Current System—the seasonal variability of its physical characteristics. *J Geophys Res* **92**: 12947–12966.

Ma Y, Jiao NZ, Zeng YH. (2004). Natural community structure of cyanobacteria in the South China Sea as revealed by *rpoC1* gene sequence analysis. *Lett Appl Microbiol* **39**: 353–358.

Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. (2005). The microbial pan-genome. *Curr Opin Genet Dev* **15**: 589–594.

Moore LR, Ostrowski M, Scanlan DJ, Feren K, Sweetsir T. (2005). Ecotypic variation in phosphorus acquisition mechanisms within marine picocyanobacteria. *Aquat Microb Ecol* **39**: 257–269.

Moore LR, Post AF, Rocap G, Chisholm SW. (2002). Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnol Oceanogr* **47**: 989–996.

Muhling M, Fuller NJ, Millard A, Somerfield PJ, Marie D, Wilson WH *et al.* (2005). Genetic diversity of marine *Synechococcus* and co-occurring cyanophage communities: evidence for viral control of phytoplankton. *Environ Microbiol* **7**: 499–508.

Muhling M, Fuller NJ, Somerfield PJ, Post AF, Wilson WH, Scanlan DJ *et al.* (2006). High resolution genetic diversity studies of marine *Synechococcus* isolates using rpoC1-based restriction fragment length polymorphism. *Aquat Microb Ecol* **45**: 263–275.

Palacios DM, Bograd SJ, Mendelssohn R, Schwing FB. (2004). Long-term and seasonal trends in stratification in the California Current, 1950-1993. *J Geophys Res* **109**: C10016.

Palenik B. (1994). Cyanobacterial community structure as seen from RNA polymerase gene sequence analysis. *Appl Environ Microbiol* **60**: 3212–3219.

Palenik B, Brahamsha B, Larimer FW, Land M, Hauser L, Chain P *et al.* (2003). The genome of a motile marine *Synechococcus*. *Nature* **424**: 1037–1042.

Palenik B, Haselkorn R. (1992). Multiple evolutionary origins of prochlorophytes, the chlorophyll B-containing prokaryotes. *Nature* **355**: 265–267.

Palenik B, Ren QH, Dupont CL, Myers GS, Heidelberg JF, Badger JH *et al.* (2006). Genome sequence of *Synechococcus* CC9311: insights into adaptation to a coastal environment. *Proc Natl Acad Sci USA* **103**: 13555–13559.

Palenik B, Ren Q, Tai V, Paulsen IT. (2009). Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environ Microbiol* **11**: 349–359.

Penno S, Lindell D, Post AF. (2006). Diversity of *Synechococcus* and *Prochlorococcus* populations determined from DNA sequences of the N-regulatory gene *ntcA*. *Environ Microbiol* **8**: 1200–1211.

Ramette A, Tiedje JM. (2007). Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. *Microb Ecol* **53**: 197–207.

Richardson TL, Jackson GA. (2007). Small phytoplankton and carbon export from the surface ocean. *Science* **315**: 838–840.

Rocap G, Distel DL, Waterbury JB, Chisholm SW. (2002). Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* **68**: 1180–1191.

Smith EM, Kemp WM. (2001). Size structure and the production/respiration balance in a coastal plankton community. *Limnol Oceanogr* **46**: 473–485.

Strom SL. (1991). Growth and grazing rates of the herbivorous dinoflagellate *Gymnodinium* sp from the open subarctic Pacific Ocean. *Mar Ecol Prog Ser* **78**: 103–113.

Suttle CA, Chan AM. (1993). Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*—abundance, morphology, cross-infectivity and growth characteristics. *Mar Ecol Prog Ser* **92**: 99–109.

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL *et al.* (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial 'pan-genome'. *Proc Natl Acad Sci USA* **102**: 13950–13955.

Toledo G, Palenik B. (1997). *Synechococcus* diversity in the California current as seen by RNA polymerase (rpoC1) gene sequences of isolated strains. *Appl Environ Microbiol* **63**: 4298–4303.

Toledo G, Palenik B. (2003). A *Synechococcus* serotype is found preferentially in surface marine waters. *Limnol Oceanogr* **48**: 1744–1755.

Toledo G, Palenik B, Brahamsha B. (1999). Swimming marine *Synechococcus* strains with widely different photosynthetic pigment ratios form a monophyletic group. *Appl Environ Microbiol* **65**: 5247–5251.

Waterbury JB, Valois FW. (1993). Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Appl Environ Microbiol* **59**: 3393–3399.

Waterbury JB, Watson SW, Guillard RRL, Brand LE. (1979). Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* **277**: 293–294.

Waterbury JB, Watson SW, Valois FW, Franks DG. (1986). Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus*. *Can Bull Fish Aquat Sci* **214**: 71–120.

Wilson WH, Joint IR, Carr NG, Mann NH. (1993). Isolation and molecular characterization of five marine cyanophages propagated on *Synechococcus* sp. strain WH7803. *Appl Environ Microbiol* **59**: 3736–3743.

Winant CD, Bratkovich AW. (1981). Temperature and currents on the southern California shelf—a description of the variability. *J Phys Oceanogr* **11**: 71–86.

Worden AZ, Nolan JK, Palenik B. (2004). Assessing the dynamics and ecology of marine picophytoplankton: the importance of the eukaryotic component. *Limnol Oceanogr* **49**: 168–179.

Zeidner G, Preston CM, Delong EF, Massana R, Post AF, Scanlan DJ *et al.* (2003). Molecular diversity among marine picophytoplankton as revealed by *psbA* analyses. *Environ Microbiol* **5**: 212–216.

Zubkov MV, Fuchs BM, Tarran GA, Burkill PH, Amann R. (2003). High rate of uptake of organic nitrogen compounds by *Prochlorococcus* cyanobacteria as a key to their dominance in oligotrophic oceanic waters. *Appl Environ Microbiol* **69**: 1299–1304.

Zwirglmaier K, Heywood JL, Chamberlain K, Woodward EMS, Zubkov MV, Scanlan DJ. (2007). Basin-scale distribution patterns lineages in the Atlantic Ocean. *Environ Microbiol* **9**: 1278–1290.

Zwirglmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaulot D *et al.* (2008). Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* **10**: 147–161.

Supplementary information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)

Chapter 1, in full, is a reprint of the material as it appears in Tai, Vera and Palenik, Brian. 2009. *The ISME Journal*. Volume 3, pages 903-915. The dissertation author was the primary investigator and first author of this paper.

CHAPTER 2

Temporal and spatial distributions of coastal marine *Synechococcus* assessed using Luminex bead-arrays

Abstract

The Luminex bead-array, a multi-analyte profiling method, was developed for assessing marine *Synechococcus* diversity. Four probes were designed based on sequence diversity of a gene encoding a subunit of the RNA polymerase (*rpoC1*) to enumerate *Synechococcus* clades I, II, III, and IV from environmental samples. The probes were used simultaneously in a multiplexed assay to follow *Synechococcus* diversity over a one-year from the SIO pier, a Pacific Ocean coastal environment. The Luminex assay successfully demonstrated that *Synechococcus* from clades I and IV were the dominant types at this site throughout the year. *Synechococcus* from clades II and III were not detected except during the late summer or early winter. To further examine the biogeography of taxa within clades I and IV, Luminex probes were designed to target phylogenetically defined sub-clades. By examining samples taken from a coastal to open ocean transect, it was found that in coastal waters, two sub-clade IV probes were dominant at the surface, whereas two sub-clade I probes and a third sub-clade IV probe had increased signals in deeper water near the chlorophyll maximum. In mesotrophic

waters, this third sub-clade IV probe dominated deep water at the chlorophyll maximum whereas all other sub-clade probes were below detection limits. The Luminex assay was successful in showing that sub-clades below the level of 16S rRNA resolution have different spatial distributions and thus likely different ecological strategies and adaptations.

Introduction

Marine *Synechococcus* are a major cyanobacterial group that dominates coastal environments and are found throughout the world's oceans (Johnson and Sieburth, 1979; Waterbury *et al.*, 1979). They are important contributors to primary production and are also very diverse. At least 10 different clades of *Synechococcus* have been identified (Ferris and Palenik, 1998; Rocap *et al.*, 2002; Fuller *et al.*, 2003; Ahlgren and Rocap, 2006; Penno *et al.*, 2006). However, little is known about the physiological and ecological characteristics that distinguish each clade, the abundance and distribution of these clades, and how these factors might affect rates of primary production or other biogeochemical cycles.

To understand the ecology of *Synechococcus* communities, diversity and abundance measures are fundamental pieces of information. However, due to their small size, these measures are technically challenging to obtain. Traditionally, microbes are often distinguished by physiological characteristics but since only a small fraction of microbes are easily cultured, culture-independent and molecular methods have become the primary means to investigate microbial diversity (Fuhrman *et al.*, 2002; Nocker *et al.*, 2007).

Synechococcus biogeography has been examined using clade-specific probes for Southern hybridizations (Fuller *et al.*, 2005; Fuller *et al.*, 2006; Zwirglmaier *et al.*, 2007; Zwirglmaier *et al.*, 2008) and serotyping (Campbell and Carpenter, 1987; Toledo and Palenik, 2003). A more precise measure of abundance using quantitative PCR (qPCR) has also been developed for marine *Synechococcus* clades (Tai and Palenik, 2009). In addition, sequencing and fingerprinting methods, such as denaturing gradient gel electrophoresis (DGGE) or terminal restriction fragment length polymorphism (T-RFLP), of a marker gene (often the 16S rRNA gene) are among the most common ways to describe or document changes in microbial diversity (see for eg. Giovannoni *et al.*, 1990; Fuhrman *et al.*, 2006; Muhling *et al.*, 2006; Huber *et al.*, 2007; Nocker *et al.*, 2007). In particular, DNA sequences from cultured isolates and environmental libraries of a gene encoding a subunit of the RNA polymerase (*rpoC1*) have demonstrated the diversity of *Synechococcus* within the Southern California Bight (Toledo and Palenik, 1997; Ferris and Palenik, 1998; Toledo *et al.*, 1999). While these methods have provided a wealth of information concerning microbial ecology and evolution, there is a need to develop faster, less expensive, but quantitative means to interrogate microbial diversity.

This study investigates the use of Luminex bead-array technology to assess the diversity and abundance of marine *Synechococcus*. The Luminex bead-array is based on DNA hybridization to oligonucleotide probes and is designed for rapid, high-throughput, customizable, multiplex investigations. Luminex hybridizations have been successfully developed for detecting microbial pathogens from clinical (for eg. Bandyopadhyay *et al.*, 2007) and environmental samples (Baums *et al.*, 2007). Applications of this technology have also been used to determine eukaryotic phytoplankton abundances from the marine environment (Ellison and Burton, 2005; Scorzetti *et al.*, 2009).

For this study, we developed probes based *rpoC1* for use in Luminex assays to assess changes in the diversity and abundance of marine *Synechococcus* from a one-year time series of coastal samples and a coastal to open ocean transect from the Southern California Bight in the northeastern Pacific Ocean. Several probes and samples could be analyzed together in a single assay providing the ability to rapidly examine microscale diversity patterns of *Synechooccus* biogeography.

Materials and Methods

Environmental samples

Microorganisms were collected from surface waters at the end of the Scripps Institution of Oceanography (SIO) pier, La Jolla, CA, located at the southern end of the Southern California Bight (32.8672 N, 117.2583 W). Seawater samples were filtered onto 0.2 μm Supor filters as previously described (Tai and Palenik, 2009). Approximately monthly samples from January 5, 2006 to February 26, 2007 were analyzed for this study.

Samples were also collected from 10 stations of a coastal to open ocean transect from July 17 to August 4, 2007 for DNA extractions and to determine total *Synechococcus* abundances using flow cytometry as previously described (Tai and Palenik, 2009). The transect was equivalent to Line 93.3 of the California Cooperative Oceanic Fisheries Investigations (CalCOFI) sampling grid and located just off-shore of the SIO pier in the Southern California Bight (from 32.9564 N, 117.3054 W to 29.8464 N, 123.5866 W). Samples were taken from stations 26, 35, 40, 45, 50, 65, 80, 90, 110, and 120, station coordinates can be found at www.calcofi.org) (Figure 2.1). Station 65 is not part of the CalCOFI sampling grid and was located between stations 60 and 70 (31.6776 N, 119.9025 W). The stations are designated by the line, a period, then the station; for example 93.26 for Line 93.3, station 26. At each station, samples from the surface, the *Synechococcus* abundance maxima, and chlorophyll maxima were analyzed with a few exceptions (Table 2.1). At stations 93.26, 93.40, and 93.65, the depth of the *Synechococcus* and chlorophyll maxima coincided, but for all other stations, the *Synechococcus* maxima were above the chlorophyll maxima. For stations 110 and 120, samples from the *Prochlorococcus* maxima at depths of 100 and 90 m, which were 25 and 18 m above the chlorophyll maxima, respectively, were analyzed instead of the chlorophyll maxima. For station 93.35, the chlorophyll maximum at 23 m was not analyzed. At this station, the *Synechococcus* maximum was at 20 m.

Strains and environmental clones used as Luminex standards

Synechococcus strains used as standards were grown in artifical seawater as previously described (Su *et al.*, 2006) with 9.0 mM $NaNO_3$. Vitamins (0.2 mg l^{-1} thiamin-HCl, 1 μg l^{-1} d-biotin, and 1 μg l^{-1} B_{12}) were also added. Strains representing 4 different clades of *Synechococcus* were used: *Synechococcus* CC9311 (clade I), *Synechococcus* CC9605 (clade II), *Synechococcus* WH8102 (clade III) and *Synechococcus* CC9902 (clade IV). Batch cultures were grown in glass flasks, gently stirred, at 25 °C under 30 μEinstein$\cdot m^{-2} \cdot s^{-1}$ continuous white light. Cells were harvested by filtration onto 0.2 μm Supor filter discs and stored at –20 °C.

Plasmid DNA of *Synechococcus rpoC1* clones from environmental libraries were used as standards for sub-clade probes. The clones were obtained by PCR amplification, ligation into a plasmid vector, and transformation into *Escherichia coli* as described in Tai and Palenik (2009). Phylogenetic trees calculated from DNA sequence alignments of these environmental clones were used to determine which clones to use as standards (see below).

DNA Extraction

Genomic DNA from cultured strains and environmental samples were extracted as previously described (Tai and Palenik, 2009). Briefly, lysozyme, proteinase k, and SDS were used to lyse cells collected on a 0.2 μm Supor filter. Phenol:chloroform extractions were used to isolate the DNA away from lipids and proteins, and the DNA was cleaned using the DNeasy Blood and Tissue kit (Qiagen).

To prepare plasmid DNA of *rpoC1* clones, the QIAprep kit (Qiagen) was used following the manufacturer's protocol.

Biotinylation of *Synechococcus rpoC1* gene fragments

To generate biotinylated DNA targets, *Synechococcus*-specific PCR primers that flank a variable region of the *rpoC1* gene were designed (RPOC493F = 5' biotin-CTN ACN GAR GAY GAR TGG YT 3' and SAC-1039R = 5' CYT GYT TNC CYT CDA TDA TRT C 3'). The forward primer (RPOC493F) was biotinylated on the 5'end for detection using streptavidin-phycoerythrin (strep-PE) following hybridization to the Luminex beads. The PCR consisted of 25 μl of Platinum® PCR SuperMix (Invitrogen), 200 pmoles each of RPOC493F and SAC-1039R, and template DNA. The reaction was subjected to 94 °C for 2 min, followed by 35 cycles of 94 °C for 30 s, 52 °C for 30 s, and 72 °C for 30 s and ending with a final extension at 72 °C for 5 minutes.

To test probe specificity and for use as hybridization standards, PCR reactions were conducted with genomic DNA from cultured strains and environmental *rpoC1* clones. For strain-specific probes, serial dilutions of genomic DNA from *Synechococcus* CC9311 (clade I), *Synechococcus* CC9605 (clade II), *Synechococcus* WH8102 (clade III) and *Synechococcus* CC9902 (clade IV) ranging from approximately 0.0002 to 2 ng were used in the PCR reactions. For environmental *rpoC1* clones, 50 ng of plasmid DNA was used. For environmental samples, typically 1 to 8 ng of DNA was used in the PCR reactions. The PCRs and subsequent hybridizations were replicated 5 to 6 times.

Probe design

Two sets of probes were designed, one set to target *Synechococcus* at the clade level and another to target sub-clades within *Synechococcus* clades I and IV. Probes were designed to hybridize to a variable region (nucleotides

604 to 624 from the start of the *rpoC1* gene) within the *rpoC1* fragment amplified by PCR as described above. The probe sequences were designed to be complementary to the biotinylated strand amplified by PCR and contained an amine-12 carbon linker on the 5' end for coupling to the carboxylated Luminex beads. Care was taken to include mismatches near the middle of the probe as this enhances hybridization specificity.

For the clade probes, 4 probes were designed to exactly match the *rpoC1* variable region of cultured strains representing clades I, II, III, and IV to avoid using degenerate oligonucleotides (Table 2.2). In all cases, the strain-specific probe sequence was the most common of the clade based on environmental libraries of *rpoC1* gene sequences.

Sub-clade probes were designed based on a phylogenetic tree calculated using Neighbor-joining from a 460 bp alignment of all available *Synechococcus rpoC1* sequences (paup v 4.10b). Neighbor-joining trees were re-calculated with statistical support from 1000 bootstrap replicates for all *Synechococcus* sequences belonging to clade I (Figure 2.2a) and all *Synechococcus* sequences belonging to clade IV (Figure 2.2.b). These clade I and IV sequences were examined for variability within the probing region. A sub-clade probe was designed if 3 or more environmental clones were found for this specific probe sequence. This resulted in 4 clade I probes in addition to the one specific to CC9311 and 2 clade IV probes in addition to the one specific to CC9902 (Table 2.2). For probe IV-C, the probe was shifted 6 nucleotides towards the 5'end to enhance hybridization specificity. The *rpoC1* sequences targeted by these sub-clade probes generally clustered phylogenetically (Figure 2.2).

Probe coupling to beads

Probes were coupled to carboxylated Microplex microspheres (Luminex and MiraiBio), refered to here as beads, following the procedure of Yang *et al*. (2001) except using 0.2 nmoles of oligonucleotide probe. Each probe was coupled to a different stock of beads that are identified by a unique fluorescent stain. The coupled beads (bead-probes) were stored in 10 mM Tris and 1 mM EDTA, pH 8.0 at 4 °C in the dark.

Hybridization

The hybridization protocols were developed using a 96-well plate to allow for high-throughput sample processing.

To test the hybridization specificity of the bead-probes, a range of hybridization temperatures from 50 to 65 °C was tested. For the clade probes (probes I-A, II, III, and IV-A), 55 °C gave the best probe signal with the lowest non-specific hybridization. For the sub-clade probes (probes I-A, I-B, I-C, I-D, I-E, IV-A, IV-B, and IV-C), a hybridization temperature of 58 °C was chosen.

Initial experiments using DNA from cultured strains for testing the clade probe set were conducted using the following hybridization procedure: 10 μl of the biotinylated PCR product was mixed with 5000 of each bead-probe in 40 μl 1.5 X TMAC hybridization buffer (4.5 M tetramethyl ammonium chloride, 0.1 % SDS, 4 mM EDTA, 50 mM Tris, pH 8.0) and 10 μl water. The biotinylated PCR product was hybridized to the bead-probes by first denaturing the DNA at 95 °C for 5 minutes followed by hybridization at 55 °C for 30 minutes. The bead-probes were pelleted by centrifugation for 3 min at 3050 × g, the supernatent removed, then the bead-probes were incubated with 4 μg/ml of streptavidin-phycoerythrin (strep-PE) in 1X TMAC hybridization buffer (a 2/3 dilution of 1.5 X TMAC buffer) for 10 minutes at 55 °C. The bead-probes were then analyzed on a Luminex-100 instrument following the manufacturer's protocol.

For all other experiments including environmental samples and clones, the protocol was modified to enhance the hybridization signal. The biotinylated PCR products were cleaned and concentrated by ethanol precipitation. Twenty μl of the PCR product was mixed with 2 μl of 3 M sodium acetate, pH 5.2 and 40 μl of 100% ethanol. The PCR product was allowed to precipitate at –20 °C for 60 minutes then pelleted by centrifugation at 3050 × g for 60 min at 4 °C. The supernatent was removed and the DNA pellet was washed with 150 μl of 70 % ethanol and centrifuged at 3050 × g for

10 min. The supernatent was removed and the pellet was left to air-dry. The DNA pellet was resuspended in 15 μl or 30 μl of 1X TMAC buffer (except with 0.1 % Sarkosyl instead of SDS) containing 2000 of each clade bead-probe or 3000 of each sub-clade bead-probe to be tested. Denaturation, hybridization, detection, and analysis were the same as described above except with some modifications for the sub-clade probes. Hybridization was performed at 58 °C for the sub-clade probes and after incubating in strep-PE, the strep-PE was replaced with 1X TMAC buffer and incubated for 5 minutes at 58 °C before analysis.

Analysis

Successful hybridization of the biotinylated PCR amplicons to a particular bead-probe was assessed by the amount of phycoerythrin flurorecence recorded on an arbitrary scale from each bead-probe. One hundred of each bead-probe was evaluated for a given sample. The trimmed mean fluorescence relative to background from PCR negative controls (signal:background) was used to evaluate the hybridizations. When using DNA from cultured isolates, the average amount of fluorescence from non-specific targets was used as the background fluorescence. A reliable signal (detection limit) was considered to be signal:background greater than 2.

Statistics

The results from the clade probe set using samples from the SIO pier time series were compared to results obtained from a previous study using quantitative PCR (qPCR) (Tai and Palenik, 2009). The Pearson product-moment correlation coefficients (r) were calculated using JMP v. 5.0. Pair-wise correlations were also determined between the sub-clade probe results and total *Synechococcus* abundances from the SIO pier time series and the coastal to open-ocean transect.

Results

Development of the Luminex assay

The PCR primers used here successfully amplified a 546 bp fragment of the *rpoC1* gene. These primers were designed to amplify the *rpoC1* gene from marine cyanobacteria (i.e *Synechococcus* and *Prochlorococcus*). The forward primer (RPOC493F) anneals to the *rpoC1* gene in between primers *rpoC1*-39F and *rpoC1*-462R that have previously been used to document *Synechococcus* diversity (Muhling *et al.*, 2006). The reverse primer (SAC-1039R) has been used to generate *rpoC1* clone libraries from environmental *Synechococcus* (Tai and Palenik, 2009). The RPOC493 – SAC-1039R primer pair produced an appropriately sized fragment encompassing a variable region of the *rpoC1* gene suitable as a target for probe hybridization. The length of the PCR amplicon and the location of the target sequence within the PCR amplicon have been shown to affect the success of the hybridization (Diaz and Fell, 2004).

Two sets of probes were used in Luminex assays for assessing the diversity and abundance of marine *Synechococcus* (Table 2.2). The first set consisted of 4 strain-specific probes (I-A, II, III, and IV-A) designed to hybridize to *Synechococcus* from clades I, II, III, and IV (Table 2.2). The second set consisted of 5 probes (I-A, I-B, I-C, I-D, and I-E) and 3 probes (IV-A, IV-B, and IV-C) targeting sub-clades within clades I and IV, respectively (Table 2.2). Probes I-A and IV-A were common to both sets. These probes all target the same region of the *rpoC1* gene. Initially, we tested probes that hybridized to different regions of the PCR fragment or were degenerate, but hybridization signals were not obtained. Similarly, a set of 16S rRNA probes designed to evaluate *Synechococcus* diversity (Fuller *et al.*, 2003) did not produce consistent results with the Luminex assay. The 16S rRNA probes hybridized to different regions of the 16S rRNA gene and those that did not produce a Luminex signal targeted the same region of the gene (positions 927-944). This suggests that the location of the probe sequence in Luminex assays is crucial, likely due to secondary structure of the target PCR amplicon.

For the clade probes, each bead-probe hybridized only to their intended target and not to PCR amplicons from other strains when using a hybridization temperature of 55 °C (Figure 2.3a). For the sub-clade probe set, a higher hybridization temperature of 58 °C was required to reduce non-specific hybridization (Figure 2.3b). Using this hybridization temperature and by placing polymorphisms in the middle of the probe sequence, single nucleotide specificity was achieved. All the probes could not be combined into a single set because the higher hybridization signal required for specificity of the sub-clade probes resulted in a reduction of signal for the clade II and III strain-specific probes.

Using the protocol developed here, the hybridization signal increased log-linearly over 4 orders of magnitude ranging from 0.2 to 2000 ng of starting template DNA (data not shown). The overall signal increased if the PCR amplicons were ethanol precipitated before hybridization (data not shown). Ethanol precipitation removed excess primers that may have been interfering with the hybridization. It also concentrated the biotinylated PCR product resulting in a higher hybridization signal per bead.

Each bead-probe did not result in the same magnitude of hybridization signal given the same amount of starting template. Probe IV-A gave the best signal, while the I-E probe gave the weakest signal (Figure 2.4). As PCR efficiency was similar for control reactions, the differences in hybridization signal are likely due to the ability of a given probe to hybridize efficiently to its target, which could be affected by secondary structure of the target or probe, or the density of the probes coupled onto the beads (Diaz and Fell, 2004). Although the hybridization signal increased with increasing amounts of template in the PCR reactions, the signal did not span over a very large range. The average recorded PE signal (on an arbitrary scale) for probe IV-A ranged from 50, for at least a 2-fold signal above background, to 1250 with a signal:background ratio of 50. Large errors were also associated with the hybridization signals preventing precise abundances to be calculated. However, relative abundances and the observed trends in the Luminex signals are accurate.

ENVIRONMENTAL SAMPLES

The PCR primers successfully amplified and therefore biotinylated cyanobacterial *rpoC1* fragments from DNA extracted from samples collected over one-year from the SIO pier in 2006 and from a coastal to open ocean transect in the summer of 2007 (data not shown). The clade probe set was used to assay the composition of the *Synechococcus* population and their dynamics from the SIO pier time-series. These samples and the samples from the coastal to open ocean transect were evaluated with the sub-clade probe set.

SIO pier time series

From approximately monthly samples over one year, hybridization signals were consistently high at the SIO pier for probes I-A and IV-A indicating a dominance of *Synechococcus* clades I and IV throughout the year (Figure 2.5a). Signals to probes II and III were absent or very low. These hybridization results are consistent with previously reported data using qPCR (Tai and Palenik, 2009). The Luminex hybridization signals were significantly correlated with the qPCR results for clades II, III, and IV (Table 2.3). The probe IV-A Luminex hybridization signal also correlated significantly with total *Synechococcus* abundances determined by flow cytometry (Table 2.3).

It was expected that the dynamics of the I-A probe would also correlate with clade I abundances determined using qPCR. They did not, however (Table 2.3). This result is surprising as the relatively high Luminex signals from the I-A probe suggested that *Synechococcus* from this sub-clade were dominant in this environment. The lack of correlation may reside in the diversity of *Synechococcus* from clade I. Although the I-A probe hybridizes to most of the clade I *Synechococcus* based on *rpoC1* environmental libraries, during certain times of the year, other variants may form a significant component of the total clade I abundance.

The sub-clade probe set was used to test this hypothesis and examine the dynamics of *Synechococcus* microdiversity from the SIO pier time series

(Figure 2.5b, c, and d). Even though a higher hybridization temperature was used for this probe set, the results for probes I-A and IV-A were very similar (correlation coefficients of 0.780 and 0.931, respectively). Several of the sub-clade probes showed similar dynamics. Luminex signals from probes I-A and I-B; I-C and I-D; I-C and IV-B; IV-A and IV-C; and IV-A and total *Synechococcus* were correlated very significantly (Table 2.4). In addition, the Luminex results for I-C and IV-B also significantly correlated with clade I abundances determined using qPCR (Tai and Palenik, 2009). Therefore, sub-clade I-C may have been the dominant clade I type at the SIO pier. Unfortunately, the signal for this probe was generally low, so further confirmation, perhaps with clone libraries, is needed. It was not expected, however, that a clade IV probe would also correlate with clade I abundances. These results suggest that *Synechococcus* targeted by probes I-C and IV-B may have similar ecological niches.

Coastal to open ocean transect

The sub-clade probe set was also used to assess the spatial distribution of *Synechococcus* from the coast 10 km away from the SIO pier to oligotrophic waters 687 km off-shore. From the coast to open ocean, the Luminex signals for all sub-clade probes consistently decreased from both surface samples and at depth although a distinct peak was observed at station

93.45 (Figure 2.6). A sharp decrease in total *Synechococcus* abundances occurred at station 93.80 coinciding with a drop in the nitricline, thermocline, and chlorophyll maximum depths (C. Dupont, R. Bundy, K. Barbeau, unpublished data). Sub-clade abundances at depth also became very low at station 93.80 while at the surface, low sub-clade abundances began at station 93.65. The pattern of decreasing total *Synechococcus* abundances from the coast to off-shore is consistent with previous descriptions of the Southern California Bight (Collier and Palenik, 2003).

For each station, the change in sub-clade composition from the surface to the *Synechococcus* maximum and chlorophyll maximum was examined (Figure 2.7). The results are grouped based on the Luminex hybridization results of the sub-clade probes that showed similar trends. Luminex signals to sub-clade probes I-A, I-B, and IV-B showed similar trends with depth with the exception of stations 93.80 and 93.90. This is in contrast to the SIO pier time series in which the results for sub-clade probe IV-B correlated with sub-clade probe I-C.

For stations 93.26, 93.35, and 93.40 hybridization to sub-clade probes I-A, I-B, and IV-B increased from the surface to the *Synechococcus* and chlorophyll maxima in parallel to the trend observed with total *Synechococcus* abundances. In contrast, hybridization to sub-clades IV-A and IV-C decreased with depth.

At station 93.45, the Luminex signals to subclades IV-A and IV-C again decreased with depth. For sub-clades I-A, I-B, and IV-B, the signals decreased from the surface to the *Synechococcus* maximum, then slightly increased at the chlorophyll maximum. At this station, a reliable signal (> 2 signal:background) could be detected at the surface for sub-clades I-C and I-D and the signal decreased with depth. A reliable signal for sub-clade I-E could not be detected.

At stations 93.50 and 93.65, all the sub-clade probes and total *Synechococcus* showed similar trends. For station 93.50, the Luminex signals increased slightly from the surface to the *Synechococcus* maximum, then decreased slightly or remained the same at the chlorophyll maximum. For station 93.65, the Luminex signals increased from the surface to the *Synechococcus*/chlorophyll maximum.

Stations 93.80 and 93.90 showed a very different pattern. The signal for probe IV-B consistently increased from the surface to the chlorophyll maximum whereas reliable signals were not detected for any of the other sub-clade probes. The magnitude of the sub-clade IV-B signal, however, was lower than from the more coastal stations. Total *Synechococcus* abundances did not exactly match the sub-clade IV-B trend even though this was the only *Synechococcus* sub-clade detected at these stations. For station 93.80, *Synechococcus* abundances increased with depth but were low at the chlorophyll maximum.

Over the transect, the distribution of *Synechococcus* targeted by probe IV-B was clearly distinct in mesotrophic waters (stations 93.80 and 93.90) (Figure 2.8). *Synechococcus* that were dominant in coastal waters (targeted by probes I-A and IV-A) were below detection limits at these stations.

Hybridization to the sub-clade probes was not detected from the two oligotrophic stations, 93.110 and 93.120 except for probe IV-A (Figure 2.7). At station 93.110, the signal for probe IV-A increased slightly from the surface to the *Synechococcus* maximum, then decreased at 100 m following the trend in total *Synechococcus*. The abundance of *Synechococcus* was very low at these stations, so it is not surprising that little was detected with the sub-clade probes.

Discussion

The Luminex bead-array was successfully developed for high-throughput multiplex detection of marine *Synechococcus* from environmental samples. Using mixtures of 4 and 8 probes, the temporal and spatial distributions of clades and sub-clades of *Synechococcus* were determined. The assay can be easily augmented to include more probes, either from other taxa within the marine *Synechococcus* or different taxa altogether as the

Luminex bead-array can accommodate up to 100 different beads. However, the success of the Luminex assay depends on a combination of many factors, including the size of the PCR product or target DNA, secondary structure of the target DNA, relative location of the probe sequence on the target DNA, and the subsequent hybridization and detection conditions. These factors become even more critical when dealing with environmental DNA samples that may contain PCR inhibitors or low amounts of target DNA. Thus careful evaluation of the assay is vital.

Evaluation of the Luminex assay for *Synechococcu*s

The 4 *Synechococcu*s clades evaluated for the Luminex assay were chosen because they were known to be the most abundant types in this Pacific coastal environment. Over a one-year time series from the SIO pier, the Luminex assay corroborated previous results using qPCR to quantify *Synechococcus* clades, especially for clades II, III and IV (Tai and Palenik, 2009). Changes in relative abundances were readily observed and strong correlations of the mean signal:background with qPCR data shows that both methods concur and the Luminex assay accurately represented the relative temporal trends. The advantage of using Luminex hybridizations over qPCR is that only a single assay was required to obtain data for all 4 clades.

Quantification, however, was less precise using Luminex hybridizations. For Luminex hybridizations, the signal was log-linear over approximately 4 orders of magnitude of target template DNA, whereas it spanned 6 orders of magnitude in qPCR. Especially for targets of low abundance, the Luminex assay was not as sensitive as qPCR. For example, *Synechococcus* from clade III were of very low abundance at the SIO pier. Using the Luminex bead-probes, a reliable signal (signal:background > 2) was never obtained for the clade III probe whereas low clade III abundances were detected in the winter and late summer using qPCR (Tai and Palenik, 2009).

The Luminex assay has the advantage over qPCR for evaluating *Synechococcus* microdiversity. The *rpoC1* gene provided sufficient variability to examine diversity within clades also defined by 16S rRNA and internal transcribed spacer (ITS) sequences (Rocap *et al.*, 2002; Fuller *et al.*, 2003). Using this variation, many probes can be evaluated simultaneously with the Luminex assay. Eight probes were used here to examine diversity within clades I and IV, whereas qPCR can have a maximum of 4 targets in a multiplex reaction.

Biogeography of *Synechococcus* microdiversity

Synechococcus from clades I and IV are often found together in coastal environments including the coastal Southern California Bight (Brown and

Fuhrman, 2005; Brown *et al.*, 2005; Zwirglmaier *et al.*, 2007; Zwirglmaier *et al.*, 2008; Tai and Palenik, 2009). The factors that regulate their co-existence and relative dynamics, however, are not well understood. The Luminex bead-array was used to examine the distribution of sub-clades within clades I and IV using representative sub-clade probes.

The sub-clade probes generally matched all of the *rpoC1* sequences within a phylogenetic cluster, but some exceptions occurred particularly for the clade I probes (Figure 2.2). In addition, the probes did not capture all of the microdiversity. The clade I-A and IV-A probes, for example targeted more than one phylogenetic cluster. The probes do represent diversity beyond the clade level, however. Many of the sub-clade probes have varying temporal and spatial dynamics suggesting that *Synechococcus* sub-clades occupy different ecological niches. Others have highly correlated dynamics - specifically, sub-clade probes I-A and I-B; and IV-A and IV-C. Sub-clade probes I-A and I-B generally targeted *Synechococcus* belonging to the same larger phylogenetic cluster, as did sub-clade probes IV-A and IV-C (Figure 2.2). That these related pairs of sub-clade probes shared similar biogeographic patterns suggest that there is a strong genetic component regulating their distributions.

The composition of the *Synechococcus* communities from the coast to open ocean transect generally changed with the nitricline, thermocline, and chlorophyll maximum depths. As the stations progressed off-shore, the nitricline, thermocline, and chlorophyll maxima deepened and there were distinct drops at stations 93.80 (= mesotrophic) and at 93.110 (= oligotrophic) (C. Dupont, R. Bundy, K. Barbeau, unpublished data). These drops were matched by noticeable changes in the *Synechococcus* community composition. The coastal stations were dominated by Luminex signals to sub-clade probes IV-A and IV-C at the surface and sub-clades probes I-A, I-B, and IV-B at depth. In mesotrophic waters (stations 93.80 and 93.90), only the signals to sub-clade probe IV-B were detected, especially at depth, suggesting that *Synechococcus* targeted by this probe were dominant. At oligotrophic stations (93.110 and 93.120), low signals to sub-clade probe IV-A were detected. In oligotrophic environments, however, *Synechococcus* clades other than I and IV may be dominant, such as clade II (Toledo and Palenik, 2003).

The Luminex signals for probes I-C, I-D, and I-E were generally very low across this transect, but the general trends were similar to probes I-A and I-B. Therefore, the depth preference observed for probes I-A and I-B at the coastal stations might apply for all *Synechococcus* from clade I. Subtle changes in the distribution of *Synechococcus* targeted by probes I-C, I-D, and I-E, however, would not have been detected using the Luminex assay.

Using flow cytometry, marine *Synechococcus* also show distinct fluorescence and scattering changes from coastal to open-ocean environments and with depth in the Southern California Bight (Collier and Palenik, 2003). It is thought that these changes might be tied to genetic differences in the *Synechococcus* community and not solely due to physiological conditions, such as light or nutrient levels. Some of the patterns could also be grouped into coastal, mesotrophic, and off-shore/open-ocean types but the season of sampling and the sampling stations were not the same. Coordinated sampling, following both the fluorescence and scattering characteristics and the genetic diversity using Luminex would confirm whether these patterns are correlated.

The Luminex signals for sub-clade probe IV-B had a very distinct spatial distribution being dominant at depth in mesotrophic waters. In these environments, the nitricline and thermocline depths deepened to below 50 m relative to the near coast (~ 35 m). At depth, *Synechococcus* targeted by probe IV-B may be better at photosynthesizing under low light conditions while gaining access to higher nutrient concentrations. The sequences that match the sub-clade IV-B probe group together in a taxonomic cluster separate from CC9902, BL107, and other clade IV *Synechococcus* (Figure 2.2). The majority of these sequences were obtained from a near-shore station at the chlorophyll maximum. These data show for the first time a distinct spatial distribution for marine *Synechococcus* of taxa below the clade level typically used to describe their diversity.

Two distinct clade IV clusters were also observed using environmental sequences of the nitrate reductase gene (*narB*) from Monterey Bay, California (Jenkins *et al.*, 2006). It is likely that these clade IV clusters are from the same *Synechococcus* types observed with the *rpoC1* sequence phylogeny. However, since the *rpoC1* and *narB* clade IV phylogenies are comprised mainly of environmental sequences, the linkage between the *rpoC1* and *narB* clusters remains to be proved.

While some *Synechococcus* targeted by sub-clade probes shared a similar distribution, others were distinct. The factors regulating the distribution of *Synechococcus* sub-clades is not well understood, but even within a clade, *Synechococcus* strains are not genetically uniform. Genome sequences from two clade IV strains, CC9902 and BL107, have been sequenced and compared (Dufresne *et al.*, 2008). While the majority of the genes are shared and largely syntenic between the two genomes, each genome also contains genes that the other does not. These unique genes often have functions in transport or cell-wall modification implying their use in nutrient acquisition or grazer and phage avoidance. It has yet to be demonstrated, but because *Synechococcus* targeted by probe IV-B have a distinct biogeographical distribution, they must also have a distinct set of genes to take advantage of mesotrophic environments at depth. In contrast, *Synechococcus* targeted by probes IV-A and IV-C likely share genetic strategies to occupy similar ecological niches.

Conclusions

The Luminex assay provided a rapid and accurate tool for tracking multiple groups of microbes simultaneously. At a taxonomic level below that which can be resolved using 16S rRNA sequences, distinct distributions for marine *Synechoccoccus* taxa were found. But as these distributions were based on polymorphisms within a single gene marker, the genes and the genetic polymorphisms that play a role in regulating their distribution are not yet known.

At least 10 different genome sequences are available for marine *Synechococcus* strains (Palenik *et al.*, 2003; Palenik *et al.*, 2006; Dufresne *et al.*, 2008). These sequencing efforts have focused on diversity across the breadth of marine *Synechococcus*. However, as we have shown here, biogeographic distributions of *Synechococcus* also vary within clades.

To understand the genetic and environmental factors driving the ecology and population dynamics of the *Synechococcus*, sampling efforts should be centered around relevant temporal and spatial scales. Using the Luminex assay, a picture of *Synechococcus* biogeography emerged indicating that *Synechococcus* with unique ecological niches are distributed within the scale of the Southern California Bight. With continued and coordinated sampling of the Southern California Bight including genomic, metagenomic (Palenik *et al.*, 2009), physical and chemical approaches, this environment may prove to be a model ecosystem for understanding marine microbial adaptation and responses to environmental change.

Acknowledgements

Many thanks to Chris Dupont for collecting samples during the Deep Chlorophyll Maximum 2007 (DCM07) cruise and, in addition to Randie Bundy and Kathy Barbeau, providing access to unpublished data. This project was funded by a National Science Foundation grant (NSF OCE05-26979) to Brian Palenik and Ron Burton that also supported the graduate research of the dissertation author. The research was conducted in association with the California Current Ecosystem – Long Term Ecological Research (CCE-LTER) program.

Table 2.1. Sample stations and depths analyzed using *Synechococcus* sub-clade probes from a coastal to open ocean transect. [a] These samples were not analyzed. [b] The *Prochlorococcus* at 100 m was analyzed at this station. [c] The *Prochlorococcus* maximum at 90 m was analyzed at this station.

Station	distance from shore (km)	Surface sample (m)	*Synechococcus* maximum (m)	Chlorophyll maximum (m)
93.26	10.8	3	18	18
93.35	61.1	5	20	23[a]
93.40	97.6	5	30	30
93.45	134.4	5	20	40
93.50	171.2	5	15	30
93.65	281.9	5	30	30
93.80	392.6	5	40	70
93.90	466.4	10	43	53
93.110	614.0	5	50	125[a,b]
93.120	687.8	10	50	108[a,c]

Table 2.2. Oligonucleotide probes used in the Luminex assay. Probes I-A and IV-A are common to both sets, but used at different hybridization temperatures. The clade probe set was used at a hybridization temperature of 55°C and the sub-clade probe set was used at 58°C. Nucleotides that differ between the strain-specific probes are in bolded in upper case. Nucleotides that differentiate the sub-clade probes are bolded in lower case.

Probe	Strain	Sequence (5' to 3')
Clade probe set		
I-A	CC9311	**AG**C AT**T** GAG **TTG** GAG ATC CTC
II	CC9605	TTC ATC GAG G**G**T **A**AG ATC TTC
III	WH8102	TTC **C**TC **C**AG **A**TT GAG **G**TC TTC
IV-A	CC9902	TTC **T**TC **A**AG G**C**T GAG ATC CTC
Sub-clade probe set		
I-A	CC9311	**AG**C AT**T** GAG **TTG** GAG ATC CTC
I-B		**AG**C AT**T** GAG **TTG** **a**AG ATC CTC
I-C		**AG**C **g**T**T** GAG **TTG** GAG **g**TC **t**TC
I-D		**AG**C **g**T**T** GAG **TTG** **a**AG **g**TC **t**TC
I-E		**AG**C AT**T** GAG **TTG** **a**AG ATC **t**TC
IV-A	CC9902	TTC **T**TC **A**AG G**C**T GAG ATC CTC
IV-B		**c**TC **g**T**t** **A**AG G**C**T GAG ATC CT**g**
IV-C		CGC GAC TTC **c**TC **A**AG G**C**T GAG

Table 2.3. Pairwise correlation coefficients of Luminex hybridization signals between clade probes, total *Synechococcus* abundance, and clade I, II, III and IV abundances determined using qPCR from a one year time series of surface samples from the SIO pier. Correlation coefficients in bold have a p-value less than 0.05. * indicates correlation coefficients with a p-value less than 0.005. qPCR abundances are from Tai and Palenik (2009).

	probe I-A	probe II	probe III	probe IV-A	total Syn	qPCR clade I	qPCR clade II	qPCR clade III
probe II	-0.050							
probe III	0.036	**0.545**						
probe IV-A	0.030	0.264	-0.045					
total Syn	-0.202	-0.095	-0.314	**0.809***				
qPCR clade I	0.005	-0.315	-0.438	**0.562**	**0.874***			
qPCR clade II	-0.210	**0.699***	**0.526**	0.264	0.083	-0.143		
qPCR clade III	-0.093	0.359	**0.669***	0.281	0.054	-0.231	0.358	
qPCR clade IV	-0.120	-0.112	-0.262	**0.776***	**0.935***	**0.848***	0.241	-0.003

Table 2.4. Pairwise correlation coefficients of Luminex hybridization signals between sub-clade probes, total *Synechococcus* abundance, and clade I and IV abundances determined using qPCR from a one year time series of surface samples from the SIO pier. Correlation coefficients in bold have a p-value less than 0.05. * indicates correlation coefficients with a p-value less than 0.005. qPCR abundances are from Tai and Palenik (2009).

	probe I-A	probe I-B	probe I-C	probe I-D	probe I-E	probe IV-A	probe IV-B	probe IV-C	total Syn	qPCR clade I
probe I-B	**0.876***									
probe I-C	0.246	**0.519**								
probe I-D	-0.128	0.159	**0.692***							
probe I-E	0.199	0.357	**0.512**	**0.636**						
probe IV-A	0.215	0.160	0.243	0.087	0.231					
probe IV-B	0.171	0.373	**0.893***	**0.530**	0.223	0.091				
probe IV-C	0.339	0.214	-0.028	-0.055	-0.017	**0.825***	-0.127			
total Syn	0.005	-0.038	**0.517**	0.194	0.176	**0.710***	**0.521**	0.340		
qPCR clade I	0.153	0.176	**0.715***	0.291	0.240	0.425	**0.744***	0.005	**0.874***	
qPCR clade IV	0.042	-0.084	0.336	0.045	0.113	**0.661**	0.363	0.299	**0.935***	**0.848***

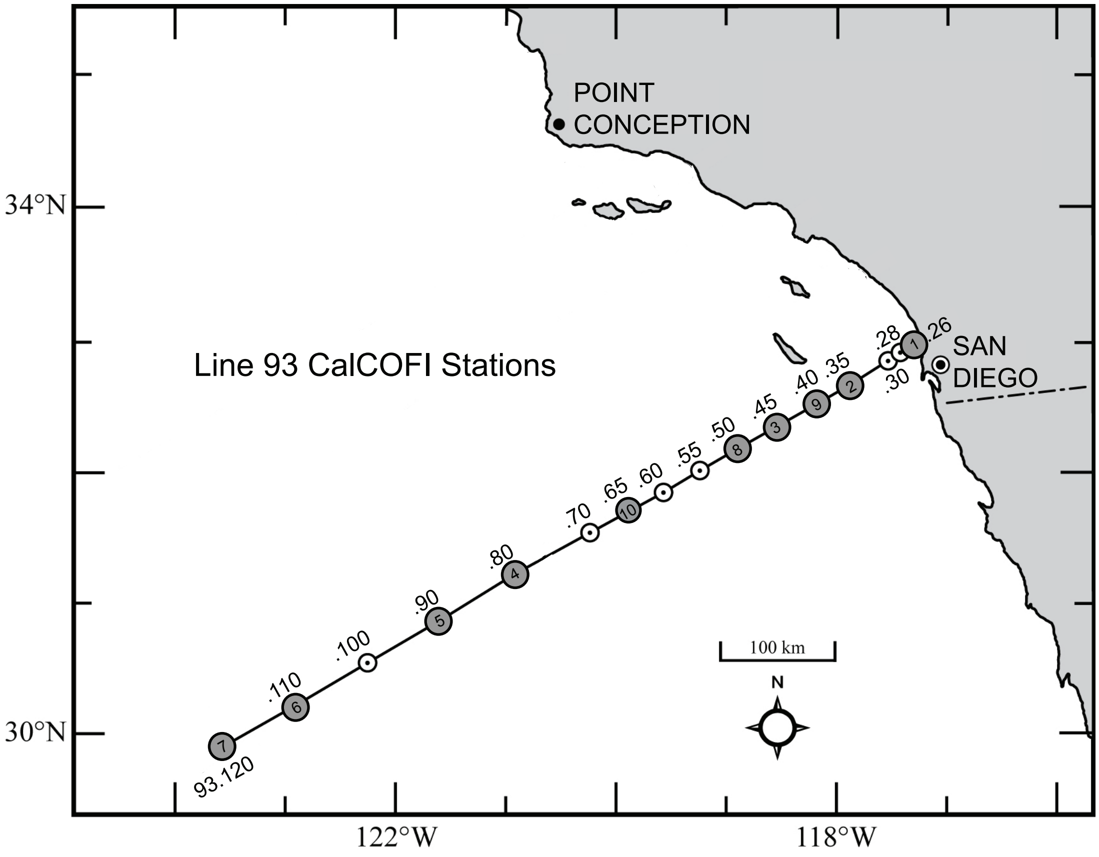


Figure 2.1 Location of stations sampled from a coastal to open-ocean transect. The transect was along Line 93 of the CalCOFI sampling grid. Stations sampled are colored in grey and numbered in the order that they were sampled. Station 93.65 is not part of the CalCOFI sampling grid, but was sampled for this transect.

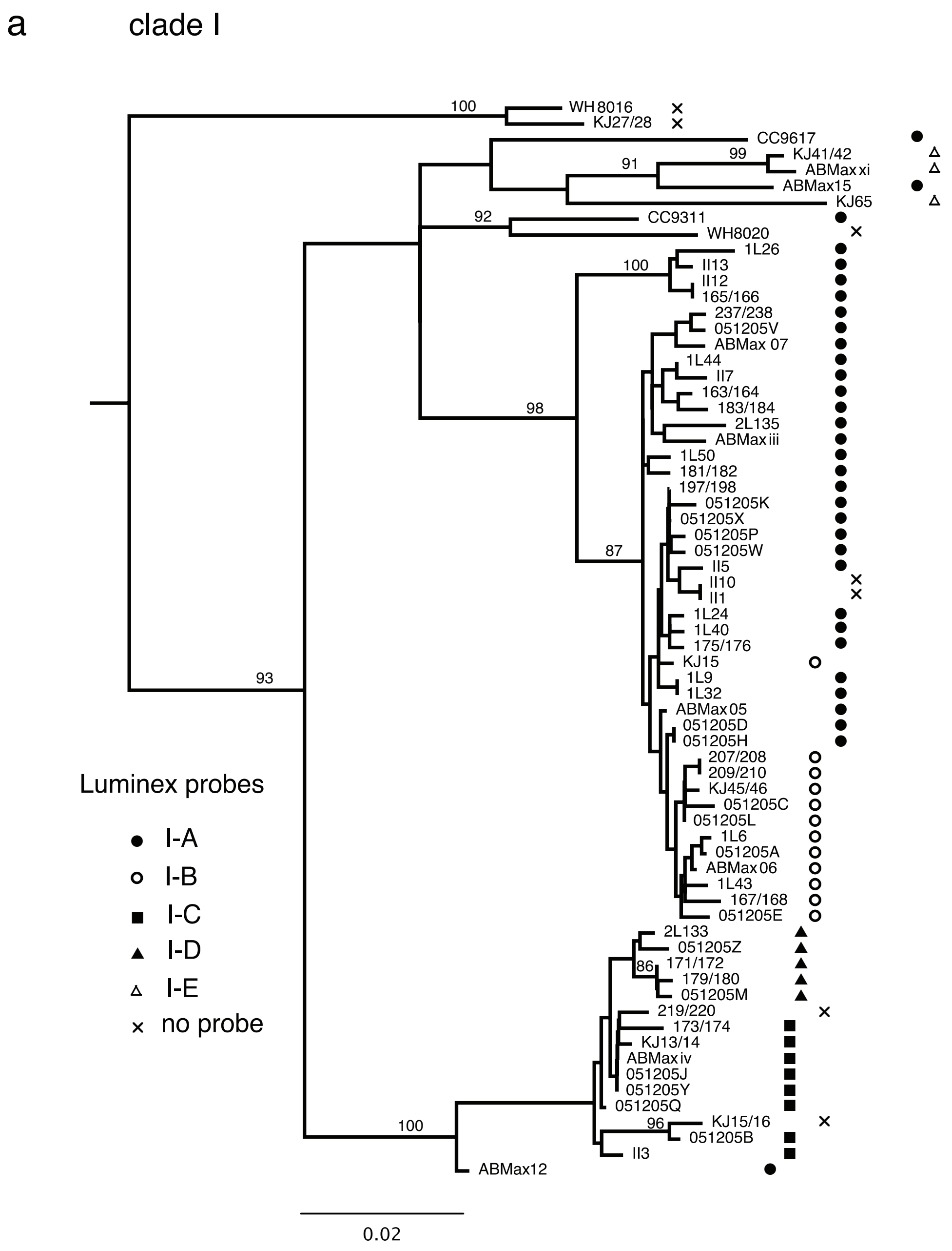


Figure 2.2 Neighbor-joining tree of *rpoC1* sequences belonging to a) *Synechococcus* from clade I and b) *Synechococcus* from clade IV. The Luminex sub-clade probe that targets each sequence is denoted by a symbol next to the sequence name. Bootstrap values are shown on branches with 85 % or greater support. The clade I and clade IV trees were rooted with the CC9902 and CC9311 *rpoC1* sequences, respectively.

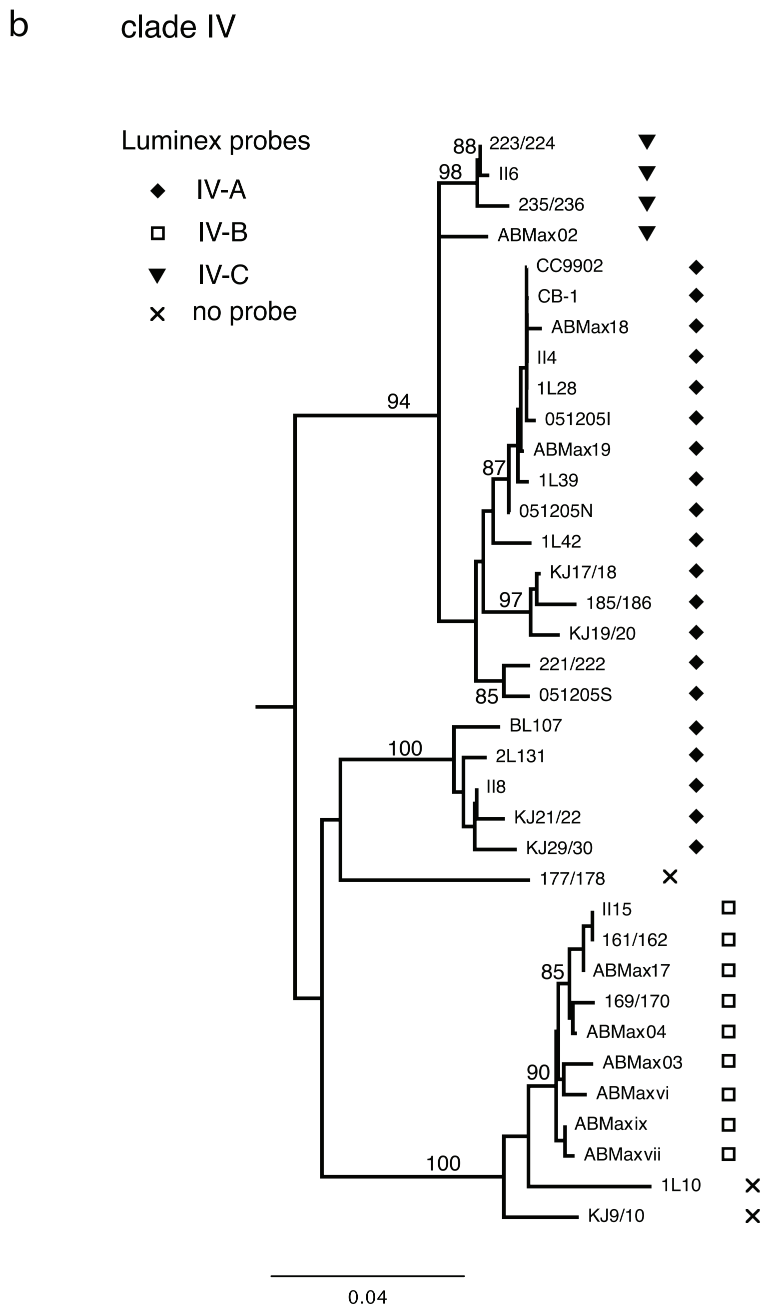


Figure 2.2. Neighbor-joining tree of *rpoC1* sequences, continued.

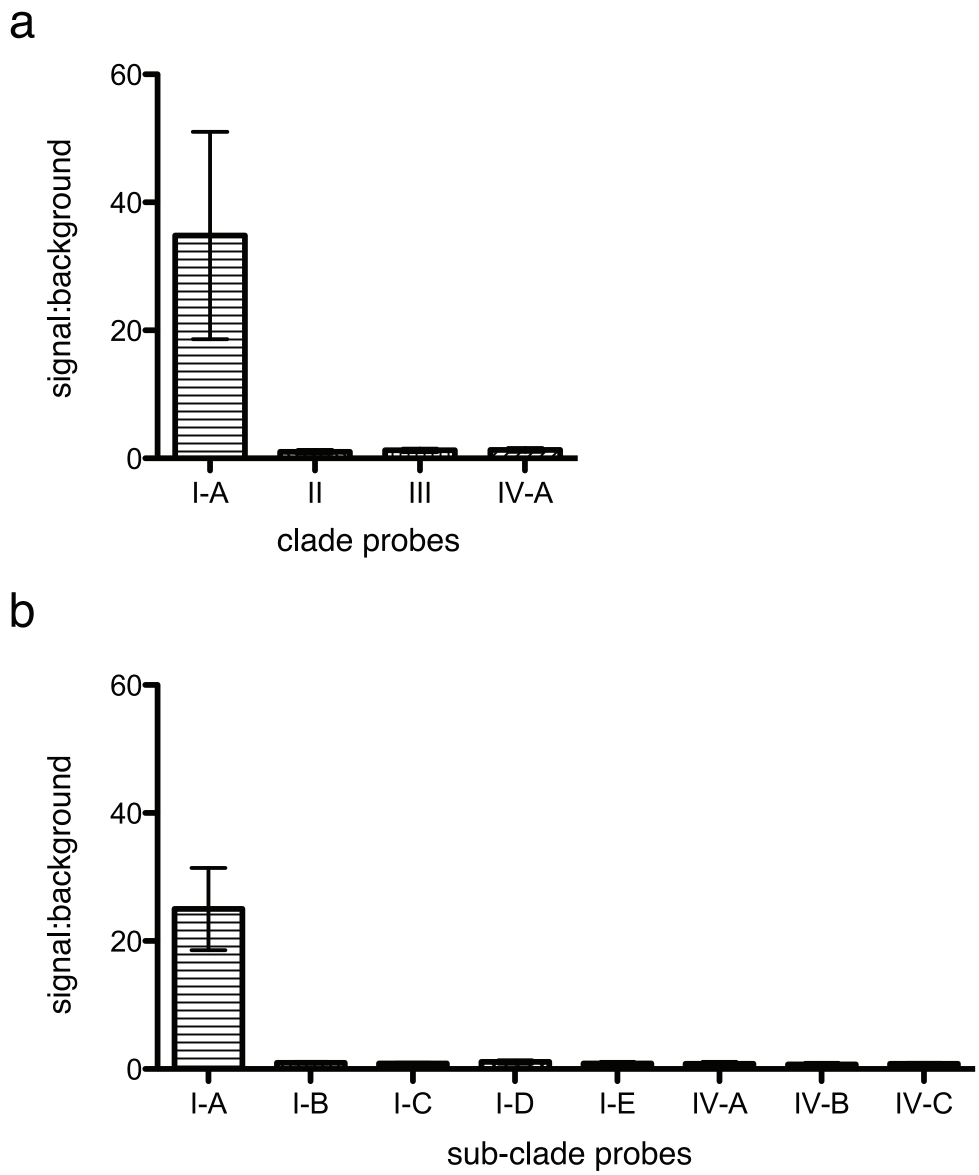


Figure 2.3. Luminex signal:background from a) the clade probe set hybridized to CC9311, a cultured representative of clade I and b) the sub-clade probe set hybridized to an environmental *rpoC1* clade I clone. In both cases, the Luminex signal is specific for probe I-A. All other probes have near background levels of signal (signal:background = 1). The mean signal:background from 6 replicates ± the standard deviation is shown.

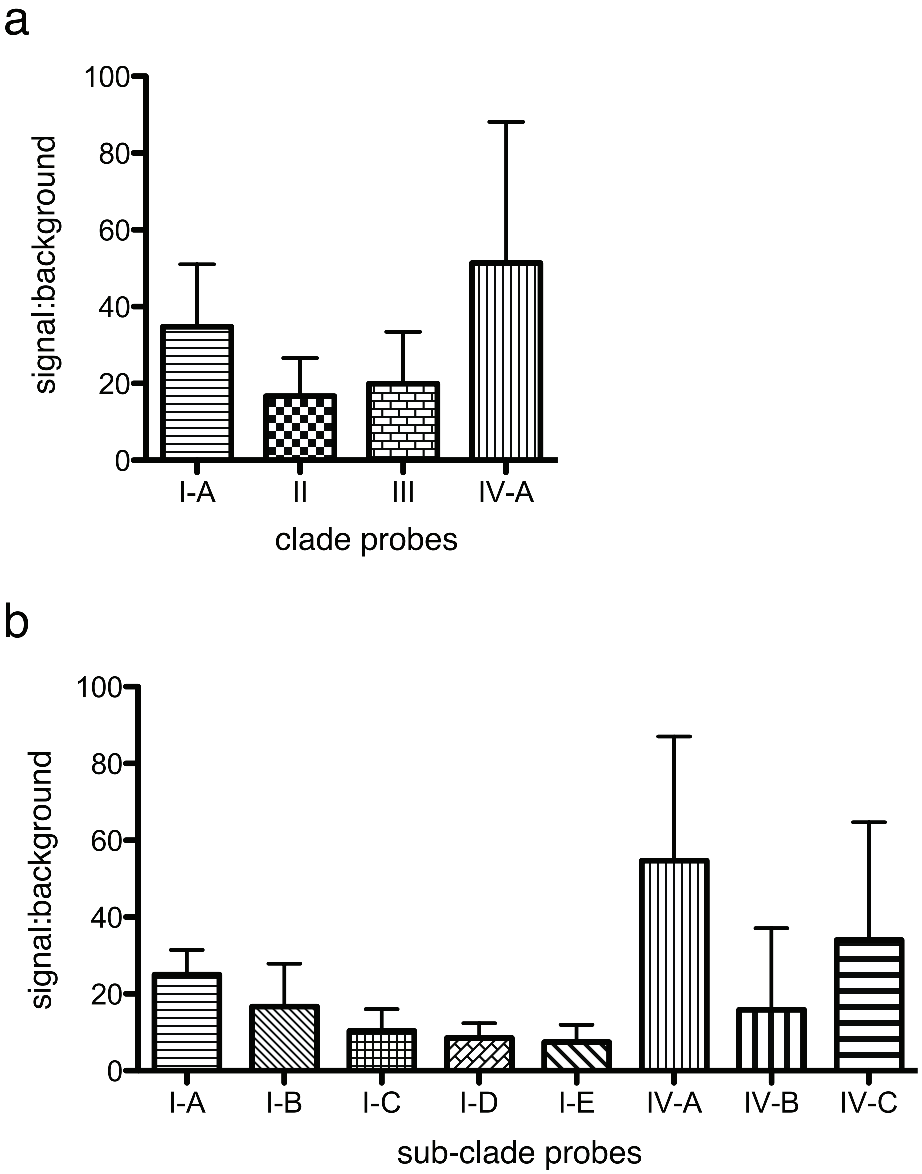
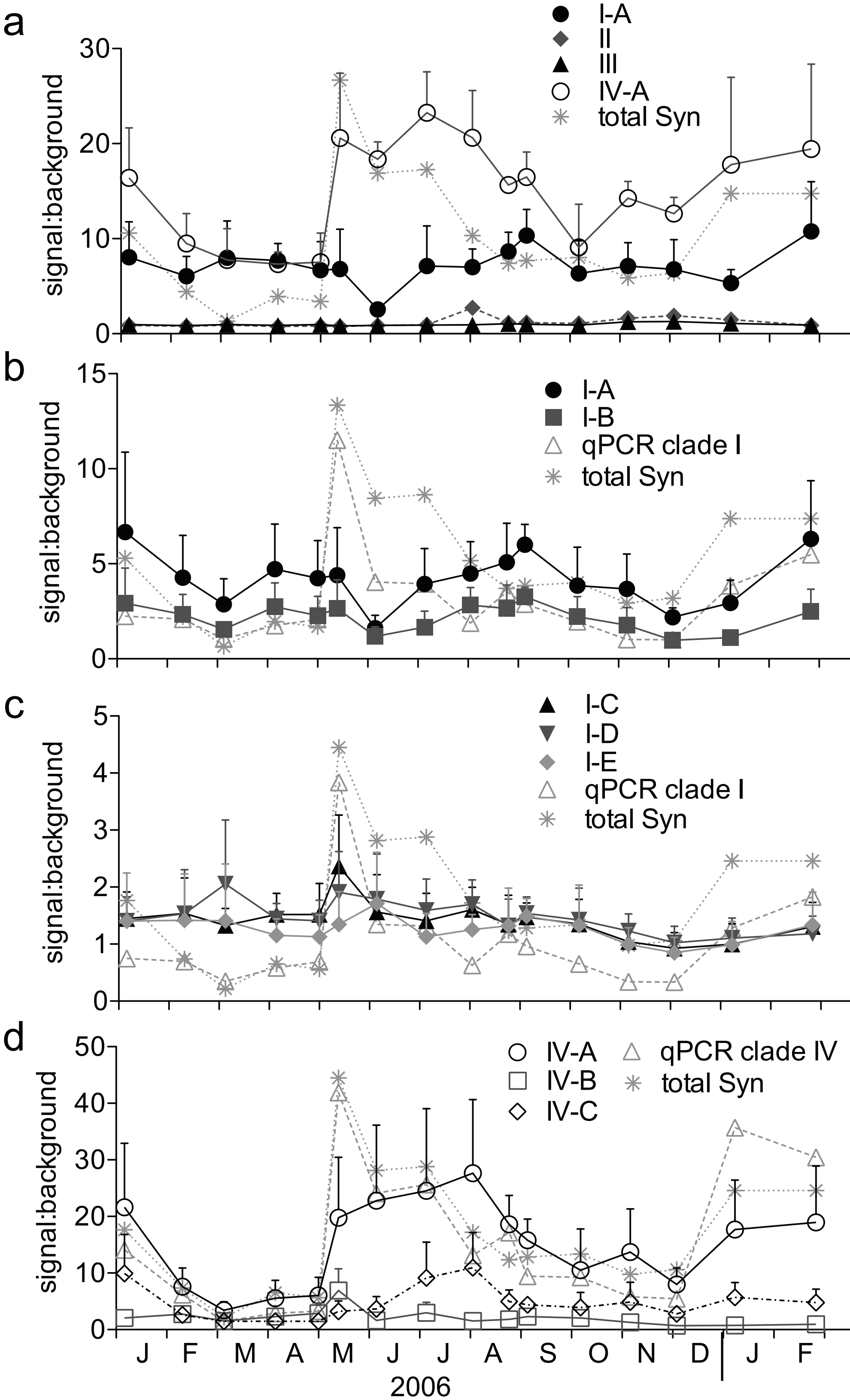
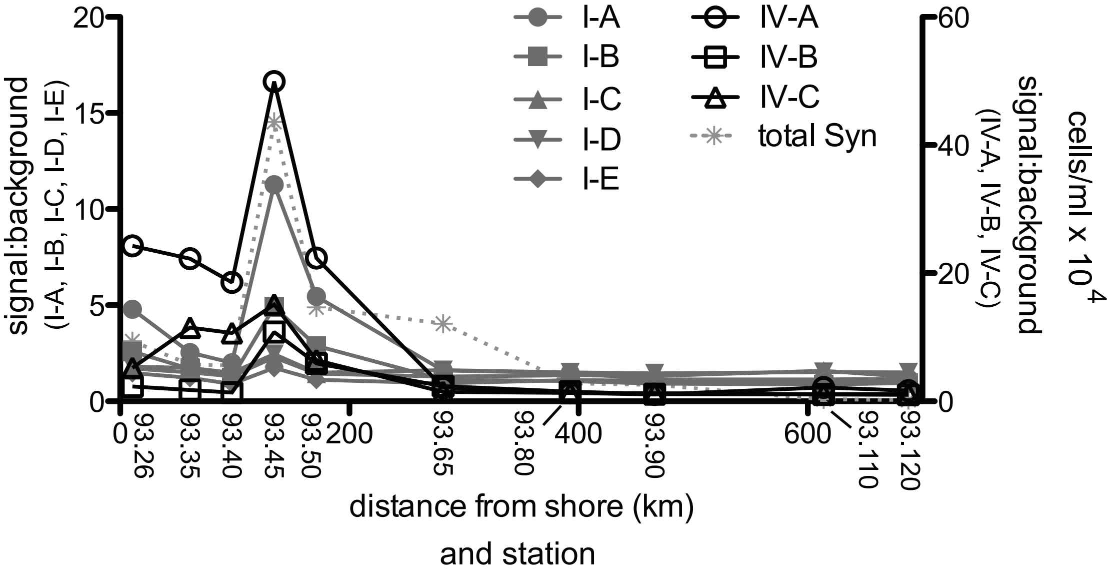


Figure 2.4. Luminex signal:background levels obtained for a) the clade probes and b) the sub-clade probes using DNA standards from cultured *Synechococcus* strains or environmental *rpoC1* clones. The mean signal:background from 6 replicates + the standard deviation is shown.

Figure 2.5. Temporal dynamics of *Synechococcus* clades and sub-clades over 1 year from the SIO pier using the Luminex bead-array. a) clade probes I-A, II, III, and IV-A. b) sub-clade probes I-A and I-B. c) sub-clade probes I-C, I-D, and I-E. d) sub-clade probes IV-A, IV-B, and IV-C. The mean signal:background from 5 or 6 replicates + the standard deviation is shown. The error bar is shown only above the mean for clarity.


a
I-A
II
III
IV-A
total Syn
signal:background
cells/ml x 10^5
30
20
10
0
2.5
2.0
1.5
1.0
5.0
0
b
I-A
I-B
qPCR clade I
total Syn
15
10
5
0
c
I-C
I-D
I-E
qPCR clade I
total Syn
5
4
3
2
1
0
d
IV-A
IV-B
IV-C
qPCR clade IV
total Syn
50
40
30
20
10
0
J
F
M
A
M
J
J
A
S
O
N
D
J
F
2006

a surface



b *Synechococcus* maximum

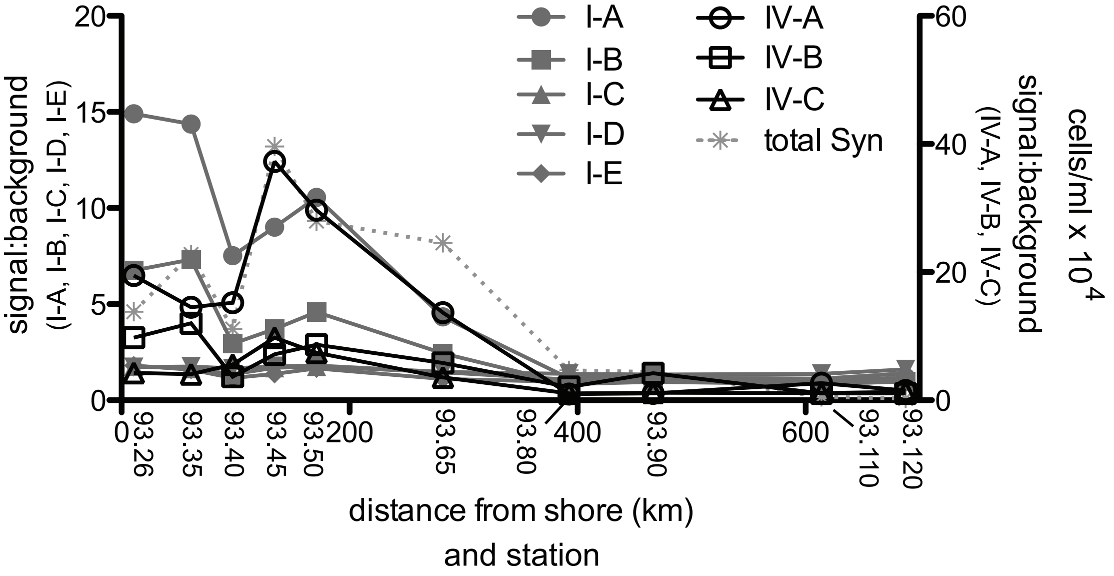


Figure 2.6. Sub-clade Luminex signals and total *Synechococcus* abundances from a) the surface and b) depths of the *Synechococcus* maxima along the Line 93 transect.

Figure 2.7. Sub-clade Luminex signals with depth for each station sampled along the Line 93 transect. Each row of graphs is from one station. The upper x-axis shows the Luminex signal:background while the lower x-axis shows the total *Synechococcus* abundance in cells/ml x 10^5. The left panel of graphs shows the results for sub-clade probes I-A, I-B, and IV-B; the middle, sub-clade probes IV-A and IV-B; the right, sub-clade probes I-C, I-D, and I-E.

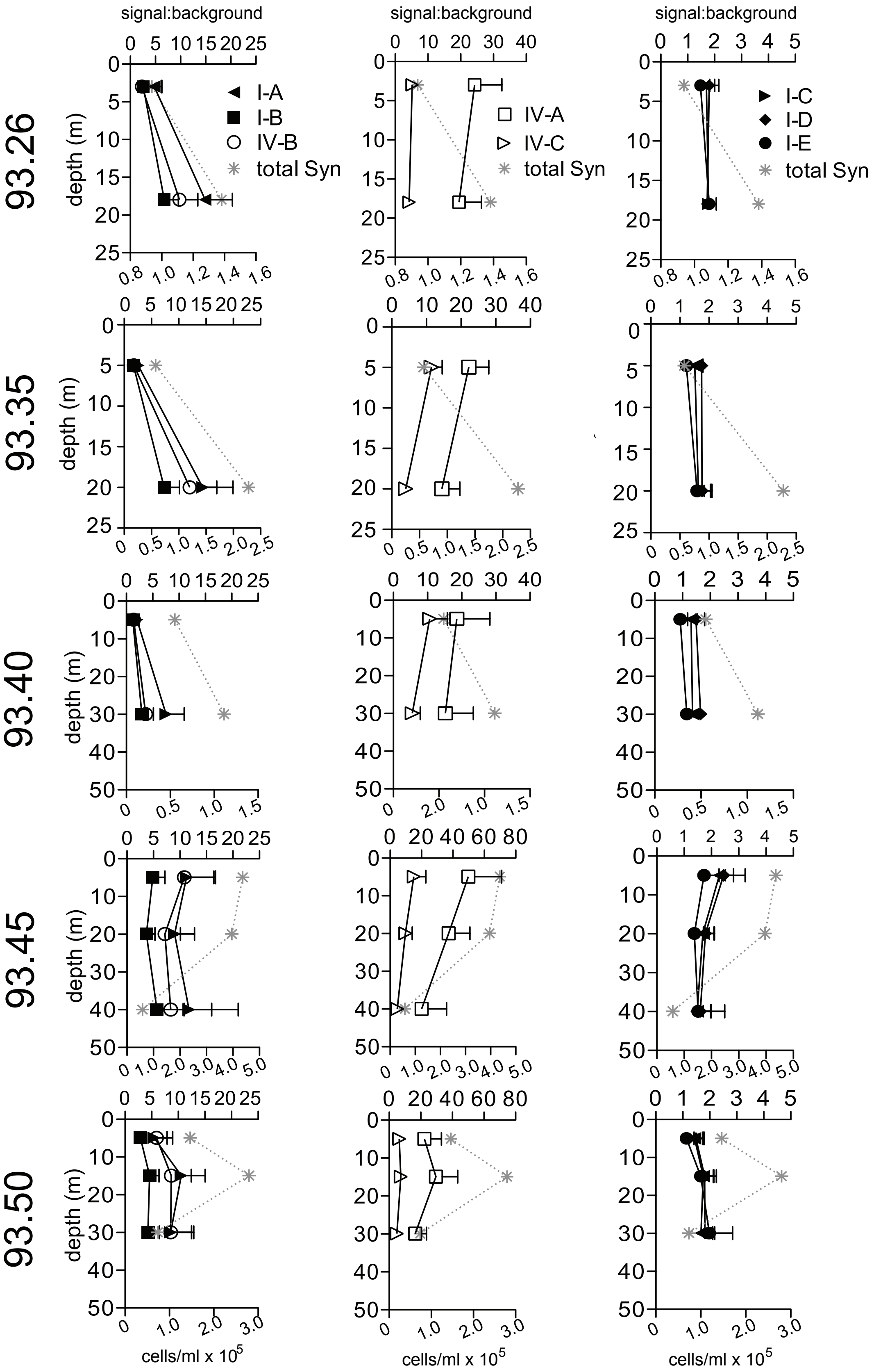
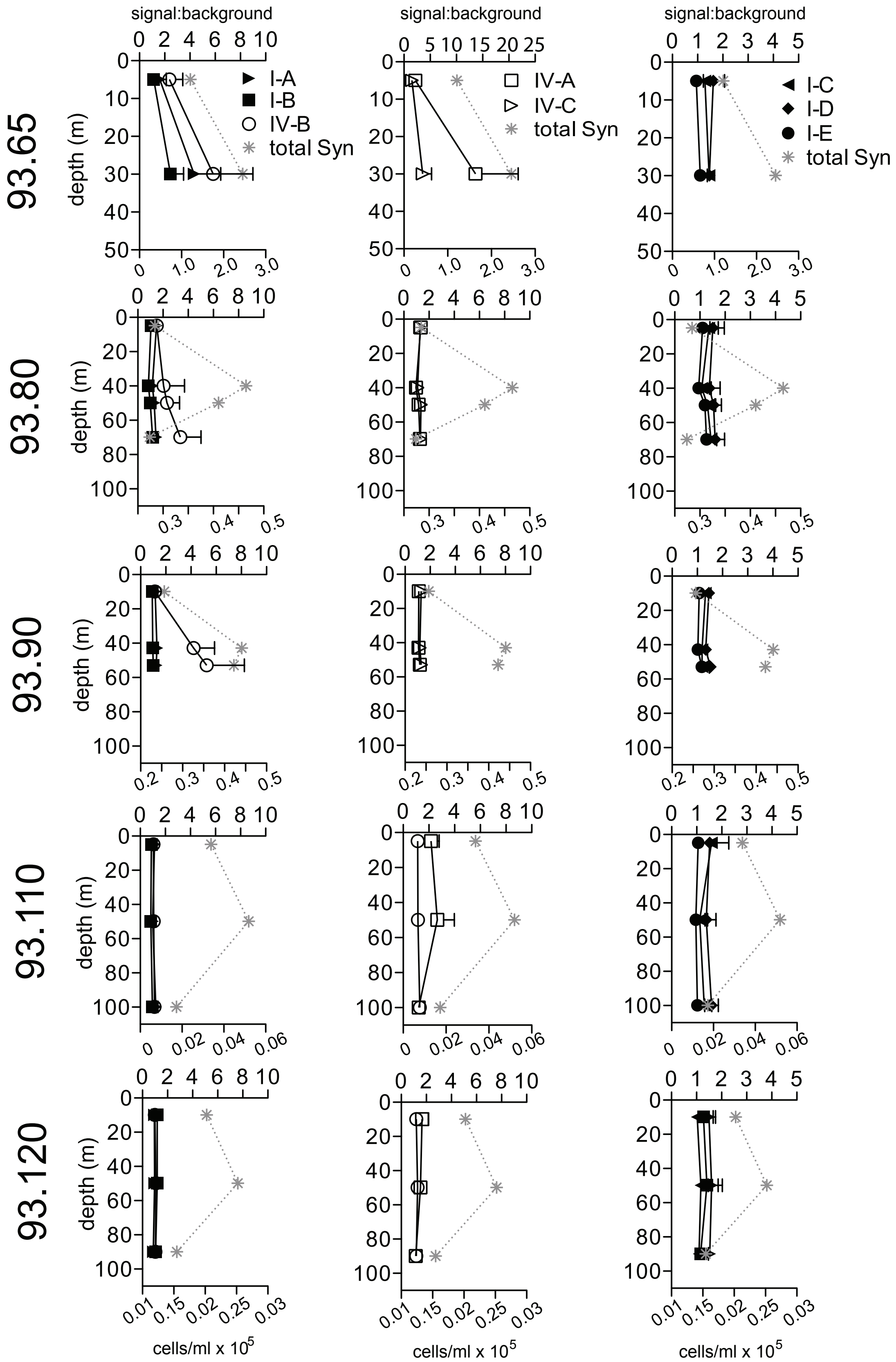

signal:background
signal:background
signal:background
I-A
I-B
IV-B
total Syn
IV-A
IV-C
total Syn
I-C
I-D
I-E
total Syn
93.26
93.35
93.40
93.45
93.50
depth (m)
depth (m)
depth (m)
depth (m)
depth (m)
cells/ml x 10^5
cells/ml x 10^5
cells/ml x 10^5

Figure 2.7. Sub-clade Luminex signals with depth for each station sampled along the Line 93 transect, continued.


signal:background
0 2 4 6 8 10
0 5 10 15 20 25
0 1 2 3 4 5
I-A
I-B
IV-B
total Syn
IV-A
IV-C
total Syn
I-C
I-D
I-E
total Syn
93.65
93.80
93.90
93.110
93.120
depth (m)
0 10 20 30 40 50
0 20 40 60 80 100
0 1.0 2.0 3.0
0.3 0.4 0.5
0.2 0.3 0.4 0.5
0 0.02 0.04 0.06
0.01 0.15 0.02 0.25 0.03
cells/ml x 10^5

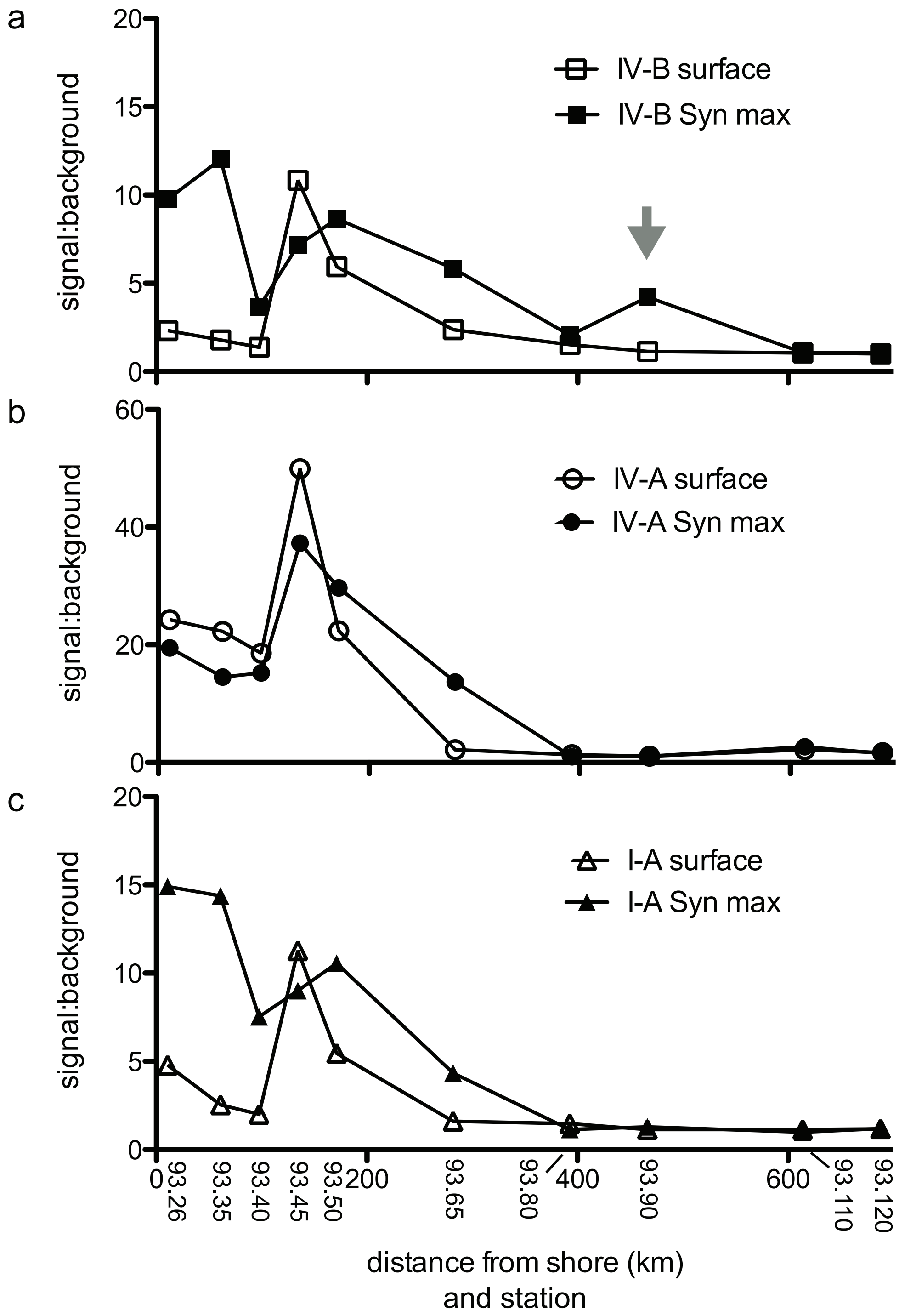


Figure 2.8. Sub-clade Luminex signals from the surface and the *Synechococcus* maxima for probes a) IV-B, b) IV-A, and c) I-A along the Line 93 transect. The arrow points to the mesotrophic station 93.90 where the signal for probe IV-B was dominant at depth.

References

Ahlgren, N.A. and Rocap, G. 2006. Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies. *Appl Environ Microbiol* 72:7193-7204.

Bandyopadhyay, K., Kellar, K.L., Moura, I., Cristina, M., Carollo, C., Graczyk, T.K., Slemenda, S., Johnston, S.P., and da Silva, A.J. 2007. Rapid microsphere assay for identification of *Cryptosporidium hominis* and *Cryptosporidium parvum* in stool and environmental samples. *Journal of Clinical Microbiology* 45:2835-2840.

Baums, I.B., Goodwin, K.D., Kiesling, T., Wanless, D., Diaz, M.R., and Fell, J.W. 2007. Luminex detection of fecal indicators in river samples, marine recreational water, and beach sand. *Mar Pollut Bull* 54:521-536.

Brown, M.V. and Fuhrman, J.A. 2005. Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. *Aquat Microb Ecol* 41:15-23.

Brown, M.V., Schwalbach, M.S., Hewson, I., and Fuhrman, J.A. 2005. Coupling 16S-ITS rDNA clone libraries and automated ribosomal intergenic spacer analysis to show marine microbial diversity: development and application to a time series. *Environ Microbiol* 7:1466-1479.

Campbell, L. and Carpenter, E.J. 1987. Characterization of phycoerythrin-containing *Synechococcus* spp. populations by immunofluorescence. *J Plankton Res* 9:1167-1181.

Collier, J.L. and Palenik, B. 2003. Phycoerythrin-containing picoplankton in the Southern California Bight. *Deep Sea Res Part II* 50:2405-2422.

Diaz, M.R. and Fell, J.W. 2004. High-throughput detection of pathogenic yeasts of the genus *Trichosporon*. *Journal of Clinical Microbiology* 42:3696-3706.

Dufresne, A., Ostrowski, M., Scanlan, D.J., Garczarek, L., Mazard, S., Palenik, B.P., Paulsen, I.T., de Marsac, N.T., Wincker, P., Dossat, C., Ferriera, S., Johnson, J., Post, A.F., Hess, W.R., and Partensky, F. 2008.

Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* 9:R90.

Ellison, C.K. and Burton, R.S. 2005. Application of bead array technology to community dynamics of marine phytoplankton. *Mar Ecol Prog Ser* 288:75-85.

Ferris, M.J. and Palenik, B. 1998. Niche adaptation in ocean cyanobacteria. *Nature* 396:226-228.

Fuhrman, J.A., Griffith, J.F., and Schwalbach, M.S. 2002. Prokaryotic and viral diversity patterns in marine plankton. *Ecol Res* 17:183-194.

Fuhrman, J.a., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M.V., and Naeem, S. 2006. Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci U S A* 103:13104-13109.

Fuller, N.J., Marie, D., Partensky, F., Vaulot, D., Post, A.F., and Scanlan, D.J. 2003. Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ Microbiol* 69:2430-2443.

Fuller, N.J., West, N.J., Marie, D., Yallop, M., Rivlin, T., Post, A.F., and Scanlan, D.J. 2005. Dynamics of community structure and phosphate status of picocyanobacterial populations in the Gulf of Aqaba, Red Sea. *Limnol Oceanogr* 50:363-375.

Fuller, N.J., Tarran, G.A., Yallop, M., Orcutt, K.M., and Scanlan, D.J. 2006. Molecular analysis of picocyanobacterial community structure along an Arabian Sea transect reveals distinct spatial separation of lineages. *Limnol Oceanogr* 51:2515-2526.

Giovannoni, S.J., Britschgi, T.B., Moyer, C.L., and Field, K.G. 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 345:60-63.

Huber, J.A., Mark Welch, D., Morrison, H.G., Huse, S.M., Neal, P.R., Butterfield, D.A., and Sogin, M.L. 2007. Microbial population structures in the deep marine biosphere. *Science* 318:97-100.

Jenkins, B.D., Zehr, J.P., Gibson, A., and Campbell, L. 2006. Cyanobacterial assimilatory nitrate reductase gene diversity in coastal and oligotrophic marine environments. *Environ Microbiol* 8:2083-2095.

Johnson, P.W. and Sieburth, J.M. 1979. Chroococcoid cyanobacteria in the sea: a ubiquitous and diverse phototropic biomass. *Limnol Oceanogr* 24:928-935.

Muhling, M., Fuller, N.J., Somerfield, P.J., Post, A.F., Wilson, W.H., Scanlan, D.J., Joint, I., and Mann, N.H. 2006. High resolution genetic diversity studies of marine *Synechococcus* isolates using rpoC1-based restriction fragment length polymorphism. *Aquat Microb Ecol* 45:263-275.

Nocker, A., Burr, M., and Camper, A.K. 2007. Genotypic microbial community profiling: A critical technical review. *Microb Ecol* 54:276-289.

Palenik, B., Brahamsha, B., Larimer, F.W., Land, M., Hauser, L., Chain, P., Lamerdin, J., Regala, W., Allen, E.E., McCarren, J., Paulsen, I., Dufresne, A., Partensky, F., Webb, E.A., and Waterbury, J. 2003. The genome of a motile marine *Synechococcus*. *Nature* 424:1037-1042.

Palenik, B., Ren, Q.H., Dupont, C.L., Myers, G.S., Heidelberg, J.F., Badger, J.H., Madupu, R., Nelson, W.C., Brinkac, L.M., Dodson, R.J., Durkin, A.S., Daugherty, S.C., Sullivan, S.A., Khouri, H., Mohamoud, Y., Halpin, R., and Paulsen, I.T. 2006. Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. *Proc Natl Acad Sci U S A* 103:13555-13559.

Palenik, B., Ren, Q., Tai, V., and Paulsen, I.T. 2009. Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environmetal Microbiology* 11:349-359.

Penno, S., Lindell, D., and Post, A.F. 2006. Diversity of *Synechococcus* and *Prochlorococcus* populations determined from DNA sequences of the N-regulatory gene *ntcA*. *Environ Microbiol* 8:1200-1211.

Rocap, G., Distel, D.L., Waterbury, J.B., and Chisholm, S.W. 2002. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* 68:1180-1191.

Scorzetti, G., Brand, L.E., Hitchcock, G.L., Rein, K.S., Sinigalliano, C.D., and Fell, J.W. 2009. Multiple simultaneous detection of Harmful Algal Blooms (HABs) through a high throughput bead array technology, with potential use in phytoplankton community analysis. *Harmful Algae* 8:196-211.

Su, Z.C., Mao, F.L., Dam, P., Wu, H.W., Olman, V., Paulsen, I.T., Palenik, B., and Xu, Y. 2006. Computational inference and experimental validation of the nitrogen assimilation regulatory network in cyanobacterium *Synechococcus* sp. WH 8102. *Nucleic Acids Res* 34:1050-1065.

Tai, V. and Palenik, B. 2009. Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site. *ISME J* 3:903-915.

Toledo, G. and Palenik, B. 1997. *Synechococcus* diversity in the California current as seen by RNA polymerase (rpoC1) gene sequences of isolated strains. *Appl Environ Microbiol* 63:4298-4303.

Toledo, G., Palenik, B., and Brahamsha, B. 1999. Swimming marine *Synechococcus* strains with widely different photosynthetic pigment ratios form a monophyletic group. *Appl Environ Microbiol* 65:5247-5251.

Toledo, G. and Palenik, B. 2003. A *Synechococcus* serotype is found preferentially in surface marine waters. *Limnol Oceanogr* 48:1744-1755.

Waterbury, J.B., Watson, S.W., Guillard, R.R.L., and Brand, L.E. 1979. Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* 277:293-294.

Yang, L., Tran, D.K., and Wang, X. 2001. BADGE, BeadsArray for the Detection of Gene Expression, a high-throughput diagnostic bioassay. *Genome Res* 11:1888-1898.

Zwirglmaier, K., Heywood, J.L., Chamberlain, K., Woodward, E.M.S., Zubkov, M.V., and Scanlan, D.J. 2007. Basin-scale distribution patterns lineages in the Atlantic Ocean. *Environ Microbiol* 9:1278-1290.

Zwirglmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vaulot, D., Not, F., Massana, R., Ulloa, O., and Scanlan, D.J. 2008. Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* 10:147-161.

CHAPTER 3

dN/dS ratios from coastal *Synechococcus* populations

Abstract

Metagenomics has provided an unprecedented resource of information ripe for population genetics analyses. Current analytical methods, however, are poorly equipped to handle metagenomic data, particularly of short, unlinked sequences. A custom analytical pipeline was developed to examine the role of selection in *Synechococcus* populations by calculating dN/dS ratios using data derived from 454 metagenome sequences. Genome-wide polymorphism data was analyzed based on sequence alignments to *Synechococcus* genomes representing the dominant clades in the Southern California Bight. The large majority of genes have evolved under purifying selection (dN/dS < 1). The metagenome sequence coverage of the reference genomes was not uniform and a negative correlation was observed between the depth of read coverage per gene and dN/dS ratios. Therefore genes that were highly represented and may be essential for the *Synechococcus* population in this environment tended to be more evolutionarily conserved. Genes that were more rare were diversifying (dN/dS > 1) and were mostly hypothetical genes of the accessory genome. Many of the highly represented genes in the environment with low dN/dS ratios were also hypothetical.

Although these hypothetical genes may play important roles in defining the ecological niches of *Synechococcus* populations, only scant hints of their probable functions could be ascertained.

Introduction

Circumventing the need for cultured isolates, environmental metagenomics provides genetic sequences from microorganisms of entire communities. The collections of sequences reveal simultaneously the breadth of gene content and genetic diversity of these communities. Environmental metagenomes have emerged from a wide diversity of ecosystems, all driven by the biochemistry of the microbial community. These data have been used to examine the biogeochemistry, the taxonomic diversity, and the evolutionary history of the microbes from these ecosystems (see for eg. Gill *et al.*, 2006; Lo *et al.*, 2007; Rusch *et al.*, 2007; Warnecke *et al.*, 2007; Gianoulis *et al.*, 2009; Hewson *et al.*, 2009).

The sampling and analyses of environmental metagenomes can be customized to examine taxonomic groupings at any phylogenetic or ecological scale. This is an unprecedented resource, especially for microbial population biology. As each metagenome sequence is derived from an individual microbe, the diversity of native microbial populations exists within the metagenome datasets. Population analyses of microbes were previously limited to sequences of marker genes or genomes of cultured isolates. Environmental metagenomes contain a census of the true population diversity, genome-wide, without bias with respect to culturability.

The population is the fundamental unit upon which selection acts. Individuals in a population have varying phenotypes, but those with fittest phenotypes have a greater chance of survival and reproduction. Variations in genetic sequences, acquired through mutation or gene transfer, provide the raw material for phenotypic variants of a given trait. While mutation increases variation, selection and genetic drift act to reduce it.

Population genetics provides the theoretical framework to examine the effects of mutation, selection, and drift on the diversity, evolution, and demographic history of a population. One method of investigating the role of selection in the evolution of a gene in a population is to determine the dN/dS ratio (Perler *et al.*, 1980; Li *et al.*, 1985; Hughes and Nei, 1988; Nielsen and Yang, 1998). This ratio tests whether the nucleotide variants (polymorphisms) of a gene in a population contribute to an excess of non-synonymous mutations (mutations that change the amino acid sequence) than expected by chance ($dN/dS > 1$). This would imply that the gene is undergoing positive selection as natural selection is favoring more diversification at the amino acid level. Purifying selection is signified by fewer amino acid changes than expected by chance ($dN/dS < 1$). Most genes are thought to have evolved this way as most mutations are deleterious and removed by selection, and the occurrence and fixation of beneficial mutations is rare (Mukai, 1964; Mukai *et al.*, 1972; Kimura and Ohta, 1974; Eyre-Walker and Keightley, 1999).

Only a handful of studies have used environmental metagenomic data to investigate the population genetics of a microbial communities (Allen *et al.*, 2007; Simmons *et al.*, 2008; Hu and Blanchard, 2009; Konstantinidis *et al.*, 2009). These investigations all demonstrate the predominance of purifying selection in the evolution of microbial populations. By analyzing Sanger-based sequences, these previous studies could take advantage of methods already established for population genetics analysis. Several approaches to calculate dN/dS rely on pairwise alignments or multiple sequence alignments to first infer a phylogenetic tree from which the evolutionary path of mutational changes is assessed (Nei and Gojobori, 1986; Yang, 1998; Yang and Nielsen, 2002; Kosakovsky Pond and Frost, 2005; Librado and Rozas, 2009). These methods, however, are ill equipped to handle the short sequence reads (< 500 bp) of next-generation sequencing technologies, such as Roche 454 or Illumina. With short sequence reads, especially if linkage data is unavailable such as from environmental metagenomes, obtaining lengthy sequence alignment blocks (all sequences aligned to the same region) across a gene is not possible.

This chapter examines the population diversity of *Synechococcus* from coastal samples of the Southern California Bight – an environment dominated by *Synechococcus* from clades I and IV (Tai and Palenik, 2009). New analytical methodologies were developed to calculate dN/dS ratios using Roche 454 sequences and examine the role of selection in the evolution and diversification of *Synechococcus* populations from clades I and IV.

Materials and methods

Environmental sampling, sorting, and sequencing

Samples were primarily collected from surface waters at the end of the Scripps Institution of Oceanography (SIO) pier, La Jolla, CA (32.8672 N, 117.2583 W). As previously described for a metagenomic sample from October 10, 2006 (Palenik *et al.*, 2009), samples were similarly collected and sorted on May 17, 2007, March 6, 2008, and April 17, 2008. On May 15, 2008, samples were collected from the surface and at 20 m depth using a niskin bottle from a station approximately 1.7 km off-shore of the SIO pier (32.8733 N, 117.2750 W). Chlorophyll and nutrient concentrations were determined for each sample as previously described (Palenik *et al.*, 2009) (Table 3.1).

For each sample, approximately 10 liters of seawater were pre-filtered through a 2 μm polycarbonate filter (GE Osmonics) then the microorganisms were collected onto a 0.2 μm Supor filter disc (Pall). The microorganisms were resuspended from the filter by vortexing in approximately 10 ml of 0.2 μm filtered seawater.

Synechococcus were enriched using flow cytometry sorting and collected onto a 0.2 μm Supor filter disc and stored at –80 °C as previously described (Palenik et al. 2009). DNA was extracted, amplified, and sequenced as previously described (Palenik *et al.*, 2009) except using a GS-FLX instrument (454 Sequencing, Roche) providing reads averaging 250 bp in length. The sequences previously published from 10/10/06 averaged 100 bp in length (Palenik *et al.*, 2009) and were also included in this study.

Sequence tiling to CC9311 and CC9902 genomes

The metagenomic sequences from the 6 samples were pooled and analyzed together. Using the Reference Assembly tool from the CLC Genomics Workbench (version 3.2), the pooled environmental metagenomic sequences were tiled (a.k.a. aligned, recruited) to the *Synechococcus* CC9311 (clade I) and *Synechococcus* CC9902 (clade IV) genomes simultaneously. These genome sequences represent the two dominant *Synechococcus* clades in the coastal Southern California Bight and tilings to *Synechococcus* genomes from other clades resulted in very poor sequence recuitment (Palenik *et al.*, 2009; Palenik, Ren, Tai, Paulsen unpublished data).

Different tiling parameters were tested. The % identity and % coverage (length fraction) were varied between 100 and 50 %, while using alignment penalties of mismatch = 1, insertion = 2, and deletion = 2. For each tiling, two alignments resulted comprising separate, non-overlapping sets of metagenome sequences, one with sequence reads aligned to the CC9311 genome and another to the CC9902 genome. As the reads were tiled to the CC9311 and CC9902 genomes simultaneously, reads that would have hit both genomes with the given parameters were recruited to the genome with the best match. Equal hits were assigned randomly. The number of new sequences tiled to the reference genomes decreased as the % identity parameter was lowered below 80 % (Figure 3.1). Thus the tiling resulting from 80 % identity and 80 % coverage was used for all subsequent analyses.

To confirm that the reads tiling to the CC9311 and CC9902 genomes with 80 % identity and 80 % coverage encompassed a *Synechococcus* clade I or IV population, a BLASTn search of the non-redundant nucleotide database from GenBank was performed (Altschul *et al.*, 1990, http://blast.ncbi.nlm.nih.gov./Blast.cgi). For all reads that tiled to either the CC9311 or CC9902 genomes, the best BLASTn hit was determined. It was expected that reads tiling to the CC9311 genome would have this genome as the top hit and similarly with the CC9902 tiling. If reads did not have the expected top hit, the taxon of the top hit was noted.

Simulated metagenome sequences were generated to test analytical methods. 300 000 sequence reads were simulated from the CC9311 and CC9902 genomes using MetaSim (Richter *et al.*, 2008, http://www-ab.informatik.uni-tuebingen.de/software/metasim). Systematic sequencing errors known to occur from 454 sequencing were introduced into the simulated reads. These sequences were analyzed as for the actual 454 sequence reads from the environment.

Analysis of tiled reads and dN/dS calculation

The CC9311 and CC9902 alignments resulting from the tiling using the criteria of 80 % identity and 80 % coverage were exported from the CLC Genomics Workbench. Custom Python scripts were implemented to analyze the alignments. For each nucleotide position of the reference genomes, the depth of read coverage and the occurrence of polymorphisms were recorded. When examining the alignments for polymorphisms, nucleotides from the sequence reads were only considered if they had a quality score above 20 and the average quality score of the surrounding 11 nucleotides was greater than 15. A nucleotide position was considered polymorphic if the aligned nucleotides were not identical (but not considering Ns). A majority-rule consensus sequence was determined from the metagenome reads aligning to the reference genomes. Thus, the polymorphisms and the consensus sequences were derived solely from the environmental *Synechococcus* community because they were determined from the metagenome sequences only and the reference genome sequences were excluded.

The gene annotations from the reference genomes were used to define the start and stop coordinates of genes in the consensus sequences. For each gene, dN/dS was calculated as:

$$\frac{dN}{dS} = \frac{\dfrac{\text{the number of observed non-synonymous mutations}}{\text{the number of non-synonymous sites}}}{\dfrac{\text{the number of observed synonymous mutations}}{\text{the number of synonymous sites}}}$$

The consensus sequence was used to calculate the number of non-synonymous and synonymous sites for each nucleotide of a gene. A site is a possible change/mutation for each nucleotide position of a codon. If the possible mutation to the codon would result in an amino acid change, this is a non-synonymous site. If there is no amino acid change, it is a synonymous site. Each possible mutation was considered to determine the number of sites. For each observed polymorphic nucleotide position, if the alignment contained a nucleotide that differed from the consensus, the type of mutation (non-synonymous or synonymous) was determined. The type of mutation was counted only for positions that were covered by a depth of at least 5 metagenomic sequence reads.

dN/dS was also calculated using the reference genome (i.e. CC9311 or CC9902) instead of the consensus sequence as the basis for determining whether the polymorphisms were synonymous or non-synonymous.

Analysis of hypothetical genes with high read coverage and dN/dS > 1

Experimental data was examined to determine if hypothetical genes are expressed and bioinformatics tools were used to determine possible functions. A tBLASTn search of the non-redundant GenBank database was used to find the best hit to the hypothetical genes (Altschul *et al.*, 1990, http://blast.ncbi.nlm.nih.gov/Blast.cgi). Hits to genes with a low e-value provided evidence that the gene is real and annotated correctly. The protein products of these genes were searched for in proteomes of CC9311 and CC9902 cultures (Palenik et al. unpublished data). For microarray expression data existing for CC9311 under stress conditions, i.e. exposed to toxins (ethidium bromide and mitomycin C), salt shock, and copper shocks (Palenik, Stuart, Tetu, Dupont, Johnson, Paulsen et al. unpublished data), it was noted if these genes showed significant changes in expression. Possible structures for these genes were determined using the Superfamily SCOP classification (http://supfam.org/SUPERFAMILY/) and HHpred (http://toolkit.tuebingen.mpg.de/hhpred).

Statistical analyses

All statistical analyses were conducted with Prism (version 5.0b, GraphPad Software Inc.). To compare dN/dS ratios from the clade I and IV populations, the dN/dS ratios were square-root transformed to obtain a more normal distribution of values.

Results

Defining the reads belonging to clade I and IV *Synechococcus* populations

Pooling the metagenomes together, a total of 2 856 850 sequence reads equivalent to approximately 530 Mb were tested for alignment with the CC9311 and CC9902 genomes simultaneously using varying identity and coverage criteria. After tiling, two sets of sequence reads were obtained, one set aligning to CC9311, the other to CC9902.

With the % identity and % coverage lowered from 100 to 80 %, a high rate of sequence reads were recruited to the reference genomes (Figure 3.1). Below 80 % identity and 80 % coverage, the number of additional sequence reads recruited to the genomes diminished. This suggests that the metagenome sequences were not uniformly divergent and there is likely a biological reason for this observed pattern of read recruitment.

As the number of additional sequences recruited to the reference genomes drastically decreased at less than 80 % identity and 80 % coverage, this change in sequence recruitment likely represents a threshold of genetic diversity that delimits a taxonomic boundary within *Synechococcus*, specifically below clade-level. This is by no means a sharp boundary as some genes necessitate strong conservation and tolerate less variation than others. There may be sequences with greater than 80 % identity that do not belong to these particular *Synechococcus* clades and vice versa. However, this was a guideline that was used for all subsequent population genetic analyses. Because *Synechococcus* from clades I and IV were dominant in this environment, it is unlikely that sequences tiled to the reference genomes do not actually belong to clades I or IV.

Using the criteria of 80 % identity and 80 % coverage, 138 734 (4.9% of the total number of sequence reads) and 131 261 (4.6 %) reads tiled to the CC9311 and CC9902 genomes, respectively. 543 of the recruited reads hit both genomes equally. Most of these reads hit ribosomal RNA genes (85.3 %) or other highly conserved genes.

These tiling criteria worked well to define the reads comprising a *Synechococcus* clade I population and a clade IV population. The majority of reads tiling to the CC9311 and CC9902 genomes were consistent with a BLASTn search to the non-redundant GenBank database. Of the reads recruited to the CC9311 genome, 96.9 % had the same genome as their top BLASTn hit. Of the reads that tiled to CC9311 but did not have CC9311 as their top BLASTn hit, 76.5 % of these blasted to environmental sequences, likely from clade I *Synechococcus* (Figure 3.2a). Similarly 97.2 % of the reads tiling to CC9902 also had CC9902 as their top BLASTn hit. Top hits to environmental sequences comprised 67.0 % of the remainder of the reads (Figure 3.2b). Only 57 reads (1.6 %) had BL107, another clade IV *Synechococcus* isolate, as their top hit, all hitting copies of the photosystem II reaction centre protein PsbA/D1. Excluding these reads, only 0.7 % and 0.9 % of the sequences that tiled to CC9311 or CC9902 might not belong to either clade I or IV, respectively. These reads mostly hit strains from other *Synechococcus* clades (68.8 and 81 %), *Prochlorococcus* (4.7 and 7.0 %), photosynthetic genes found in cyanophages (4.5 and 7.8 %), or the rRNA genes of bacteria other than *Synechococcus* (10.2 and 6.5 %). Therefore, using the criteria of 80 % identity and 80 % coverage, with some minor exceptions, the metagenome sequences that tiled to either the CC9311 and CC9902 genomes effectively represented population diversity within clades I and IV, respectively.

dN/dS

A dN/dS ratio was calculated for every gene as annotated in the reference genomes at nucleotide positions with at least 5-fold coverage of metagenome sequence reads. The average depth of read coverage for a gene ranged from 0 to 68.7 and 72.6 for tilings to the CC9311 and CC9902 genomes, respectively with means of 11.1 and 12.3 (Figure 3.3a). For core genes (i.e. orthologous genes found commonly in *Synechococcus* strains CC9311, CC9902, CC9605, and WH8102), the average depth of read coverage was 12.4 and 12.1 for the CC9311 and CC9902 tilings, respectively. For accessory genes (i.e. genes that are found in some but not all sequenced genomes of *Synechococcus* strains), it was 8.6 and 13.3, respectively. 50.5 and 64.7 % of the genes were covered > 90 % of their length at least 5-fold (Figure 3.3b). 18.5 and 4.7 % of the genes were not well covered by metagenome sequences (< 10 %). Of these genes, 79.1 and 75.2 are part of the accessory genome and likely highly variable components that were rare in the environmental populations (Palenik *et al.*, 2009). A greater number of the CC9311 genes were covered < 10 % from the metagenome (535 genes versus 109 CC9902 genes) as CC9311 has a higher number of accessory genes than CC9902 (Dufresne *et al.*, 2008).

Based on the polymorphisms in the metagenome sequences tiled to the CC9311 and CC9902 genomes, the median dN/dS was 0.118 and 0.117, respectively. The frequency distribution of the dN/dS ratios indicates that the large majority of genes have a mean dN/dS ratio < 1 and have evolved under purifying selection (Figures 3.4a,b). Some sites of the gene may still have experienced positive selection, but the overall trend is purifying selection. The dN/dS ratios for each gene as well as read coverage information are provided in Appendix 5 (Tables A5.1 and A5.2).

Simulated 454 sequences of the CC9311 and CC9902 genomes, resulted in median dN/dS ratios of 0.940 and 1.05 (Figures 3.4c,d). The dN/dS ratios were also more evenly distributed than the dN/dS ratios derived from natural populations indicating that the mutations were not biased towards synonymous or non-synonymous types. The polymorphisms from these simulated data occurred exclusively through random mutation introduced by sequencing error. Therefore, the method used to calculate dN/dS accurately portrayed the simulated data as evolving neutrally without selection pressure.

To examine the relationship of read coverage to dN/dS ratios, the distribution of dN/dS ratios were compared between genes with depths of read coverage greater and less than the mean (Figure 3.5). The mean dN/dS ratios of genes with greater and less than the average read coverage were statistically significantly different for the clade IV population (t-test, $p = 0.009$), but not for the clade I population ($p = 0.592$). However, as the distributions and means were very similar, this difference is not considered to be biologically significant. For both clade I and clade IV populations, the depth of read coverage did not significantly impact the distribution of dN/0dS ratios.

The individual dN/dS ratios were compared to the average depth of coverage for each gene. There was a strong negative correlation between dN/dS and depth of coverage. To better assess this correlation, the analysis was repeated using the difference dN - dS (Figure 3.6) because ratio statistics (i.e. dN/dS) are susceptible to excessive variation when the denominator takes small values. Where dN - dS = 0 implies dN=dS and is therefore equivalent to the null hypothesis (no selection, dN/dS = 1). The slopes of the linear regressions were significantly different for the core compared to the accessory genes ($p < 0.0001$ and $p = 0.013$ for the clade I and IV populations, respectively). Because relatively fewer core genes had dN/dS > 1, the slopes were slightly flatter. No relationships were observed using simulated data.

Genes with high read coverage are likely abundant components of *Synechococcus* genomes in the environment and may have important roles in defining the ecological niches of *Synechococcus* populations. Many of the genes with high read coverage are part of the accessory genomes of CC9311 and CC9902. The accessory genome includes the genes that are found in some but not all of the sequenced *Synechococcus* genomes. A partial list of these genes is provided in Tables 3.2 and 3.3. This list includes “unique” genes, i.e. genes unique to the genome based on reciprocal BLAST searches comparing the genomes of *Synechococcus* strains CC9311, CC9605, WH8102, and CC9902. However, as this analysis was performed prior to the availability of the genomes for *Synechococcus* strains BL107, WH7803, WH7805, RS9916, RS9917, and RCC307, many of these genes are not truly “unique” to a genome, but occur in some but not all sequenced genomes. Highly represented genes with greater than 20X read coverage are listed. 20X is approximately 1 standard deviation greater than the mean depth of coverage.

Most (73 and 88 %, respectively) of the highly covered “unique” genes from the CC9311 and CC9902 genomes are hypothetical with no known function (Tables 3.2 and 3.3). Some of these genes, however, have very well characterized functions such as the CC9311 genes sync_0681 and sync_0682 that encode components of a ferrous iron transporter (Table 3.2). These genes have atypical GC content and may have been horizontally transferred into an ancestor of CC9311(Palenik *et al.*, 2006). As their dN/dS ratios are low, this indicates that the genes have evolved under purifying selection and conservation of function has been maintained.

Genes with dN/dS ratios > 1

From the clade I and clade IV populations, 75 and 15 genes, respectively, had dN/dS ratios > 1 which suggests that they may have evolved under positive selection. Again, many of these are hypothetical genes with no known function and are part of the accessory genome of *Synechococcus* (Tables 3.4 and 3.5). These hypothetical genes may not be essential to the core functions of *Synechococcus*, but may provide unique functions that differentiate them from other *Synechococcus*. If these genes provide an adaptive trait, then they are more likely to be positively selected.

Only a few of the genes with dN/dS > 1 are not hypothetical. For example, only non-synonymous polymorphisms were observed for the ribosomal protein L35 (sync_0059) (Table 3.4). A putative lipoprotein (sync_2711) has a dN/dS < 1, but relative to the CC9311 genome, the dN/dS = 2.878. Otherwise, other functionally annotated genes with dN/dS > 1 suffer from low metagenome sequence coverage and their dN/dS ratios are likely not accurate.

Assessing the function of hypothetical genes

As most of the highly covered "unique" genes and genes with a dN/dS > 1 are annotated as hypothetical (Tables 3.2, 3.3, 3.4, and 3.5), it is difficult to assess what functions they may provide or if in fact these are real genes. Different sources of information were used to determine if these genes were annotated correctly and what their probable function might be. tBLASTn searches typically found no significant hit or the best hit was to genes from other *Synechococcus* or *Prochlorococcus* genomes. Hits to other hypothetical genes at least indicate that these open reading frames are conserved within the *Synechococcus/Prochlorococcus* group, are likely true genes, and should be re-annotated as conserved hypothetical proteins.

Very few of the hypothetical genes had predicted folds homologous to protein fold databases. From the list of genes in Tables 3.2, 3.3, 3.4, and 3.5, the only hypothetical genes with homology to known structures were sync_1772 (alpha/beta fold family of hydrolases), sync_2816 (protein kinase-like), and sync_2817 (periplasmic binding protein-like) (Table 3.2); sync_0002 (PRC-barrel domain) (Table 3.4); and syncc9902_0240 (RmlC-like cupin fold) and syncc9902_2037 (UDP glycosyltransferase-like) (Table 3.5).

Most of the CC9311 hypothetical genes, however, did show differential expression in microarrays (Tables 3.2 and 3.4). Microarray data is available for CC9311 under stress conditions (Palenik, Stuart, Tetu, Dupont, Johnson, Paulsen, et al. unpublished data). Under these conditions, 58 % of the highly covered CC9311 hypothetical genes increased in expression and 62 % decreased under one or more of the physiological shocks. For the hypothetical genes with dN/dS ratios > 1, 44 % increased in expression while 66 % of the genes decreased. This confirms that under certain physiological circumstances, mRNA corresponding to these genes was expressed, but it did not confirm the reading frame. Genome-wide expression data is not available for CC9902 so a similar analysis could not be performed.

dN/dS ratios of homologous genes

The dN/dS ratios from the clade I and IV populations were compared. For homologous genes, the dN/dS ratios were similar in both populations (Figure 3.7). Based on a linear regression, the dN/dS ratio for each homologous gene was slightly higher in the clade IV population for ratios < 0.35, and lower for ratios > 0.35. Lower dN/dS ratios suggest that these genes experienced stronger purifying selection. However, dN/dS ratios can change over time and the rate of change is sensitive to population size and the "hitchhiking" of slightly deleterious mutations (Rocha *et al.*, 2006). Thus, the comparison of dN/dS ratios between homologous genes from different populations is valid only if these parameters have been similar in the two populations since they diverged.

For genes homologous between CC9311 and CC9902, only 11 genes had dN/dS ratios > 1 (Table 3.6). Generally the homologous gene had dN/dS < 1. The genes with dN/dS ratios > 1, tended to have low metagenome read coverage, so the estimates of dN/dS are not accurate over the whole gene. Sync_0059, sync_0523, and syncc9902_1875 are genes with good sequence coverage and dN/dS > 1. Sync_0523 and syncc9902_1875 are homologous genes based on being reciprocal best tBLASTn hits. However, syncc9902_1875 (251 amino acids) is a considerably larger open reading frame (ORF) than sync_0523 (61 amino acids) indicating that nucleotide deletion has contributed to the evolution of these genes.

Discussion

By developing methods to calculate dN/dS ratios from Roche 454 sequence reads, the predominance of purifying selection was demonstrated from environmental metagenomes of two co-existing coastal *Synechococcus* populations.

Defining *Synechococcus* populations

Flow cytometry sorting was used to enrich for *Synechococcus*, but the enrichment was not sufficient to assemble entire genomes or very large contigs for population analysis as has been done previously (Simmons *et al.*, 2008). Alternatively, completed *Synechococcus* genomes of strains known to represent the dominant clades in the environment were used as scaffolds to collect and align sequences representing *Synechococcus* populations.

Two sets of sequences were obtained: one representing a clade I population as the sequences were aligned to the genome of a clade I strain, CC9311 and another representing a clade IV population, aligned to CC9902, a clade IV strain. *Synechococcus* from clades I and IV dominate the coastal Southern California Bight, but these sets of sequences may not include all of the sequences belonging to clades I and IV nor do they represent the only *Synechococcus* populations in the environment. For instance, if the BL107 draft genome was used as a reference genome instead of CC9902 (like CC9902, BL107 is also a clade IV *Synechococcus* strain), 19 898 (15.5 %) sequences were recruited to the BL107 genome that were not recruited to CC9902. Thus the two sets of sequences used for population analyses, either tiled to CC9311 or CC9902, represent populations within, and not of, clades I and IV.

It is not a trivial matter to define the boundaries of a microbial population. Even to define a microbial species is a controversial issue (Konstantinidis and Tiedje, 2005; Cohan, 2006; Doolittle and Zhaxybayeva, 2009). With only genetic data available, can microbial populations be characterized without further physiological and ecological data? The criteria that define a microbial population or species have yet to be resolved. The amount of sequencing effort required to accurately describe the microbial population structure is also not known (Huber *et al.*, 2007). Regardless, the criteria used here worked well to align the metagenomic sequences and loosely describe populations within a clade of *Synechococcus*.

Roche 454 sequencing for use in environmental population genetics

As 454 sequencing technology is new relative to Sanger-based sequencing, the reliability of 454 sequencing has only recently been tested. These studies have shown that 454 sequences are over 99.9 % accurate and can detect variant sequences with 95 % accuracy (Huse *et al.*, 2007; Harismendy *et al.*, 2009). In addition, strict quality criteria were used to remove poorly called nucleotides from the sequence alignments.

The 454 sequences analyzed here were typically 100 or 250 bp in length. Thus, individual sequence reads could not cover the breadth of a gene and the aligned sequence reads were staggered across in varying depths of coverage. Furthermore, the 454 reads could not be linked genetically. Therefore, pairwise approaches to calculating dN/dS ratios could not be used practically as the sequences are required to align to the same genetic region, *i.e.* an alignment block (Nei and Gojobori, 1986). This also prevented the use of more sophisticated phylogeny based calculations that can determine site or lineage specific dN/dS ratios (Yang, 1998; Yang and Nielsen, 2002; Kosakovsky Pond and Frost, 2005).

Instead, an alternative method was developed where polymorphisms in the population were determined based on differences to a population consensus sequence. With this method, the 454 sequences and their associated polymorphisms are assumed to have equally diverged from the consensus sequence. It was not necessary to trim sequences to obtain alignment blocks and the polymorphism data from all aligned sequences could be used. Using this method, gene alignments from simulated 454 sequences were correctly shown to be under no selection pressure and the mean dN/dS ratio of all genes was close to 1 (Figure 3.4).

The 454 read coverage per gene varied widely (Figure 3.3). To evaluate the role of selection in genes that were more rare in the *Synechococcus* populations, a minimum read coverage of 5X was used. This may be considered low for determining polymorphisms, but the depth of sequence coverage did not affect the overall distribution of dN/dS ratios. The disribution of dN/dS ratios of genes with greater than average depth of coverage were similar to those with less than average depth of coverage (Figure 3.5).

Read coverage, gene conservation, and evolutionary rate

Read coverage per gene varied because the environmental populations of *Synechococcus* were not uniform in gene content. The *Synechococcus* community in the Southern California Bight is dominated by two clades, but tremendous diversity exists within these clades. Comparative genomics and genome hybridizations have demonstrated that strains within a clade do not share the same genetic make-up (Dufresne *et al.*, 2008; Tai, Paulsen, Palenik et al. unpublished data). Numerous genomic regions exist whose genetic content and occurrences are highly variable from strain to strain, and likely individual to individual. In addition many genomic regions belonging to *Synechococcus* are not represented by cultured isolates (Palenik et al. unpublished data). Thus, genes from the reference genomes that were rare in the environmental populations would not have high sequence representation in the metagenomes. In contrast, genes that were common within the *Synechococcus* populations were covered to a greater depth by the metagenome sequences. Genes that may have been duplicated or in greater copy number in environmental genomes, compared to the reference genomes, would also result in relatively higher sequence representation.

Using metagenomic data, a negative correlation was observed between read coverage and dN - dS indicating that *Synechococcus* genes with high representation in the environment were more conserved evolutionarily (Figure 3.6). It was expected that core genes would be well covered by the metagenome sequences and be conserved with dN/dS ratios $<<$ 1 (or dN – dS < 0). However, many accessory genes were also well represented by the metagenomes. High sequence representation suggests that these genes are important in this environment regardless if they are from the core or accessory genomes, and the correlation with low dN - dS indicates that these genes are also evolutionarily conserved.

It has often been assumed that essential genes that are indispensable for an organism to function are more conserved and evolve more slowly than non-essential ones (Wilson *et al.*, 1977). Recent studies that directly test this hypothesis, however, have been equivocal (Hurst and Smith, 1999; Hirsh and Fraser, 2001; Jordan *et al.*, 2002; Wang and Zhang, 2009) as it is difficult to identify which genes are essential in a given environment. By examining this question in a different way, if high sequence representation in the environment indicates the essentiality or necessity of a gene in an ecosystem, then the data presented here shows that these genes do evolve more slowly. This rationale would imply that, for example, ferrous iron transport is an essential function for a clade I *Synechococcus* population in the coastal Southern California Bight (Table 3.2). Unfortunately, the function of many of the highly represented genes that are likely important in defining the ecological niches of *Synechococcus* populations were poorly studied hypothetical genes (Tables 3.2 and 3.3). Additionally, the reasons for the variation in read coverage, such as gene duplication, horizontal gene transfer, or the presence of phage genomes containing *Synechococcus* genes, need to be examined further.

Rare genetic components may also have a role in defining the ecological niche of a population because they provide unique functions. Unfortunately, metagenomics is not well suited for investigating the biology of these genes. Obtaining increased read coverage of these components, especially from a diverse community, would require exponentially greater sequencing effort. These regions may be better analyzed by targeted sequencing efforts rather than through metagenomics (see for eg. Appendix 1).

Selection in the evolution of *Synechococcus*

The majority of the genes from the two populations of *Synechococcus* have evolved under purifying selection. Thus, most of the observed polymorphisms were selectively neutral (or nearly neutral). By comparing genome sequences between strains of *Synechococcus* from different clades, at this scale of taxonomy the dN/dS ratios of core genes were also well below 1 (Hu and Blanchard, 2009). Purifying selection is the dominant mode of selection in other microbial populations as well (Allen *et al.*, 2007; Simmons *et al.*, 2008).

Purifying selection for slightly deleterious mutations is more efficient when N_e is large (Ohta, 1976). For *Synechococcus*, the assumption that N_e is large is thought to be appropriate given the vast number of these microorganisms in the coastal Southern California Bight. Both clade I and clade IV *Synechococcus* can reach abundances of 10^5/ml (Tai and Palenik, 2009). Given a conservative estimate of 10^3/ml over the entire region (~170

000 km^2), their population sizes are on the order of 10^{16}. But how the census population size compares to N_e is not clear (Frankham, 1995).

Many aspects of microbial population genetics, especially of environmental populations are very poorly understood, including the mutation rate, the selection coefficients on different types of mutations, the rate of migration, and N_e for different microbial populations (Mes, 2008). Even the ecological and taxonomic scales that define a microbial population and thus N_e are controversial (Lynch and Conery, 2003; Daubin and Moran, 2004; Lynch and Conery, 2004). In addition, the combined effect of generation time, mutation rate, and rate of environmental change on the diversity of a population are not clearly understood (Ohta and Gillespie, 1996). These parameters will be necessary to obtain an accurate portrayal of the forces that shape microbial population diversity and evolution.

Positive selection in *Synechococcus* populations

A small number of genes had dN/dS ratios > 1 suggesting that they evolved under positive selection. These genes are especially interesting as they may provide clues to adaptive phenotypes in *Synechococcus*. Many of these genes, however, are covered poorly by metagenome reads. Even though the distribution of dN/dS ratios was not greatly affected by the depth of read coverage (Figure 3.5), because positive selection has important implications, all of these observations should be confirmed with greater sequencing depth.

The majority of genes with dN/dS > 1 encode small hypothetical proteins (< 100 amino acids) with no known homologs and a probable function could not be determined from bioinformatics analyses alone. For the genes homologous to CC9311, the genes do appear to be differentially expressed, so they are likely functional (Table 3.4). A few genes with high dN/dS ratios have annotated functions (Tables 3.4 and 3.5). However, with the exception of the ribosomal protein L35, these genes all suffer from low sequence coverage. A putative lipoprotein from the clade I population had a dN/dS > 1 but only relative to the CC9311 genome. The high dN/dS for this gene relative to CC9311 may represent positive selection in either the environmental population or the cultured strain.

The ribosomal protein L35 is one of the smallest proteins associated with the large subunit (23S) of the ribosome. Four non-synonymous polymorphisms were observed, histidine (positively charged) to glutamine (polar), histidine to lysine (positively charged), proline (nonpolar) to alanine (nonpolar), and methionine (nonpolar) to isoleucine (nonpolar). L35 has not been studied as intensively as other ribosomal proteins probably due to its small size, so the effect of these polymorphisms is not known (Klein *et al.*, 2004). L35 is part of a dicistronic operon with L20, also protein of the ribosomal large subunit. The expression of L35 and L20 is auto-regulated by L20 which acts as a repressor protein that binds upstream of the L35 gene (Haentjens-Sitri *et al.*, 2008). Many ribosomal proteins, including L20 but not L35, may have been transferred amongst the *Synechococcus/Prochlorococcus* group (Zhaxybayeva *et al.*, 2009). As L35 and L20 form an operon, if the ribosomal proteins are being transferred separately, then L35 may be diversifying to adapt to L20 genes originating from a different genetic environment.

Conclusions

Population genetics provides a theoretical framework to understand diversification, adaptation, and evolution. While the population genetics of microorganisms has long been studied experimentally with cultured isolates, environmental metagenomics provides a tremendous resource of data from native microbial communities that is ripe for microbial population genetics analysis. These new sources of data, however, require new analytical methodologies. Using 454 sequences, the analysis described here provided the first insight into the role of selection in the evolution of *Synechococcus*, a globally significant primary producer.

As more metagenomes become available, more opportunities arise to investigate the factors that have given rise to the observed diversity. Further methods are in need of development, however, to take advantage of these data and derive fundamental evolutionary parameters, such as mutation, recombination, and migration rates, for understanding microbial evolution. Hopefully, as these analytical challenges are met, we will gain a better understanding of the ecological and evolutionary scales under which microbes adapt.

Acknowledgements

A National Science Foundation grant (UCSD 20072923) supported this project and the graduate research of the dissertation author. Many thanks to Ian Paulsen, Andy Allen, and the staff at the J. Craig Venter Institute (JCVI) for sequencing and analytical support. Thanks as well to Alan Saluk, Cody Fine, and Alex Ilic from The Scripps Research Institute (TSRI) Flow Cytometry Core Facility.

Table 3.1. Temperature, chlorophyll, and nutrient concentrations of the seawater collected for flow cytometry sorting and metagenome sequencing.

Metagenome	Sample	temp (°C)	chlorophyll (whole) (mg/m^3)	chlorophyll (< 2 μm) (mg/m^3)	nitrate (NO^{3-}) (μM)	nitrite (NO^{2-}) (μM)	ammonium (NH^{4+}) (μM)	phosphate (PO^{4-}) (μM)	silicate (μM)
1	10/10/06	17.8	0.63	0.33	0.19	0.27	3.1	0.19	0.32
2	5/17/07	17.0	20.28	3.3	0.43	0.09	5.1	0.5	1.15
3	3/6/08	15.0	1.82	0.47	0.61	0.35	4.05	0.05	0.32
4	4/17/08	15.0	6.7	0.58	1.07	0.39	5.08	0.07	0.16
5	5/15/08 20 m	12.5	0.94	0.43	16.04	1.49	17.90	0.25	0.73
6	5/15/08 surface	19.0	6.23	0.98	0.00	0.05	1.11	0.03	0.15

Table 3.2. dN/dS ratios of genes that are “unique” to CC9311 with greater than 20X metagenome sequence coverage. A gene that is “unique” to CC9311 was based on the absence of reciprocal BLAST hits with *Synechococcus* strains WH8102, CC9605, and CC9902 prior to the availability of the BL107, WH7803, WH7805, RS9916, RS9917, and RCC307 genomes. 20X coverage is approximately 1 standard deviation greater than the mean depth of coverage per gene. The taxon containing the best tBLASTn hit with an e-value < 0.0001 is shown. Predicted folds using SCOP and HHpred searches are listed. For these searches, ns indicates that a significant hit was not found and x indicates the search was not performed because a probable function has already been determined. Y indicates if the gene showed significant gene expression changes under one or more of the stress conditions tested in microarray experiments (Palenik, Stuart, et al. unpublished data). N indicates that none of the stress conditions resulted in a significant change in gene expression. * indicates that the annotation of the gene was changed from “hypothetical protein” to “conserved hypothetical protein” because a significant tBLASTn hit to an ORF in another cyanobacterium was found.

gene	description	dN/dS	dN/dS relative to CC9311	% of gene covered 5X /avg. read depth	best tBLASTn hit	SCOP or HHpred predicted fold	microarray gene expression changes	
							increased	decreased
sync_0494	hypothetical protein	0.983	0.889	100/37.1	*Synechococcus* sp.	ns	N	Y
sync_0573	two-component hybrid sensor and regulator alr4878	0.184	0.188	100/29.9	*Nostoc* PCC7120	x	N	Y
sync_0574	Two-component response regulator, CheY-like receiver and wHTH DNA-binding domains	0.233	0.226	100/30.6	*Prochlorococcus* SS120	x	N	Y
sync_0625	hypothetical protein	0.241	0.252	100/26.9	ns	ns	Y	N
sync_0649	Glycerol dehydrogenase (GLDH)	0.195	0.185	100/24.1	*Synechococcus* RCC307	x	Y	N
sync_0650	conserved hypothetical protein*	0.396	0.369	100/23.3	*Synechococcus* CC9605	ns	Y	N
sync_0673	hypothetical	0.278	0.292	100/22.1	*Prochlorococcus* MIT9303	ns	Y	N
sync_0679	TM2 domain family	0.103	0.093	100/24.2	*Microcystis aeruginosa* NIES-843	x	Y	Y
sync_0681	ferrous iron transport protein B	0.071	0.067	100/21.9	*Synechococcus* PCC6803	x	N	Y
sync_0682	ferrous iron transport protein A	0.093	0.109	100/21.5	*Synechococcus* PCC6803	x	N	N
sync_0683	hypothetical protein	0.179	0.198	100/20.8	ns	ns	Y	N
sync_0684	membrane protein, putative	0.154	0.155	100/21.1	*Methylobacterium extorquens* AM1 plasmid p1META1	ns	Y	Y

Table 3.2. dN/dS ratios “unique” to CC9311 genes, continued.

gene	description	dN/dS	dN/dS relative to CC9311	% of gene covered 5X /avg. read depth	best tBLASTn hit	SCOP or HHpred predicted fold	microarray gene expression changes	
							increased	decreased
sync_0685	hypothetical protein	0.150	0.144	100/30.1	ns	ns	Y	Y
sync_0710	conserved hypothetical protein*	0.469	0.482	100/21.8	*Synechococcus* WH7803	ns	N	N
sync_0830	conserved hypothetical protein*	0.280	0.282	100/21.8	*Synechococcus* WH7803	ns	N	Y
sync_0839	conserved hypothetical protein*	0.495	0.477	66/20.9	*Synechococcus* WH7803	ns	N	Y
sync_0862	hypothetical protein	0.193	0.188	100/38.0	ns	ns	N	Y
sync_0867	conserved hypothetical protein*	0.050	0.066	100/26.8	*Synechococcus* RCC307	ns	Y	Y
sync_0868	conserved hypothetical protein*	0.430	0.465	86/21.4	*Synechococcus* CC9605	ns	Y	Y
sync_0871	GrpE	0.113	0.112	100/65.4	ns	x	N	
sync_0877	hypothetical protein	0.170	0.149	100/22.9	ns	ns	N	Y
sync_0949	hypothetical protein	0.161	0.161	100/31.9	ns	ns	Y	N
sync_0951	hypothetical protein	1.046	1.056	100/21.7	ns	ns	Y	N
sync_0957	hypothetical protein	0.113	0.112	100/26.7	ns	ns	Y	N
sync_0958	hypothetical protein	0.156	0.159	100/28.6	ns	ns	Y	N
sync_1041	putative RND family outer membrane efflux protein	0.169	0.179	100/21.8	*Synechococcus* WH7803	x	Y	N
sync_1059	transporter, bile acid/Na+ symporter family	0.175	0.194	100/30.1	*Synechococcus* WH7803	x	Y	N
sync_1060	hypothetical protein	0.296	0.296	100/38.5	ns	ns	Y	N
sync_1061	hypothetical protein	0.219	0.244	100/38.3	ns	ns	Y	Y
sync_1066	conserved hypothetical protein*	0.079	0.098	100/29.0	*Synechococcus* CC9902	ns	Y	N
sync_1112	hypothetical protein	0.261	0.271	95/20.1	ns	ns	Y	N
sync_1155	hypothetical protein	1.021	1.172	100/25.8	ns	ns	Y	Y
sync_1156	conserved hypothetical protein*	0.165	0.163	100/27.6	*Synechococcus* CC9902	ns	Y	Y
sync_1174	RecF/RecN/SMC N terminal domain protein	0.239	0.250	100/26.7	*Synechococcus* WH7803	x	Y	Y
sync_1237	conserved hypothetical protein*	0.390	0.377	100/31.1	*Synechococcus* CC9902	ns	Y	N
sync_1244	oxidoreductase, short-chain dehydrogenase/reductase family	0.198	0.190	100/26.8	*Synechococcus* RCC307	x	N	Y
sync_1253	glutamine synthetase	0.062	0.062	100/33.8	*Synechococcus* WH7803	x	Y	Y
sync_1257	conserved hypothetical protein*	0.160	0.129	100/34.7	*Synechococcus* RCC307	ns	N	Y
sync_1259	conserved hypothetical protein*	0.147	0.156	100/26.6	*Synechococcus* WH7803	ns	N	Y
sync_1291	conserved hypothetical protein	0.096	0.098	100/61.9	*Synechococcus* WH7803	ns	N	Y
sync_1292	hypothetical protein	0.381	0.345	100/28.8	ns	ns	Y	N

Table 3.2. dN/dS ratios “unique” to CC9311 genes, continued.

gene	description	dN/dS	dN/dS relative to CC9311	% of gene covered 5X /avg. read depth	best tBLASTn hit	SCOP or HHpred predicted fold	microarray gene expression changes	
							increased	decreased
sync_1421	hypothetical protein	0.103	0.105	100/22.0	ns	ns	N	Y
sync_1424	transporter, major facilitator family	0.079	0.081	100/27.8	*Prochlorococcus* MIT9211	x	Y	N
sync_1449	similar to flotillin	0.074	0.080	100/37.2	*Synechococcus* PCC6301	x	N	Y
sync_1464	conserved hypothetical protein	0.108	0.110	100/21.3	*Synechococcus* CC9902	ns	N	Y
sync_1515	long-chain-fatty-acid-CoA ligase	0.275	0.273	100/24.7	*Solibacter usitatus* Ellin6076	x	Y	N
sync_1568	hypothetical protein	0.727	0.961	100/20.0	ns	ns	Y	Y
sync_1666	hypothetical protein	0.978	0.951	100/20.5	ns	ns	Y	N
sync_1740	hypothetical protein	0.015	0.015	100/24.2	ns	ns	N	Y
sync_1744	conserved hypothetical protein*	0.334	0.345	100/37.7	*Synechococcus* WH7803	ns	Y	N
sync_1745	possible tRNA synthetases class I	0.218	0.238	81/22.1	*Prochlorococcus* MIT9313	x	Y	N
sync_1762	conserved hypothetical protein*	0.057	0.068	100/27.3	*Synechococcus* WH7803	ns	Y	N
sync_1772	conserved hypothetical protein	0.231	0.246	100/28.4	*Synechococcus* WH7803	alpha/beta fold family of hydrolases	Y	Y
sync_1773	conserved hypothetical protein*	0.120	0.105	100/25.7	*Synechococcus* CC9902	ns	N	Y
sync_1777	conserved hypothetical protein*	0.205	0.217	100/36.0	*Synechococcus* CC9605	ns	Y	Y
sync_1778	hypothetical protein	0.934	1.070	100/47.7	ns	ns	N	Y
sync_1782	hypothetical protein	3.235	4.079	100/35.3	ns	ns	N	Y
sync_1783	hypothetical protein	0.256	0.244	100/59.0	ns	ns	Y	Y
sync_1789	conserved hypothetical protein*	0.173	0.177	100/31.3	*Synechococcus* WH8102	ns	N	Y
sync_1792	conserved hypothetical-related protein	0.149	0.132	100/40.1	*Synechococcus* WH7803	ns	N	Y
sync_1799	hypothetical protein	0.440	0.459	100/20.0	ns	ns	Y	N
sync_1806	conserved hypothetical protein	0.191	0.204	100/28.5	ns	ns	N	Y
sync_1826	hypothetical protein	0.139	0.151	100/23.4	ns	ns	N	Y
sync_1830	Predicted enzyme of the cupin superfamily	0.235	0.222	100/20.9	*Prochlorococcus* MIT9313	x	N	Y
sync_1835	aldehyde dehydrogenase family protein	0.182	0.186	100/27.0	*Synechococcus* RCC307	x	Y	N

Table 3.2. dN/dS ratios “unique” to CC9311 genes, continued.

gene	description	dN/dS	dN/dS relative to CC9311	% of gene covered 5X /avg. read depth	best tBLASTn hit	SCOP or HHpred predicted fold	microarray gene expression changes	
							increased	decreased
sync_1872	possible Arginine repressor, C-terminal domain	0.240	0.234	100/29.0	*Synechococcus* WH7803	x	N	N
sync_1873	conserved hypothetical protein	0.084	0.090	100/24.2	*Synechococcus* WH7803	ns	N	Y
sync_1874	conserved hypothetical protein	0.143	0.140	100/27.7	*Synechococcus* WH7803	ns	N	Y
sync_1875	putative membrane protein	0.063	0.060	100/35.5	*Synechococcus* WH7803	x	N	Y
sync_1915	dehydrogenase subunit-like protein	0.074	0.073	100/20.3	*Synechococcus* WH7803	x	N	Y
sync_1916	hypothetical conserved protein	0.125	0.111	100/24.4	*Synechococcus* WH7803	ns	Y	Y
sync_1918	Glucose 1-dehydrogenase.	0.074	0.071	99/29.8	*Synechococcus* WH7803	x	N	Y
sync_2059	conserved hypothetical protein*	0.217	0.227	100/27.4	*Synechococcus* WH7803	ns	N	Y
sync_2063	hypothetical-related protein	0.162	0.189	80/21.5	*Synechococcus* CC9902	ns	Y	N
sync_2064	hypothetical protein	0.190	0.189	100/48.0	ns	ns	Y	Y
sync_2066	hypothetical protein	0.467	0.419	100/38.0	ns	ns	N	Y
sync_2071	conserved hypothetical protein*	0.180	0.178	100/23.4	*Synechococcus* CC9902	ns	N	N
sync_2076	hypothetical protein	1.370	1.215	82/23.4		ns	Y	Y
sync_2077	conserved hypothetical protein	0.130	0.138	100/41.6	*Synechococcus* WH8102	ns	Y	N
sync_2078	conserved hypothetical protein*	0.225	0.223	100/38.6	*Synechococcus* CC9605	ns	N	Y
sync_2238	hypothetical protein	0.135	0.132	100/31.6	ns	ns	Y	Y
sync_2297	hypothetical protein	0.229	0.252	100/29.7	ns	ns	N	Y
sync_2300	hypothetical protein	0.079	0.082	100/27.3	ns	ns	N	Y
sync_2305	hypothetical protein	0.571	0.613	100/39.6	ns	ns	Y	N
sync_2408	conserved hypothetical protein*	0.133	0.147	100/29.9	*Synechococcus* CC9605	ns	Y	N
sync_2409	ParB nuclease domain protein	0.197	0.207	100/30.4	*Nostoc* PCC73102	x	N	Y
sync_2410	conserved hypothetical protein*	0.587	0.575	88/22.1	*Synechococcus* CC9902	ns	N	Y
sync_2816	conserved hypothetical protein*	0.383	0.409	100/30.1	*Synechococcus* CC9605	protein kinase-like	Y	N
sync_2817	conserved hypothetical protein*	0.437	0.469	100/49.0	*Synechococcus* WH7803	periplasmic binding protein-like	Y	N
sync_2937	hypothetical protein	0.908	0.711	100/30.6	ns	ns	Y	N

Table 3.3. dN/dS ratios of genes that are “unique” to CC9902 with greater than 20X metagenome sequence coverage. A gene that is “unique” to CC9902 was based on the absence of reciprocal BLAST hits with *Synechococcus* strains WH8102, CC9605, and CC9311 prior to the availability of the BL107, WH7803, WH7805, RS9916, RS9917, and RCC307 genomes. 20X coverage is approximately 1 standard deviation greater than the mean depth of coverage per gene. The taxon containing the best tBLASTn hit with an e-value < 0.0001 is shown. Predicted folds using SCOP and HHpred searches were not found. For these searches, ns indicates that a significant hit was not found and x indicates the search was not performed because a probable function has already been determined. * indicates that the annotation of the gene was changed from “hypothetical protein” to “conserved hypothetical protein” because a significant tBLASTn hit to an ORF in another cyanobacterium was found.

gene	description	dN/dS	dN/dS relative to CC9902	% of gene covered 5X /avg. read depth	best tBLASTn hit	SCOP or HHpred predicted fold
Syncc9902_0286	conserved hypothetical protein*	0.298	0.331	100/20.0	*Synechococcus* WH8102	ns
Syncc9902_0441	hypothetical protein	0.619	0.657	100/27.1	ns	ns
Syncc9902_0450	hypothetical protein	1.093	1.059	100/60.2	ns	ns
Syncc9902_0509	hypothetical protein	0.219	0.218	100/21.9	ns	ns
Syncc9902_0810	conserved hypothetical protein*	0.370	0.326	100/23.4	*Synechococcus* WH8102	ns
Syncc9902_0968	hypothetical protein	0.335	0.367	100/25.5	ns	ns
Syncc9902_0981	conserved hypothetical protein*	0.175	0.169	100/29.5	*Synechococcus* WH8102	ns
Syncc9902_1022	hypothetical protein	0.177	0.179	100/24.0	ns	ns
Syncc9902_1024	hypothetical protein	0.253	0.250	100/31.5	ns	ns
Syncc9902_1027	hypothetical protein	0.350	0.371	100/28.4	ns	ns
Syncc9902_1032	hypothetical protein	0.421	0.424	100/26.6	ns	ns
Syncc9902_1033	hypothetical protein	0.137	0.140	100/52.2	ns	ns
Syncc9902_1035	conserved hypothetical protein*	0.411	0.403	100/29.0	*Synechococcus* CC9605	ns
Syncc9902_1174	hypothetical protein	2.962	2.962	100/22.3	ns	ns
Syncc9902_1352	oxidoreductase	0.097	0.095	100/28.4	*Synechococcus* WH7803	x
Syncc9902_1464	membrane protein-like	0.149	0.158	100/21.4	*Prochlorococcus* MIT9303	x
Syncc9902_1467	hypothetical protein	0.272	0.262	100/21.1	ns	ns
Syncc9902_1557	conserved hypothetical protein*	0.185	0.182	100/25.8	*Prochlorococcus* MIT9303	ns
Syncc9902_1598	hypothetical protein	0.542	0.567	100/25.2	ns	ns
Syncc9902_1801	conserved hypothetical protein*	1.156	1.181	100/28.2	*Synechococcus* CC9605	ns
Syncc9902_1805	conserved hypothetical protein*	0.194	0.203	100/21.4	*Synechococcus* CC9311	ns
Syncc9902_1889	conserved hypothetical protein*	0.135	0.141	100/22.2	*Synechococcus* RCC307	ns

Table 3.3. dN/dS ratios “unique” to CC9902 genes, continued.

gene	description	dN/dS	dN/dS relative to CC9902	% of gene covered 5X /avg. read depth	best tBLASTn hit	SCOP or HHpred predicted fold
Syncc9902_1899	hypothetical protein	0.319	0.296	100/25.6	*Synechococcus* sp.	ns
Syncc9902_2008	hypothetical protein	1.356	1.270	100/37.1	ns	ns
Syncc9902_2196	transcriptional regulator, AraC family	0.309	0.321	100/28.7	*Synechococcus* CC9311	ns

Table 3.4. Genes with dN/dS ratios > 1 based on polymorphisms from metagenomic sequences tiled to the CC9311 genome. Inf indicates genes that had non-synonymous polymorphisms but no synonymous polymorphisms resulting in an infinite value for dN/dS. The last 7 genes have dN/dS < 1 relative to the consensus sequence, but have dN/dS > 1 relative to the CC9311 genome. Whether the gene is part of the core (C) or accessory (A) genomes of *Synechococcus* is provided. The taxon containing the best tBLASTn hit with an e-value < 0.0001 is shown. Predicted folds for hypothetical genes using SCOP and HHpred searches are listed. For these searches, ns indicates that a significant hit was not found and x indicates the search was not performed because a probable function has already been determined. Y indicates if the gene showed significant gene expression changes under one or more of the stress conditions tested in microarray experiments (Palenik, Stuart, et al. unpublished data). N indicates that none of the stress conditions resulted in a significant change in gene expression. * indicates that the annotation of the gene was changed from "hypothetical protein" to "conserved hypothetical protein" because a significant tBLASTn hit to an ORF in another cyanobacterium was found. $^{\#}$ indicates genes with homologues in the CC9902 genome, but the metagenome tiling to these CC9902 genes results in dN/dS < 1. ^ sync_0523 best hits syncc9902_1875 from the CC9902 genome which has a dN/dS = 1.192 (Table 3.5).

gene	description	dN/dS	dN/dS relative to CC9311	% of gene covered 5X /avg. read depth	core (C) or accessory (A)	best tBLASTn hit	SCOP or HHpred predicted fold	microarray gene expression changes	
								increased	decreased
sync_1519	Breast cancer type 1 susceptibility protein homolog	1.018	1.051	40/4.4	A	Uncultured cyanobacterium	x	N	N
sync_1155	hypothetical protein	1.021	1.172	100/25.8	A	ns	ns	Y	Y
sync_1164	hypothetical protein	1.033	1.223	82/6.9	A	ns	ns	Y	N
sync_0951	hypothetical protein	1.046	1.056	100/21.7	A	ns	ns	N	Y
sync_2406	hypothetical protein	1.050	0.947	80/10.7	A	ns	ns	Y	N
sync_1822	hypothetical protein	1.061	1.207	90/8.1	A	ns	ns	Y	N
sync_1128	hypothetical protein	1.071	1.076	38/17.6	A	ns	ns	Y	Y
sync_2849	hypothetical protein	1.153	1.674	57/4.2	A	ns	ns	N	Y
sync_1648	hypothetical protein	1.159	1.367	100/11.0	A	ns	ns	Y	N
sync_2944	hypothetical protein	1.167	1.109	45/3.4	A	ns	ns	N	Y
sync_0645	hypothetical protein	1.177	1.049	81/9.1	A	ns	ns	Y	N
sync_2175	hypothetical protein	1.179	1.300	47/2.9	A	ns	ns	N	Y
sync_2834	hypothetical protein	1.235	0.951	57/5.2	A	ns	ns	N	Y

Table 3.4. CC9311 genes with dN/dS ratios > 1, continued.

gene	description	dN/dS	dN/dS relative to CC9311	% of gene covered 5X /avg. read depth	core (C) or accessory (A)	best tBLASTn hit	SCOP or HHpred predicted fold	microarray gene expression changes	
								increased	decreased
sync_2810	hypothetical protein	1.248	1.576	81/9.7	A	ns	ns	N	Y
sync_0864	hypothetical protein	1.273	1.574	38/7.2	A	ns	ns	N	Y
sync_2186	hypothetical protein	1.314	0.678	100/6.5	A	ns	ns	N	Y
sync_1049	hypothetical protein	1.335	1.231	68/8.1	A	ns	ns	N	N
sync_2435	conserved hypothetical protein*	1.356	1.356	39/3.6	A	*Prochorococcus* MIT9313	ns	Y	N
sync_2076	hypothetical protein	1.370	1.215	82/23.4	A	ns	ns	Y	Y
sync_0552	hypothetical protein	1.397	1.397	44/4.2	A	ns	ns	Y	N
sync_2383	hypothetical protein	1.427	0.803	22/1.8	A	ns	ns	Y	Y
sync_1859	hypothetical protein	1.482	1.582	100/18.5	A	ns	ns	N	N
sync_1481	hypothetical protein	1.500	1.312	6/1.0	A	ns	ns	Y	N
sync_1947	hypothetical protein	1.513	1.513	11/2.7	A	ns	ns	N	Y
sync_2428	hypothetical protein	1.533	1.287	98/7.6	A	ns	ns	N	Y
sync_2278	hypothetical protein	1.552	1.552	15/2.6	A	ns	ns	Y	Y
sync_1402#	ABC transporter for sugars, solute-binding protein	1.591	1.512	2/1.9	A	*Synechococcus* WH7803	x	N	Y
sync_1647	hypothetical protein	1.736	1.067	33/3.7	A	ns	ns	Y	N
sync_1791	hypothetical protein	1.738	1.863	100/18.9	A	ns	ns	Y	Y
sync_1348	hypothetical protein	1.759	1.874	100/11.8	A	ns	ns	Y	Y
sync_0588	hypothetical protein	1.774	1.398	100/7.8	A	ns	ns	N	Y
sync_2178	hypothetical protein	1.826	1.490	100/15.2	A	ns	ns	N	Y
sync_0523^	conserved hypothetical protein*	1.923	1.571	84/8.2	A	*Synechococcus* CC9902	ns	N	Y
sync_1626	hypothetical protein	1.944	2.208	55/3.9	A	ns	ns	N	Y
sync_1047	hypothetical protein	1.959	1.316	63/6.7	A	ns	ns	N	Y
sync_1761	hypothetical protein	2.013	1.452	9310.8	A	ns	ns	Y	N
sync_2833	hypothetical protein	2.046	3.161	100/14.9	A	ns	ns	N	Y
sync_0324	hypothetical protein	2.225	2.309	30/3.8	A	ns	ns	Y	Y
sync_2273	hypothetical protein	2.259	2.198	100/17.0	A	ns	ns	N	Y
sync_1784	hypothetical protein	2.329	1.390	93/6.7	A	ns	ns	Y	N
sync_2115	hypothetical protein	2.474	2.474	100/18.4	A	ns	ns	Y	Y
sync_2912	hypothetical protein	2.545	1.622	13/2.3	A	ns	ns	N	Y
sync_0585	hypothetical protein	2.752	1.174	28/4.0	A	ns	ns	N	Y
sync_1828	hypothetical protein	2.79	2.097	74/8.6	A	ns	ns	N	N

Table 3.4. CC9311 genes with dN/dS ratios > 1, continued.

gene	description	dN/dS	dN/dS relative to CC9311	% of gene covered 5X /avg. read depth	core (C) or accessory (A)	best tBLASTn hit	SCOP or HHpred predicted fold	microarray gene expression changes	
								increased	decreased
sync_2359	hypothetical protein	2.844	1.232	81/5.4	A	ns	ns	N	Y
sync_1472	possible transcriptionsl regulator protein	2.916	3.238	4/0.9	A	*Prochlorococcus* MIT9313	x	Y	Y
sync_1782	hypothetical protein	3.235	4.079	100/35.3	A	ns	ns	N	Y
sync_1085	hypothetical protein	3.428	2.368	42/4.5	A	ns	ns	N	N
sync_1306	hypothetical protein	3.678	5.884	100/11.4	A	ns	ns	N	Y
sync_2079	hypothetical protein	3.818	1.364	21/2.8	A	ns	ns	N	Y
sync_0800	hypothetical protein	5.509	2.981	100/12.6	A	ns	ns	Y	N
sync_1796	hypothetical protein	7.558	2.105	100/13.0	A	ns	ns	N	N
sync_0059$^{\#}$	ribosomal protein L35	Inf	Inf	72/8.1	C	*Synechococcus* CC9605	x	N	N
sync_0536	photosystem II reaction center protein PsbM	Inf	Inf	43/4.4	C	*Synechococcus* WH7803	x	N	Y
sync_0637	conserved hypothetical protein*	Inf	Inf	0.5/0.5	A	*Prochorococcus* MIT9313	ns	N	Y
sync_0780	hypothetical protein	Inf	Inf	32/3.8	A	ns	ns	Y	N
sync_0838	conserved hypothetical protein*	Inf	Inf	3/1.0	A	*Synechococcus* CC9605	ns	N	Y
sync_0889$^{\#}$	conserved hypothetical protein*	Inf	Inf	26/3.7	C	*Synechococcus* CC9605	ns	N	Y
sync_0925	hypothetical protein	Inf	Inf	6/3.0	A	ns	ns	N	N
sync_1021	hypothetical protein	Inf	Inf	7/1.1	A	ns	ns	Y	N
sync_1490	4-methyl-5(B-hydroxyethyl)-thiazole monophosphate biosynthesis enzyme	Inf	Inf	29/2.9	A	*Theileria annulata*	x	N	Y
sync_1603	hypothetical protein	Inf	Inf	16/1.3	A	ns	ns	N	Y
sync_2225	hypothetical protein	Inf	0.444	16/1.4	A	ns	ns	Y	N
sync_2247	hypothetical protein	Inf	Inf	100/9.9	A	ns	ns	N	Y
sync_2506	hypothetical protein	Inf	Inf	17/3.7	A	ns	ns	Y	Y
sync_2513	hypothetical protein	Inf	Inf	19/2.4	A	ns	ns	Y	Y
sync_2649$^{\#}$	conserved hypothetical protein*	Inf	Inf	0.5/3.0	C	*Synechococcus* WH8102	ns	Y	N
sync_2902	hypothetical protein	Inf	Inf	98/7.9	A	ns	ns	N	Y

Table 3.4. CC9311 genes with dN/dS ratios > 1, continued.

gene	description	dN/dS	dN/dS relative to CC9311	% of gene covered 5X /avg. read depth	core (C) or accessory (A)	best tBLASTn hit	SCOP or HHpred predicted fold	microarray gene expression changes	
								increased	decreased
sync_2590	hypothetical protein	0.605	1.004	100/15.6	A	ns	ns	Y	N
sync_0002	conserved hypothetical protein*	0.980	1.005	11.4/1.5	C	*Synechococcus* WH7803	PRC-barrel domain	N	Y
sync_1778	hypothetical protein	0.935	1.070	100/47.7	A	ns	ns	N	Y
sync_1255	hypothetical protein	0.634	1.147	53/5.6	A	ns	ns	N	Y
sync_2070	hypothetical protein	0.923	1.466	87/10.6	A	ns	ns	Y	Y
sync_2477	hypothetical protein	0.927	1.849	64/4.2	A	ns	ns	Y	Y
sync_2711	lipoprotein, putative	0.303	2.878	75/5.4	A	ns	ns	N	N

Table 3.5. Genes with dN/dS ratios > 1 based on polymorphisms from metagenomic sequences tiled to the CC9902 genome. Inf indicates genes that had non-synonymous polymorphisms but no synonymous polymorphisms resulting in an infinite value for dN/dS. Whether the gene is part of the core (C) or accessory (A) genomes of *Synechococcus* is provided. The taxon containing the best tBLASTn hit with an e-value < 0.0001 is shown. Predicted folds using SCOP and HHpred searches are listed. For these searches, ns indicates that a significant hit was not found and x indicates the search was not performed because a probable function has already been determined. * indicates that the annotation of the gene was changed from "hypothetical protein" to "conserved hypothetical protein" because a significant tBLASTn hit to an ORF in another cyanobacterium was found. # indicates genes with homologues in the CC9311 genome, but the metagenome tiling to these CC9311 genes results in dN/dS < 1. ^ Syncc9902_1875 best hit sync_0523 from the CC9311 genome which has a dN/dS = 1.923 (Table 3.3). $ Syncc9902_0240 best hit sync_1115 with dN/dS = 0/0, i.e. no polymorphisms observed (Table A5.1). There were no additional genes with dN/dS > 1 when calculated relative to the CC9902 genome rather than the consensus sequence.

gene	description	dN/dS	dN/dS relative to CC9902	% of gene covered 5X /avg. read depth	core (C) or accessory (A)	best tBLASTn hit	SCOP or HHpred predicted fold
Syncc9902_0450	hypothetical protein	1.093	1.059	100/60.2	A	ns	ns
Syncc9902_1801	conserved hypothetical protein*	1.156	1.181	100/28.2	A	*Synechococcus* CC9605	ns
Syncc9902_1875^	conserved hypothetical protein*	1.192	1.220	96/7.9	A	*Synechococcus* CC9311	ns
Syncc9902_2037	conserved hypothetical protein*	1.206	1.206	5/1.7	A	*Synechococcus* WH8102	UDP glycosyltransferase-like
Syncc9902_2008	hypothetical protein	1.356	1.270	100/37.1	A	ns	ns
Syncc9902_0240$	conserved hypothetical protein*	1.656	1.656	32/3.5	A	*Synechococcus* CC9311	RmlC-like cupin fold
Syncc9902_1954	hypothetical protein	2.112	2.139	89/7.4	A	ns	ns
Syncc9902_1026	hypothetical protein	2.430	1.973	100/12.6	A	ns	ns
Syncc9902_1874	hypothetical protein	2.493	2.901	23/3.8	A	ns	ns
Syncc9902_1174	hypothetical protein	2.962	2.962	100/22.3	A	ns	ns
Syncc9902_0923#	conserved hypothetical protein*	Inf	Inf	20/3.4	C	*Synechococcus* CC9605	ns
Syncc9902_1576#	Methylated-DNA-(protein)-cysteine S-methyltransferase	Inf	Inf	9/3.5	C	*Synechococcus* CC9605	x

Table 3.5. CC9902 genes with dN/dS ratios > 1, continued.

gene	description	dN/dS	dN/dS relative to CC9902	% of gene covered 5X /avg. read depth	core (C) or accessory (A)	best tBLASTn hit	SCOP or HHpred predicted fold
Syncc9902_2149#	hypothetical protein	Inf	Inf	8/2.0	A	*Desulfotalea psychrophila* LSv54	ns
Syncc9902_2225	hypothetical protein	Inf	Inf	4/1.2	A	ns	ns
Syncc9902_2241#	amino acid permease family protein	Inf	Inf	15/2.0	A	*Prochlorococcus* MIT9303	x

Table 3.6. Genes homologous between CC9311 and CC9902 with dN/dS ratios > 1. Inf indicates genes that had non-synonymous polymorphisms but no synonymous polymorphisms resulting in an infinite value for dN/dS. NaN indicates that no polymorphisms were observed so a dN/dS ratio could not be calculated.

Description	CC9311 (clade I) gene	Clade I dN/dS	% of gene covered 5X /avg. read depth	CC9902 (clade IV) gene	Clade IV dN/dS	% of gene covered 5X /avg. read depth
hypothetical protein	sync_2649	Inf	0.5/3.0	Syncc9902_2114	0.236	100/11.9
Photosystem II reaction center M protein (PSII-M)-related protein	sync_0536	Inf	43/4.4	Syncc9902_1862	0.147	100/7.6
ribosomal protein L35	sync_0059	Inf	72/8.1	Syncc9902_0055	0.14	100/9.5
conserved hypothetical protein	sync_0889	Inf	26/3.7	Syncc9902_1688	0.127	100/11.4
ABC transporter for sugars, solute-binding protein	sync_1402	1.591	2/1.9	Syncc9902_1078	0.184	100/18.5
hypothetical protein	sync_2716	0.195	50/4.4	Syncc9902_2149	Inf	8/2
putative methylated-DNA-[protein]-cysteine S-methyltransferase	sync_0693	0.192	100/17.4	Syncc9902_1576	Inf	9/3.5
amino acid permease family protein	sync_2867	0.140	100/17.4	Syncc9902_2241	Inf	15/2.0
Uncharacterized conserved protein	sync_1882	0.137	97/9.6	Syncc9902_0923	Inf	20/3.4
hypothetical protein	sync_0523	1.923	83/8.2	Syncc9902_1875	1.192	96/7.9
hypothetical protein	sync_1115	NaN	0/0.3	Syncc9902_0240	1.656	32/3.4

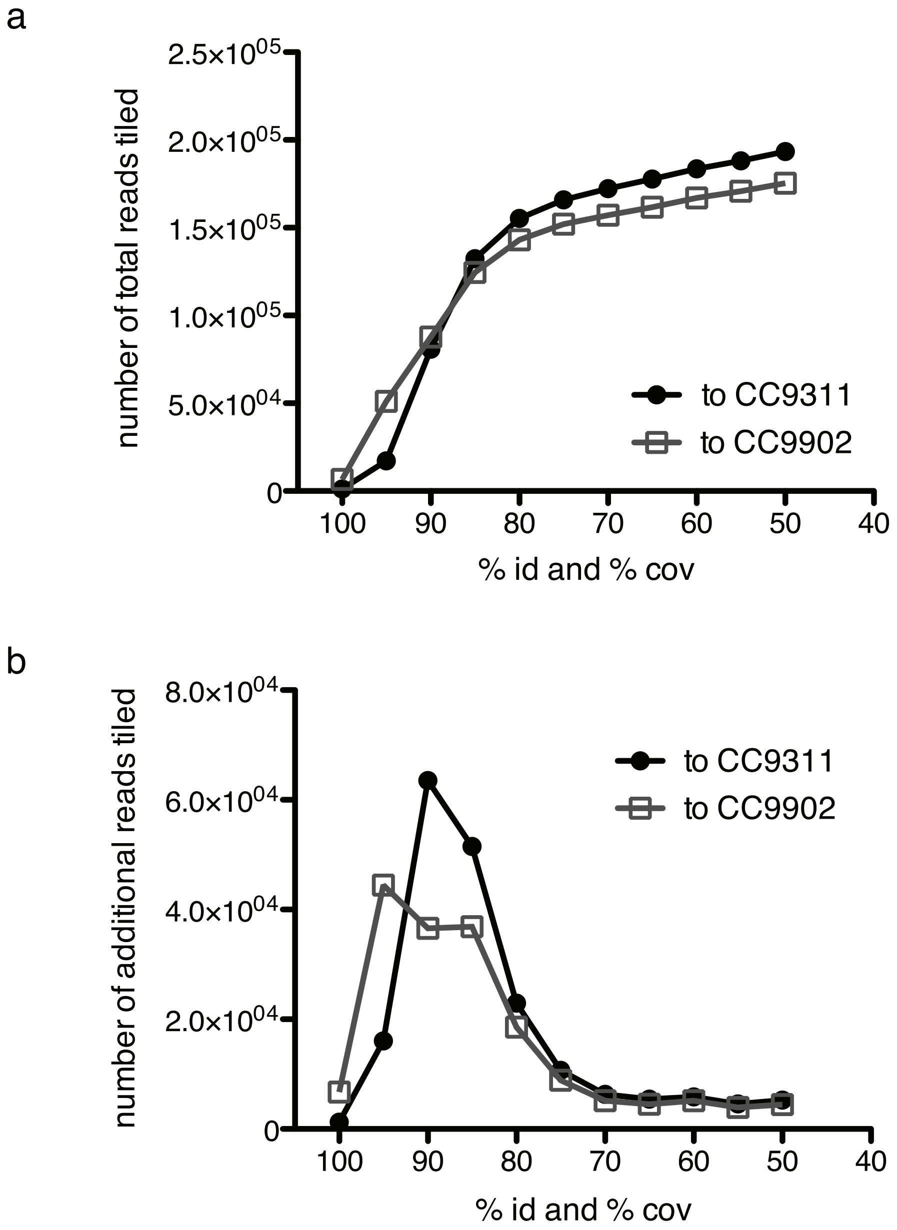
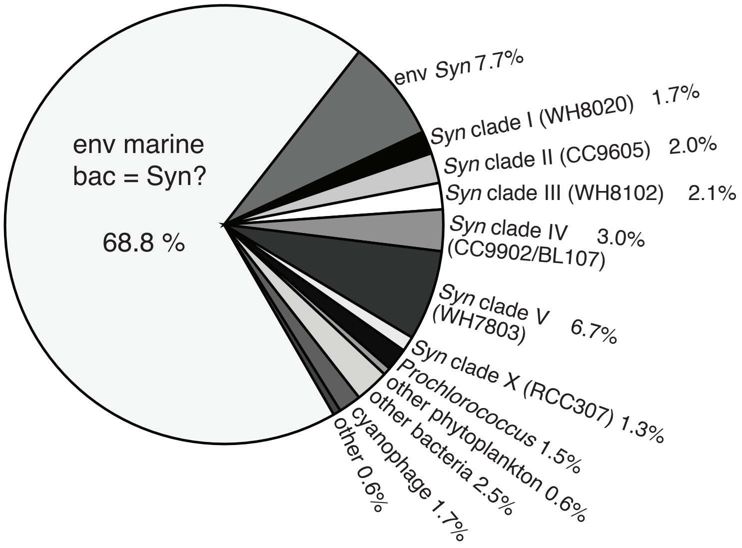


Figure 3.1. Number of a) total and b) additional metagenome sequence reads tiled to the CC9311and CC9902 genomes as the % identity (id) and % coverage (cov) was lowered.

a clade I % out of 4950 reads



b clade IV % out of 4082 reads

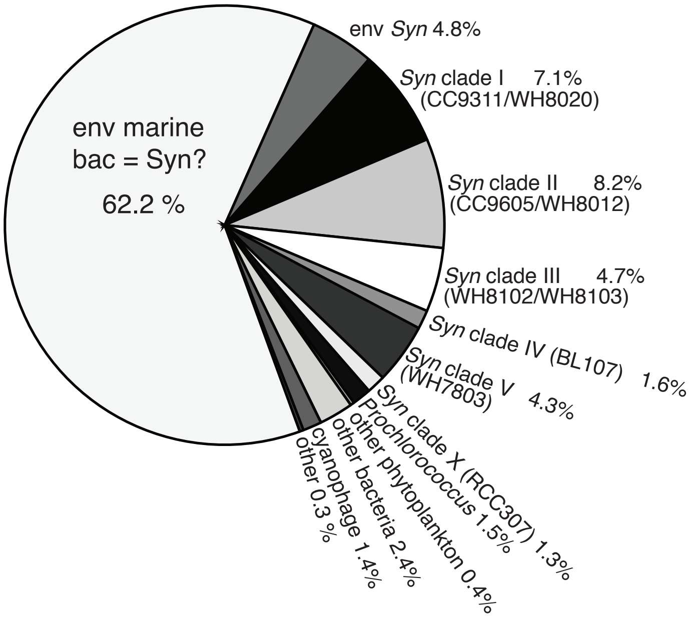


Figure 3.2. Taxon of the best BLASTn hit for tiled metagenome reads that did not have the a) CC9311 or b) CC9902 reference genomes as the top hit. Hits to environmental marine bacteria sequences are likely *Synechococcus* (env marine bac = Syn?).

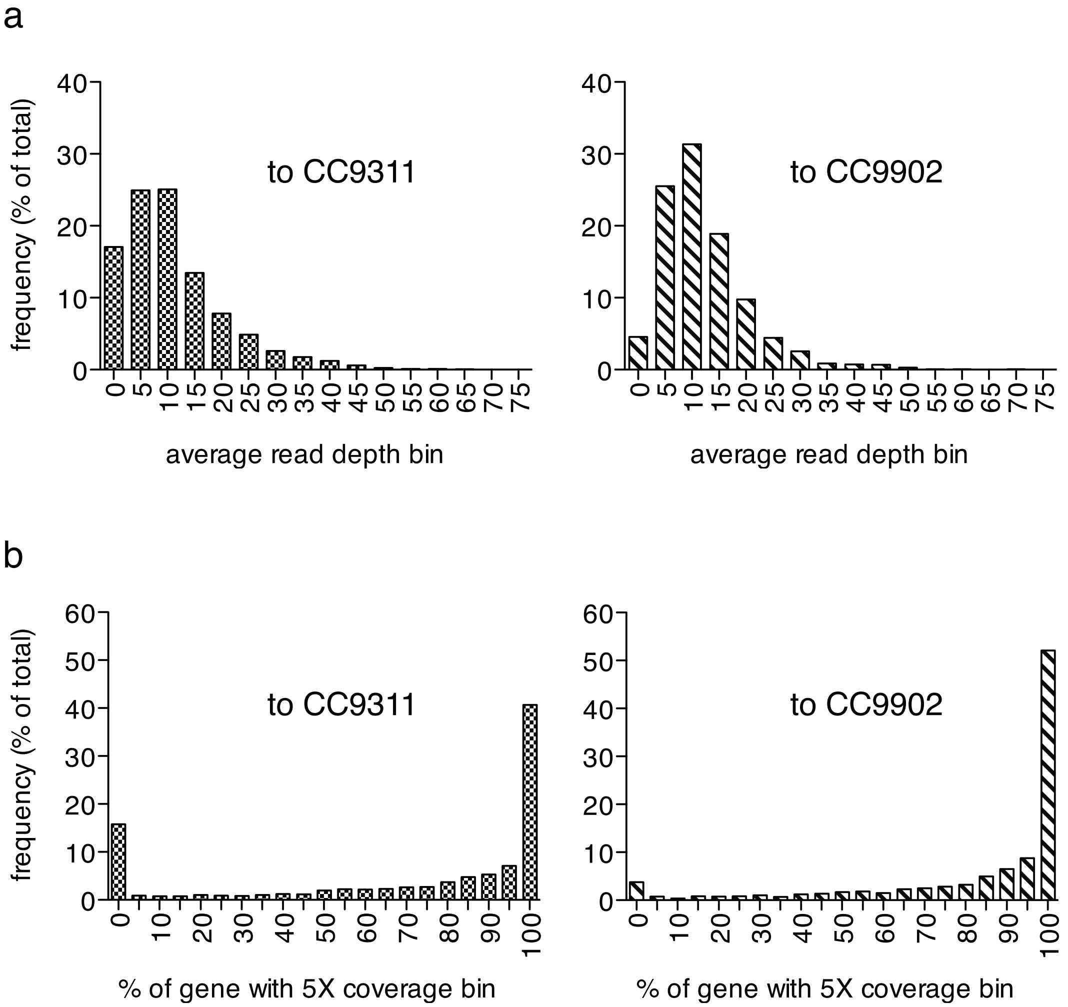


Figure 3.3 Histograms of a) average read depth per gene and b) % of gene covered with a minimum of 5X read depth based on tiling of metagenome reads to the CC9311 and CC9902 reference genomes. The tilings were based on criteria of 80 % identity and 80 % coverage.

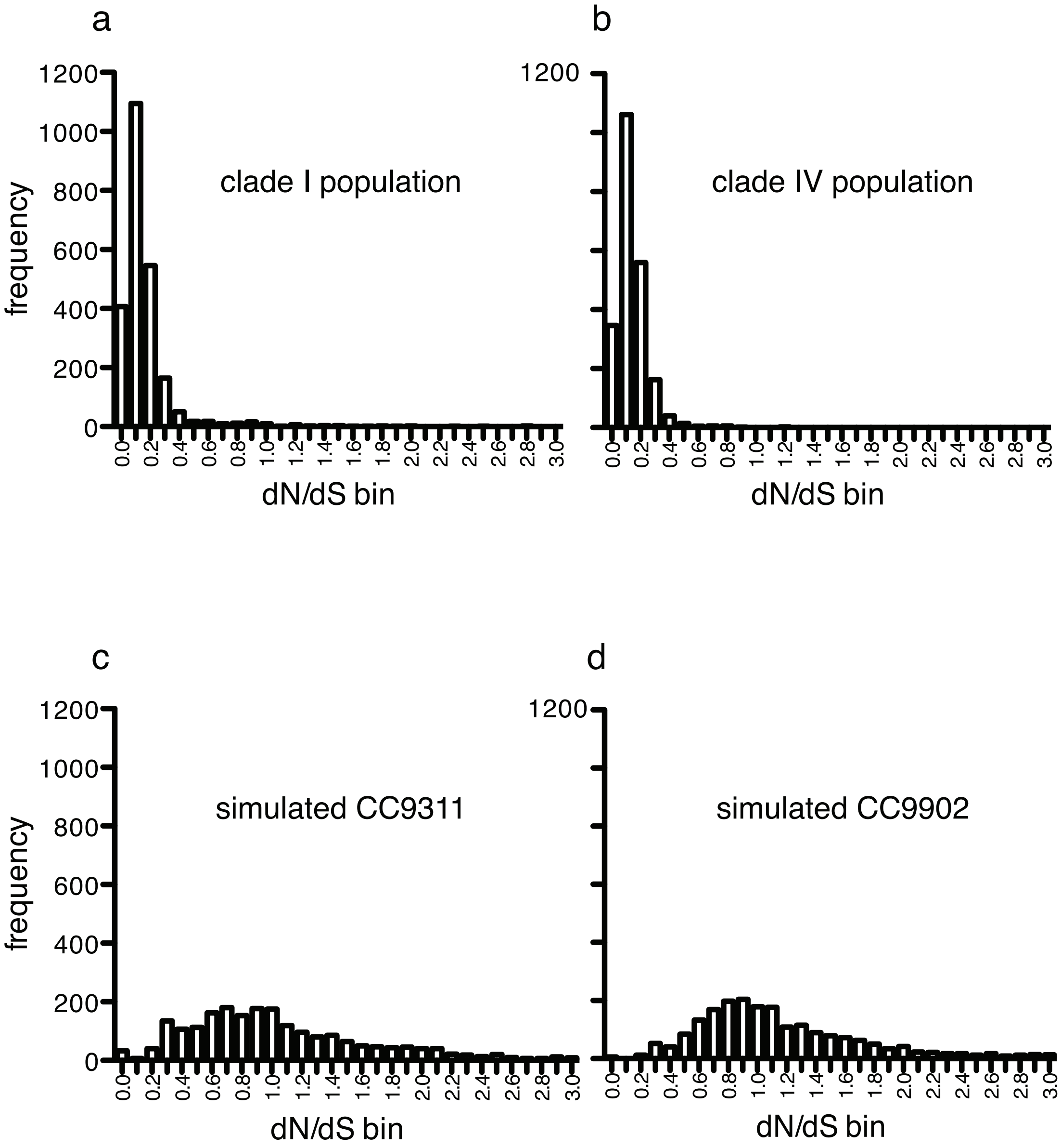


Figure 3.4. Histograms of dN/dS ratios from gene alignments using a) metagenomic sequences from a *Synechococcus* clade I population, b) metagenomic sequences from a *Synechococcus* clade IV population, c) simulated CC9311 sequences and d) simulated CC9902 sequences.

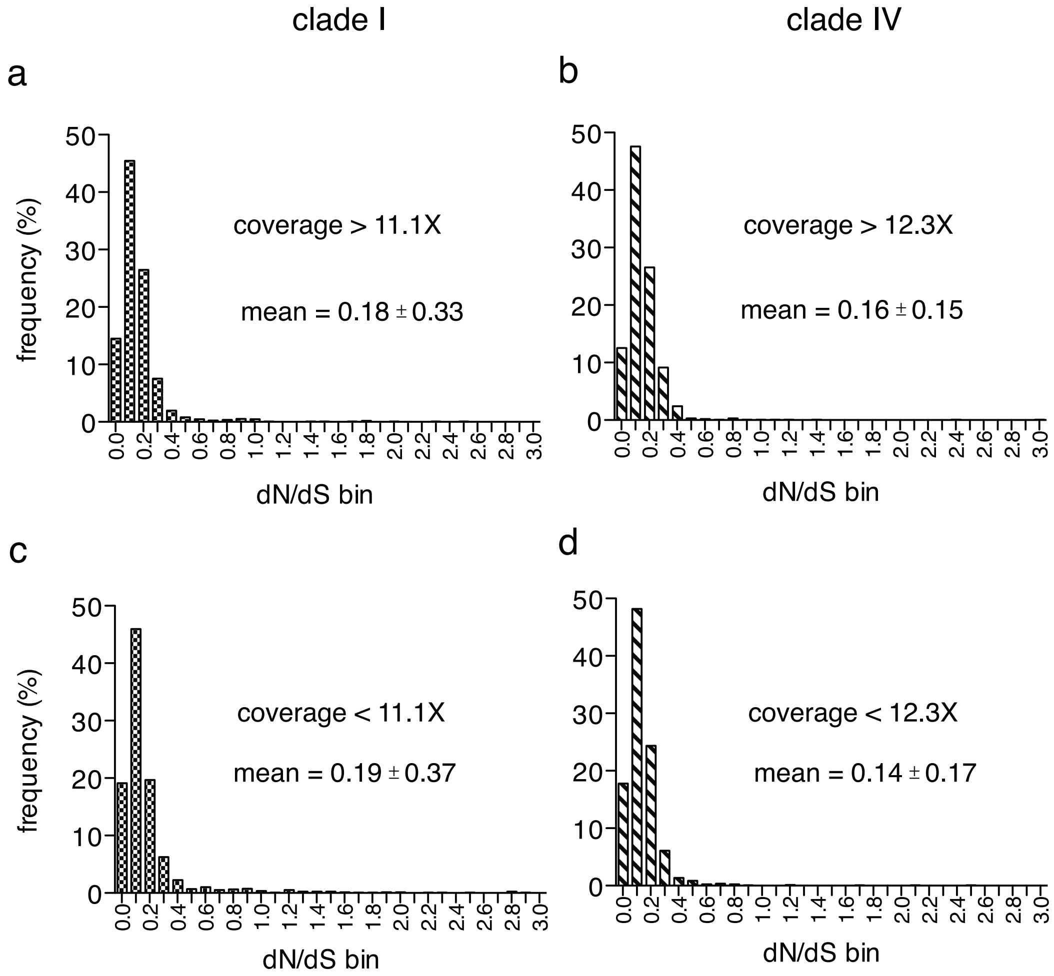


Figure 3.5. Histograms of dN/dS ratios for genes with greater than the mean depth of read coverage (a and b) and less than the mean depth of read coverage (c and d). For each histogram, the mean ± the standard deviation is provided.

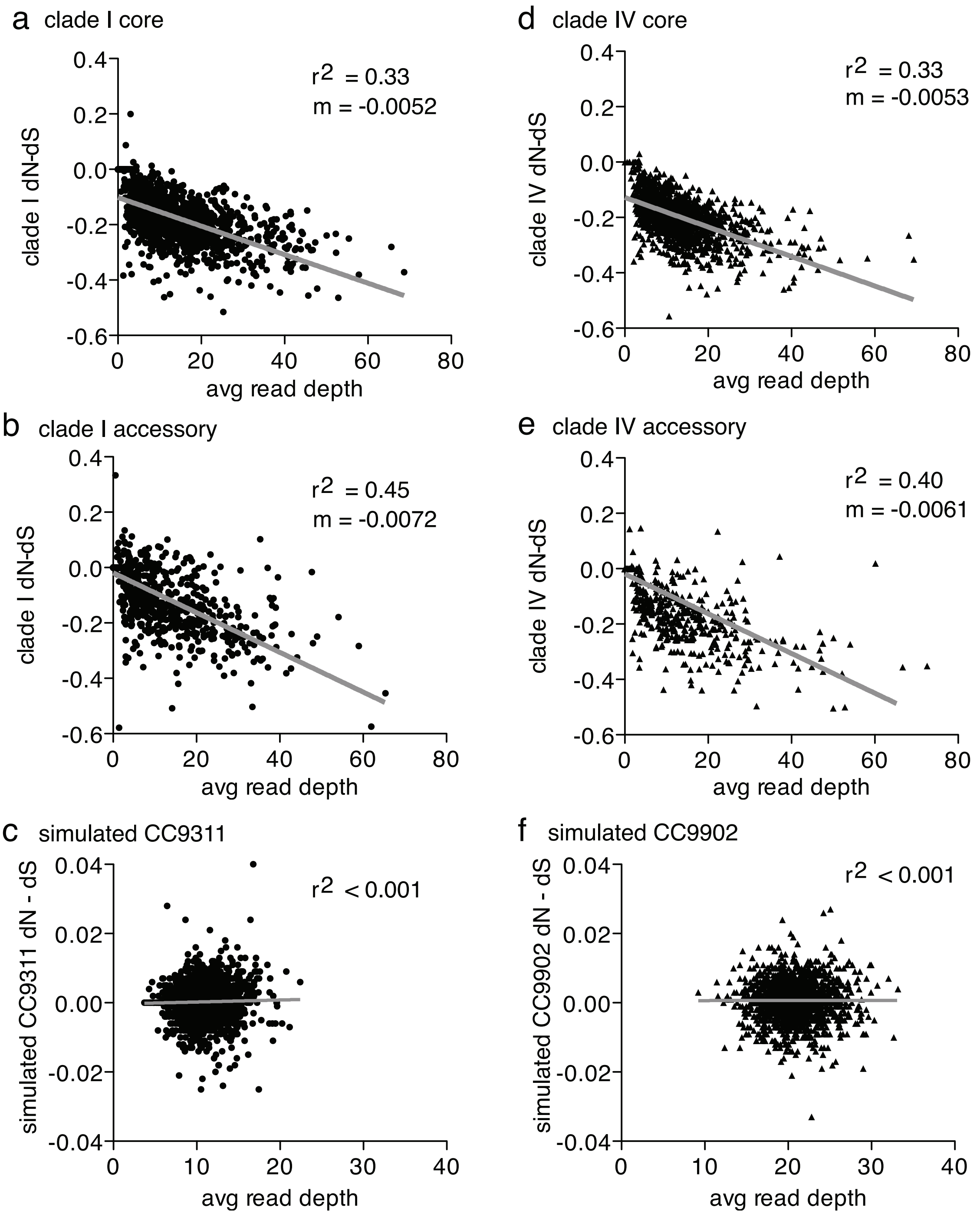


Figure 3.6. dN - dS versus the average depth of read coverage for each gene from a) the core genome of a clade I population, b) the accessory genome of a clade I population, c) simulated CC9311 sequences, d) the core genome of a clade IV population, e) the accessory genome of a clade IV population, and f) simulated CC9902 sequences. The gray line is the linear regression and m is the slope of the line. Note that dN - dS is plotted instead of dN/dS.

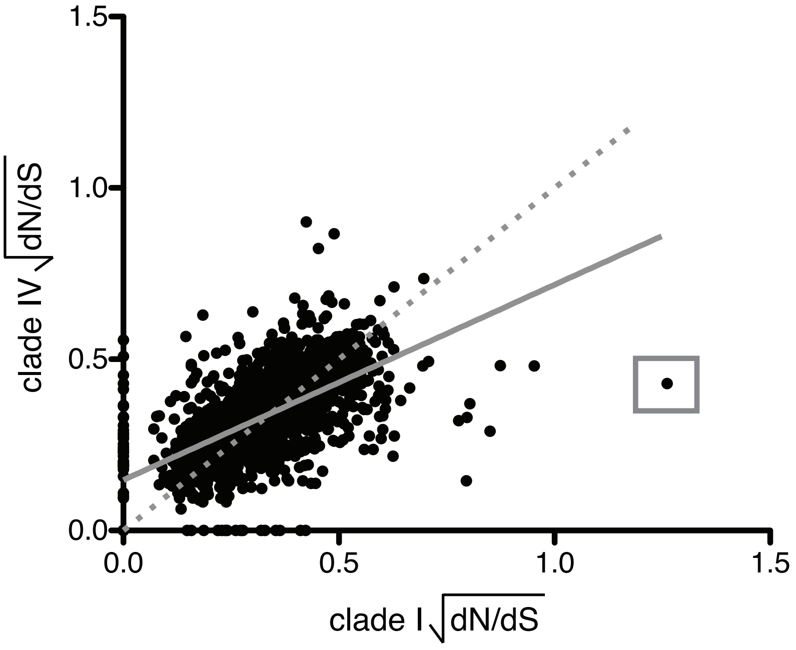


Figure 3.7. Scatter plot of dN/dS ratios of orthologous genes from a clade IV population versus a clade I population. The dN/dS ratios were square root transformed prior to plotting to obtain a more normal distribution of values. The solid gray line is the linear regression (r2 = 0.37, y = 0.57x + 0.15). The dotted line has a slope of 1. The boxed outlier corresponds to the binding protein for an ABC transporter for sugars (sync_1402 and syncc9902_1078). Not plotted are 98 genes with no observed polymorphisms and 8 genes with no observed synonymous polymorphisms based on metagenome tilings to either the CC9311 or CC9902 genomes.

References

Allen, E.E., Tyson, G.W., Whitaker, R.J., Detter, J.C., Richardson, P.M., and Banfield, J.F. 2007. Genome dynamics in a natural archaeal population. *Proc Natl Acad Sci U S A* 104:1883-1888.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic Local Alignment Search Tool. *J Mol Biol* 215:403-410.

Cohan, F.M. 2006. Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. *Philosophical Transactions of the Royal Society B-Biological Sciences* 361:1985-1996.

Daubin, V. and Moran, N.A. 2004. Comment on "The origins of genome complexity". *Science* 306:978a.

Doolittle, W.F. and Zhaxybayeva, O. 2009. On the origin of prokaryotic species. *Genome Res* 19:744-756.

Dufresne, A., Ostrowski, M., Scanlan, D.J., Garczarek, L., Mazard, S., Palenik, B.P., Paulsen, I.T., de Marsac, N.T., Wincker, P., Dossat, C., Ferriera, S., Johnson, J., Post, A.F., Hess, W.R., and Partensky, F. 2008. Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* 9:R90.

Eyre-Walker, A. and Keightley, P.D. 1999. High genomic deleterious mutation rates in hominids. *Nature* 397:344-347.

Frankham, R. 1995. Effective population size/adult population size ratios in wildlife: a review. *Genetics Research* 66:95-107.

Gianoulis, T.A., Raes, J., Patel, P.V., Bjornson, R., Korbel, J.O., Letunic, I., Yamada, T., Paccanaro, A., Jensen, L.J., Snyder, M., Bork, P., and Gerstein, M.B. 2009. Quantifying environmental adaptation of metabolic pathways in metagenomics. *Proc Natl Acad Sci U S A* 106:1374-1379.

Gill, S.R., Pop, M., DeBoy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., Gordon, J.I., Relman, D.A., Fraser-Liggett, C.M., and Nelson, K.E. 2006. Metagenomic analysis of the human distal gut microbiome. *Science* 312:1355-1359.

Haentjens-Sitri, J., Allemand, F., Springer, M., and Chiaruttini, C. 2008. A competition mechanism regulates the translation of the *Escherichia coli* operon encoding ribosomal proteins L35 and L20. *J Mol Biol* 375:612-625.

Harismendy, O., Ng, P.C., Strausberg, R.L., Wang, X., Stockwell, T., Beeson, K., Schork, N.J., Murray, S.S., Topol, E.J., Levy, S., and Frazer, K.A. 2009. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol* 10:R32.

Hewson, I., Paerl, R.W., Tripp, H.J., Zehr, J.P., and Karl, D.M. 2009. Metagenomic potential of microbial assemblages in the surface waters of the central Pacific Ocean tracks variability in oceanic habitat. *Limnol Oceanogr* 54:1981-1994.

Hirsh, A.E. and Fraser, H.B. 2001. Protein dispensability and rate of evolution. *Nature* 411:1046-1049.

Hu, J. and Blanchard, J.L. 2009. Environmental sequence data from the Sargasso Sea reveal that characteristics of genome reduction in *Prochlorococcus* are not a harbingr for an escalation in genetic drift. *Mol Biol Evol* 26:5-13.

Huber, J.A., Mark Welch, D., Morrison, H.G., Huse, S.M., Neal, P.R., Butterfield, D.A., and Sogin, M.L. 2007. Microbial population structures in the deep marine biosphere. *Science* 318:97-100.

Hughes, A.L. and Nei, M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335:167-170.

Hurst, L.D. and Smith, N.G.C. 1999. Do essential genes evolve slowly? *Curr Biol* 9:747-750.

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., and Mark Welch, D. 2007. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* 8:R143.

Jordan, I.K., Rogozin, I.B., Wolf, Y.I., and Koonin, E.V. 2002. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res* 12:962-968.

Kimura, M. and Ohta, T. 1974. On some principles governing molecular evolution. *Proc Natl Acad Sci U S A* 71:2848-2852.

Klein, D.J., Moore, P.B., and Steintz, T.A. 2004. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J Mol Biol* 340:141-177.

Konstantinidis, K.T. and Tiedje, J.M. 2005. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* 102:2567-2572.

Konstantinidis, K.T., Braff, J., Karl, D.M., and Delong, E.F. 2009. Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific subtropical gyre. *Appl Environ Microbiol* 75:5345-5355.

Kosakovsky Pond, S.L. and Frost, S.D.W. 2005. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. *Mol Biol Evol* 22:478-485.

Li, W.-H., Wu, C.-I., and Luo, C.-C. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol* 2:150-174.

Librado, P. and Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451-1452.

Lo, I., Denef, V.J., VerBerkmoes, N.C., Shah, M.B., Goltsman, D., DiBartolo, G., Tyson, G.W., Allen, E.E., Ram, R.J., Detter, J.C., Richardson, P., Thelen, M.P., Hettich, R.L., and Banfield, J.F. 2007. Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. *Nature* 446:537-541.

Lynch, M. and Conery, J.S. 2003. The origins of genome complexity. *Science* 302:1401-1404.

Lynch, M. and Conery, J.S. 2004. Response to Comment on "The origins of genome complexity". *Science* 306:978b.

Mes, T.H.M. 2008. Microbial diversity - insights from population genetics. *Environ Microbiol* 10:251-264.

Mukai, T. 1964. The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* 50:1-19.

Mukai, T., Chigusa, S.I., Mettler, L.E., and Crow, J.F. 1972. Mutatioin rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* 72:335-355.

Nei, M. and Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418-426.

Nielsen, R. and Yang, Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929-936.

Ohta, T. 1976. Role of very slightly deleterious mutations in molecular evolution and polymorphism. *Theor Popul Biol* 10:254-275.

Ohta, T. and Gillespie, J.H. 1996. Development of neutral and nearly neutral theories. *Theor Popul Biol* 49:128-142.

Palenik, B., Ren, Q.H., Dupont, C.L., Myers, G.S., Heidelberg, J.F., Badger, J.H., Madupu, R., Nelson, W.C., Brinkac, L.M., Dodson, R.J., Durkin, A.S., Daugherty, S.C., Sullivan, S.A., Khouri, H., Mohamoud, Y., Halpin, R., and Paulsen, I.T. 2006. Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. *Proc Natl Acad Sci U S A* 103:13555-13559.

Palenik, B., Ren, Q., Tai, V., and Paulsen, I.T. 2009. Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environmetal Microbiology* 11:349-359.

Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R., and Dodgson, J. 1980. The evolution of genes: the chicken preproinsulin gene. *Cell* 20:555-566.

Richter, D.C., Ott, F., Auch, A.F., Schmid, R., and Huson, D.H. 2008. MetaSim - a sequencing simulator for genomics and metagenomics. *PLoS ONE* 3:e3373.

Rocha, E.P.C., Smith, J.M., Hurst, L.D., Holden, M.T.G., Cooper, J.E., Smith, N.H., and Feil, E.J. 2006. Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J Theor Biol* 239:226-235.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D.Y., Eisen, J.A., Hoffman, J.M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, H., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J.E., Li, K., Kravitz, S., Heidelberg, J.F., Utterback, T., Rogers, Y.H., Falcon, L.I., Souza, V., Bonilla-Rosso, G., Eguiarte, L.E., Karl, D.M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M.R., Strausberg, R.L., Nealson, K., Friedman, R., Frazier, M., and Venter, J.C. 2007. The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* 5:398-431.

Simmons, S.L., DiBartolo, G., Denef, V.J., Goltsman, D.S.A., Thelen, M.P., and Banfield, J.F. 2008. Population genomic analysis of strain variation in *Leptospirillum* group II bacteria involved in acid mine drainage formation. *PLoS Biol* 6:e177.

Tai, V. and Palenik, B. 2009. Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site. *ISME J* 3:903-915.

Wang, Z. and Zhang, J. 2009. Why is the correlation between gene importance and gene evolutionary rate so weak? *PLoS Genet* 5:e1000329.

Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., Cayouette, M., McHardy, A.C., Djordjevic, G., Aboushadi, N., Sorek, R., Tringe, S.G., Podar, M., Martin, H.G., Kunin, V., Dalevi, D., Madejska, J., Kirton, E., Platt, D., Szeto, E., Salamov, A., Barry, K., Mikhailova, N., Kyrpides, N.C., Matson, E.G., Ottesen, E.A., Zhang, X., Hernandez, M., Murillo, C., Acosta, L.G., Rigoutsos, I., Tamayo, G., Green, B.D., Chang, C., Rubin, E.M., Mathur, E.J., Robertson, D.E., Hugenholtz, P., and Leadbetter, J.R. 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. 450:560-569.

Wilson, A.C., Carlson, S.S., and White, T.J. 1977. Biochemical evolution. *Annu Rev Biochem* 46:573-639.

Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 15:568-573.

Yang, Z. and Nielsen, R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* 19:908-917.

Zhaxybayeva, O., Doolittle, W.F., Papke, R.T., and Gogarten, J.P. 2009. Intertwined evolutionary histories of marine *Synechococcus* and *Prochlorococcus marinus*. *Genome Biology and Evolution* 1:325-339.

APPENDIX 1

The variability and evolution of horizontally transferred genes

Bacterial evolution is shaped not only by vertical descent, but perhaps more importantly by the acquisition of horizontally transferred genes (Koonin *et al*., 2001; Boucher *et al*., 2003; Dobrindt *et al*., 2004; Gogarten and Townsend, 2005). Once acquired, there are several possible outcomes: the gene is immediately beneficial and provides novel functionality to its host; it accumulates some mutations before becoming beneficial; it is neutral or deleterious thus accumulates mutations freely; it is neutral or deleterious and the gene is deleted. The evolutionary history of horizontally acquired genes in marine *Synechococcus* was investigated here.

Genes that were potentially horizontally transferred were identified from the *Synechococcus* CC9311 and CC9902 genomes based on atypical gene content and a low lineage probability index (LPI) (Podell et al. 2008). Candidate regions for analysis were identified if conserved genes flanked the variable genes and the flanking genes could be used to design PCR primers to amplify across the variable region.

Table A1.1 lists the PCR primers designed for this purpose. The PCR primers were specific to either CC9311 or CC9902 depending on the genome from which variable regions were found. The variable regions were expected to be at least clade-specific, if not strain-specific.

The variable region between CC9311 genes sync_1607 to sync_1610 was analyzed further. This variable region contained two genes in CC9311 – an FAD-dependent monooxygenase and an auxin efflux carrier superfamily protein. PCR products were amplified and sequenced from 4 *Synechococcus* clade I strains (CC9311, CC9617, WH8016, and WH8020) and from three surface samples from the SIO pier (on 5/14/06, 10/10/06, and 5/15/07). The PCR products from the environmental samples were of various sizes (data not shown) and were cloned prior to sequencing. Selecting a diversity of insert sizes, sixteen environmental clones were sequenced.

The sequences were aligned using clustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html). From the default settings, the gap open penalty was increased to 50 and the gap extension penalty was decreased to 0.5 to favor alignments with large gaps. The result was then manually aligned using Se-Al (v2.0a11, http://tree.bio.ed.ac.uk/software/seal/). The alignment indicated that the sequences were of 6 distinct types (Figure A1.1). CC9311 was the only sequence to have both genes of the variable region. WH8016 and two environmental clones had the FAD-dependent monooxygenease, but not the auxin efflux carrier superfamily protein. Seven environmental clones carried a section of the FAD-dependent monooxygenase gene, but not the entire gene. WH8020 contained small fragments of the variable region at either end. Four environmental clones also only contained the ends of the variable region, but these sequences were distinct from WH8020. Lastly, CC9617 and 3 environmental clones did not contain any portion of the variable region. The date of sampling was not correlated with the type of sequences obtained from environmental clones but more samples should be analyzed.

These results suggest that the FAD-dependent monooxygenase and auxin efflux carrier superfamily genes were inserted into an ancestor of the clade I *Synechococcus*. Subsequently, in some lineages the genes were lost entirely, whereas in other lineages only fragments of the genes remain. Only in CC9311 were both genes retained.

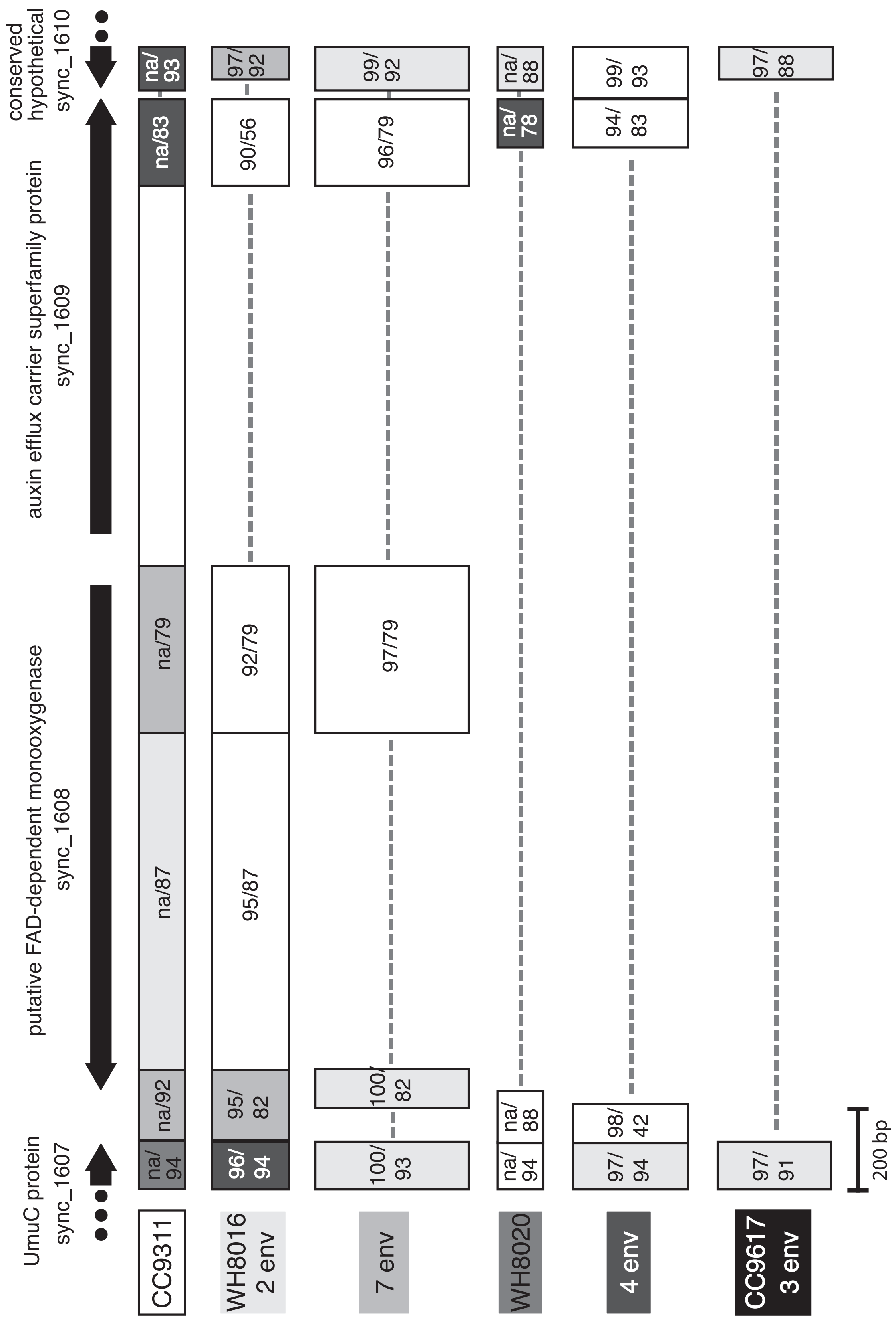
Acknowledgements

The dissertation author would like to thank Kim Tsai and Rudy Urbano for technical contributions to the research in this appendix.

Table A1.1. Primers designed to amplify variable regions from the CC9311 and CC9902 genomes. The primers anneal to conserved genes flanking the variable region. sync genes are from the CC9311 genome and syncc9902 genes are from the CC9902 genome. The primers were tested on DNA from the strain for which the primers were designed. The size of the PCR product is provided.

Flanking genes	Forward primer	Reverse primer	size (bp)
sync_0261-0266	TGG ATT TAG TTT GGA ATG GAG CA	CTT CCG CTA GCA ATC CAA G	CC9311 = 7 528
sync_0346-0351	TCC TCT CAC TTC ACC CGC TT	GAT CGC TAA TGC ACG CAA CAG A	CC9311 = 3 604
sync_0652-0668	CGT AAG GAT GTG TTG GCC A	TGA TTG TCA CGA CGC TGG A	CC9311 = 13 426
sync_0924-0928	CGT GAC GCC GCA GCA GTA AA	TTA CAA GTG CAT GAT GAG CT	CC9311 = 2 388
sync_1607-1610	AAT CAC CGC TAT GGG CGA GG	CGA GAA CAG ACA GGC CTT AA	CC9311 = 2 669
sync_2703-2706	GTG GCG AGT GGC TTC TGC CA	CAG TTG GGC CTC GAT CAT CCG AT	CC9311 = 2 331
syncc9902_0059-0061	TCC CCT TGR CTG AAC TCT GA	AGG TAG TAT CGA RGA AAT CAT GCT C	CC9902 = 1 787
syncc9902_0819-0821	CGA TTS GGA AAC TGG GCY TGA A	TTC AGG GGA ACG CAG CAT C	CC9902 = 2 516
syncc9902_1380-1385	GGA TAC CAC TTC TCA TWC CRG G	CCT TGA ATT GGG CCA TGC G	CC9902 = 8 336
syncc9902_2053-2056	GTG GGC GGA TAT TTG TTC AAA GT	TGT TGC TCC GTG GAT CCC A	CC9902 = 1 761
syncc9902_2180-2183	TTC TTG GGC TGC CGC GTT GA	GCT CTA AGG CGT ACT CCA GTT C	CC9902 = 3 329

Figure A1.1 Representation of an alignment of sequences between genes homologous to sync_1607 and sync_1610 from cultured strains and environmental samples. Sequences were aligned in segments that were clearly similar to each other. Sequences with a similar alignment were grouped into blocks. The sequences included in each block are shown on the right and are labeled on a gray scale. The fill color of the blocks indicates the next most similar group of sequences. The values within the blocks are the percent similarity among the sequences within the block and with the next most similar group, respectively. Dashes indicate alignment gaps. Genes annotated in this region are shown at the top of the figure, represented by arrows. The length of the CC9311 sequence is 2630 base pairs. Sequences from cultured strains and the environment (env) were all shorter than CC9311.


UmuC protein
sync_1607
putative FAD-dependent monooxygenase
sync_1608
auxin efflux carrier superfamily protein
sync_1609
conserved
hypothetical
sync_1610
CC9311
na/94
na/92
na/87
na/79
na/83
na/93
WH8016
2 env
96/94
95/82
95/87
92/79
90/56
97/92
7 env
100/93
100/82
97/79
96/79
99/92
WH8020
na/94
na/88
na/78
na/88
4 env
97/94
98/42
94/83
99/93
CC9617
3 env
97/91
97/88
200 bp

References

Boucher, Y., Douady, C. J., Papke, R. T., Walsh, D. A., Boudreau, M. E. R., Nesbo, C. L., Case, R. J., Doolittle, W. F. 2003. Lateral gene transfer and the origins of prokaryotic groups. *Annu Rev Genet* 37:283-328

Dobrindt, U., Hochut, B., Hentschel, U., and Hacker, J. 2004. Genomic islands in pathogenic and environmental microorganisms. *Nature Reviews Microbiology* 2:414-424.

Gogarten, J. P. and Townsend, J. P. 2005. Horizontal gene transfer, genome innovation and evolution. *Nature Reviews Microbiology* 3: 679-687.

Koonin, E. V., Makarove, K. S., and Aravind, L. 2001. Horizontal gene transfer in prokaryotes: Quantification and classification. *Annu Rev Microbiol* 55:709-742.

Podell, S., Gaasterland, T., and Allen, E. 2008. A database of phylogenetically atypical genes in archael and bacterial genomes, identified using the DarkHorse algorithm. *BMC Bioinformatics* 9:419.

APPENDIX 2

Assessing *Synechococcus* enrichment from flow cytometry sorting

Flow cytometry sorting was used to enrich seawater samples for *Synechococcus*. DNA was extracted from these sorted samples, then amplified using whole-genome amplification before being used for environmental metagenomic sequencing (Chapter 3). To assess the success of flow cytometry sorting to enrich for *Synechococcus*, a quality control method was developed to determine the relative abundance of *Synechococcus* DNA to total bacterial DNA extracted from the sorted samples. This method can be used to assess the relative abundance or enrichment of *Synechococcus* from environmental samples prior to proceeding with metgenomic sequencing.

Methods

Two pairs of primers for use in quantitative PCR (qPCR) were designed to amplify the 16S rRNA gene from cyanobacteria only (cyano) or from most bacteria (univ) (Table A2.1). Using the primer pairs in separate qPCR reactions, the flow sorted, whole-genome amplified samples were tested (see Chapter 3). Genomic DNA from *Synechococcus* sp. strain CC9311 and two heterotrophic bacterial strains, TW7 (Alteromonadaceae) and BBFL7

(Flavobacteria), were used as standards. Serial dilutions ranging from 10 to 0.00001 ng of DNA were used.

qPCR reactions consisted of 1X Brilliant SybrGreen Master Mix (Stratagene), 0.4 μM of each primer, 0.1 mg/ml BSA, and 1 to 5 ng of template DNA. The same temperature cycling parameters were used for both primer pairs. The reactions were incubated at 95 °C for 10 minutes followed by 40 cycles of 95 °C for 30 s, 56 °C for 30 s, and 72 °C for 30 s.

Using standard curves, the amount of cyanobacterial DNA and total bacterial DNA in the samples was determined from the qPCR reactions using the cyano and univ primer pairs, respectively. As *Synechococcus* are the dominant cyanobacteria in the coastal Southern California Bight, the relative amount of *Synechococcus* DNA in a sample was calculated as the ratio of the ng of cyanobacterial DNA to the ng of total bacterial DNA.

Results

Using cyanobacteria-specific and universal primers in qPCR, a ratio of the amount of *Synechococcus* DNA relative to total bacterial DNA was determined for each sorted sample (Table A2.2). The ratios ranged from 0.47 to 1.98. For pure *Synechococcus* CC9311 genomic DNA, the ratio was 2.13 ± 0.56 (standard deviation). The ratio was not 1 (*Synechococcus* DNA = total bacterial DNA) because the standard curve using the univ primers was based on bacteria with different copy numbers of the 16S rRNA gene.

For each sample the qPCR ratio of cyanobacterial to total bacterial DNA was compared to the percentage of metagenome sequences that tiled to the CC9311 and CC9902 genomes (Table A2.2, Figure A2.1). As CC9311 and CC9902 are *Synechococcus* strains representing the dominant *Synechococcus* clades in the coastal Southern California Bight, the percentage of metagenome reads tiling to these genomes provides an estimate of the *Synechococcus* abundance in the sample. This estimate and the qPCR ratio were highly correlated and best fit by an exponential curve (Figure A2.1). Thus, the qPCR method successfully predicted the relative number of *Synechococcus* sequences obtained from 454 sequencing.

Table A2.1. qPCR primers used for determining the amount of cyanobacterial DNA (cyano) and total bacterial DNA (univ) from flow sorted samples.
* from Sogin *et al.* (2006).

primer name	sequence (5' – 3')
cyano16S_1231F	GTA CTA CAA TGC TAC GGA CAA AG
cyano16S_1433R	AGG GTT GGA GTA ACG GCT T
univ16S_907F	AAA CTC AAA KGA ATT GAC GGG G
univ16S_1046R*	CGA CAG CCA TGC ANC ACC T

Table A2.2. *Synechococcus* abundance indices from flow cytometry sorted seawater samples.

sample	qPCR cyano DNA : total DNA	% of metagenome reads tiling to the CC9311 & CC9902 genomes
10/10/06	1.98	13.7
5/17/07	0.17	1.1
3/6/08	1.59	11.6
4/17/08	0.47	1.7
5/15/08 surface	0.54	0.9
5/15/08 20m	1.30	3.9

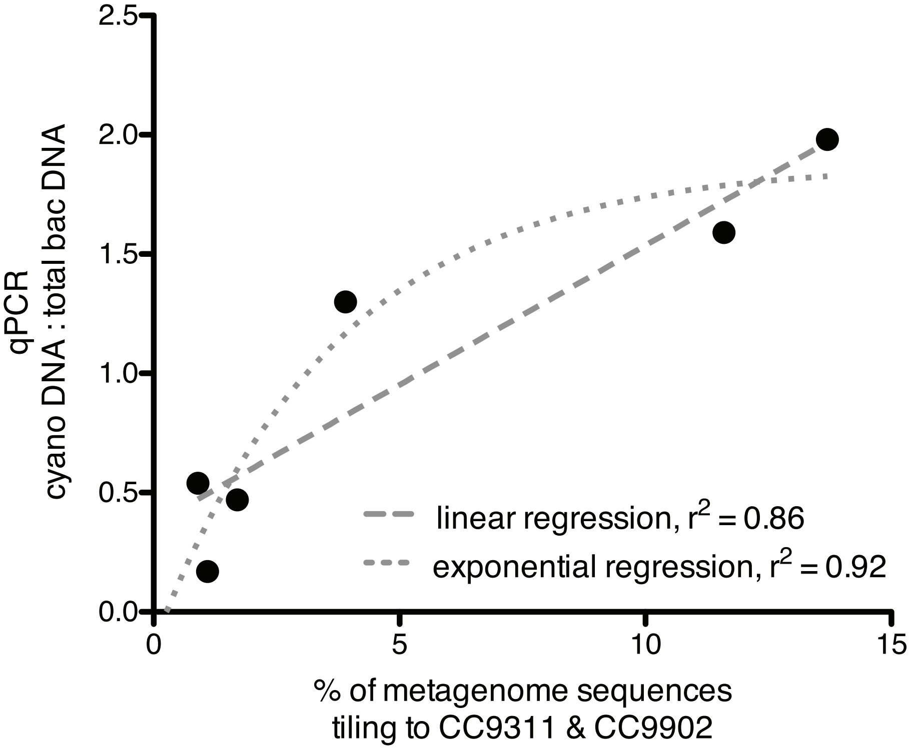


Figure A2.1. Relative abundance of *Synechococcus* (cyano) DNA from flow cytometry sorted samples determined by qPCR versus the percentage of metagenome sequences tiled to the CC9311 and CC9902 genomes. The metagenome sequences were tiled to the reference genomes using criteria of 90 % identity and 90 % coverage.

References

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M., Herndl, G. J. 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *PNAS* 103: 12115-12110.

APPENDIX 3

Luminex hybridization signals from environmental samples

Table A3.1. Luminex hybridization signals for clade probes from a one-year time series of surface samples at the SIO pier. The mean signal:background (S:B) and the standard deviation (sd) from 5 or 6 technical replicates is shown.

sample date	I-A		II		III		IV-A		total *Syn* (10^4 cells/ml)
	S:B	sd	S:B	sd	S:B	sd	S:B	sd	
1/5/06	8.05	3.72	0.88	0.16	0.98	0.17	16.39	5.27	8.83
2/9/06	6.06	2.08	0.79	0.07	0.86	0.09	9.50	3.16	3.69
3/9/06	8.00	3.87	0.91	0.11	0.97	0.09	7.75	3.30	1.08
4/6/06	7.69	1.80	0.76	0.12	0.88	0.13	7.33	1.28	3.26
5/2/06	6.70	3.00	0.84	0.17	0.96	0.24	7.52	3.08	2.82
5/14/06	6.82	4.17	0.78	0.14	0.84	0.15	20.60	6.82	22.25
6/6/06	2.55	0.80	0.87	0.14	0.90	0.10	18.35	1.83	14.07
7/6/06	7.13	4.23	0.95	0.24	0.92	0.09	23.26	4.31	14.40
8/3/06	7.00	1.92	2.72	0.45	0.95	0.06	20.63	4.97	8.60
8/25/06	8.66	2.03	1.16	0.28	1.05	0.13	15.66	0.75	6.17
9/5/06	10.34	2.73	1.18	0.23	1.00	0.11	16.48	2.64	6.41
10/3/06	6.35	3.43	1.07	0.15	0.90	0.13	9.08	4.55	6.35
11/6/06	7.13	2.45	1.64	0.35	1.26	0.15	14.27	1.74	7.13
12/4/06	6.77	3.13	1.90	0.57	1.28	0.14	12.64	1.70	6.77
1/8/07	5.32	1.43	1.49	0.48	1.09	0.14	17.79	9.21	5.32
2/26/07	10.78	5.23	0.87	0.12	0.90	0.08	19.43	8.94	10.78

Table A3.2. Luminex hybridization signals for sub-clade probes from a one-year time series of surface samples at the SIO pier. The mean signal:background (S:B) and the standard deviation (sd) from 5 or 6 technical replicates is shown.

sample date	I-A		I-B		I-C		I-D		I-E		IV-A		IV-B		IV-C		total *Syn* (10^4 cells/ml)
	S:B	sd	S:B	sd	S:B	sd	S:B	sd	S:B	sd	S:B	sd	S:B	sd	S:B	sd	
1/5/06	6.67	4.20	2.93	1.85	1.45	0.47	1.40	0.43	1.40	0.84	21.63	11.33	2.05	1.22	9.94	6.88	8.83
2/9/06	4.27	2.23	2.34	1.06	1.54	0.62	1.54	0.77	1.42	0.81	7.60	3.30	2.72	1.09	2.58	1.05	3.69
3/9/06	2.87	1.34	1.55	0.36	1.33	0.30	2.05	1.12	1.41	1.00	3.42	1.50	1.52	0.50	1.48	0.28	1.08
4/6/06	4.71	2.38	2.74	1.26	1.52	0.37	1.44	0.27	1.16	0.29	5.58	3.11	2.29	0.97	1.49	0.45	3.26
5/2/06	4.24	1.98	2.28	1.00	1.52	0.54	1.41	0.36	1.13	0.38	6.05	3.20	2.89	1.80	1.49	0.48	2.82
5/14/06	4.40	2.50	2.66	1.50	2.37	0.89	1.91	0.71	1.34	0.58	19.76	10.70	6.90	3.84	3.24	1.81	22.25
6/6/06	1.61	0.68	1.18	0.50	1.56	0.65	1.79	0.79	1.71	0.90	22.77	13.38	1.63	0.79	3.65	2.20	14.07
7/6/06	3.94	1.87	1.67	0.84	1.41	0.48	1.59	0.55	1.13	0.41	24.53	14.53	2.98	1.84	9.15	6.33	14.40
8/3/06	4.49	1.68	2.82	0.93	1.61	0.39	1.70	0.43	1.25	0.27	27.67	12.99	1.52	0.35	10.95	6.21	8.60
8/25/06	5.07	2.06	2.66	1.22	1.35	0.51	1.32	0.47	1.32	0.66	18.61	5.12	1.82	0.99	4.99	2.04	6.17
9/5/06	6.00	1.07	3.27	0.49	1.48	0.25	1.54	0.29	1.49	0.31	15.82	3.71	2.29	0.40	4.39	0.78	6.41
10/3/06	3.85	2.01	2.22	1.05	1.35	0.43	1.42	0.56	1.33	0.70	10.50	7.28	2.09	0.92	3.92	2.65	6.35
11/6/06	3.67	1.84	1.77	0.92	1.04	0.27	1.23	0.30	1.00	0.29	13.71	7.62	1.27	0.53	4.85	3.54	7.13
12/4/06	2.19	0.48	0.98	0.26	0.93	0.27	1.02	0.29	0.85	0.27	8.00	2.91	0.71	0.17	2.79	0.93	6.77
1/8/07	2.94	1.20	1.12	0.40	1.00	0.35	1.11	0.35	0.99	0.40	17.69	8.74	0.75	0.22	5.69	2.65	5.32
2/26/07	6.31	3.07	2.49	1.17	1.31	0.42	1.18	0.31	1.34	0.57	18.97	9.95	0.95	0.29	4.80	2.34	10.78

Table A3.3. Luminex hybridization signals for sub-clade probes from a coastal to open ocean transect along CalCOFI Line 93. Samples were taken from July 17 to August 4, 2007. The mean signal:background (S:B) and the standard deviation (sd) from 5 or 6 technical replicates is shown.

station /depth	I-A		I-B		I-C		I-D		I-E		IV-A		IV-B		IV-C		total *Syn* (10^4 cells/ml)
	S:B	sd	S:B	sd	S:B	sd	S:B	sd	S:B	sd	S:B	sd	S:B	sd	S:B	sd	
93.26 3m	4.79	1.49	2.58	0.67	1.70	0.30	1.81	0.35	1.48	0.34	24.27	8.32	2.32	0.65	5.23	1.66	9.37
93.26 18m	14.91	5.41	6.76	2.79	1.78	0.20	1.73	0.17	1.80	0.26	19.49	6.86	9.77	3.69	4.25	1.26	13.82
93.35 5m	2.55	0.64	1.69	0.41	1.51	0.28	1.75	0.11	1.22	0.22	22.28	5.77	1.78	0.34	11.51	3.06	5.78
93.35 20m	14.38	5.57	7.33	2.79	1.64	0.45	1.76	0.29	1.59	0.51	14.52	5.15	12.02	4.95	4.08	1.38	22.78
93.40 5m	2.00	0.52	1.25	0.44	1.31	0.23	1.49	0.31	0.91	0.27	18.56	9.67	1.37	0.40	10.66	5.15	5.51
93.40 30m	7.53	3.47	2.94	1.02	1.36	0.22	1.66	0.16	1.15	0.21	15.23	8.20	3.67	1.45	5.58	2.30	11.15
93.45 5m	11.27	5.44	4.90	2.28	2.29	0.52	2.45	0.79	1.74	0.55	49.92	21.08	10.84	5.55	15.12	7.75	43.64
93.45 20m	9.00	3.76	3.70	1.57	1.69	0.41	1.78	0.34	1.38	0.33	37.27	13.67	7.15	2.95	9.77	4.33	39.64
93.45 40m	11.71	9.32	5.68	5.10	1.48	0.49	1.60	0.40	1.51	0.98	20.16	15.89	8.26	7.74	4.58	3.46	5.88
93.50 5m	5.46	2.45	2.89	1.15	1.43	0.35	1.49	0.26	1.13	0.28	22.38	10.81	5.93	3.00	6.44	3.32	14.69
93.50 15m	10.56	4.49	4.60	1.78	1.76	0.48	1.82	0.34	1.66	0.47	29.70	13.74	8.67	3.77	7.41	3.16	27.99
93.50 30m	8.58	3.91	4.29	2.16	1.68	0.44	1.82	0.36	1.99	0.86	16.47	7.18	8.57	4.36	4.93	2.17	7.36

Table A3.3. Luminex signals for sub-clade probes from Line 93, continued.

station /depth	I-A		I-B		I-C		I-D		I-E		IV-A		IV-B		IV-C		total *Syn* (10^4 cells/ml)
	S:B	sd	S:B	sd	S:B	sd	S:B	sd	S:B	sd	S:B	sd	S:B	sd	S:B	sd	
93.65 5m	1.61	0.52	1.13	0.32	1.29	0.22	1.62	0.45	0.94	0.28	2.19	1.01	2.37	1.06	1.50	0.30	12.12
93.65 30m	4.33	2.11	2.45	1.05	1.46	0.21	1.47	0.19	1.11	0.28	13.66	8.12	5.83	3.16	3.59	1.68	24.58
93.80 5m	1.48	0.48	1.05	0.34	1.35	0.37	1.51	0.46	1.11	0.29	1.35	0.40	1.52	0.46	1.31	0.29	2.84
93.80 40m	1.15	0.44	0.84	0.30	1.08	0.31	1.38	0.42	0.95	0.50	0.96	0.39	2.05	1.65	1.09	0.31	4.65
93.80 50m	1.29	0.27	0.98	0.21	1.30	0.33	1.54	0.31	1.20	0.40	1.18	0.36	2.32	1.00	1.36	0.34	4.11
93.80 70m	1.44	0.36	1.17	0.32	1.44	0.20	1.60	0.38	1.26	0.40	1.28	0.46	3.38	1.65	1.32	0.36	2.74
93.90 10m	1.14	0.08	0.93	0.25	1.33	0.19	1.47	0.15	1.08	0.13	1.10	0.26	1.14	0.09	1.28	0.25	2.57
93.90 43m	1.29	0.17	0.96	0.15	1.19	0.17	1.34	0.21	1.03	0.14	1.07	0.27	4.22	1.66	1.18	0.13	4.41
93.90 53m	1.24	0.22	1.00	0.20	1.20	0.17	1.53	0.14	1.18	0.30	1.18	0.29	5.25	2.98	1.28	0.12	4.22
93.110 5m	1.14	0.19	0.88	0.16	1.59	0.69	1.52	0.24	1.06	0.18	2.16	0.64	1.06	0.17	1.13	0.13	0.34
93.110 50m	0.99	0.31	0.84	0.33	1.12	0.32	1.38	0.39	0.97	0.36	2.65	1.33	1.07	0.48	1.15	0.38	0.52
93.120 10m	1.17	0.16	0.96	0.21	1.28	0.39	1.50	0.26	1.03	0.19	1.66	0.53	1.00	0.18	1.19	0.28	0.20
93.120 50m	1.19	0.41	0.98	0.25	1.41	0.45	1.60	0.43	1.17	0.29	1.53	0.43	1.07	0.28	1.31	0.41	0.25
93.120 90m	1.06	0.14	0.86	0.27	1.16	0.22	1.53	0.23	1.10	0.38	1.16	0.34	1.03	0.31	1.21	0.30	0.16

APPENDIX 4

Demarcating ecotypes by simulation for *Synechococcus* from clades I and IV

Microbial ecotypes have been described as lineages with ecologically distinct roles (Koeppel *et al.* 2008). Because it has been recognized that microbes within traditionally defined "species" are physiologically and genetically diverse, the concept of a microbial ecotype helps to define phylogenetic and phenotypic boundaries that are ecologically and evolutionarily relevant.

From *rpoC1* gene sequence alignments, Ecotype Simulation (Koeppel *et al.* 2008, http://fcohan.web.wesleyan.edu/) was used to identify phyogenetic clusters that are ecotypes within clade I and clade IV. The simulation determines the number of ecotypes in a group based on maximum likelihood estimates of net ecotype formation, periodic selection, and drift rates. The ecotypes are identified on phylogenetic trees calculated using Neighbor-joining as described in Chapter 2. These trees were derived from *rpoC1* sequence alignments. All available clade I and IV *rpoC1* sequences from cultured strains and environmental libraries were used in the alignments, as described in Chapter 2.

Results

For *Synechococcus* from clade I, 11 ecotypes were identified (Figure A4.1a). Ten ecotypes were estimated within clade IV (Figure A4.1b). For clade IV, the strains for which genome sequences are available, CC9902 and BL107, are in separate ecotypes.

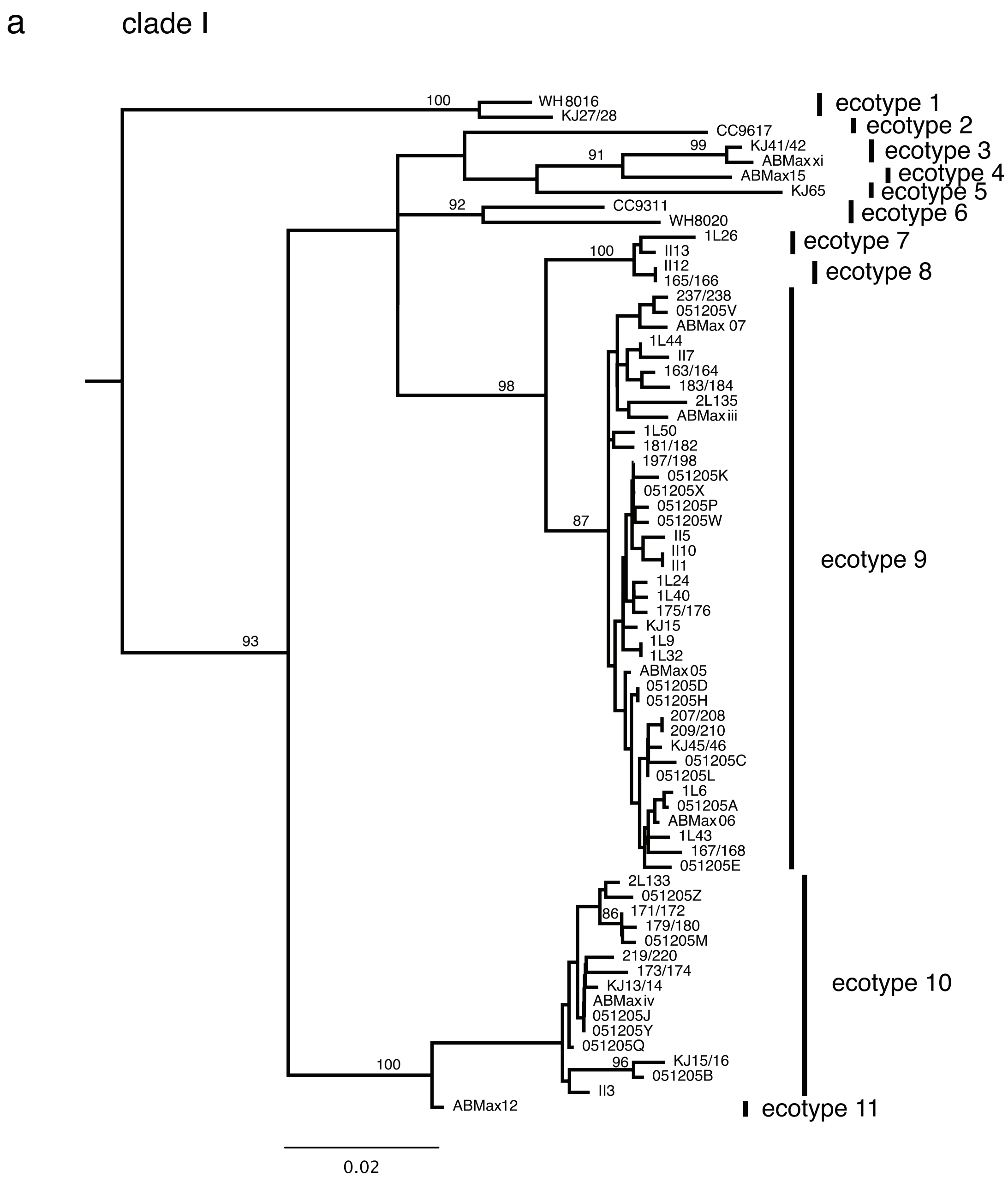


Figure A4.1. Ecotypes, as determined using Ecotype Simulation, labeled on Neighbor-joining trees of *rpoC1* sequences belonging to a) *Synechococcus* from clade I and b) *Synechococcus* from clade IV. Bootstrap values are shown on branches with 85% or greater support. The clade I and clade IV trees were rooted with the CC9902 and CC9311 *rpoC1* sequences, respectively.

b clade IV

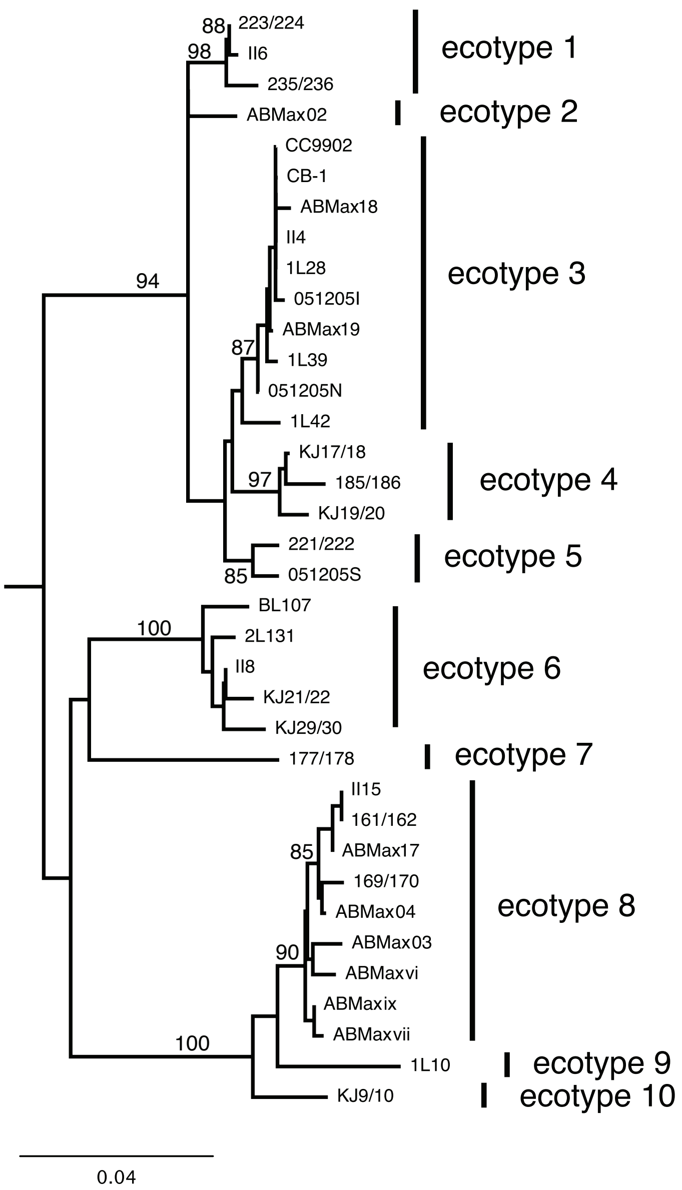


Figure A4.1. Ecotypes, continued.

References

Koeppel, A., Perry, E.B., Sikorski, J., Krizanc, D., Warner, W.A., Ward, D.M., Rooney, A.P., Brambilla, E., Connor, N., Ratcliff, R.M., Nevo, E., and Cohan, F.M. 2008. Identifying the fundamental units of bacterial diversity: a paradigm shif to incorporate ecology into bacterial systematics. *Proc Natl Acad Sci U S A* 105:2504-2509.

APPENDIX 5

dN/dS ratios from gene alignments using *Synechococcus* metagenomes

Table A5.1. dN/dS ratios calculated from the tiling of metagenome sequences to the *Synechococcus* CC9311 genome (see Chapter 3). dN/dS relative to the CC9311 genome is also provided (dN/dS CC9311). NaN indicates that there were no polymorphisms found in the regions aligned with at least 5-fold read coverage Inf indicates that only non-synonymous polymorphisms were found.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_0001	DNA polymerase III, beta subunit (EC:2.7.7.7)	0.000	0.094	0.000	0.000	98.5	21.1	3.5
sync_0002	hypothetical protein	0.010	0.016	0.605	1.005	86.5	11.4	1.5
sync_0003	phosphoribosylformylglycinamidine synthase II (EC:6.3.5.3)	0.021	0.137	0.154	0.156	93.4	56.7	5.7
sync_0004	amidophosphoribosyltransferase (EC:2.4.2.14)	0.005	0.127	0.043	0.046	100.0	100.0	10.2
sync_0005	DNA gyrase subunit A (EC:5.99.1.3)	0.011	0.175	0.064	0.068	99.7	93.2	11.5
sync_0006	tetratricopeptide repeat protein	0.013	0.211	0.063	0.066	100.0	100.0	14.2
sync_0007	iron-sulfur cluster binding protein, putative	0.008	0.128	0.059	0.058	100.0	82.2	8.3
sync_0008	hypothetical protein	0.009	0.185	0.050	0.058	100.0	100.0	13.2
sync_0009	hypothetical protein	0.002	0.140	0.015	0.028	100.0	88.8	8.8
sync_0010	putative transcription antitermination factor NusB	0.003	0.135	0.020	0.034	100.0	83.6	9.3
sync_0011	signal recognition particle-docking protein FtsY	0.011	0.146	0.073	0.083	90.9	71.0	6.8
sync_0012	SpoIIE domain protein	0.007	0.149	0.048	0.049	97.8	83.6	8.1
sync_0013	argininosuccinate lyase (EC:4.3.2.1)	0.011	0.141	0.079	0.078	100.0	99.3	10.4
sync_0014	RNA-binding region RNP-1 (RNA recognition motif)	0.011	0.165	0.067	0.059	100.0	92.7	12.6
sync_0015	tRNA-dihydrouridine synthase A (EC:1.-)	0.011	0.183	0.062	0.064	100.0	100.0	15.9
sync_0016	methionine-R-sulfoxide reductase (EC:1.8.4.-)	0.014	0.120	0.117	0.127	100.0	100.0	17.9
sync_0017	conserved hypothetical protein TIGR00275	0.040	0.224	0.178	0.175	100.0	99.2	20.1
sync_0018	hypothetical protein	0.000	0.000	NaN	NaN	14.2	0.0	0.2
sync_0019	type IV pilus assembly protein PilC	0.029	0.316	0.092	0.096	100.0	100.0	15.9
sync_0020	twitching motility protein	0.017	0.266	0.063	0.064	100.0	100.0	15.8
sync_0021	type IV pilus assembly protein PilB	0.014	0.223	0.064	0.071	100.0	79.4	10.6
sync_0022	co-chaperone GrpE	0.020	0.134	0.148	0.143	100.0	100.0	11.1
sync_0023	DnaJ protein	0.007	0.244	0.027	0.023	100.0	98.7	10.1
sync_0024	hypothetical protein	0.030	0.093	0.327	0.485	100.0	55.6	4.7
sync_0025	ribosome-associated GTPase YjeQ	0.015	0.134	0.112	0.107	100.0	90.4	7.2
sync_0026	conserved hypothetical protein TIGR00103	0.016	0.146	0.106	0.102	100.0	94.4	6.9
sync_0027	UDP-N-acetylenolpyruvoylglucosamine reductase (EC:1.1.1.158)	0.012	0.116	0.102	0.101	100.0	74.8	5.9
sync_0028	UDP-N-acetylmuramate--alanine ligase (EC:6.3.2.8)	0.032	0.181	0.174	0.177	98.8	79.3	6.6
sync_0029	glyceraldehyde-3-phosphate dehydrogenase, type I (EC:1.2.1.-)	0.010	0.241	0.040	0.039	100.0	27.2	4.0
sync_0030	thiamine-monophosphate kinase (EC:2.7.4.16)	0.017	0.168	0.104	0.094	94.4	24.8	3.4
sync_0031	peptidyl-prolyl cis-trans isomerase, cyclophilin-type (EC:5.2.1.8)	0.009	0.144	0.065	0.069	100.0	66.9	6.5
sync_0032	translation elongation factor P	0.004	0.061	0.060	0.061	100.0	87.1	5.3
sync_0033	acetyl-CoA carboxylase, biotin carboxyl carrier protein	0.006	0.083	0.074	0.068	100.0	79.4	6.7
sync_0034	4-hydroxythreonine-4-phosphate dehydrogenase (EC:1.1.1.262)	0.019	0.095	0.202	0.196	94.1	47.8	4.3
sync_0035	NAD dependent epimerase/dehydratase	0.026	0.133	0.198	0.173	100.0	85.2	7.0
sync_0036	possible Squash family serine protease inhibit	0.028	0.156	0.178	0.189	100.0	100.0	9.6
sync_0037	HNH endonuclease:HNH nuclease	0.025	0.158	0.156	0.138	100.0	89.1	5.8
sync_0038	possible helicase	0.010	0.176	0.058	0.069	100.0	83.6	6.8
sync_0039	possible Penicillin amidase	0.030	0.256	0.118	0.079	91.3	23.3	3.2
sync_0040	hypothetical protein	0.001	0.171	0.006	0.017	100.0	94.6	6.9
sync_0041	possible methyl-accepting chemotaxis protein	0.034	0.076	0.441	0.270	100.0	57.4	5.1
sync_0042	hypothetical protein	0.036	0.245	0.149	0.172	100.0	100.0	12.3
sync_0043	hypothetical protein	0.052	0.145	0.358	0.325	100.0	100.0	10.2
sync_0045	soluble hydrogenase, tritium exchange subunit, putative	0.016	0.187	0.088	0.083	100.0	92.5	8.5
sync_0046	cobalamin biosynthesis protein CbiD	0.037	0.179	0.208	0.230	97.4	58.4	4.9
sync_0047	GMP synthase, glutamine-hydrolyzing (EC:6.3.5.2)	0.015	0.311	0.047	0.053	100.0	99.1	19.0
sync_0048	hypothetical protein	0.000	0.000	NaN	NaN	23.9	0.0	0.2
sync_0049	Predicted membrane protein	0.012	0.165	0.074	0.071	100.0	100.0	10.2
sync_0050	hypothetical protein	0.025	0.145	0.171	0.198	100.0	100.0	10.2

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_0051	Putative penicillin-binding protein	0.014	0.225	0.064	0.069	100.0	100.0	15.5
sync_0052	glycosyl transferase, group 1 family protein (EC:2.4.1.-)	0.008	0.185	0.041	0.044	100.0	99.8	10.7
sync_0053	sulfolipid biosynthesis protein SqdB	0.014	0.244	0.057	0.053	100.0	100.0	16.0
sync_0054	possible high light inducible protein	0.011	0.070	0.159	0.119	100.0	100.0	11.9
sync_0055	thiamin biosynthesis protein (EC:1.5.3.-)	0.006	0.110	0.050	0.055	100.0	68.7	6.1
sync_0056	hypothetical protein	0.013	0.094	0.136	0.140	100.0	90.2	7.9
sync_0057	putative photosystem I assembly related protein Ycf37	0.011	0.162	0.071	0.069	100.0	71.9	5.1
sync_0058	ribosomal protein L20	0.000	0.088	0.000	0.022	100.0	67.5	4.8
sync_0059	ribosomal protein L35	0.013	0.000	Inf	Inf	99.5	71.7	8.1
sync_0060	SpoIID/LytB domain protein	0.020	0.091	0.215	0.224	100.0	73.7	6.3
sync_0061	glycosyl transferase, group 2 family protein (EC:2.4.1.-)	0.010	0.095	0.108	0.092	100.0	74.2	6.1
sync_0062	DNA polymerase III, subunits gamma and tau (EC:2.7.7.7)	0.016	0.168	0.095	0.090	89.5	75.2	7.1
sync_0063	hypothetical protein	0.023	0.150	0.152	0.151	100.0	100.0	9.7
sync_0064	ATP-dependent Clp protease, ATP-binding subunit ClpX	0.009	0.196	0.045	0.043	100.0	100.0	12.6
sync_0065	ATP-dependent Clp protease, proteolytic subunit ClpP (EC:3.4.21.92)	0.011	0.191	0.056	0.054	100.0	100.0	13.6
sync_0066	trigger factor (EC:5.2.1.8)	0.005	0.103	0.050	0.052	100.0	98.0	10.5
sync_0067	aspartate-semialdehyde dehydrogenase (EC:1.2.1.11)	0.014	0.175	0.081	0.089	100.0	99.7	11.8
sync_0068	dihydrodipicolinate synthase (EC:4.2.1.52)	0.014	0.147	0.093	0.091	100.0	93.7	7.2
sync_0069	metallo-beta-lactamase family protein	0.002	0.104	0.019	0.019	97.7	62.3	5.8
sync_0070	hypothetical protein	0.012	0.333	0.037	0.036	75.3	22.3	3.0
sync_0071	possible MesJ homolog	0.037	0.155	0.239	0.237	97.6	71.6	6.5
sync_0072	hypothetical protein	0.016	0.173	0.091	0.075	87.6	72.9	5.9
sync_0073	hypothetical protein	0.007	0.169	0.039	0.036	100.0	35.0	4.1
sync_0074	excinuclease ABC, B subunit	0.011	0.272	0.039	0.039	100.0	94.5	8.5
sync_0075	asparate kinase, monofunctional class (EC:2.7.2.4)	0.007	0.179	0.039	0.048	92.1	77.5	6.6
sync_0076	Curli production assembly/transport component CsgG subfamily protein	0.024	0.163	0.148	0.156	96.0	16.5	2.5
sync_0077	DNA polymerase III, delta subunit (EC:2.7.7.7)	0.009	0.209	0.043	0.059	98.3	23.6	3.2
sync_0078	precorrin-8X methylmutase (EC:5.4.1.2)	0.008	0.110	0.073	0.088	100.0	42.7	4.1
sync_0079	DNA mismatch repair protein MutS	0.016	0.218	0.075	0.078	96.8	52.6	5.4
sync_0080	photosystem II core protein PsbZ	0.031	0.218	0.141	0.135	100.0	73.0	6.1
sync_0081	6,7-dimethyl-8-ribityllumazine synthase (EC:6.3.3.-)	0.023	0.228	0.102	0.101	100.0	99.2	9.2
sync_0083	glyoxalase family protein family	0.032	0.107	0.295	0.272	100.0	66.3	5.6
sync_0084	hypothetical protein	0.005	0.086	0.059	0.054	100.0	60.8	6.1
sync_0085	glycosyl transferase, group 1 family protein (EC:2.4.1.-)	0.027	0.162	0.168	0.142	93.2	23.2	2.9
sync_0086	preprotein translocase, SecA subunit	0.029	0.480	0.061	0.059	100.0	94.7	12.5
sync_0087	serine O-acetyltransferase (EC:2.3.1.30)	0.009	0.416	0.022	0.023	100.0	70.8	8.6
sync_0088	transcriptional regulator, GntR family protein	0.007	0.301	0.024	0.028	100.0	100.0	13.0
sync_0089	Dienelactone hydrolase (EC:3.1.1.45)	0.050	0.183	0.272	0.282	100.0	92.8	9.5
sync_0090	translation initiation factor IF-3	0.009	0.209	0.041	0.034	100.0	100.0	8.2
sync_0091	tRNA delta(2)-isopentenylpyrophosphate transferase (EC:2.5.1.8)	0.030	0.177	0.169	0.169	100.0	86.1	8.1
sync_0092	DNA gyrase, B subunit (EC:5.99.1.3)	0.006	0.286	0.022	0.025	100.0	100.0	14.0
sync_0093	hypothetical protein	0.045	0.354	0.127	0.132	100.0	100.0	11.4
sync_0094	Protein CrcB homolog 1	0.030	0.167	0.180	0.172	100.0	100.0	9.5
sync_0095	crcB protein	0.054	0.287	0.189	0.188	100.0	97.0	8.5
sync_0096	Glutathione peroxidase (EC:1.11.1.9)	0.025	0.234	0.106	0.106	100.0	96.7	17.5
sync_0097	magnesium transporter	0.017	0.238	0.070	0.072	100.0	100.0	13.5
sync_0098	Type II alternative RNA polymerase sigma factor, sigma-70 family protein	0.015	0.313	0.048	0.047	100.0	100.0	16.2
sync_0099	Predicted hydrolase	0.047	0.342	0.137	0.136	100.0	100.0	14.0
sync_0100	acriflavin resistance protein acrF	0.027	0.268	0.100	0.099	100.0	100.0	15.2
sync_0101	efflux transporter, RND family, MFP subunit subfamily	0.030	0.249	0.120	0.121	100.0	100.0	17.4
sync_0102	putative A/G-specific adenine glycosylase	0.050	0.315	0.157	0.153	99.7	40.7	4.8
sync_0103	Sugar kinase, ribokinase family protein (EC:2.7.1.4)	0.040	0.241	0.167	0.194	82.2	51.4	4.5
sync_0104	probable sugar kinase YPO1816 , putative	0.000	0.000	NaN	NaN	90.6	0.0	1.7
sync_0105	putative glutamine synthetase	0.000	0.000	NaN	NaN	82.3	4.8	2.0
sync_0106	transporter, solute:sodium symporter (SSS) family protein	0.032	0.179	0.177	0.184	95.4	70.0	6.6
sync_0107	conserved hypothetical protein TIGR00150	0.035	0.157	0.219	0.205	99.8	67.1	5.9
sync_0108	adenosylhomocysteinase (EC:3.3.1.1)	0.017	0.218	0.077	0.075	100.0	92.5	11.3
sync_0109	Uncharacterized DedA family conserved membrane protein	0.010	0.192	0.052	0.054	100.0	100.0	11.4
sync_0110	putative single-stranded DNA-binding protein	0.006	0.202	0.030	0.031	100.0	100.0	8.6
sync_0111	Rod shape determining protein	0.007	0.126	0.059	0.057	99.5	59.4	4.7
sync_0112	Cell shape-determining protein	0.016	0.151	0.102	0.104	99.9	69.6	6.9
sync_0113	hypothetical protein	0.008	0.126	0.064	0.062	100.0	70.2	6.5
sync_0114	Bacterial extracellular solute-binding protein, family 1	0.020	0.207	0.098	0.093	100.0	80.4	6.5
sync_0115	two-component response regulator	0.000	0.000	NaN	NaN	85.0	0.0	1.8

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_0116	lysyl-tRNA synthetase (EC:6.1.1.6)	0.011	0.122	0.086	0.071	96.1	65.5	5.9
sync_0117	hypothetical protein	0.000	0.000	NaN	NaN	100.0	15.2	3.0
sync_0118	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	2.2
sync_0119	hypothetical protein	0.000	0.082	0.000	0.000	100.0	25.6	3.9
sync_0120	hypothetical protein	0.026	0.082	0.320	0.274	100.0	25.1	3.3
sync_0121	Domain of unknown function (DUF323) family protein	0.040	0.213	0.186	0.192	90.0	19.9	3.0
sync_0122	serine/threonine protein kinase (EC:2.7.11.1)	0.019	0.171	0.109	0.109	100.0	42.4	3.9
sync_0123	SsrA-binding protein	0.001	0.131	0.005	0.019	100.0	73.5	6.4
sync_0124	Holliday junction DNA helicase RuvB	0.005	0.104	0.044	0.047	100.0	48.5	5.2
sync_0125	tetratricopeptide repeat protein	0.012	0.173	0.069	0.064	100.0	71.7	5.2
sync_0126	peptidase, M20D family protein (EC:3.4.17.-)	0.011	0.144	0.073	0.074	100.0	71.3	6.4
sync_0127	hypothetical protein	0.006	0.133	0.047	0.056	100.0	100.0	10.9
sync_0128	CotB mutant	0.041	0.219	0.185	0.187	100.0	98.1	11.3
sync_0129	thiamine biosynthesis protein ThiC	0.021	0.481	0.044	0.047	100.0	100.0	22.3
sync_0130	transketolase (EC:2.2.1.1)	0.012	0.328	0.036	0.035	100.0	99.6	10.6
sync_0131	3-oxoacyl-[acyl-carrier-protein] synthase II (EC:2.3.1.41)	0.011	0.181	0.060	0.058	100.0	97.9	10.4
sync_0132	acyl carrier protein	0.003	0.148	0.022	0.022	100.0	100.0	17.3
sync_0133	photosystem I iron-sulfur protein PsaC	0.002	0.345	0.005	0.005	100.0	100.0	14.2
sync_0134	glucosamine--fructose-6-phosphate aminotransferase, isomerizing (EC:2.6.1.16)	0.016	0.285	0.057	0.058	100.0	82.2	7.8
sync_0135	Mycobacterium tuberculosis PIN domain family subfamily	0.059	0.196	0.303	0.341	100.0	53.7	5.4
sync_0136	hypothetical protein	0.099	0.187	0.531	0.449	100.0	47.6	4.0
sync_0137	Di/tricarboxylate transporter	0.000	0.000	NaN	0.044	79.8	4.3	2.2
sync_0138	hypothetical protein	0.000	0.000	NaN	NaN	5.4	0.0	0.1
sync_0139	putative glycosyl transferase	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0140	hypothetical protein	0.000	0.000	NaN	NaN	27.9	0.0	0.3
sync_0141	Glycosyl transferase family 11	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0142	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0143	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0144	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0145	GDP-L-fucose synthetase (EC:1.1.1.271)	0.115	0.725	0.159	0.164	89.4	78.9	12.5
sync_0146	GDP-mannose 4,6-dehydratase (EC:4.2.1.47)	0.056	0.564	0.100	0.101	100.0	98.2	14.2
sync_0147	possible sugar transferase (EC:2.7.-)	0.000	0.000	NaN	NaN	26.5	0.0	0.3
sync_0148	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0149	gumB protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0150	Chain length determinant protein family protein	0.000	0.000	NaN	NaN	3.4	0.0	0.1
sync_0151	putative phosphatase	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0152	hypothetical protein	0.000	0.000	NaN	NaN	43.5	0.0	0.6
sync_0153	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0154	hypothetical protein	0.000	0.000	NaN	NaN	19.0	0.0	0.2
sync_0155	hypothetical protein	0.000	0.000	NaN	NaN	61.2	0.0	1.0
sync_0156	polyketide synthase, putative	0.000	0.000	NaN	NaN	17.3	0.0	0.2
sync_0157	hypothetical protein	0.000	0.000	NaN	NaN	29.6	0.0	0.5
sync_0158	hypothetical protein	0.000	0.000	NaN	NaN	33.7	0.0	0.4
sync_0159	dTDP-glucose 4-6-dehydratase-like protein (EC:4.2.1.46)	0.045	0.507	0.089	0.088	99.1	74.8	11.1
sync_0160	mannose-1-phosphate guanylyltransferase/mannose-6-phosphate isomerase	0.028	0.381	0.074	0.086	99.8	80.1	7.4
sync_0161	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	1.0
sync_0162	Chain length determinant protein family protein	0.000	0.000	NaN	NaN	2.2	0.0	0.0
sync_0163	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	1.3
sync_0164	hypothetical protein	0.000	0.000	NaN	NaN	98.6	0.0	3.9
sync_0165	hypothetical protein	0.037	0.222	0.167	0.147	100.0	100.0	8.7
sync_0166	nucleotidyltransferase domain protein	0.059	0.178	0.332	0.307	98.8	43.9	4.4
sync_0167	hypothetical protein	0.000	0.000	NaN	NaN	42.1	0.0	0.8
sync_0168	hypothetical protein	0.000	0.000	NaN	NaN	76.3	0.0	0.8
sync_0169	UDP-N-acetylglucosamine 2-epimerase (EC:5.1.3.14)	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0170	CMP-N-acetylneuraminic acid synthetase (EC:2.7.7.43)	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0171	integral membrane protein MviN	0.000	0.000	NaN	NaN	17.5	0.0	0.2
sync_0172	imidazole glycerol phosphate synthase, glutamine amidotransferase subunit (EC:2.4.2.-)	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0173	pilin glycosylation protein PglB NMB1820	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0174	N-acetylneuraminic acid condensing enzyme (EC:4.1.3.-)	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0175	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0176	Imidazole glycerol phosphate synthase subunit hisF (EC:4.1.3.-)	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0177	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0178	asparagine synthase (glutamine-hydrolyzing) (EC:6.3.5.4)	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0179	DegT/DnrJ/EryC1/StrS aminotransferase family protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_0180	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0181	NeuB family protein	0.000	0.000	NaN	NaN	4.6	0.0	0.0
sync_0182	Posttranslational flagellin modification protein B (EC:2.7.7.43)	0.000	0.000	NaN	NaN	42.0	0.0	0.4
sync_0183	undecaprenyl-phosphate glucosephosphotransferase (EC:2.7.-)	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0184	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0185	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0186	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0187	short chain dehydrogenase (EC:1.-)	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0188	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0189	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0190	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0191	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0192	Putative nucleotide sugar epimerase/dehydratase	0.000	0.000	NaN	NaN	23.2	0.0	0.2
sync_0193	hypothetical protein	0.300	1.000	0.300	0.267	62.6	47.1	3.3
sync_0194	Di/tricarboxylate transporter	0.029	0.413	0.071	0.075	100.0	94.9	11.7
sync_0195	16S rRNA processing protein RimM	0.056	0.271	0.208	0.195	100.0	10.6	2.7
sync_0196	hypothetical protein	0.014	0.240	0.059	0.069	100.0	100.0	7.4
sync_0197	cation transporter, voltage-gated ion channel (VIC) family protein	0.021	0.258	0.082	0.109	76.9	43.0	3.8
sync_0198	PIN domain, putative	0.000	0.000	NaN	NaN	26.4	0.0	0.5
sync_0199	prevent-host-death family protein subfamily, putative	0.000	0.000	NaN	NaN	34.6	0.0	0.5
sync_0200	ribonuclease III (EC:3.1.26.3)	0.040	0.297	0.136	0.159	99.5	55.2	5.1
sync_0202	large conductance mechanosensitive channel protein	0.026	0.199	0.130	0.156	100.0	55.0	4.6
sync_0203	hypothetical protein	0.011	0.179	0.063	0.072	88.6	82.8	5.1
sync_0204	hypothetical protein	0.007	0.205	0.032	0.030	100.0	90.6	6.3
sync_0205	phosphorylase (EC:2.4.1.1)	0.013	0.212	0.062	0.063	100.0	58.4	5.5
sync_0206	transporter, monovalent cation:proton antiporter-2 (CPA2) family protein	0.005	0.163	0.033	0.040	100.0	62.8	5.2
sync_0207	hypothetical protein	0.042	0.226	0.187	0.149	70.4	39.7	3.0
sync_0208	hydrolase, alpha/beta fold family protein	0.007	0.135	0.053	0.047	99.7	36.0	3.8
sync_0209	Aldose 1-epimerase subfamily protein	0.000	0.000	NaN	NaN	46.9	0.0	0.8
sync_0210	glycolate oxidase chain (EC:1.1.3.15)	0.000	0.000	NaN	NaN	91.0	0.0	1.8
sync_0211	hypothetical protein	0.031	0.136	0.229	0.196	87.1	16.4	2.6
sync_0212	glyoxalase family protein family	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0213	Fe-S oxidoreductase (EC:1.1.3.15)	0.000	0.000	NaN	NaN	89.6	5.3	2.5
sync_0214	isocitrate dehydrogenase, NADP-dependent (EC:1.1.1.42)	0.006	0.189	0.032	0.033	95.7	54.5	5.1
sync_0215	hypothetical protein	0.050	0.260	0.191	0.211	100.0	85.1	12.6
sync_0216	Heme oxygenase (EC:1.14.99.3)	0.013	0.324	0.041	0.046	100.0	100.0	13.8
sync_0217	hypothetical protein	0.037	0.117	0.320	0.364	17.5	8.7	0.8
sync_0218	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0219	ATP-binding ABC transporter family protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0220	possible glycosyltransferase family 1	0.000	0.000	NaN	NaN	4.5	0.0	0.0
sync_0221	possible glycosyltransferase group 1	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0222	glycosyl transferase, group 1 family protein (EC:2.4.1.-)	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0223	Glycosyl transferase, group 1 (EC:2.-)	0.000	0.000	NaN	NaN	12.3	0.0	0.1
sync_0224	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0225	hypothetical protein	0.000	0.000	NaN	NaN	32.8	0.0	0.7
sync_0226	hypothetical protein (EC:2.8.2.-)	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0227	ATP-dependent DNA helicase PcrA (EC:3.6.1.-)	0.011	0.324	0.034	0.038	100.0	83.5	7.5
sync_0228	polyA polymerase family protein	0.044	0.197	0.222	0.219	78.0	17.9	2.3
sync_0229	hypothetical protein	0.000	0.000	NaN	NaN	51.2	0.0	0.7
sync_02NaN	Selenide,water dikinase	0.000	0.000	NaN	NaN	33.3	0.0	0.3
sync_0231	UDP-glucose 4-epimerase (EC:5.1.3.2)	0.056	0.309	0.181	0.194	100.0	49.5	5.6
sync_0232	histidyl-tRNA synthetase (EC:6.1.1.21)	0.031	0.237	0.132	0.138	99.1	83.2	8.2
sync_0233	UDP-glucose dehydrogenase (EC:1.1.1.22)	0.030	0.325	0.092	0.086	98.6	78.2	8.3
sync_0234	WbnF (EC:5.1.3.-)	0.036	0.310	0.118	0.107	100.0	87.1	9.5
sync_0235	hypothetical protein	0.040	0.251	0.157	0.200	87.6	50.3	4.4
sync_0236	photosystem II reaction center protein PsbJ	0.000	0.200	0.000	0.000	100.0	56.7	6.8
sync_0237	photosystem II reaction center protein PsbL	0.008	0.150	0.051	0.048	100.0	100.0	16.4
sync_0238	cytochrome b559, beta subunit	0.003	0.125	0.028	0.028	100.0	100.0	12.3
sync_0239	cytochrome b559, alpha subunit	0.002	0.126	0.015	0.014	100.0	100.0	13.0
sync_0240	photosystem II stability/assembly factor HCF136	0.009	0.237	0.039	0.042	100.0	100.0	11.1
sync_0241	Rubredoxin	0.002	0.147	0.012	0.012	100.0	57.8	5.6
sync_0242	NADH dehydrogenase I chain 3 (or A) (EC:1.6.5.-)	0.008	0.137	0.056	0.058	100.0	99.8	8.9
sync_0243	NADH dehydrogenase I chain B or NdhK (EC:1.6.5.-)	0.011	0.197	0.057	0.057	100.0	89.6	7.2
sync_0244	NADH dehydrogenase I chain J (EC:1.6.5.-)	0.025	0.179	0.140	0.123	100.0	9.8	3.3

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_0245	Uncharacterized conserved membrane-associated protein	0.015	0.202	0.073	0.065	100.0	45.1	4.6
sync_0246	possible ABC transporter, ATP-binding component	0.016	0.179	0.089	0.092	96.4	56.2	4.9
sync_0247	possible ABC transporter	0.028	0.225	0.123	0.135	100.0	45.3	3.9
sync_0248	magnesium chelatase, ATPase subunit D (EC:6.6.1.1)	0.007	0.124	0.060	0.067	100.0	54.3	5.2
sync_0249	2-amino-4-hydroxy-6- hydroxymethyldihydropteridine pyrophosphokinase (EC:2.7.6.3)	0.000	0.053	0.000	0.000	100.0	9.2	2.0
sync_0250	NUDIX hydrolase	0.015	0.117	0.131	0.209	100.0	55.2	4.5
sync_0251	NUDIX hydrolase (EC:3.6.1.13)	0.000	0.000	NaN	NaN	58.5	0.0	0.7
sync_0252	deoxyribodipyrimidine photolyase (EC:4.1.99.3)	0.013	0.164	0.082	0.058	99.3	27.5	3.8
sync_0253	DegT/DnrJ/EryC1/StrS aminotransferase family protein	0.025	0.191	0.133	0.133	100.0	84.3	9.3
sync_0254	Thioredoxin family protein	0.044	0.209	0.211	0.209	100.0	88.7	11.0
sync_0255	lipoprotein, putative	0.039	0.229	0.169	0.162	100.0	89.3	11.3
sync_0256	hypothetical protein	0.022	0.190	0.115	0.103	100.0	78.1	6.5
sync_0257	blue light photoreceptor cryptochrome (EC:4.1.99.3)	0.033	0.216	0.151	0.158	100.0	93.8	11.4
sync_0258	enoyl-[acyl-carrier-protein] reductase (EC:1.3.1.9)	0.003	0.167	0.017	0.015	97.8	64.9	6.5
sync_0259	hypothetical protein	0.034	0.147	0.232	0.239	74.8	14.4	2.7
sync_0260	imidazoleglycerol-phosphate dehydratase (EC:4.2.1.19)	0.010	0.155	0.064	0.058	100.0	52.5	4.6
sync_0261	Retinal pigment epithelial membrane protein	0.014	0.239	0.060	0.071	77.2	63.0	5.2
sync_0262	hypothetical protein	0.000	0.000	NaN	NaN	51.6	0.0	0.9
sync_0263	two component sensor histidine kinase, putative	0.000	0.000	NaN	NaN	4.0	0.0	0.0
sync_0264	probable transcriptional regulator, AraC family, putative	0.000	0.000	NaN	NaN	39.8	0.0	0.8
sync_0265	two-component response regulator	0.000	0.000	NaN	NaN	23.8	0.0	0.2
sync_0266	hypothetical protein	0.000	0.000	NaN	NaN	75.6	0.0	1.4
sync_0267	possible transporter, MFS family protein	0.019	0.136	0.143	0.138	100.0	90.8	7.2
sync_0268	Glycine betaine transporter	0.015	0.153	0.100	0.118	100.0	44.7	4.8
sync_0269	putative sarcosine oxidase	0.007	0.133	0.055	0.050	100.0	81.0	6.2
sync_0270	possible leucine dehydrogenase (EC:1.4.1.9)	0.009	0.132	0.069	0.067	96.2	39.7	3.9
sync_0271	hypothetical protein	0.012	0.131	0.088	0.076	94.6	30.2	4.8
sync_0272	non-canonical purine NTP pyrophosphatase	0.000	0.062	0.000	0.000	66.0	3.7	1.4
sync_0273	Phosphoglucomutase/phosphomannomutase family protein (EC:5.4.2.8)	0.017	0.280	0.061	0.037	83.1	9.4	2.3
sync_0274	hypothetical protein	0.018	0.137	0.133	0.130	95.4	41.7	3.9
sync_0275	Glycine cleavage T-protein (aminomethyl transferase) superfamily protein	0.021	0.072	0.289	0.295	100.0	65.3	5.6
sync_0276	orotate phosphoribosyltransferase (EC:2.4.2.10)	0.010	0.064	0.159	0.216	100.0	23.7	3.7
sync_0277	possible Occludin/ELL family protein	0.007	0.140	0.048	0.057	100.0	45.5	4.6
sync_0279	Conserved membrane protein containing two CBS domains	0.008	0.132	0.061	0.064	95.4	32.2	3.6
sync_0280	Predicted dehydrogenase (EC:1.1.1.-)	0.003	0.153	0.016	0.031	100.0	73.8	5.6
sync_0281	phosphorylase kinase alpha chain (EC:2.7.11.19)	0.012	0.128	0.096	0.102	97.6	24.6	3.0
sync_0282	Glycosyl transferase WecB/TagA/CpsF family protein (EC:2.4.1.-)	0.020	0.093	0.210	0.270	100.0	18.2	3.2
sync_0283	photosystem II reaction center protein PsbK	0.000	0.000	NaN	NaN	100.0	38.2	3.3
sync_0284	queuine tRNA-ribosyltransferase (EC:2.4.2.29)	0.007	0.107	0.064	0.061	100.0	77.5	7.4
sync_0285	cobalamin 5'-phosphate synthase (EC:2.7.8.26)	0.016	0.125	0.126	0.153	100.0	47.7	4.5
sync_0286	two-component sensor histidine kinase	0.021	0.194	0.107	0.123	92.8	62.4	6.0
sync_0287	hypothetical protein	0.005	0.027	0.190	0.257	100.0	60.2	5.3
sync_0288	Predicted esterase (EC:3.1.2.-)	0.029	0.171	0.169	0.168	100.0	80.8	10.6
sync_0289	bifunctional purine biosynthesis protein PurH (EC:2.1.2.3,EC:3.5.4.10)	0.013	0.126	0.101	0.095	100.0	73.0	7.8
sync_0290	hypothetical protein	0.004	0.122	0.036	0.032	97.5	89.0	9.1
sync_0291	hypothetical protein	0.011	0.219	0.052	0.047	100.0	86.9	8.9
sync_0292	4-hydroxy-3-methylbut-2-enyl diphosphate reductase (EC:1.17.1.2)	0.010	0.172	0.057	0.062	100.0	100.0	12.0
sync_0293	ammonium transporter	0.018	0.192	0.096	0.082	100.0	60.2	5.9
sync_0294	sugar fermentation stimulation protein	0.017	0.206	0.083	0.087	92.8	80.0	7.6
sync_0295	integral membrane protein MviN	0.012	0.114	0.110	0.118	100.0	97.2	9.2
sync_0296	possible Cytochrome c oxidase subunit Va	0.002	0.168	0.011	0.010	100.0	93.2	7.0
sync_0297	hypothetical protein	0.007	0.124	0.054	0.045	100.0	100.0	9.7
sync_0298	Uncharacterized membrane protein	0.003	0.092	0.038	0.038	100.0	52.2	5.0
sync_0300	serine hydroxymethyltransferase (EC:2.1.2.1)	0.008	0.170	0.048	0.045	100.0	99.8	12.3
sync_0301	glycosyl transferase, group 4 family protein (EC:2.7.8.-)	0.015	0.210	0.074	0.070	100.0	100.0	10.1
sync_0302	CinA-like protein	0.024	0.158	0.151	0.137	100.0	77.3	7.1
sync_0303	3-isopropylmalate dehydratase, large subunit (EC:4.2.1.33)	0.010	0.192	0.051	0.052	100.0	99.9	11.6
sync_0304	3-isopropylmalate dehydratase, small subunit (EC:4.2.1.33)	0.021	0.163	0.131	0.122	100.0	100.0	16.0
sync_0305	Pentapeptide repeats	0.025	0.192	0.129	0.124	99.4	99.4	13.1
sync_0306	hypothetical protein	0.022	0.406	0.055	0.051	57.4	3.3	1.3
sync_0307	photosystem II reaction center I protein	0.000	0.086	0.000	0.000	100.0	100.0	5.8
sync_0308	Glycoside hydrolase family 38 (EC:3.2.1.24)	0.029	0.174	0.167	0.163	100.0	86.2	7.6
sync_0309	photosystem II reaction center protein PsbN	0.000	0.097	0.000	0.000	100.0	100.0	8.0

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_0310	photosystem II reaction center protein PsbH	0.003	0.148	0.019	0.020	100.0	78.7	6.6
sync_0311	twin-arginine translocation protein TatA	0.055	0.335	0.165	0.163	100.0	100.0	11.9
sync_0312	peptidyl-tRNA hydrolase (EC:3.1.1.29)	0.016	0.197	0.080	0.071	100.0	93.9	7.5
sync_0313	hypothetical protein	0.004	0.162	0.026	0.021	100.0	87.2	8.2
sync_0314	hypothetical protein	0.023	0.083	0.280	0.245	100.0	69.1	4.9
sync_0315	hypothetical protein	0.015	0.175	0.084	0.078	100.0	83.6	8.2
sync_0316	nitrogen-responsive regulatory protein	0.001	0.143	0.008	0.007	100.0	58.2	5.7
sync_0317	ribonuclease PH (EC:2.7.7.56)	0.018	0.144	0.123	0.105	100.0	100.0	9.9
sync_0318	putative cob(I)alamin adenosyltransferase	0.016	0.206	0.079	0.072	100.0	93.9	7.8
sync_0319	deoxycytidine triphosphate deaminase (EC:3.5.4.13)	0.011	0.243	0.043	0.037	100.0	64.7	8.8
sync_0320	thymidylate synthase, flavin-dependent (EC:2.1.1.148)	0.004	0.210	0.021	0.021	100.0	56.1	6.5
sync_0321	thioredoxin	0.014	0.141	0.103	0.106	100.0	91.0	6.8
sync_0322	hypothetical protein	0.002	0.080	0.025	0.022	100.0	47.2	4.1
sync_0323	transglycosylase, SLT family protein	0.028	0.135	0.210	0.229	99.4	37.4	4.3
sync_0324	hypothetical protein	0.040	0.018	2.225	2.309	100.0	30.2	3.8
sync_0325	Phosphotransferase superclass (EC:5.4.2.3)	0.009	0.095	0.097	0.125	100.0	59.4	4.9
sync_0326	BadF/BadG/BcrA/BcrD ATPase family superfamily	0.027	0.077	0.350	0.287	100.0	86.4	7.7
sync_0327	putative inorganic carbon transporter/0-antigen polymerase (ICT/OAP) family protein	0.015	0.080	0.191	0.163	100.0	74.6	5.8
sync_0328	tRNA (guanine-N(7)-)-methyltransferase (EC:2.1.1.33)	0.019	0.086	0.221	0.214	100.0	87.8	7.9
sync_0329	hypothetical protein	0.012	0.148	0.080	0.085	100.0	81.3	8.2
sync_0330	hypothetical protein	0.012	0.170	0.073	0.084	100.0	100.0	11.4
sync_0331	hypothetical protein	0.015	0.116	0.131	0.192	74.7	48.4	3.7
sync_0332	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0333	isoleucyl-tRNA synthetase (EC:6.1.1.5)	0.013	0.141	0.095	0.095	99.9	95.2	9.9
sync_0334	hypothetical protein	0.032	0.212	0.149	0.137	88.7	61.8	6.5
sync_0336	beta-carotene hydroxylase	0.019	0.323	0.060	0.060	99.0	65.7	5.7
sync_0337	glutamyl-tRNA(Gln) amidotransferase, C subunit (EC:6.3.5.-)	0.013	0.169	0.075	0.081	100.0	100.0	9.2
sync_0338	creatininase (EC:3.5.2.10)	0.015	0.231	0.064	0.056	100.0	100.0	16.6
sync_0339	hypothetical protein	0.006	0.210	0.031	0.031	100.0	100.0	18.1
sync_0340	Transporter	0.017	0.273	0.062	0.068	100.0	100.0	15.7
sync_0341	hypothetical protein	0.017	0.064	0.272	0.184	100.0	93.0	9.4
sync_0342	possible Methylpurine-DNA glycosylase	0.068	0.288	0.236	0.240	100.0	89.8	20.1
sync_0343	aspartate carbamoyltransferase (EC:2.1.3.2)	0.031	0.279	0.112	0.117	100.0	100.0	21.1
sync_0344	NADH dehydrogenase (EC:1.6.99.3)	0.015	0.212	0.073	0.070	100.0	99.9	14.6
sync_0345	Uncharacterized secreted or membrane protein	0.019	0.224	0.087	0.082	100.0	100.0	16.1
sync_0346	hypothetical protein	0.009	0.275	0.034	0.040	100.0	100.0	15.0
sync_0347	hypothetical protein	0.046	0.176	0.263	0.190	100.0	100.0	7.9
sync_0348	hypothetical protein	0.020	0.152	0.130	0.134	91.8	72.9	4.7
sync_0349	beta-lactamase, putative	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0351	hypothetical protein	0.032	0.220	0.147	0.132	100.0	91.8	10.4
sync_0352	phosphopantothenoylcysteine decarboxylase/phosphopantothenate--cysteine ligase (EC:4.1.1.36,EC:6.3.2.5)	0.021	0.188	0.114	0.115	100.0	84.7	6.2
sync_0353	photosystem II manganese-stabilizing protein	0.011	0.195	0.057	0.056	100.0	100.0	13.7
sync_0354	sulfate adenylyltransferase (EC:2.7.7.4)	0.007	0.159	0.042	0.040	100.0	93.8	9.3
sync_0355	cell division protein FtsH (EC:3.4.24.-)	0.002	0.165	0.015	0.019	100.0	90.7	8.9
sync_0356	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase (EC:4.1.2.14,EC:4.1.3.16)	0.020	0.105	0.186	0.178	100.0	90.6	6.5
sync_0357	Predicted protein family PM-3	0.006	0.100	0.056	0.073	100.0	53.0	4.8
sync_0358	chorismate synthase (EC:4.2.3.5)	0.008	0.158	0.052	0.044	100.0	91.6	10.6
sync_0359	Cupin superfamily (DUF985) superfamily	0.014	0.102	0.142	0.196	100.0	53.5	5.7
sync_0360	possible Vng0271c	0.005	0.090	0.054	0.053	100.0	85.8	7.3
sync_0361	Glycosidase (EC:2.4.1.7)	0.017	0.117	0.142	0.146	100.0	98.7	8.1
sync_0362	mannosyl-3-phosphoglycerate phosphatase homolog (EC:3.1.3.-)	0.030	0.146	0.205	0.190	100.0	85.0	9.5
sync_0363	Predicted redox protein	0.026	0.165	0.159	0.155	100.0	100.0	8.2
sync_0364	PyrF/pyrE bifunctional enzyme (EC:4.1.1.23,EC:2.4.2.10)	0.014	0.038	0.371	0.412	50.6	2.3	1.0
sync_0365	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0366	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0367	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0368	photosystem II protein D1	0.039	0.459	0.084	0.087	100.0	87.3	15.6
sync_0369	radical SAM domain protein	0.015	0.232	0.063	0.060	99.4	89.2	8.1
sync_0370	CAAX amino terminal protease family protein	0.000	0.000	NaN	NaN	7.8	0.0	0.1
sync_0371	ATP-dependent Clp protease adaptor protein ClpS	0.000	0.108	0.000	0.033	100.0	44.9	4.2
sync_0372	aminotransferase, classes I and II (EC:2.6.1.-)	0.006	0.076	0.076	0.067	100.0	77.9	6.1
sync_0373	radical SAM domain protein	0.012	0.125	0.096	0.094	100.0	94.0	9.7

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_0374	ribonuclease, Rne/Rng family protein	0.007	0.142	0.047	0.047	97.8	70.8	6.1
sync_0375	ribonuclease HII (EC:3.1.26.4)	0.024	0.142	0.170	0.159	100.0	88.9	9.9
sync_0376	hypothetical protein	0.014	0.067	0.202	0.214	100.0	89.6	6.5
sync_0377	prephenate dehydratase (EC:4.2.1.51)	0.004	0.114	0.033	0.033	98.2	30.0	3.9
sync_0378	cyclopropane-fatty-acyl-phospholipid synthase family protein	0.007	0.122	0.059	0.065	100.0	74.0	7.2
sync_0379	ATP-dependent protease La (EC:3.4.21.53)	0.007	0.086	0.081	0.064	100.0	90.1	9.1
sync_0380	ribosomal protein S10	0.013	0.289	0.045	0.050	100.0	77.6	8.6
sync_0381	translation elongation factor Tu	0.008	0.137	0.055	0.062	100.0	100.0	9.0
sync_0382	translation elongation factor G	0.004	0.143	0.028	0.031	100.0	54.1	5.8
sync_0383	ribosomal protein S7	0.000	0.089	0.000	0.000	100.0	35.0	3.8
sync_0384	ribosomal protein S12	0.000	0.128	0.000	0.000	100.0	99.7	7.1
sync_0385	Uncharacterized HesB family conserved protein	0.000	0.090	0.000	0.000	100.0	25.7	3.0
sync_0386	hypothetical protein	0.021	0.126	0.165	0.164	99.8	29.5	3.9
sync_0387	Ferredoxin-dependent glutamate synthase, Fd-GOGAT (EC:1.4.7.1)	0.011	0.148	0.077	0.079	100.0	76.0	6.4
sync_0388	uncharacterized ycii family conserved protein	0.015	0.146	0.105	0.101	100.0	84.5	5.8
sync_0389	lipoic acid synthetase (EC:2.8.1.-)	0.008	0.128	0.065	0.057	100.0	75.5	5.7
sync_0391	cbiG protein/precorrin-3B C17-methyltransferase (EC:2.1.1.131)	0.028	0.137	0.206	0.220	100.0	93.0	9.6
sync_0392	Proline-rich region	0.035	0.202	0.173	0.179	82.8	46.8	4.5
sync_0393	photosystem I core protein PsaA	0.003	0.082	0.037	0.042	100.0	89.8	8.0
sync_0394	photosystem I core protein PsaB	0.005	0.093	0.054	0.053	89.0	56.7	5.1
sync_0395	hypothetical protein	0.020	0.152	0.133	0.126	99.2	47.7	4.6
sync_0396	possible fatty acid desaturase	0.030	0.221	0.136	0.132	100.0	83.6	8.8
sync_0397	transporter, hydrophobe/amphiphile efflux-1 (HAE1) family protein	0.016	0.175	0.089	0.086	100.0	92.8	8.7
sync_0398	photosystem I reaction center subunit XI	0.001	0.081	0.016	0.014	100.0	72.8	6.9
sync_0399	photosystem I reaction center subunit VIII	0.004	0.022	0.179	0.120	100.0	100.0	5.5
sync_0400	hypothetical protein	0.023	0.114	0.200	0.284	100.0	58.1	5.4
sync_0401	Glycosyl transferase, family 2 (EC:2.-)	0.007	0.160	0.045	0.042	100.0	97.1	9.1
sync_0402	Cell wall-associated hydrolase	0.033	0.171	0.193	0.177	100.0	100.0	12.6
sync_0403	Beta-lactamase superfamily protein	0.030	0.135	0.221	0.202	100.0	83.9	8.3
sync_0404	hypothetical protein	0.000	0.000	NaN	NaN	84.4	0.0	2.2
sync_0406	alanine racemase (EC:5.1.1.1)	0.018	0.166	0.111	0.107	100.0	100.0	13.0
sync_0407	HNH endonuclease family protein (EC:3.1.21.-)	0.006	0.231	0.028	0.027	100.0	100.0	16.7
sync_0408	peptide chain release factor 1	0.005	0.116	0.041	0.045	100.0	96.8	7.9
sync_0409	ribosomal protein L31	0.014	0.218	0.065	0.067	100.0	94.0	6.1
sync_0410	ribosomal protein S9	0.000	0.063	0.000	0.000	100.0	21.1	2.8
sync_0411	ribosomal protein L13	0.006	0.106	0.057	0.058	100.0	50.8	5.5
sync_0412	tRNA pseudouridine synthase A (EC:4.2.1.70)	0.008	0.106	0.072	0.081	100.0	93.6	7.3
sync_0413	ribosomal protein L17	0.000	0.113	0.000	0.000	100.0	49.6	5.9
sync_0414	DNA-directed RNA polymerase, alpha subunit (EC:2.7.7.6)	0.001	0.169	0.009	0.013	100.0	67.9	6.6
sync_0415	ribosomal protein S11	0.008	0.115	0.068	0.062	100.0	92.9	7.8
sync_0416	ribosomal protein S13	0.007	0.101	0.066	0.063	100.0	78.1	6.2
sync_0417	ribosomal protein L36	0.000	0.056	0.000	0.000	100.0	100.0	7.9
sync_0418	adenylate kinase (EC:2.7.4.3)	0.007	0.130	0.053	0.044	100.0	86.1	6.1
sync_0419	preprotein translocase, SecY subunit	0.001	0.180	0.007	0.009	100.0	81.2	7.2
sync_0420	ribosomal protein L15	0.005	0.129	0.036	0.036	100.0	89.2	5.9
sync_0421	ribosomal protein S5	0.025	0.116	0.214	0.205	100.0	95.7	7.8
sync_0422	ribosomal protein L18	0.004	0.040	0.093	0.112	100.0	100.0	10.1
sync_0423	ribosomal protein L6	0.006	0.094	0.064	0.071	100.0	99.8	8.4
sync_0424	ribosomal protein S8	0.005	0.035	0.131	0.130	100.0	100.0	8.4
sync_0425	ribosomal protein L5	0.003	0.130	0.020	0.026	100.0	100.0	12.3
sync_0426	ribosomal protein L24	0.010	0.113	0.092	0.085	100.0	100.0	15.1
sync_0427	ribosomal protein L14	0.002	0.147	0.017	0.017	100.0	100.0	6.6
sync_0428	ribosomal protein S17	0.017	0.144	0.119	0.119	100.0	100.0	9.9
sync_0429	ribosomal protein L29	0.005	0.038	0.135	0.135	100.0	85.2	6.4
sync_0430	ribosomal protein L16	0.003	0.084	0.035	0.033	100.0	67.7	4.8
sync_0431	ribosomal protein S3	0.002	0.112	0.018	0.017	100.0	95.8	8.6
sync_0432	ribosomal protein L22	0.003	0.072	0.044	0.042	100.0	80.1	7.4
sync_0433	ribosomal protein S19	0.002	0.065	0.025	0.025	100.0	100.0	7.4
sync_0434	ribosomal protein L2	0.006	0.095	0.060	0.050	100.0	85.7	7.4
sync_0435	ribosomal protein L23	0.002	0.025	0.068	0.068	100.0	92.1	8.0
sync_0436	ribosomal protein L4	0.006	0.101	0.061	0.062	100.0	82.7	8.3
sync_0437	ribosomal protein L3	0.010	0.112	0.089	0.080	100.0	98.9	7.4
sync_0438	hypothetical protein	0.007	0.051	0.132	0.132	93.1	15.4	2.8
sync_0439	Fe-S cluster containing protein	0.017	0.138	0.122	0.117	100.0	75.1	6.7
sync_0440	R3H domain protein	0.006	0.142	0.045	0.043	100.0	40.1	4.8

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_0442	HAD hydrolase homolog	0.013	0.096	0.140	0.155	95.0	55.0	4.7
sync_0443	recA protein	0.004	0.131	0.031	0.040	100.0	49.6	4.4
sync_0444	hypothetical protein	0.034	0.143	0.237	0.230	100.0	51.0	5.4
sync_0445	hypothetical protein	0.035	0.181	0.193	0.193	100.0	50.8	4.4
sync_0446	hypothetical protein	0.011	0.114	0.101	0.097	100.0	57.1	5.0
sync_0447	prephenate dehydrogenase (EC:1.3.1.12)	0.019	0.154	0.121	0.116	100.0	77.9	7.1
sync_0448	C-3',4' desaturase CrtD (EC:1.-)	0.027	0.130	0.206	0.208	100.0	98.4	10.4
sync_0449	fructosamine kinase family protein	0.050	0.199	0.253	0.255	100.0	100.0	9.0
sync_0450	Predicted membrane protein	0.003	0.199	0.015	0.011	100.0	100.0	9.2
sync_0451	hypothetical protein	0.039	0.129	0.301	0.351	100.0	100.0	8.2
sync_0452	putative molybdopterin biosynthesis protein MoeB	0.021	0.168	0.127	0.132	100.0	99.5	8.3
sync_0453	cob(I)alamin adenosyltransferase (EC:2.5.1.17)	0.009	0.285	0.033	0.031	100.0	95.0	10.4
sync_0454	PP-loop family protein	0.023	0.201	0.112	0.110	100.0	62.6	5.6
sync_0455	glutamate decarboxylase (EC:4.1.1.15)	0.016	0.252	0.063	0.059	98.9	76.4	8.5
sync_0456	DUF206	0.018	0.168	0.108	0.103	100.0	84.2	7.5
sync_0457	DNA replication and repair protein RecF	0.014	0.186	0.076	0.077	100.0	88.1	6.8
sync_0459	GNAT family acetyltransferase	0.012	0.083	0.151	0.151	100.0	48.5	4.9
sync_0460	phosphoenolpyruvate carboxylase (EC:4.1.1.31)	0.005	0.183	0.027	0.026	93.6	54.9	5.4
sync_0461	putative glutamate--cysteine ligase	0.014	0.132	0.105	0.106	100.0	58.1	6.7
sync_0462	putative anthranilate synthase component I	0.007	0.116	0.057	0.058	98.2	66.3	6.1
sync_0463	photosystem I reaction center subunit II	0.000	0.053	0.000	0.000	100.0	59.5	5.4
sync_0464	sensor histidine kinase	0.016	0.142	0.114	0.139	100.0	50.1	4.9
sync_0465	putative rod shape-determining protein RodA	0.013	0.129	0.104	0.091	100.0	99.0	9.0
sync_0466	MRP protein homolog	0.012	0.144	0.085	0.090	100.0	71.4	6.3
sync_0467	coproporphyrinogen III oxidase, aerobic (EC:1.3.3.3)	0.010	0.124	0.077	0.071	100.0	83.7	7.9
sync_0468	possible N-acetylmuramoyl-L-alanine amidase	0.036	0.193	0.186	0.212	100.0	75.7	6.7
sync_0469	hypothetical protein	0.006	0.093	0.061	0.056	100.0	90.6	8.4
sync_0470	possible Pex protein	0.004	0.070	0.051	0.049	100.0	66.7	6.6
sync_0471	3'-5' exonuclease family protein	0.017	0.141	0.123	0.125	100.0	97.5	8.7
sync_0472	hypothetical protein	0.018	0.099	0.178	0.217	100.0	97.8	7.3
sync_0473	hypothetical protein	0.032	0.183	0.176	0.174	100.0	100.0	15.0
sync_0475	hypothetical protein	0.042	0.102	0.415	0.442	89.9	21.2	2.7
sync_0476	cation transporter, voltage-gated ion channel (VIC) family protein	0.021	0.149	0.144	0.127	100.0	95.9	11.5
sync_0477	hypothetical protein	0.062	0.199	0.311	0.303	100.0	100.0	13.3
sync_0478	hypothetical protein	0.055	0.138	0.394	0.375	100.0	100.0	23.1
sync_0479	Predicted oxidoreductase	0.047	0.256	0.185	0.173	100.0	100.0	25.1
sync_0480	possible Hpt domain	0.010	0.225	0.046	0.063	100.0	100.0	19.9
sync_0481	phosphoribosylformylglycinamidine cyclo-ligase (EC:6.3.3.1)	0.032	0.165	0.192	0.183	100.0	100.0	19.7
sync_0482	possible rare lipoprotein A	0.022	0.145	0.154	0.122	100.0	79.0	9.2
sync_0483	pantoate--beta-alanine ligase/cytidylate kinase (EC:6.3.2.1,EC:2.7.4.14)	0.028	0.146	0.190	0.202	100.0	100.0	15.6
sync_0484	low molecular weight phosphotyrosine protein phosphatase family protein	0.015	0.126	0.118	0.131	100.0	100.0	19.2
sync_0485	phycocyanin alpha subunit phycocyanobilin lyase, CpcF subunit (EC:4.-)	0.037	0.196	0.189	0.183	100.0	100.0	23.0
sync_0486	phycocyanin alpha subunit phycocyanobilin lyase, CpcE subunit (EC:4.-)	0.049	0.202	0.241	0.251	100.0	100.0	20.5
sync_0487	Protein of unknown function (DUF1001) superfamily protein	0.044	0.293	0.151	0.161	100.0	100.0	20.5
sync_0488	phycocyanin, alpha subunit	0.011	0.186	0.061	0.062	100.0	100.0	29.3
sync_0489	phycocyanin, beta subunit	0.007	0.140	0.051	0.047	100.0	100.0	30.2
sync_0490	Phycoerythrobilin:ferredoxin oxidoreductase (EC:1.3.7.3)	0.033	0.241	0.137	0.143	100.0	100.0	41.1
sync_0491	15,16-dihydrobiliverdin:ferredoxin oxidoreductase (EC:1.3.7.2)	0.031	0.270	0.113	0.132	100.0	100.0	50.1
sync_0492	hypothetical protein	0.033	0.283	0.116	0.110	100.0	100.0	55.4
sync_0493	hypothetical protein	0.036	0.181	0.199	0.187	100.0	100.0	39.1
sync_0494	hypothetical protein	0.054	0.055	0.983	0.889	100.0	100.0	37.1
sync_0495	C-phycoerythrin class I beta chain	0.014	0.168	0.085	0.079	100.0	100.0	41.5
sync_0496	C-phycoerythrin class I alpha chain	0.012	0.288	0.042	0.037	100.0	100.0	30.4
sync_0497	Bilin biosynthesis protein mpeV	0.048	0.245	0.198	0.191	100.0	98.8	22.5
sync_0498	hypothetical protein	0.051	0.234	0.218	0.225	100.0	100.0	25.8
sync_0499	Bilin biosynthesis protein cpeY	0.061	0.224	0.270	0.283	100.0	100.0	19.2
sync_0500	Bilin biosynthesis protein cpeZ	0.046	0.252	0.182	0.200	100.0	100.0	30.1
sync_0501	Bilin biosynthesis protein mpeU	0.053	0.266	0.199	0.201	100.0	100.0	31.8
sync_0502	Phycoerythrin class II gamma chain, linker polypeptide	0.037	0.356	0.105	0.115	85.1	51.2	14.7
sync_0503	hypothetical protein	0.011	0.108	0.099	0.107	100.0	100.0	14.0
sync_0504	C-phycoerythrin class II alpha chain	0.021	0.313	0.067	0.068	100.0	100.0	46.4
sync_0505	C-phycoerythrin class II beta chain	0.023	0.256	0.089	0.089	100.0	100.0	44.4
sync_0506	CpeY protein	0.049	0.275	0.177	0.177	100.0	100.0	44.1

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_0507	hypothetical protein	0.039	0.250	0.158	0.177	100.0	100.0	35.9
sync_0508	hypothetical protein	0.090	0.239	0.378	0.399	100.0	100.0	45.4
sync_0509	hypothetical protein	0.062	0.257	0.241	0.226	100.0	100.0	35.1
sync_0510	phycoerythrin linker protein CpeS	0.038	0.287	0.131	0.131	100.0	100.0	38.7
sync_0511	possible phycobilisome linker polypeptide	0.035	0.276	0.127	0.140	100.0	91.3	26.7
sync_0512	phycobilisome linker polypeptide	0.020	0.311	0.065	0.067	100.0	100.0	46.8
sync_0513	phycobilisome linker polypeptide	0.020	0.298	0.067	0.080	100.0	100.0	45.6
sync_0514	pentapeptide repeat family protein	0.055	0.229	0.240	0.255	100.0	99.0	22.1
sync_0515	possible phycobilisome rod-core linker polypeptide (L-RC 28.5)	0.042	0.360	0.118	0.120	100.0	100.0	25.5
sync_0516	Possible phycobilisome linker polypeptide	0.047	0.404	0.116	0.111	100.0	100.0	32.1
sync_0517	possible Phycobilisome polypeptide	0.087	0.313	0.278	0.304	87.7	76.7	17.1
sync_0518	hypothetical protein	0.000	0.000	NaN	NaN	61.8	0.0	0.9
sync_0519	hypothetical protein	0.031	0.149	0.208	0.147	71.3	49.4	4.5
sync_0520	hypothetical protein	0.078	0.222	0.353	0.341	100.0	90.8	16.2
sync_0521	hypothetical protein	0.096	0.374	0.256	0.206	80.2	45.9	5.4
sync_0522	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0523	hypothetical protein	0.093	0.049	1.923	1.571	100.0	83.9	8.2
sync_0525	hypothetical protein	0.056	0.238	0.234	0.226	80.8	62.9	5.1
sync_0526	hypothetical protein	0.015	0.093	0.167	0.191	100.0	100.0	13.6
sync_0527	hypothetical protein	0.009	0.135	0.063	0.064	100.0	100.0	17.1
sync_0528	Carbohydrate kinase, FGGY family protein	0.029	0.152	0.188	0.195	100.0	100.0	13.5
sync_0529	S-adenosylmethionine synthetase (EC:2.5.1.6)	0.008	0.141	0.058	0.059	100.0	100.0	10.7
sync_0530	Haloacid dehalogenase/epoxide hydrolase family protein	0.031	0.134	0.232	0.217	100.0	96.1	10.1
sync_0531	ribosomal protein S1	0.002	0.161	0.012	0.015	100.0	100.0	11.6
sync_0532	transcriptional regulator, NrdR family protein	0.006	0.170	0.034	0.061	100.0	83.0	7.8
sync_0533	Photosystem II reaction centre T protein superfamily protein	0.000	0.015	0.000	0.000	100.0	100.0	9.6
sync_0534	photosystem II P680 chlorophyll A apoprotein	0.003	0.135	0.022	0.027	100.0	87.9	9.0
sync_0535	possible ferredoxin (2Fe-2S)	0.019	0.093	0.203	0.247	100.0	66.8	7.2
sync_0536	photosystem II reaction center protein PsbM	0.010	0.000	Inf	Inf	100.0	43.2	4.4
sync_0537	universal stress protein family protein	0.002	0.117	0.017	0.033	100.0	56.4	5.1
sync_0538	Predicted thioesterase	0.035	0.209	0.168	0.165	100.0	75.8	5.1
sync_0539	putative DNA protecting protein DprA	0.039	0.155	0.254	0.232	100.0	99.3	12.1
sync_0540	methyltransferase, HemK family protein	0.034	0.179	0.190	0.201	100.0	100.0	8.9
sync_0541	yrdC domain, putative	0.023	0.116	0.196	0.195	100.0	84.3	9.4
sync_0543	possible transcriptional regulator, luxR family protein	0.041	0.175	0.236	0.215	100.0	100.0	11.9
sync_0544	cell division topological specificity factor MinE	0.013	0.122	0.111	0.104	100.0	83.2	8.4
sync_0545	septum site-determining protein MinD	0.003	0.129	0.021	0.021	100.0	63.4	6.0
sync_0546	septum site-determining protein MinC	0.025	0.164	0.152	0.137	100.0	52.4	4.5
sync_0547	HD domain protein	0.008	0.143	0.059	0.066	100.0	72.0	6.5
sync_0548	carboxyl-terminal processing proteinase (EC:3.4.21.102)	0.008	0.117	0.067	0.073	100.0	87.0	8.3
sync_0549	cytochrome b6 (EC:1.10.99.1)	0.001	0.129	0.011	0.010	100.0	94.5	7.2
sync_0550	cytochrome b6-f complex, subunit IV	0.000	0.016	0.000	0.000	100.0	20.3	3.6
sync_0551	neutral invertase like protein	0.007	0.092	0.074	0.068	100.0	44.8	4.4
sync_0552	hypothetical protein	0.086	0.061	1.397	1.397	100.0	43.6	4.2
sync_0558	hypothetical protein	0.072	0.245	0.295	0.304	69.0	53.7	7.2
sync_0559	Sulfate permease	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0560	hypothetical protein	0.000	0.000	NaN	NaN	12.7	0.0	0.1
sync_0561	formamidopyrimidine-DNA glycosylase (EC:3.2.2.23)	0.028	0.239	0.119	0.121	100.0	80.8	9.1
sync_0562	photosystem I reaction center subunit IV	0.004	0.054	0.083	0.082	100.0	100.0	9.4
sync_0563	hypothetical protein	0.006	0.141	0.044	0.034	100.0	72.2	5.7
sync_0564	ATP-dependent DNA helicase RecQ (EC:3.6.1.-)	0.021	0.197	0.107	0.101	98.0	57.1	5.2
sync_0565	possible LysM domain	0.031	0.147	0.214	0.218	100.0	72.4	7.8
sync_0566	NAD-dependent aldehyde dehydrogenase (EC:1.2.1.3)	0.025	0.184	0.137	0.142	100.0	89.9	6.9
sync_0567	trehalose synthase (EC:5.4.99.16)	0.005	0.144	0.032	0.037	100.0	84.9	7.5
sync_0568	glycerol dehydrogenase homolog (EC:1.1.99.5)	0.020	0.128	0.158	0.156	100.0	97.8	11.0
sync_0569	Glycerol kinase (EC:2.7.1.30)	0.021	0.172	0.123	0.115	100.0	100.0	13.6
sync_0570	Glycoside hydrolase family protein (EC:3.2.1.135)	0.044	0.269	0.162	0.170	100.0	94.8	12.1
sync_0571	hypothetical protein	0.048	0.253	0.190	0.207	100.0	100.0	14.6
sync_0572	hypothetical protein	0.104	0.293	0.354	0.331	100.0	92.4	15.9
sync_0573	two-component hybrid sensor and regulator alr4878, putative	0.070	0.379	0.184	0.188	100.0	100.0	29.9
sync_0574	Two-component response regulator, CheY-like receiver and wHTH DNA-binding domains	0.064	0.276	0.233	0.226	100.0	100.0	30.6
sync_0575	hypothetical protein	0.179	0.217	0.823	0.831	100.0	34.7	5.2
sync_0578	3-dehydroquinate dehydratase (EC:4.2.1.10)	0.049	0.241	0.202	0.189	100.0	100.0	30.2
sync_0579	tRNA-(ms[2]io[6]A)-hydroxylase (EC:1.-)	0.035	0.320	0.108	0.119	100.0	100.0	36.0
sync_0580	hypothetical protein	0.000	0.000	NaN	NaN	23.8	0.0	0.2

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_0581	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0582	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0583	hypothetical protein	0.000	0.000	NaN	NaN	88.3	0.0	1.6
sync_0584	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0585	hypothetical protein	0.131	0.048	2.752	1.174	48.0	27.8	4.0
sync_0586	possible DNA-binding response regulator	0.061	0.336	0.181	0.167	100.0	100.0	26.9
sync_0587	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0588	hypothetical protein	0.058	0.033	1.774	1.398	100.0	100.0	7.8
sync_0589	ARD/ARD' family protein	0.086	0.232	0.369	0.364	100.0	100.0	14.7
sync_0590	precorrin-2 C20-methyltransferase (EC:2.1.1.130)	0.026	0.187	0.137	0.168	100.0	100.0	15.2
sync_0591	hypothetical protein	0.025	0.176	0.140	0.125	100.0	95.4	14.0
sync_0592	Putative glycosyltransferase	0.025	0.152	0.164	0.170	100.0	100.0	11.4
sync_0593	dihydrouridine synthase family protein	0.019	0.117	0.161	0.166	100.0	93.9	7.9
sync_0594	hypothetical protein	0.015	0.149	0.099	0.090	100.0	95.3	10.4
sync_0595	ErfK/YbiS/YcfS/YnhG superfamily protein	0.006	0.132	0.042	0.011	100.0	65.9	6.2
sync_0596	GTP-binding protein EngA	0.005	0.153	0.033	0.031	100.0	87.6	7.6
sync_0597	possible cobalt transport protein	0.011	0.122	0.091	0.083	100.0	66.8	6.1
sync_0598	hypothetical protein	0.002	0.118	0.016	0.016	100.0	91.4	6.8
sync_0599	conserved hypothetical protein TIGR00044	0.023	0.127	0.181	0.173	100.0	62.1	5.5
sync_0600	hypothetical protein	0.004	0.225	0.017	0.016	100.0	60.1	5.1
sync_0601	pyrroline-5-carboxylate reductase (EC:1.5.1.2)	0.019	0.173	0.109	0.104	100.0	74.5	6.1
sync_0602	Glycosyl transferases group 1	0.014	0.135	0.101	0.108	99.4	49.2	4.7
sync_0603	transporter, major facilitator family protein	0.007	0.110	0.068	0.060	100.0	31.6	4.2
sync_0604	DNA repair protein RecO	0.019	0.157	0.123	0.109	96.7	46.8	4.9
sync_0605	deoxyribose-phosphate aldolase (EC:4.1.2.4)	0.006	0.055	0.118	0.115	100.0	10.6	3.0
sync_0606	ribosomal subunit interface protein	0.000	0.125	0.000	0.000	100.0	51.2	4.7
sync_0607	lipoate-protein ligase B	0.023	0.104	0.219	0.197	100.0	51.5	4.5
sync_0608	Long-chain acyl-CoA synthetase (EC:6.2.1.3)	0.006	0.109	0.052	0.052	72.0	4.3	1.8
sync_0609	hypothetical protein	0.000	0.000	NaN	NaN	53.3	0.0	0.6
sync_0610	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0611	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0612	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0613	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0614	hypothetical protein	0.000	0.000	NaN	NaN	29.5	0.0	0.3
sync_0615	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0616	hypothetical protein	0.000	0.000	NaN	NaN	93.0	0.0	2.7
sync_0617	2-oxo acid dehydrogenases acyltransferase (catalytic domain) protein	0.016	0.253	0.061	0.064	97.3	63.5	6.2
sync_0618	Amino acid permease	0.017	0.114	0.147	0.133	93.5	45.6	4.2
sync_0619	S-adenosylmethionine:tRNA ribosyltransferase-isomerase(Queuosine biosynthesis protein queA) (EC:5.-)	0.012	0.131	0.092	0.096	100.0	84.4	7.0
sync_0620	cysteine synthase A (EC:2.5.1.47)	0.006	0.124	0.051	0.051	100.0	97.9	9.2
sync_0621	hypothetical protein	0.000	0.000	NaN	NaN	81.0	2.4	1.7
sync_0622	possible cystathionine gamma-synthase (EC:2.5.1.48)	0.037	0.199	0.185	0.188	100.0	93.7	11.6
sync_0623	Cystathionine beta-lyase/cystathionine gamma-synthase (EC:2.5.1.48)	0.022	0.219	0.102	0.104	100.0	99.1	12.9
sync_0624	hypothetical protein	0.029	0.386	0.076	0.082	100.0	100.0	22.5
sync_0625	hypothetical protein	0.079	0.330	0.241	0.252	100.0	100.0	26.9
sync_0626	hypothetical protein	0.060	0.238	0.252	0.246	100.0	100.0	25.9
sync_0627	hypothetical protein	0.038	0.257	0.147	0.151	100.0	100.0	22.4
sync_0628	hypothetical protein	0.039	0.267	0.144	0.120	100.0	100.0	25.7
sync_0629	ribosomal protein S4	0.012	0.227	0.051	0.046	100.0	100.0	30.9
sync_0630	conserved hypothetical protein TIGR00278	0.039	0.314	0.124	0.148	100.0	100.0	25.9
sync_0631	ribonucleotide reductase (Class II)	0.072	0.216	0.335	0.300	100.0	100.0	17.9
sync_0632	UDP-N-acetylmuramoylalanyl-D-glutamate--2, 6-diaminopimelate ligase (EC:6.3.2.13)	0.055	0.329	0.168	0.167	100.0	100.0	23.4
sync_0633	calcium/proton exchanger	0.060	0.418	0.143	0.139	100.0	100.0	25.5
sync_0634	hydrophobic amino acid ABC transporter (HAAT) family, periplasmic amino acid-binding protein	0.046	0.252	0.183	0.181	100.0	82.2	8.5
sync_0635	hypothetical protein	0.022	0.226	0.095	0.094	100.0	60.1	6.9
sync_0636	L-cysteine/cystine lyase homolog	0.066	0.303	0.217	0.205	100.0	99.5	23.1
sync_0637	hypothetical protein	0.333	0.000	Inf	Inf	39.5	0.5	0.6
sync_0638	hypothetical protein	0.000	0.000	NaN	NaN	0.8	0.0	0.0
sync_0639	possible Early Protein (E6)	0.048	0.303	0.158	0.159	97.2	91.7	19.8

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_0640	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0641	hypothetical protein	0.007	0.182	0.038	0.088	100.0	22.5	4.1
sync_0642	hypothetical protein	0.046	0.314	0.148	0.141	100.0	100.0	22.6
sync_0643	lipase/esterase family protein (EC:3.1.1.-)	0.044	0.081	0.542	0.515	100.0	36.9	5.0
sync_0644	hypothetical protein	0.033	0.261	0.125	0.138	100.0	100.0	21.6
sync_0645	hypothetical protein	0.112	0.095	1.177	1.049	100.0	81.2	9.1
sync_0646	hypothetical protein	0.000	0.000	NaN	NaN	62.5	0.0	1.3
sync_0647	NifU domain protein	0.000	0.325	0.000	0.006	100.0	100.0	20.8
sync_0648	malate:quinone-oxidoreductase (EC:1.1.99.16)	0.026	0.257	0.101	0.102	100.0	100.0	22.0
sync_0649	Glycerol dehydrogenase (GLDH) (EC:1.1.1.6)	0.028	0.145	0.195	0.185	100.0	100.0	24.1
sync_0650	hypothetical protein	0.085	0.215	0.396	0.369	100.0	100.0	23.3
sync_0651	possible drug resistance protein	0.042	0.211	0.201	0.206	100.0	100.0	15.3
sync_0652	GTP-binding protein LepA	0.016	0.436	0.038	0.042	100.0	99.8	32.0
sync_0653	hypothetical protein	0.020	0.070	0.293	0.286	68.5	4.6	1.5
sync_0654	hypothetical protein	0.000	0.000	NaN	NaN	98.5	0.0	2.2
sync_0655	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	1.4
sync_0656	hypothetical protein	0.031	0.609	0.052	0.083	44.4	13.5	1.4
sync_0657	possible pilin	0.081	0.148	0.548	0.719	100.0	18.1	2.5
sync_0658	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0659	possible pilin	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0660	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0661	hypothetical protein	0.000	0.000	NaN	NaN	23.2	0.0	0.3
sync_0662	hypothetical protein	0.000	0.000	NaN	NaN	36.0	0.0	0.4
sync_0663	hypothetical protein	0.000	0.000	NaN	NaN	5.0	0.0	0.1
sync_0664	hypothetical protein	0.000	0.000	NaN	NaN	55.4	0.0	0.6
sync_0665	hypothetical protein	0.000	0.000	NaN	NaN	32.1	0.0	0.3
sync_0666	prepilin-type N-terminal cleavage/methylation domain protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0667	possible pilin	0.000	0.000	NaN	NaN	31.2	0.0	0.7
sync_0668	two component sensor histidine kinase, putative	0.000	0.000	NaN	NaN	44.6	0.0	0.6
sync_0669	two-component response regulator	0.000	0.000	NaN	NaN	48.0	0.0	0.5
sync_0670	hypothetical protein	0.047	0.173	0.273	0.275	81.6	27.2	2.5
sync_0671	hypothetical protein	0.000	0.000	NaN	NaN	43.6	0.0	0.9
sync_0672	hypothetical protein	0.053	0.200	0.266	0.220	100.0	41.4	5.2
sync_0673	hypothetical protein	0.095	0.340	0.278	0.292	100.0	100.0	22.1
sync_0674	hypothetical protein	0.008	0.182	0.042	0.058	100.0	100.0	10.9
sync_0675	sensor histidine kinase	0.045	0.184	0.247	0.266	100.0	96.7	21.5
sync_0676	ABC-type peptide transporter permease component	0.009	0.161	0.058	0.058	100.0	100.0	18.8
sync_0677	hypothetical protein	0.004	0.245	0.015	0.071	100.0	96.2	16.8
sync_0678	possible porin	0.017	0.147	0.115	0.127	97.1	89.2	17.0
sync_0679	TM2 domain family protein	0.023	0.222	0.103	0.093	100.0	100.0	24.2
sync_0680	Ferritin	0.008	0.136	0.061	0.053	100.0	100.0	17.9
sync_0681	ferrous iron transport protein B	0.011	0.161	0.071	0.067	100.0	100.0	21.9
sync_0682	ferrous iron transport protein A	0.007	0.079	0.093	0.109	100.0	100.0	21.5
sync_0683	hypothetical protein	0.042	0.232	0.179	0.198	100.0	100.0	20.8
sync_0684	hypothetical protein	0.031	0.200	0.154	0.155	100.0	100.0	21.1
sync_0685	hypothetical protein	0.045	0.297	0.150	0.144	100.0	100.0	30.1
sync_0686	cation efflux system protein	0.054	0.244	0.221	0.220	100.0	97.5	14.6
sync_0687	ferritin	0.020	0.281	0.072	0.064	73.5	53.6	11.1
sync_0688	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	2.0
sync_0689	peroxiredoxin 2 family protein	0.043	0.331	0.131	0.132	100.0	83.4	14.1
sync_0690	tRNA/rRNA methyltransferase (EC:2.1.1.34)	0.026	0.244	0.105	0.103	100.0	98.1	10.9
sync_0691	putative surface antigen	0.050	0.236	0.212	0.227	100.0	100.0	12.1
sync_0692	hypothetical protein	0.040	0.161	0.252	0.244	100.0	100.0	10.3
sync_0693	putative methylated-DNA-[protein]-cysteine S-methyltransferase	0.038	0.197	0.192	0.195	100.0	100.0	17.4
sync_0694	sun protein (EC:2.1.1.-)	0.021	0.199	0.105	0.105	100.0	99.2	14.1
sync_0695	Membrane carboxypeptidase (penicillin-binding protein)	0.013	0.183	0.072	0.072	100.0	96.6	12.1
sync_0696	chlorophyll synthase, ChlG	0.008	0.172	0.044	0.044	100.0	100.0	14.5
sync_0697	hypothetical protein	0.006	0.083	0.077	0.146	100.0	100.0	13.9
sync_0698	imidazoleglycerol phosphate synthase, cyclase subunit	0.011	0.179	0.064	0.072	100.0	100.0	14.4
sync_0699	hypothetical protein	0.013	0.156	0.084	0.084	100.0	100.0	20.0
sync_0700	ubiquinone/menaquinone biosynthesis methyltransferase UbiE (EC:2.1.1.-)	0.026	0.203	0.128	0.119	100.0	100.0	21.2
sync_0701	hypothetical protein	0.023	0.202	0.112	0.136	100.0	100.0	24.8
sync_0702	hypothetical protein	0.067	0.306	0.219	0.236	100.0	100.0	19.0
sync_0703	hypothetical protein	0.039	0.064	0.614	0.653	49.3	34.8	2.9
sync_0704	hypothetical protein	0.000	0.000	NaN	NaN	98.8	0.0	1.0

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_0705	hypothetical protein	0.000	0.000	NaN	NaN	13.9	0.0	0.1
sync_0706	sensor histidine kinase (EC:2.7.3.-)	0.017	0.104	0.164	0.168	92.5	56.4	6.0
sync_0707	DNA-binding response regulator	0.015	0.087	0.169	0.141	71.7	22.4	4.0
sync_0708	hypothetical protein	0.088	0.097	0.908	0.943	98.7	71.1	8.3
sync_0709	type I secretion target GGXGXDXXX repeat protein (EC:3.1.-)	0.062	0.124	0.502	0.531	100.0	100.0	14.1
sync_0710	hypothetical protein	0.215	0.460	0.469	0.482	100.0	100.0	21.8
sync_0711	hypothetical protein	0.188	0.264	0.711	0.742	51.8	20.3	3.3
sync_0712	hypothetical protein	0.000	0.000	NaN	NaN	89.9	0.0	2.5
sync_0713	possible Carbamoyl-phosphate synthase L chain	0.000	0.000	NaN	NaN	67.7	0.0	0.7
sync_0714	lipoprotein, putative	0.000	0.000	NaN	NaN	100.0	0.0	2.7
sync_0715	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0716	hypothetical protein	0.000	0.000	NaN	NaN	19.4	0.0	0.2
sync_0717	glycyl-tRNA synthetase, alpha subunit (EC:6.1.1.14)	0.008	0.239	0.031	0.036	97.1	93.5	21.7
sync_0718	DNA internalization competence protein, ComEC/Rec2 family protein	0.066	0.237	0.278	0.286	100.0	100.0	25.9
sync_0720	hypothetical protein	0.000	0.000	NaN	NaN	39.3	0.0	0.5
sync_0721	hypothetical protein	0.000	0.000	NaN	NaN	97.7	0.0	1.9
sync_0722	hypothetical protein	0.038	0.289	0.131	0.125	100.0	100.0	23.8
sync_0723	galactosyl-1-phosphate transferase (EC:2.7.8.6)	0.024	0.265	0.090	0.083	100.0	100.0	15.0
sync_0724	cobalamin biosynthesis protein CobD (EC:6.3.1.10)	0.038	0.194	0.197	0.183	100.0	100.0	12.3
sync_0725	ketol-acid reductoisomerase (EC:1.1.1.86)	0.008	0.287	0.030	0.029	100.0	100.0	12.1
sync_0726	putative ATP-dependent Clp protease, proteolytic subunit ClpP	0.004	0.178	0.021	0.019	100.0	81.9	6.6
sync_0727	ATP-dependent Clp protease, proteolytic subunit ClpP (EC:3.4.21.92)	0.004	0.155	0.024	0.022	100.0	90.9	6.1
sync_0728	PIN/TRAM domain protein	0.013	0.119	0.107	0.096	98.9	62.2	6.5
sync_0729	putative oxygen-independent coproporphyrinogen III oxidase	0.023	0.132	0.173	0.153	100.0	65.2	5.3
sync_0730	3-methyl-2-oxobutanoate hydroxymethyltransferase (EC:2.1.2.11)	0.020	0.139	0.140	0.123	92.9	58.3	4.9
sync_0731	cell division protein FtsZ	0.015	0.251	0.061	0.066	100.0	55.1	5.6
sync_0732	hypothetical protein	0.028	0.180	0.156	0.148	100.0	48.2	4.5
sync_0733	Uncharacterized secreted or membrane protein	0.008	0.184	0.044	0.043	100.0	42.1	5.1
sync_0734	D-alanine--D-alanine ligase (EC:6.3.2.4)	0.019	0.211	0.091	0.096	79.6	33.4	3.2
sync_0735	tRNA-i(6)A37 thiotransferase enzyme MiaB	0.021	0.163	0.128	0.125	80.2	28.2	3.1
sync_0736	muconate cycloisomerase (EC:5.5.-)	0.022	0.175	0.126	0.131	95.1	41.0	4.4
sync_0737	hypothetical protein	0.020	0.099	0.205	0.196	100.0	47.6	4.3
sync_0739	hypothetical protein	0.014	0.099	0.138	0.136	100.0	97.3	9.0
sync_0740	amidohydrolase family protein (EC:3.5.4.1)	0.022	0.161	0.135	0.129	100.0	52.6	5.4
sync_0741	FAD linked oxidase, N-terminal	0.007	0.058	0.127	0.097	98.2	19.0	3.2
sync_0742	hypothetical protein	0.000	0.107	0.000	0.000	96.2	43.5	4.1
sync_0743	putative folylpolyglutamate synthase	0.028	0.189	0.147	0.142	83.8	12.4	2.3
sync_0744	Acetylornithine aminotransferase (ACOAT) (EC:2.6.1.11)	0.010	0.095	0.107	0.118	100.0	73.2	6.3
sync_0746	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC:2.5.1.7)	0.008	0.125	0.066	0.091	100.0	33.6	3.6
sync_0748	possible 4Fe-4S iron sulfur cluster binding pr	0.013	0.078	0.160	0.189	100.0	58.6	4.3
sync_0749	RNA methyltransferase, TrmH family protein	0.023	0.123	0.188	0.211	100.0	47.0	4.6
sync_0750	dihydrolipoamide dehydrogenase (EC:1.8.1.4)	0.002	0.145	0.017	0.016	96.9	25.7	3.2
sync_0751	indole-3-glycerol phosphate synthase (EC:4.1.1.48)	0.012	0.143	0.083	0.091	100.0	92.5	7.1
sync_0752	hypothetical protein	0.056	0.087	0.646	0.667	100.0	100.0	9.7
sync_0753	Hydrogenase accessory protein or high-affinity nickel-transport protein homolog, membrane protein	0.010	0.153	0.068	0.061	100.0	76.9	6.9
sync_0754	nickel-type superoxide dismutase maturation protease (EC:3.4.21.-)	0.056	0.159	0.350	0.335	100.0	97.7	11.9
sync_0755	superoxide dismutase, Ni (EC:1.15.1.1)	0.011	0.178	0.064	0.055	100.0	78.2	8.6
sync_0756	FKBP-type peptidyl-prolyl cis-trans isomerase (EC:5.2.1.8)	0.022	0.178	0.125	0.124	100.0	49.1	5.5
sync_0757	hypothetical protein	0.011	0.094	0.121	0.129	100.0	25.8	3.9
sync_0758	Apolipoprotein N-acyltransferase (EC:2.3.1.-)	0.024	0.120	0.200	0.207	100.0	77.5	6.6
sync_0759	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase (EC:2.1.1.61)	0.008	0.132	0.063	0.060	100.0	81.0	8.5
sync_0760	carbohydrate kinase family protein	0.027	0.155	0.174	0.190	100.0	99.7	11.8
sync_0761	Type II alternative RNA polymerase sigma factor, sigma-70 family protein	0.007	0.242	0.030	0.027	100.0	100.0	18.6
sync_0762	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0763	histone deacetylase family protein	0.016	0.213	0.076	0.088	96.5	84.9	13.2
sync_0764	Pyruvate dehydrogenase E1 alpha subunit (EC:1.2.4.1)	0.007	0.198	0.037	0.051	100.0	100.0	10.5
sync_0765	DnaJ domain protein	0.027	0.212	0.127	0.128	98.6	63.9	6.3
sync_0766	signal recognition particle protein	0.008	0.137	0.061	0.056	100.0	91.4	8.9
sync_0767	ribosomal protein S16	0.019	0.159	0.123	0.100	100.0	100.0	12.7
sync_0768	PhoH family protein	0.007	0.153	0.047	0.054	100.0	96.0	10.5
sync_0769	hypothetical protein	0.000	0.104	0.000	0.000	87.6	10.5	2.1

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_0770	unnamed protein product; glcP gene product	0.000	0.000	NaN	NaN	65.9	3.5	1.2
sync_0771	hypothetical protein	0.017	0.139	0.125	0.113	100.0	95.6	8.2
sync_0772	alpha-acetolactate decarboxylase (EC:4.1.1.5)	0.030	0.201	0.151	0.157	100.0	87.7	7.7
sync_0773	GTP-binding protein Era	0.009	0.138	0.065	0.066	100.0	96.8	9.0
sync_0774	hypothetical protein	0.014	0.136	0.106	0.102	100.0	99.0	9.1
sync_0775	tRNA (guanine-N1)-methyltransferase (EC:2.1.1.31)	0.022	0.179	0.122	0.130	100.0	100.0	15.6
sync_0776	hypothetical protein	0.098	0.117	0.834	0.715	100.0	100.0	24.4
sync_0777	hypothetical protein	0.046	0.046	0.995	0.935	100.0	100.0	9.5
sync_0778	hypothetical protein	0.000	0.000	NaN	NaN	77.1	20.3	2.7
sync_0779	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0780	hypothetical protein	0.036	0.000	Inf	Inf	100.0	31.7	3.8
sync_0781	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC:4.6.1.12)	0.009	0.213	0.044	0.047	100.0	100.0	9.1
sync_0782	hypothetical protein	0.054	0.326	0.165	0.174	100.0	100.0	15.7
sync_0783	putative circadian phase modifier CpmA homolog	0.031	0.257	0.119	0.114	100.0	93.8	8.4
sync_0784	hypothetical protein	0.078	0.287	0.272	0.252	100.0	100.0	13.4
sync_0785	hypothetical protein	0.014	0.253	0.056	0.063	100.0	95.5	11.4
sync_0786	thiamine biosynthesis protein ThiS	0.018	0.052	0.352	0.580	97.1	50.5	5.1
sync_0787	hypothetical protein	0.000	0.000	NaN	NaN	93.0	0.0	0.9
sync_0788	hypothetical protein	0.000	0.000	NaN	NaN	78.0	0.0	1.9
sync_0789	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0790	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0791	thiamine-phosphate pyrophosphorylase (EC:2.5.1.3)	0.024	0.185	0.128	0.133	100.0	96.8	12.0
sync_0792	hypothetical protein	0.016	0.233	0.067	0.105	100.0	100.0	13.5
sync_0793	riboflavin biosynthesis protein RibF (EC:2.7.1.26,EC:2.7.7.2)	0.017	0.165	0.104	0.112	100.0	76.5	6.5
sync_0794	Uncharacterized membrane protein	0.010	0.203	0.051	0.057	100.0	88.8	10.1
sync_0795	acid phosphatase SurE (EC:3.1.3.2)	0.016	0.200	0.081	0.079	100.0	93.9	15.2
sync_0796	phenylalanyl-tRNA synthetase, alpha subunit (EC:6.1.1.20)	0.010	0.219	0.046	0.043	100.0	100.0	17.8
sync_0797	hypothetical protein	0.031	0.238	0.130	0.131	100.0	99.1	14.1
sync_0798	hypothetical protein	0.000	0.000	NaN	NaN	20.8	0.0	0.4
sync_0799	hypothetical protein	0.047	0.187	0.253	0.258	100.0	97.6	16.0
sync_0800	hypothetical protein	0.065	0.012	5.509	2.981	100.0	100.0	12.6
sync_0801	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0802	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0803	hypothetical protein	0.000	0.000	NaN	NaN	20.5	0.0	0.2
sync_0804	ABC1 family protein	0.006	0.084	0.069	0.072	100.0	89.7	7.9
sync_0805	zeta-carotene desaturase	0.022	0.109	0.205	0.215	100.0	91.6	9.7
sync_0806	ATP-NAD kinase (EC:2.7.1.23)	0.007	0.100	0.069	0.062	100.0	96.2	7.9
sync_0807	possible two-component response regulator	0.031	0.090	0.341	0.337	100.0	100.0	12.5
sync_0808	hypothetical protein	0.023	0.217	0.104	0.104	100.0	82.7	12.2
sync_0809	hypothetical protein	0.008	0.089	0.091	0.092	100.0	95.2	9.6
sync_0810	precorrin-6y C5,15-methyltransferase (decarboxylating), CbiE/T subunits (EC:2.1.1.132)	0.014	0.093	0.150	0.128	100.0	98.4	11.2
sync_0811	Possible AraC-type regulatory protein, putative	0.020	0.163	0.125	0.138	84.6	57.2	5.6
sync_0812	hypothetical protein	0.026	0.107	0.243	0.236	100.0	94.4	9.2
sync_0813	hypothetical protein	0.020	0.156	0.128	0.128	98.4	80.3	7.0
sync_0814	hypothetical protein	0.004	0.083	0.048	0.037	100.0	22.4	3.2
sync_0815	hypothetical protein	0.017	0.102	0.168	0.142	100.0	98.5	7.1
sync_0816	transporter, major facilitator family protein	0.008	0.082	0.100	0.090	99.7	87.7	6.8
sync_0817	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	2.5
sync_0818	hypothetical protein	0.006	0.151	0.043	0.041	100.0	85.1	8.5
sync_0819	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0820	hypothetical protein	0.052	0.388	0.134	0.139	100.0	93.4	22.9
sync_0821	riboflavin biosynthesis protein RibD (EC:3.5.4.26,EC:1.1.1.193)	0.072	0.399	0.180	0.182	100.0	100.0	21.9
sync_0822	alcohol dehydrogenase (EC:1.1.1.-)	0.035	0.216	0.161	0.156	100.0	100.0	10.0
sync_0823	Potassium channel	0.036	0.214	0.167	0.156	100.0	100.0	15.5
sync_0824	hypothetical protein	0.000	0.000	NaN	NaN	16.7	0.0	0.4
sync_0825	cell division protein FtsH3 (EC:3.4.24.-)	0.012	0.391	0.029	0.028	100.0	100.0	16.2
sync_0826	ornithine carbamoyltransferase (EC:2.1.3.3)	0.009	0.189	0.045	0.049	100.0	100.0	14.1
sync_0827	carboxylesterase (EC:3.1.1.1)	0.030	0.172	0.173	0.161	100.0	100.0	15.7
sync_0828	hypothetical protein	0.060	0.332	0.182	0.184	100.0	100.0	16.8
sync_0829	hypothetical protein	0.047	0.323	0.147	0.136	100.0	100.0	17.6
sync_0830	hypothetical protein	0.079	0.281	0.280	0.282	100.0	100.0	21.8
sync_0831	LexA repressor (EC:3.4.21.88)	0.019	0.282	0.067	0.073	100.0	100.0	24.7
sync_0833	hypothetical protein	0.000	0.000	NaN	NaN	21.4	0.0	0.2
sync_0834	hypothetical protein	0.000	0.000	NaN	NaN	12.9	0.0	0.1
sync_0835	hypothetical protein	0.130	0.511	0.255	0.254	100.0	100.0	15.3

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_0836	hypothetical protein	0.018	0.069	0.259	0.257	92.9	11.7	2.6
sync_0837	hypothetical protein	0.052	0.257	0.203	0.293	92.9	28.8	3.0
sync_0838	hypothetical protein	0.067	0.000	Inf	Inf	79.5	3.0	1.0
sync_0839	hypothetical protein	0.195	0.395	0.495	0.477	100.0	65.8	20.9
sync_0840	hypothetical protein	0.147	0.162	0.909	1.000	100.0	100.0	14.1
sync_0841	hypothetical protein	0.016	0.048	0.336	0.336	100.0	100.0	19.0
sync_0842	hypothetical protein	0.000	0.000	NaN	NaN	5.5	0.0	0.1
sync_0843	hypothetical protein	0.000	0.000	NaN	NaN	97.7	0.0	1.2
sync_0844	hypothetical protein	0.000	0.000	NaN	NaN	72.2	0.0	0.8
sync_0845	hypothetical protein	0.000	0.000	NaN	NaN	42.2	0.0	0.4
sync_0846	hypothetical protein	0.045	0.387	0.116	0.109	65.9	43.2	3.1
sync_0847	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	1.0
sync_0848	axoneme-associated protein gasp-180	0.000	0.000	NaN	NaN	86.7	24.4	2.4
sync_0849	hypothetical protein	0.076	0.255	0.297	0.320	97.2	49.6	6.0
sync_0850	hypothetical protein	0.000	0.000	NaN	NaN	81.5	0.0	0.8
sync_0851	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0852	hypothetical protein	0.000	0.000	NaN	NaN	22.7	0.0	0.2
sync_0853	bacterial metallothionein-related protein	0.000	0.000	NaN	NaN	89.4	0.0	0.9
sync_0854	ferritin	0.025	0.190	0.130	0.120	65.0	57.8	5.4
sync_0855	hypothetical protein	0.008	0.027	0.305	0.290	100.0	96.3	12.6
sync_0856	gluconolactonase precursor (EC:3.1.1.17)	0.000	0.000	NaN	NaN	94.0	5.7	1.7
sync_0857	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0858	hypothetical protein	0.000	0.000	NaN	NaN	33.8	0.0	0.3
sync_0859	lipase/esterase (EC:3.1.1.-)	0.000	0.000	NaN	NaN	74.3	0.0	1.2
sync_0860	3-oxoacyl-(acyl-carrier protein) reductase BH0932 (EC:1.1.1.100)	0.000	0.000	NaN	NaN	27.2	0.0	0.4
sync_0861	Mandelate racemase / muconate lactonizing enzyme, N-terminal domain protein	0.000	0.000	NaN	NaN	40.6	9.7	0.9
sync_0862	hypothetical protein	0.059	0.304	0.193	0.188	100.0	100.0	38.0
sync_0863	hypothetical protein	0.038	0.238	0.158	0.150	100.0	100.0	37.0
sync_0864	hypothetical protein	0.220	0.173	1.273	1.574	86.2	37.7	7.2
sync_0865	hypothetical protein	0.047	0.429	0.110	0.110	100.0	100.0	41.6
sync_0866	hypothetical protein	0.071	0.351	0.201	0.197	100.0	100.0	65.7
sync_0867	hypothetical protein	0.017	0.345	0.050	0.066	100.0	100.0	26.8
sync_0868	hypothetical protein	0.103	0.240	0.430	0.465	100.0	86.1	21.4
sync_0869	hypothetical protein	0.057	0.413	0.137	0.144	100.0	100.0	45.0
sync_0870	possible SAP domain	0.035	0.407	0.085	0.088	100.0	100.0	68.7
sync_0871	GrpE	0.058	0.512	0.113	0.112	100.0	100.0	65.4
sync_0872	glyceraldehyde-3-phosphate dehydrogenase, type I (EC:1.2.1.-)	0.024	0.488	0.050	0.052	100.0	100.0	52.9
sync_0873	hypothetical protein	0.115	0.432	0.265	0.281	100.0	100.0	40.9
sync_0874	Photosystem I reaction center subunit psaK (Photosystem I subunit X)	0.057	0.427	0.132	0.139	100.0	100.0	41.5
sync_0875	hypothetical protein	0.098	0.154	0.637	0.684	100.0	79.1	9.9
sync_0876	hypothetical protein	0.020	0.195	0.102	0.093	93.4	33.3	3.8
sync_0877	hypothetical protein	0.049	0.291	0.170	0.149	100.0	100.0	22.9
sync_0878	proline/betaine transporter	0.027	0.265	0.102	0.115	100.0	100.0	20.0
sync_0879	hypothetical protein	0.029	0.179	0.160	0.170	100.0	100.0	15.5
sync_0880	hypothetical protein	0.020	0.163	0.123	0.156	100.0	100.0	12.1
sync_0881	hypothetical protein	0.039	0.236	0.167	0.150	84.2	73.5	9.1
sync_0882	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0883	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0884	hypothetical protein	0.000	0.000	NaN	NaN	69.2	0.0	0.7
sync_0885	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0886	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0887	hypothetical protein	0.000	0.000	NaN	NaN	11.5	0.0	0.1
sync_0888	hypothetical protein	0.000	0.000	NaN	NaN	18.9	0.0	0.2
sync_0889	hypothetical protein	0.026	0.000	Inf	Inf	100.0	25.9	3.7
sync_0890	hypothetical protein	0.029	0.202	0.142	0.151	100.0	100.0	10.5
sync_0891	hypothetical protein	0.032	0.220	0.144	0.129	100.0	100.0	13.8
sync_0892	Ferredoxin	0.045	0.241	0.186	0.234	100.0	100.0	18.0
sync_0893	cobyric acid synthase CobQ (EC:6.3.5.10)	0.025	0.188	0.131	0.129	61.2	17.2	3.0
sync_0894	hypothetical protein	0.025	0.188	0.130	0.134	99.2	65.3	6.1
sync_0895	septum formation protein MaF	0.050	0.166	0.301	0.303	98.0	82.1	6.4
sync_0896	photosystem II 44 kDa subunit reaction center protein	0.003	0.101	0.025	0.028	100.0	61.6	5.6
sync_0897	photosystem II protein D2	0.007	0.153	0.046	0.048	100.0	99.2	12.4
sync_0898	photosystem I assembly protein Ycf4	0.006	0.164	0.037	0.051	100.0	50.9	4.5

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_0899	peptidyl-prolyl cis-trans isomerase, cyclophilin-type (EC:5.2.1.8)	0.013	0.205	0.062	0.071	100.0	97.2	7.2
sync_0900	acetolactate synthase, small subunit (EC:2.2.1.6)	0.007	0.118	0.060	0.057	100.0	87.4	6.2
sync_0901	predicted alpha/beta hydrolase superfamily protein	0.011	0.111	0.099	0.128	99.9	51.0	4.7
sync_0902	N2,N2-dimethylguanosine tRNA methyltransferase	0.021	0.145	0.143	0.144	100.0	68.9	6.0
sync_0903	Cytochrome B6-F complex subunit VII-related protein	0.000	0.000	NaN	NaN	100.0	0.0	2.1
sync_0904	3-beta hydroxysteroid dehydrogenase/isomerase family protein	0.007	0.170	0.039	0.049	100.0	68.5	5.5
sync_0905	hypothetical protein	0.000	0.280	0.000	0.010	100.0	63.9	5.7
sync_0906	SAM (and some other nucleotide) binding motif:Generic methyltransferase	0.025	0.181	0.139	0.141	100.0	65.7	6.0
sync_0907	possible pseudouridine synthase (EC:4.2.1.70)	0.040	0.185	0.217	0.245	100.0	99.0	8.2
sync_0908	hypothetical protein	0.018	0.085	0.213	0.313	100.0	72.3	6.0
sync_0909	translation initiation factor IF-1	0.005	0.091	0.051	0.044	100.0	75.6	6.1
sync_0910	thioredoxin-disulfide reductase (EC:1.8.1.9)	0.010	0.178	0.058	0.075	100.0	85.1	7.5
sync_0911	hypothetical protein	0.016	0.174	0.090	0.080	100.0	100.0	9.9
sync_0912	Helicase C-terminal domain	0.007	0.176	0.042	0.045	100.0	70.5	8.5
sync_0913	hypothetical protein	0.030	0.033	0.907	0.992	100.0	100.0	8.0
sync_0914	hypothetical protein	0.000	0.101	0.000	0.000	85.5	19.6	2.7
sync_0915	NAD/NADP transhydrogenase alpha subunit (EC:1.6.1.1)	0.013	0.179	0.074	0.072	97.8	52.7	4.2
sync_0916	NAD/NADP transhydrogenase alpha subunit (EC:1.6.1.1)	0.003	0.060	0.050	0.038	100.0	50.8	3.8
sync_0917	NAD/NADP transhydrogenase beta subunit (EC:1.6.1.2)	0.001	0.113	0.013	0.020	97.8	67.6	6.5
sync_0918	Alpha/beta superfamily hydrolase	0.026	0.199	0.128	0.141	99.1	66.4	5.8
sync_0919	Ferredoxin	0.008	0.115	0.072	0.095	100.0	98.8	6.8
sync_0920	1-deoxy-D-xylulose 5-phosphate reductoisomerase (EC:1.1.1.267)	0.007	0.147	0.050	0.054	100.0	100.0	9.2
sync_0921	Na+-dependent transporter of the SNF family protein	0.019	0.244	0.076	0.083	100.0	100.0	11.6
sync_0922	possible HAMP domain	0.039	0.261	0.151	0.140	100.0	100.0	15.0
sync_0923	hypothetical protein	0.046	0.186	0.249	0.262	100.0	100.0	17.5
sync_0924	cysteinyl-tRNA synthetase (EC:6.1.1.16)	0.026	0.194	0.135	0.152	100.0	100.0	15.9
sync_0925	hypothetical protein	0.087	0.000	Inf	Inf	100.0	6.4	3.0
sync_0926	hypothetical protein	0.133	0.167	0.800	0.380	80.6	22.3	3.1
sync_0927	multidrug transporter, putative	0.034	0.238	0.144	0.156	72.1	59.2	6.5
sync_0928	DNA polymerase I (EC:2.7.7.7)	0.020	0.214	0.093	0.092	100.0	94.0	14.2
sync_0929	possible membrane fusion protein	0.035	0.219	0.159	0.155	100.0	100.0	14.1
sync_0930	GTP-binding protein YchF	0.010	0.233	0.042	0.043	100.0	100.0	16.3
sync_0931	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_0932	sensor kinase CpxA, putative	0.045	0.268	0.167	0.181	100.0	95.3	9.5
sync_0933	hypothetical protein	0.000	0.000	NaN	NaN	76.9	0.0	1.1
sync_0934	hypothetical protein	0.060	0.263	0.230	0.203	100.0	98.9	15.5
sync_0935	hypothetical protein	0.021	0.260	0.080	0.087	100.0	100.0	17.8
sync_0936	transporter, major facilitator family protein	0.024	0.206	0.116	0.118	100.0	93.3	9.7
sync_0937	hypothetical protein	0.032	0.199	0.159	0.163	100.0	94.3	11.2
sync_0938	ABC-type oligopeptide transport system permease	0.014	0.229	0.059	0.061	100.0	99.3	11.5
sync_0939	ABC transporter, substrate binding protein, possibly oligopeptides	0.031	0.216	0.145	0.139	100.0	90.1	10.0
sync_0940	hypothetical protein	0.033	0.223	0.150	0.142	100.0	100.0	17.5
sync_0941	homoserine dehydrogenase (EC:1.1.1.3)	0.012	0.262	0.046	0.045	100.0	100.0	14.7
sync_0942	Fe-S metabolism protein, SufE family protein	0.031	0.302	0.103	0.108	100.0	100.0	17.7
sync_0943	hypothetical protein	0.075	0.284	0.263	0.261	100.0	100.0	18.1
sync_0944	5-formyltetrahydrofolate cyclo-ligase family protein	0.075	0.250	0.300	0.283	100.0	100.0	16.6
sync_0945	hypothetical protein	0.088	0.092	0.954	0.778	100.0	99.5	6.9
sync_0946	hypothetical protein	0.132	0.147	0.897	0.939	100.0	100.0	12.1
sync_0947	hypothetical protein	0.120	0.133	0.903	0.807	99.2	88.3	10.5
sync_0948	carbonic anhydrase (EC:4.2.1.1)	0.043	0.276	0.154	0.141	100.0	100.0	24.2
sync_0949	hypothetical protein	0.041	0.257	0.161	0.161	100.0	100.0	31.9
sync_0950	hypothetical protein	0.008	0.221	0.034	0.033	100.0	100.0	38.1
sync_0951	hypothetical protein	0.097	0.093	1.046	1.056	100.0	100.0	21.7
sync_0952	hypothetical protein	0.071	0.113	0.630	0.777	100.0	64.3	8.9
sync_0953	hypothetical protein	0.115	0.176	0.653	0.608	100.0	100.0	18.9
sync_0954	hypothetical protein	0.039	0.313	0.123	0.128	100.0	100.0	28.4
sync_0955	hypothetical protein	0.041	0.152	0.267	0.253	100.0	77.0	6.8
sync_0956	conserved hypotheical protein	0.076	0.364	0.209	0.210	100.0	98.6	43.9
sync_0957	hypothetical protein	0.050	0.441	0.113	0.112	100.0	100.0	26.7
sync_0958	hypothetical protein	0.059	0.376	0.156	0.159	100.0	100.0	28.6
sync_0959	hypothetical protein	0.028	0.261	0.105	0.128	100.0	100.0	17.8
sync_0960	hypothetical protein	0.000	0.000	NaN	NaN	40.9	0.0	0.4

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_0961	Crossover junction endodeoxyribonuclease ruvC (Hollidayjunction nuclease ruvC) (Holliday juction resolvase ruvC) (EC:3.1.22.4)	0.008	0.244	0.034	0.034	97.1	94.6	12.7
sync_0962	magnesium chelatase, ATPase subunit I (EC:6.6.1.1)	0.011	0.167	0.068	0.062	100.0	85.0	9.3
sync_0963	hypothetical protein	0.035	0.185	0.187	0.170	100.0	68.8	5.9
sync_0964	carboxylesterase, beta-lactamase family protein	0.010	0.155	0.064	0.059	100.0	91.7	7.3
sync_0965	RNA methyltransferase, TrmH family, group 1	0.018	0.170	0.108	0.112	100.0	80.3	7.2
sync_0966	cytochrome cM	0.015	0.149	0.101	0.118	100.0	73.3	6.2
sync_0967	cytochrome b6-f complex, subunit V	0.000	0.119	0.000	0.000	100.0	78.1	5.8
sync_0968	putative methyltransferase	0.027	0.148	0.180	0.161	100.0	84.0	6.0
sync_0969	imidazole glycerol phosphate synthase, glutamine amidotransferase subunit (EC:2.4.2.-)	0.023	0.191	0.121	0.124	100.0	80.8	8.3
sync_0970	thioredoxin	0.000	0.000	NaN	NaN	100.0	5.3	2.6
sync_0971	IMP dehydrogenase family protein	0.006	0.118	0.048	0.035	99.7	64.1	5.4
sync_0972	Proline-rich region	0.064	0.725	0.089	0.112	71.5	32.9	3.1
sync_0973	DNA gyrase, A subunit (EC:5.99.1.3)	0.008	0.208	0.038	0.035	96.8	56.4	5.6
sync_0974	lycopene beta cyclase (EC:1.14.-)	0.013	0.081	0.157	0.165	100.0	58.8	5.7
sync_0975	hypothetical protein	0.030	0.332	0.090	0.090	100.0	89.7	10.1
sync_0976	2-isopropylmalate synthase (EC:2.3.3.13)	0.005	0.208	0.024	0.033	100.0	100.0	10.9
sync_0977	putative sn-glycerol-3-phosphate ABC transporter, permease protein	0.017	0.227	0.076	0.075	100.0	91.5	11.1
sync_0978	hypothetical protein	0.055	0.167	0.330	0.323	100.0	92.0	7.4
sync_0979	hypothetical protein	0.021	0.252	0.082	0.099	100.0	90.2	9.3
sync_0980	Ferredoxin	0.013	0.234	0.054	0.079	100.0	88.3	6.6
sync_0981	hypothetical protein	0.012	0.110	0.112	0.092	100.0	55.6	4.4
sync_0982	hypothetical protein	0.079	0.458	0.173	0.162	83.4	30.2	3.3
sync_0983	hypothetical protein	0.023	0.129	0.178	0.146	100.0	12.0	3.0
sync_0984	conserved hypothetical protein TIGR00106	0.024	0.232	0.105	0.139	100.0	90.4	8.3
sync_0985	hypothetical protein	0.039	0.176	0.220	0.220	100.0	78.9	6.8
sync_0986	hypothetical protein	0.003	0.149	0.021	0.030	100.0	100.0	8.4
sync_0987	possible Hepatitis C virus core protein	0.003	0.126	0.023	0.031	100.0	83.1	8.3
sync_0988	pseudouridine synthase RluA PA3246 (EC:4.2.1.70)	0.059	0.228	0.258	0.274	100.0	93.4	8.1
sync_0989	HDIG domain protein	0.013	0.182	0.071	0.069	100.0	83.8	7.8
sync_0990	putative FolD bifunctional protein	0.011	0.104	0.108	0.111	100.0	79.9	7.1
sync_0991	geranylgeranyl diphosphate synthase	0.008	0.130	0.061	0.057	100.0	93.7	8.3
sync_0992	hypothetical protein	0.013	0.217	0.059	0.062	100.0	91.8	14.0
sync_0993	hypothetical protein (EC:3.1.11.5)	0.008	0.216	0.039	0.044	100.0	97.6	10.3
sync_0994	hypothetical protein	0.009	0.020	0.423	0.437	100.0	97.6	9.5
sync_0995	hypothetical protein	0.058	0.219	0.265	0.316	100.0	100.0	17.7
sync_0996	hypothetical protein	0.059	0.254	0.234	0.219	100.0	95.7	14.7
sync_0997	hypothetical protein	0.032	0.287	0.111	0.114	100.0	100.0	17.7
sync_0998	EntD (EC:3.1.3.-)	0.043	0.231	0.186	0.174	100.0	100.0	19.1
sync_0999	acylphosphatase	0.008	0.180	0.044	0.055	100.0	87.6	7.2
sync_1000	cobyrinic acid a,c-diamide synthase (EC:6.3.1.-)	0.021	0.165	0.130	0.125	100.0	84.1	8.9
sync_1001	Possible glucose 6-phosphate dehydrogenase effector OpcA	0.022	0.176	0.126	0.116	100.0	94.9	7.6
sync_1002	glucose-6-phosphate 1-dehydrogenase (EC:1.1.1.49)	0.003	0.167	0.020	0.024	100.0	81.1	6.9
sync_1003	ferredoxin--NADP reductase (EC:1.18.1.2)	0.013	0.197	0.067	0.064	100.0	100.0	9.0
sync_1004	cyclase/dehydrase family protein	0.008	0.203	0.040	0.055	100.0	22.4	4.0
sync_1006	adaptive-response sensory histidine kinase SasA (EC:2.7.-)	0.007	0.175	0.037	0.036	100.0	98.3	8.9
sync_1007	hypothetical protein	0.013	0.252	0.052	0.083	100.0	100.0	7.6
sync_1008	possible cAMP phosphodiesterases class-II	0.022	0.161	0.136	0.139	100.0	100.0	12.1
sync_1009	aminopeptidase N (EC:3.4.11.2)	0.028	0.238	0.118	0.118	100.0	99.9	13.7
sync_1010	hypothetical protein	0.032	0.283	0.114	0.113	100.0	100.0	18.2
sync_1011	ribose-phosphate pyrophosphokinase (EC:2.7.6.1)	0.006	0.227	0.028	0.026	100.0	100.0	16.1
sync_1012	hypothetical protein	0.042	0.168	0.250	0.263	100.0	94.1	13.5
sync_1013	possible LytR-membrane bound transcriptional regulator	0.043	0.258	0.168	0.169	100.0	100.0	20.9
sync_1014	hypothetical protein	0.045	0.268	0.169	0.167	100.0	100.0	18.4
sync_1015	4-alpha-glucanotransferase (EC:2.4.1.25)	0.022	0.226	0.098	0.096	100.0	100.0	16.5
sync_1016	Helix-turn-helix motif	0.051	0.305	0.166	0.179	100.0	100.0	23.2
sync_1017	16S rRNA pseudouridylate synthase (EC:4.2.1.70)	0.038	0.205	0.185	0.188	100.0	100.0	15.0
sync_1018	Type II alternative RNA polymerase sigma factor, sigma-70 family protein	0.011	0.317	0.035	0.042	100.0	100.0	24.9
sync_1019	hypothetical protein	0.015	0.363	0.040	0.044	100.0	100.0	26.1
sync_1020	ABC1 domain protein	0.022	0.387	0.058	0.057	100.0	100.0	30.0
sync_1021	hypothetical protein	0.062	0.000	Inf	Inf	35.7	6.6	1.1
sync_1022	Orn/Lys/Arg decarboxylases family 1 (EC:4.1.1.18)	0.045	0.231	0.196	0.190	100.0	99.9	16.9
sync_1023	phosphatidate cytidylyltransferase (EC:2.7.7.41)	0.021	0.249	0.086	0.088	100.0	100.0	13.3
sync_1024	hypothetical protein	0.031	0.247	0.124	0.127	100.0	100.0	15.4

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_1025	hydrolase, alpha/beta fold family protein	0.031	0.265	0.118	0.112	100.0	100.0	16.6
sync_1026	Glycerol dehydrogenase family enzyme (EC:1.1.1.6)	0.051	0.295	0.172	0.171	100.0	100.0	16.4
sync_1027	putative Clp protease, ATP-binding subunit ClpC	0.012	0.262	0.044	0.042	100.0	95.4	8.6
sync_1028	ribosomal-protein-alanine acetyltransferase (EC:2.3.1.128)	0.000	0.000	NaN	NaN	92.5	0.0	1.3
sync_1029	diaminopimelate decarboxylase (EC:4.1.1.20)	0.008	0.237	0.033	0.039	100.0	56.4	5.5
sync_1030	conserved hypothetical protein TIGR00159	0.008	0.194	0.039	0.042	100.0	98.3	7.7
sync_1031	undecaprenyl diphosphate synthase (EC:2.5.1.31)	0.013	0.177	0.074	0.069	100.0	96.4	8.6
sync_1032	biotin synthase (EC:2.8.1.6)	0.003	0.152	0.017	0.020	97.0	91.7	9.6
sync_1033	hypothetical protein	0.000	0.000	NaN	NaN	3.2	0.0	0.1
sync_1034	rhodanese domain protein	0.024	0.204	0.117	0.117	100.0	98.5	10.8
sync_1035	Glutathione S-transferase domain protein	0.031	0.204	0.152	0.171	100.0	100.0	10.1
sync_1036	hypothetical protein	0.015	0.216	0.068	0.070	100.0	95.4	11.1
sync_1037	lipoic acid synthetase (EC:2.8.1.-)	0.023	0.238	0.096	0.092	100.0	100.0	14.3
sync_1038	recombination protein RecR	0.011	0.267	0.043	0.045	100.0	100.0	12.1
sync_1039	PsbP	0.000	0.173	0.000	0.000	100.0	100.0	13.3
sync_1040	hypothetical protein	0.023	0.228	0.102	0.101	100.0	100.0	18.9
sync_1041	putative RND family outer membrane efflux protein	0.050	0.295	0.169	0.179	100.0	100.0	21.8
sync_1042	hypothetical protein	0.023	0.285	0.081	0.102	94.5	19.4	3.0
sync_1043	hypothetical protein	0.030	0.134	0.222	0.199	86.3	17.5	2.0
sync_1044	putative cell envelope-related function transcriptional attenuator	0.064	0.389	0.165	0.172	100.0	92.9	19.6
sync_1045	transporter, monovalent cation:proton antiporter-2 (CPA2) family protein	0.027	0.287	0.093	0.096	100.0	100.0	19.0
sync_1046	ABC transporter, multidrug efflux family protein	0.023	0.238	0.095	0.100	100.0	100.0	12.4
sync_1047	hypothetical protein	0.097	0.050	1.959	1.316	100.0	63.2	6.7
sync_1048	putative RNA-binding protein	0.002	0.061	0.031	0.032	100.0	66.7	7.4
sync_1049	hypothetical protein	0.130	0.098	1.335	1.231	100.0	67.9	8.1
sync_1050	hypothetical protein	0.051	0.230	0.222	0.271	100.0	81.0	9.1
sync_1051	Superfamily II DNA/RNA helicase	0.008	0.217	0.039	0.040	100.0	90.9	9.8
sync_1052	Mannose-6-phosphate isomerase (EC:5.3.1.8)	0.025	0.217	0.115	0.130	100.0	100.0	9.5
sync_1053	exodeoxyribonuclease V, alpha subunit (EC:3.1.11.5)	0.030	0.141	0.214	0.198	100.0	87.6	6.9
sync_1054	exodeoxyribonuclease V, beta subunit (EC:3.1.11.5)	0.024	0.151	0.160	0.173	100.0	75.0	7.1
sync_1055	possible exodeoxyribonuclease V gamma chain (EC:3.1.11.5)	0.029	0.177	0.162	0.157	100.0	99.4	13.0
sync_1056	Serine/threonine specific protein phosphatase (EC:3.6.1.41)	0.070	0.278	0.253	0.268	100.0	100.0	20.5
sync_1057	hypothetical protein	0.029	0.283	0.102	0.088	82.5	53.7	5.4
sync_1058	Antibiotic biosynthesis monooxygenase domain protein	0.055	0.283	0.194	0.183	100.0	100.0	30.2
sync_1059	transporter, bile acid/Na+ symporter family protein	0.034	0.193	0.175	0.194	100.0	100.0	30.1
sync_1060	hypothetical protein	0.046	0.157	0.296	0.296	100.0	100.0	38.5
sync_1061	hypothetical protein	0.059	0.271	0.219	0.244	100.0	100.0	38.3
sync_1062	hypothetical protein	0.000	0.000	NaN	NaN	33.3	0.0	0.3
sync_1063	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1064	hypothetical protein	0.000	0.000	NaN	NaN	79.2	0.0	1.4
sync_1065	hypothetical protein	0.029	0.328	0.088	0.072	100.0	100.0	18.6
sync_1066	hypothetical protein	0.033	0.414	0.079	0.098	100.0	100.0	29.0
sync_1067	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1068	hypothetical protein	0.000	0.000	NaN	NaN	78.0	0.0	0.8
sync_1069	hypothetical protein	0.000	0.000	NaN	NaN	67.7	0.0	1.5
sync_1070	Predicted intracellular protease/amidase, ThiJ/PfpL family protein	0.000	0.000	NaN	NaN	84.6	0.0	1.2
sync_1071	hypothetical protein	0.000	0.000	NaN	NaN	58.5	0.0	1.1
sync_1072	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1073	lipoprotein, putative	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1074	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1075	hypothetical protein	0.000	0.000	NaN	NaN	46.2	0.0	0.5
sync_1076	hypothetical protein	0.135	0.145	0.932	0.952	100.0	100.0	10.2
sync_1077	ferritin	0.018	0.117	0.154	0.152	100.0	100.0	20.3
sync_1078	isochorismatase family protein	0.000	0.000	NaN	NaN	71.8	0.0	0.7
sync_1079	sensory box histidine kinase/response regulator	0.000	0.000	NaN	NaN	30.0	0.0	0.5
sync_1080	hypothetical protein	0.000	0.000	NaN	NaN	1.5	0.0	0.0
sync_1081	bacterial metallothionein-related protein	0.000	0.000	NaN	NaN	0.6	0.0	0.0
sync_1082	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1083	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1084	hypothetical protein	0.000	0.000	NaN	NaN	1.0	0.0	0.0
sync_1085	hypothetical protein	0.159	0.047	3.428	2.368	100.0	42.5	4.5
sync_1086	possible high light inducible protein	0.104	0.196	0.530	0.526	100.0	100.0	19.4
sync_1087	hypothetical protein	0.000	0.000	NaN	NaN	59.0	0.0	1.1
sync_1088	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_1089	methyltransferase, putative	0.000	0.000	NaN	NaN	44.7	0.0	0.7
sync_1090	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	1.4
sync_1091	hypothetical protein	0.000	0.000	NaN	NaN	29.4	0.0	0.3
sync_1092	FAD-dependent monooxygenase	0.000	0.000	NaN	NaN	65.3	0.0	0.9
sync_1093	hypothetical protein	0.000	0.000	NaN	NaN	73.5	0.0	0.7
sync_1094	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1095	hypothetical protein	0.045	0.148	0.303	0.315	100.0	79.8	8.1
sync_1096	putative nuclease	0.000	0.000	NaN	NaN	82.2	0.0	0.8
sync_1097	hypothetical protein	0.052	0.204	0.256	0.250	100.0	99.1	7.3
sync_1098	hypothetical protein	0.000	0.000	NaN	NaN	21.1	0.0	0.2
sync_1099	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	2.9
sync_1100	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	1.0
sync_1101	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	1.0
sync_1102	hypothetical protein	0.037	0.234	0.157	0.227	75.1	18.2	2.7
sync_1103	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1104	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	1.0
sync_1105	hypothetical protein	0.167	0.585	0.285	0.284	97.3	85.6	33.1
sync_1106	hypothetical protein	0.079	0.439	0.179	0.192	89.2	61.7	7.2
sync_1107	hypothetical protein	0.000	0.000	NaN	NaN	10.4	0.0	0.1
sync_1108	hypothetical protein	0.000	0.000	NaN	NaN	78.5	0.0	0.8
sync_1109	hypothetical protein	0.000	0.000	NaN	NaN	11.7	0.0	0.1
sync_1110	ParA family chromosome partitioning protein	0.000	0.000	NaN	NaN	33.9	0.0	0.3
sync_1111	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1112	hypothetical protein	0.084	0.322	0.261	0.271	99.6	94.9	20.1
sync_1113	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1114	cupin domain protein	0.000	0.000	NaN	NaN	70.4	0.0	1.0
sync_1115	lipoprotein, putative	0.000	0.000	NaN	NaN	26.1	0.0	0.3
sync_1116	DNA-binding response regulator	0.041	0.283	0.145	0.165	100.0	79.3	7.0
sync_1117	hypothetical protein	0.000	0.000	NaN	NaN	44.1	0.0	0.4
sync_1118	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1119	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1120	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1121	hypothetical protein	0.000	0.000	NaN	NaN	99.3	0.0	1.0
sync_1122	hypothetical protein	0.000	0.000	NaN	NaN	21.8	0.0	0.2
sync_1123	hypothetical protein	0.109	0.349	0.313	0.336	100.0	100.0	10.9
sync_1124	periplasmic serine proteinase	0.000	0.000	NaN	NaN	53.1	0.0	1.1
sync_1125	choloylglycine hydrolase (EC:3.5.1.24)	0.000	0.000	NaN	NaN	37.4	0.0	0.6
sync_1126	possible MFS family transporter, putative	0.000	0.000	NaN	NaN	71.0	0.0	0.8
sync_1127	hypothetical protein	0.085	0.250	0.338	0.314	100.0	60.5	6.1
sync_1128	hypothetical protein	0.493	0.460	1.071	1.076	37.9	37.9	17.6
sync_1129	type I secretion target GGXGXDXXX repeat protein (EC:3.1.-)	0.000	0.000	NaN	0.294	37.5	8.3	1.3
sync_1130	hypothetical protein	0.150	0.288	0.519	0.531	100.0	90.1	16.6
sync_1132	hypothetical protein	0.030	0.277	0.108	0.104	100.0	100.0	26.3
sync_1133	histidine protein kinase; sensor protein (EC:2.7.3.-)	0.066	0.255	0.259	0.254	100.0	99.9	16.0
sync_1134	Response regulator receiver domain	0.039	0.282	0.140	0.141	100.0	100.0	12.3
sync_1136	transferase hexapaptide repeat protein	0.042	0.184	0.228	0.235	100.0	100.0	18.7
sync_1137	Photosystem II protein Y-related protein	0.021	0.151	0.140	0.132	100.0	100.0	19.1
sync_1138	gid protein	0.017	0.277	0.063	0.060	100.0	100.0	16.8
sync_1139	putative cation efflux transporter (CDF family)	0.023	0.103	0.221	0.242	100.0	100.0	13.8
sync_1140	carotenoid isomerase (EC:5.-)	0.018	0.240	0.076	0.076	100.0	99.5	19.1
sync_1141	hypothetical protein	0.131	0.146	0.901	0.997	100.0	68.2	5.9
sync_1142	hypothetical protein	0.073	0.324	0.225	0.243	100.0	100.0	22.7
sync_1143	hypothetical protein	0.045	0.304	0.148	0.147	100.0	100.0	23.8
sync_1144	hypothetical protein	0.178	0.220	0.808	0.882	100.0	100.0	16.7
sync_1145	two-component response regulator	0.009	0.229	0.038	0.037	100.0	100.0	18.0
sync_1146	hypothetical protein	0.005	0.265	0.019	0.033	100.0	100.0	23.1
sync_1147	glutaredoxin homolog	0.018	0.299	0.060	0.087	100.0	100.0	21.3
sync_1148	BolA family protein	0.019	0.186	0.103	0.103	100.0	100.0	23.0
sync_1149	Uncharacterized membrane protein	0.016	0.271	0.058	0.058	100.0	100.0	17.7
sync_1150	Phospholipid and glycerol acyltransferase	0.018	0.282	0.063	0.063	100.0	100.0	19.5
sync_1151	pyridoxal phosphate biosynthetic protein PdxJ	0.017	0.176	0.094	0.092	100.0	100.0	25.2
sync_1152	hypothetical protein	0.021	0.211	0.102	0.121	100.0	100.0	23.8
sync_1153	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1154	hypothetical protein	0.061	0.342	0.178	0.180	100.0	100.0	15.8
sync_1155	hypothetical protein	0.111	0.108	1.021	1.172	100.0	100.0	25.8

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_1156	hypothetical protein	0.053	0.321	0.165	0.163	100.0	100.0	27.6
sync_1157	hypothetical protein	0.180	0.212	0.846	0.688	90.1	55.3	4.6
sync_1158	possible phosphoribosyltransferase family protein (EC:2.4.2.22)	0.125	0.412	0.303	0.292	100.0	100.0	40.5
sync_1159	Nucleoside 2-deoxyribosyltransferase	0.060	0.392	0.152	0.162	100.0	100.0	34.7
sync_1160	hypothetical protein	0.094	0.329	0.284	0.298	100.0	100.0	34.4
sync_1161	glutathione S-transferase (EC:2.5.1.18)	0.076	0.425	0.178	0.187	100.0	100.0	32.5
sync_1162	possible light-dependent protochlorophyllide oxido-reductase (EC:1.3.1.33)	0.065	0.266	0.244	0.254	100.0	100.0	21.9
sync_1163	hypothetical protein	0.043	0.173	0.249	0.199	100.0	100.0	17.5
sync_1164	hypothetical protein	0.053	0.051	1.033	1.223	100.0	81.7	6.9
sync_1165	hypothetical protein	0.059	0.288	0.204	0.214	93.9	77.2	11.9
sync_1166	hypothetical protein	0.008	0.265	0.031	0.030	100.0	100.0	33.7
sync_1167	transcriptional regulator, Fur family protein	0.021	0.298	0.070	0.071	100.0	100.0	25.0
sync_1168	hypothetical protein	0.048	0.280	0.172	0.196	100.0	100.0	22.7
sync_1169	hypothetical protein	0.022	0.195	0.113	0.110	100.0	100.0	22.1
sync_1170	possible L-Asparaginase II	0.070	0.347	0.202	0.198	100.0	100.0	25.4
sync_1171	glucosylglycerol 3-phosphatase (EC:3.1.3.69)	0.049	0.332	0.149	0.141	100.0	100.0	29.2
sync_1172	glutathione S-transferase gst (EC:2.5.1.18)	0.057	0.253	0.226	0.225	100.0	91.9	11.5
sync_1173	hypothetical protein	0.010	0.336	0.029	0.033	100.0	100.0	30.5
sync_1174	RecF/RecN/SMC N terminal domain protein	0.063	0.264	0.239	0.250	100.0	100.0	26.7
sync_1175	DNA repair exonuclease (EC:3.1.-)	0.040	0.241	0.166	0.171	100.0	100.0	18.3
sync_1176	hypothetical protein	0.037	0.222	0.169	0.163	100.0	100.0	20.4
sync_1177	hypothetical protein	0.065	0.260	0.250	0.242	100.0	100.0	17.3
sync_1178	possible Transglycosylase SLT domain	0.020	0.137	0.148	0.166	100.0	100.0	11.1
sync_1179	hypothetical protein	0.021	0.198	0.105	0.095	100.0	100.0	18.3
sync_1180	O-acetylhomoserine sulfhydrylase (EC:2.5.1.49)	0.007	0.203	0.032	0.033	100.0	99.2	15.1
sync_1181	homoserine O-succinyltransferase (EC:2.3.1.46)	0.008	0.182	0.042	0.045	100.0	100.0	10.8
sync_1182	hypothetical protein	0.033	0.164	0.201	0.190	100.0	96.4	7.1
sync_1183	possible alkylated DNA repair protein	0.036	0.193	0.188	0.166	100.0	100.0	14.6
sync_1184	transporter, auxin efflux carrier (AEC) family protein	0.052	0.252	0.207	0.208	100.0	100.0	17.7
sync_1185	hypothetical protein	0.047	0.177	0.262	0.278	93.5	93.5	13.4
sync_1186	polar amino amino acid ABC transporter (PAAT) family, ATP-binding protein	0.021	0.177	0.121	0.132	100.0	100.0	15.1
sync_1187	polar amino amino acid ABC transporter (PAAT) family, permease protein	0.032	0.229	0.141	0.146	100.0	100.0	19.1
sync_1188	hypothetical protein	0.025	0.199	0.124	0.136	100.0	100.0	25.6
sync_1189	polar amino amino acid ABC transporter (PAAT) family, permease protein	0.027	0.250	0.107	0.108	100.0	100.0	22.1
sync_1190	polar amino amino acid ABC transporter (PAAT) family, periplasmic amino acid-binding protein	0.044	0.316	0.138	0.146	100.0	100.0	30.8
sync_1191	Sulfate permease	0.024	0.337	0.073	0.067	100.0	100.0	26.3
sync_1192	hypothetical protein	0.067	0.314	0.212	0.226	100.0	100.0	16.9
sync_1193	hypothetical protein	0.025	0.120	0.208	0.257	100.0	100.0	10.0
sync_1194	possible oxidoreductase	0.090	0.284	0.316	0.310	100.0	100.0	15.7
sync_1195	sirohydrochlorin cobaltochelatase (EC:4.99.1.3)	0.027	0.238	0.115	0.116	100.0	100.0	18.1
sync_1196	hypothetical protein	0.150	0.187	0.801	0.749	100.0	100.0	17.8
sync_1197	hypothetical protein	0.050	0.329	0.151	0.139	100.0	100.0	18.2
sync_1198	Glucose-methanol-choline (GMC) oxidoreductase:NAD binding site	0.042	0.256	0.163	0.164	100.0	100.0	14.1
sync_1200	L-asparaginase, thermolabile (EC:3.5.1.1)	0.015	0.264	0.055	0.062	100.0	100.0	18.0
sync_1201	hypothetical protein	0.016	0.282	0.057	0.056	100.0	100.0	22.3
sync_1202	hypothetical protein	0.012	0.287	0.043	0.047	100.0	100.0	16.2
sync_1203	Uncharacterized membrane protein	0.004	0.236	0.018	0.024	100.0	100.0	16.1
sync_1204	carbamoyl-phosphate synthase, large subunit (EC:6.3.5.5)	0.017	0.240	0.070	0.073	100.0	99.9	18.5
sync_1205	hypothetical protein	0.015	0.211	0.073	0.078	74.5	66.4	12.4
sync_1206	Sodium:alanine symporter family protein	0.000	0.000	NaN	NaN	28.8	0.0	0.4
sync_1207	hypothetical protein	0.016	0.198	0.082	0.107	100.0	91.2	12.9
sync_1208	ABC transporter, permease/ATP-binding protein	0.020	0.279	0.072	0.076	100.0	100.0	21.4
sync_1209	S4 domain protein	0.053	0.220	0.243	0.239	100.0	100.0	32.8
sync_1210	triosephosphate isomerase (EC:5.3.1.1)	0.009	0.248	0.038	0.040	100.0	100.0	22.8
sync_1211	dihydropteroate synthase (EC:2.5.1.15)	0.024	0.256	0.092	0.119	45.9	39.6	7.7
sync_1212	hypothetical protein	0.000	0.000	NaN	NaN	78.1	0.0	1.4
sync_1213	hypothetical protein	0.000	0.000	NaN	NaN	82.1	0.0	1.3
sync_1214	ABC transporter RzcB, putative	0.000	0.000	NaN	NaN	21.9	0.0	0.3
sync_1215	ABC transporter, ATP-binding protein domain protein	0.000	0.000	NaN	NaN	27.5	0.0	0.3
sync_1216	HlyD family secretion protein domain protein	0.000	0.000	NaN	NaN	15.9	0.0	0.2
sync_1217	structural toxin protein RtxA	0.000	0.000	NaN	NaN	7.9	0.0	0.1
sync_1218	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1219	two-component response regulator	0.000	0.000	NaN	NaN	0.0	0.0	0.0

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_1220	magnesium chelatase, H subunit (EC:6.6.1.1)	0.014	0.304	0.045	0.046	100.0	99.8	28.0
sync_1221	dihydrodipicolinate reductase (EC:1.3.1.26)	0.023	0.262	0.086	0.090	100.0	100.0	19.4
sync_1222	hypothetical protein	0.027	0.321	0.084	0.093	100.0	100.0	27.8
sync_1223	CAB/ELIP/HLIP family protein	0.009	0.312	0.029	0.035	100.0	100.0	28.4
sync_1224	putative 2-octaprenyl-6-methoxyphenol hydroxylase	0.067	0.321	0.207	0.203	100.0	100.0	20.5
sync_1225	hypothetical protein	0.021	0.393	0.054	0.091	100.0	100.0	15.1
sync_1226	adenine phosphoribosyltransferase (EC:2.4.2.7)	0.042	0.304	0.136	0.141	100.0	100.0	20.8
sync_1227	hypothetical protein	0.024	0.276	0.089	0.098	100.0	100.0	31.9
sync_1228	Uncharacterized membrane protein	0.040	0.298	0.133	0.143	100.0	100.0	26.6
sync_1229	hypothetical protein	0.015	0.202	0.074	0.072	100.0	100.0	18.8
sync_1230	hypothetical protein	0.069	0.264	0.261	0.257	100.0	100.0	18.1
sync_1231	EF hand domain protein	0.045	0.266	0.168	0.174	100.0	100.0	27.3
sync_1232	two-component response regulator	0.018	0.257	0.068	0.066	100.0	100.0	28.2
sync_1233	two-component sensor histidine kinase (EC:2.7.3.-)	0.028	0.333	0.083	0.086	100.0	100.0	32.3
sync_1234	PvdS (EC:2.-)	0.044	0.399	0.111	0.114	100.0	100.0	44.4
sync_1235	hypothetical protein	0.102	0.140	0.729	0.875	79.6	37.4	3.7
sync_1236	hypothetical protein	0.045	0.241	0.188	0.133	100.0	21.6	2.7
sync_1237	hypothetical protein	0.156	0.400	0.390	0.377	100.0	100.0	31.1
sync_1238	Ferrochelatase (EC:4.99.1.1)	0.130	0.331	0.395	0.378	100.0	100.0	28.9
sync_1239	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1240	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	3.0
sync_1241	hypothetical protein	0.000	0.000	NaN	NaN	27.2	0.0	0.3
sync_1242	hypothetical protein	0.000	0.000	NaN	NaN	28.5	0.0	0.3
sync_1243	hypothetical protein	0.045	0.274	0.163	0.136	100.0	100.0	19.2
sync_1244	oxidoreductase, short-chain dehydrogenase/reductase family protein (EC:1.-)	0.029	0.145	0.198	0.190	100.0	100.0	26.8
sync_1245	possible Carbamoyl-phosphate synthase L chain	0.208	0.711	0.292	0.288	75.0	64.8	33.5
sync_1246	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1247	hypothetical protein	0.000	0.000	NaN	NaN	34.3	0.0	0.3
sync_1248	hypothetical protein	0.012	0.137	0.086	0.089	100.0	100.0	19.0
sync_1249	possible phycobilisome rod-core linker polypeptide (L-RC 28.5)	0.044	0.259	0.170	0.175	100.0	85.1	25.8
sync_1250	Predicted membrane protein-related protein	0.004	0.323	0.013	0.026	100.0	100.0	10.9
sync_1251	hypothetical protein	0.000	0.260	0.000	0.007	100.0	100.0	35.9
sync_1252	Staphylococcal nuclease homologue	0.098	0.304	0.324	0.302	100.0	100.0	38.9
sync_1253	glutamine synthetase (EC:6.3.1.2)	0.022	0.362	0.062	0.062	100.0	100.0	33.8
sync_1254	SAM-dependent methyltransferase	0.077	0.382	0.201	0.209	100.0	100.0	22.6
sync_1255	hypothetical protein	0.488	0.769	0.634	1.147	58.3	52.8	5.6
sync_1256	hypothetical protein	0.118	0.189	0.624	0.789	100.0	90.5	10.9
sync_1257	hypothetical protein	0.062	0.390	0.160	0.129	100.0	100.0	34.7
sync_1258	putative nuclease	0.039	0.119	0.326	0.335	100.0	100.0	22.3
sync_1259	hypothetical protein	0.035	0.241	0.147	0.156	100.0	100.0	26.6
sync_1260	type I secretion target GGXGXDXXX repeat protein	0.000	0.000	NaN	NaN	12.7	0.0	0.1
sync_1261	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1262	hypothetical protein	0.027	0.048	0.567	0.544	100.0	62.2	4.7
sync_1263	Type II alternative RNA polymerase sigma factor, sigma-70 family protein	0.058	0.329	0.177	0.176	100.0	99.4	25.2
sync_1265	ribonucleotide reductase (Class II) (EC:1.17.4.2)	0.026	0.429	0.061	0.061	100.0	100.0	26.2
sync_1266	hypothetical protein	0.000	0.000	NaN	NaN	59.8	1.6	1.6
sync_1267	hypothetical protein	0.000	0.000	NaN	NaN	49.9	0.0	0.6
sync_1268	SAM-dependent methyltransferase	0.089	0.295	0.303	0.302	94.2	56.8	6.5
sync_1269	peptide chain release factor 3	0.020	0.371	0.055	0.055	100.0	100.0	24.0
sync_1270	hypothetical protein	0.022	0.191	0.115	0.123	100.0	88.2	7.9
sync_1271	cyanobacteria-specific chaperone containing dnaj domain fused to a membrane domain	0.041	0.376	0.110	0.114	100.0	100.0	16.8
sync_1272	33 kDa chaperonin (Heat shock protein 33 homolog) (HSP33)	0.023	0.309	0.074	0.078	100.0	100.0	17.0
sync_1273	ABC transporter, ATP-binding protein	0.024	0.204	0.117	0.105	100.0	100.0	15.1
sync_1274	hypothetical protein	0.011	0.249	0.043	0.044	100.0	100.0	13.0
sync_1275	ribosomal protein S2	0.008	0.202	0.038	0.036	100.0	100.0	12.7
sync_1276	translation elongation factor Ts	0.006	0.231	0.028	0.030	100.0	96.4	12.8
sync_1277	hypothetical protein	0.047	0.201	0.236	0.228	100.0	100.0	11.9
sync_1278	ATP-dependent DNA helicase RecG (EC:3.6.1.-)	0.017	0.225	0.074	0.072	100.0	99.6	13.3
sync_1279	D-ala-D-ala dipeptidase (EC:3.4.13.-)	0.065	0.272	0.239	0.245	100.0	100.0	15.8
sync_1280	Ferredoxin-sulfite reductase (EC:1.8.7.1)	0.020	0.296	0.066	0.067	100.0	100.0	12.5
sync_1281	glycyl-tRNA synthetase, beta subunit (EC:6.1.1.14)	0.029	0.194	0.147	0.144	100.0	100.0	12.3
sync_1282	geranylgeranyl reductase	0.014	0.212	0.065	0.058	100.0	100.0	15.9
sync_1283	D-alanyl-D-alanine carboxypeptidase	0.014	0.158	0.091	0.086	100.0	100.0	10.7
sync_1284	collagenase (EC:3.4.-)	0.032	0.212	0.150	0.148	100.0	98.8	13.2

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_1285	GTP-binding protein TypA	0.021	0.429	0.049	0.048	100.0	100.0	29.0
sync_1286	hypothetical protein	0.068	0.402	0.169	0.161	100.0	100.0	33.4
sync_1287	hypothetical protein	0.091	0.362	0.251	0.227	100.0	100.0	37.5
sync_1288	possible ABC transporter, ATP binding component	0.029	0.485	0.060	0.061	100.0	100.0	45.7
sync_1289	putative membrane protein, YjgP/YjgQ family protein	0.038	0.483	0.078	0.086	100.0	100.0	38.9
sync_1290	cytochrome c assembly protein	0.026	0.457	0.056	0.060	100.0	100.0	47.8
sync_1291	hypothetical protein	0.061	0.635	0.096	0.098	100.0	100.0	61.9
sync_1292	hypothetical protein	0.134	0.352	0.381	0.345	100.0	100.0	28.8
sync_1293	chaperone protein dnaK2 (Heat shock protein 70-2) (Heat shock 70 kDaprotein 2) (HSP70-2)	0.042	0.339	0.125	0.120	100.0	100.0	28.3
sync_1294	hypothetical protein	0.089	0.364	0.245	0.251	100.0	97.0	19.3
sync_1295	membrane protein, TerC family protein	0.085	0.508	0.167	0.160	100.0	100.0	36.6
sync_1296	hypothetical protein	0.079	0.418	0.190	0.214	100.0	100.0	26.1
sync_1297	hypothetical protein	0.116	0.308	0.377	0.373	98.8	71.0	9.8
sync_1298	hypothetical protein	0.000	0.000	NaN	NaN	8.9	0.0	0.2
sync_1299	hypothetical protein	0.079	0.377	0.211	0.230	100.0	93.0	18.1
sync_1300	DNA and RNA helicase	0.014	0.311	0.045	0.051	100.0	100.0	24.0
sync_1301	methionyl-tRNA formyltransferase (EC:2.1.2.9)	0.043	0.292	0.146	0.161	100.0	100.0	17.2
sync_1302	tldD/pmbA family protein	0.019	0.270	0.071	0.068	100.0	100.0	14.2
sync_1303	TldD protein	0.015	0.271	0.055	0.061	100.0	97.2	12.1
sync_1304	magnesium-protoporphyrin IX monomethyl ester aerobic oxidative cyclase	0.012	0.264	0.045	0.037	100.0	84.9	12.6
sync_1305	hypothetical protein	0.013	0.217	0.060	0.060	100.0	100.0	10.4
sync_1306	hypothetical protein	0.100	0.027	3.678	5.884	100.0	100.0	11.4
sync_1307	hypothetical protein	0.070	0.178	0.393	0.357	100.0	100.0	10.1
sync_1308	decarboxylase, polyprenyl P-hydroxybenzoate/phenylacrylic acid family protein	0.008	0.189	0.044	0.048	100.0	100.0	8.2
sync_1309	acetazolamide conferring resistance protein Zam	0.027	0.217	0.125	0.127	100.0	98.3	9.2
sync_1310	hypothetical protein	0.012	0.199	0.059	0.065	100.0	100.0	7.4
sync_1311	hypothetical protein	0.006	0.169	0.034	0.045	100.0	100.0	12.6
sync_1312	Fe-S cluster protein	0.025	0.108	0.227	0.228	100.0	89.9	8.7
sync_1313	Protein of unknown function (DUF888) family protein	0.000	0.032	0.000	0.000	100.0	73.8	5.8
sync_1314	possible chloride channel	0.023	0.181	0.126	0.123	100.0	86.4	7.0
sync_1315	single-stranded-DNA-specific exonuclease RecJ (EC:3.1.-)	0.017	0.198	0.086	0.086	100.0	64.1	6.2
sync_1316	Predicted phosphatase/phosphohexomutase of HAD family protein	0.030	0.153	0.198	0.182	100.0	93.7	9.2
sync_1317	Uncharacterized secreted or membrane protein	0.044	0.287	0.154	0.210	100.0	86.0	7.2
sync_1318	Phosphoribulokinase/uridine kinase family enzyme	0.050	0.204	0.245	0.223	100.0	100.0	12.4
sync_1319	ribosomal protein L32	0.011	0.170	0.064	0.065	100.0	100.0	17.5
sync_1320	cell division protein FtsH4 (EC:3.4.24.-)	0.019	0.299	0.065	0.071	100.0	100.0	17.5
sync_1321	hypothetical protein	0.053	0.360	0.146	0.140	100.0	99.5	20.7
sync_1322	thioredoxin peroxidase	0.007	0.277	0.025	0.023	100.0	100.0	27.4
sync_1323	hypothetical protein	0.029	0.365	0.080	0.088	100.0	100.0	24.2
sync_1325	peptidase, M23 family protein	0.028	0.296	0.094	0.095	100.0	98.5	17.7
sync_1326	Predicted rRNA methylase (EC:2.1.1.-)	0.041	0.373	0.109	0.110	100.0	90.4	18.9
sync_1327	Adenosyl cobinamide kinase/adenosyl cobinamide phosphate guanylyltransferase (EC:2.7.7.-)	0.062	0.244	0.256	0.247	100.0	99.8	11.6
sync_1328	CemA family protein	0.016	0.366	0.043	0.049	100.0	100.0	34.5
sync_1329	hypothetical protein	0.036	0.247	0.145	0.157	100.0	86.6	10.8
sync_1330	hypothetical protein	0.026	0.224	0.115	0.107	100.0	100.0	24.9
sync_1331	hypothetical protein	0.038	0.251	0.152	0.163	100.0	93.6	16.9
sync_1332	hypothetical protein	0.064	0.263	0.245	0.246	96.8	78.8	10.4
sync_1333	methionyl-tRNA synthetase (EC:6.1.1.10)	0.032	0.316	0.101	0.103	100.0	100.0	19.2
sync_1334	Mo-dependent nitrogenase C-terminus family protein	0.029	0.082	0.348	0.348	100.0	53.3	5.1
sync_1335	hypothetical protein	0.045	0.207	0.216	0.217	100.0	89.5	10.4
sync_1336	exoribonuclease, VacB/RNB family protein	0.026	0.216	0.120	0.134	100.0	88.1	9.3
sync_1337	ribosomal protein S18	0.004	0.242	0.017	0.017	100.0	100.0	14.0
sync_1338	ribosomal protein L33	0.000	0.051	0.000	0.000	100.0	100.0	6.3
sync_1339	phenylalanyl-tRNA synthetase, beta subunit (EC:6.1.1.20)	0.014	0.193	0.072	0.078	100.0	63.6	6.5
sync_1340	23S rRNA (uracil-5-)-methyltransferase RumA (EC:2.1.1.-)	0.017	0.141	0.119	0.118	98.8	85.1	7.4
sync_1341	allophycocyanin alpha, B subunit	0.007	0.227	0.031	0.038	100.0	93.1	9.3
sync_1342	possible DnaJ domain	0.038	0.215	0.178	0.198	99.1	79.1	7.7
sync_1343	hypothetical protein	0.034	0.284	0.120	0.117	100.0	100.0	12.6
sync_1344	hypothetical protein	0.043	0.177	0.243	0.225	100.0	100.0	19.1
sync_1345	Possible type 3 alternative RNA polymerase sigma factor	0.065	0.273	0.238	0.256	100.0	86.5	12.8
sync_1346	possible ATP adenylyltransferase (EC:2.7.7.53)	0.047	0.196	0.239	0.256	100.0	97.1	14.4
sync_1347	SpoIID/LytB domain protein	0.048	0.235	0.203	0.206	100.0	100.0	12.3
sync_1348	hypothetical protein	0.090	0.051	1.759	1.874	100.0	100.0	11.8

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_1349	possible CsgA C-factor signaling protein	0.033	0.269	0.124	0.130	100.0	100.0	13.2
sync_1350	hypothetical protein	0.015	0.139	0.109	0.099	100.0	99.8	8.7
sync_1351	hypothetical protein	0.004	0.039	0.105	0.103	100.0	93.5	8.3
sync_1352	hypothetical protein	0.024	0.180	0.135	0.155	100.0	100.0	13.1
sync_1353	methionine synthase (EC:2.1.1.13)	0.010	0.189	0.054	0.052	100.0	83.2	8.6
sync_1354	branched-chain amino acid aminotransferase (EC:2.6.1.42)	0.006	0.108	0.053	0.053	100.0	70.4	6.1
sync_1355	cobaltochelatase, CobN subunit (EC:6.6.1.2)	0.025	0.170	0.146	0.136	100.0	68.5	6.0
sync_1356	hypothetical protein	0.024	0.194	0.123	0.122	100.0	48.3	7.6
sync_1357	PHP domain protein	0.017	0.240	0.073	0.106	83.7	41.9	7.2
sync_1358	hypothetical protein	0.015	0.143	0.102	0.089	100.0	98.2	8.7
sync_1359	deoxyribodipyrimidine photolyase family protein	0.013	0.217	0.062	0.064	93.0	42.2	4.0
sync_1360	hypothetical protein	0.003	0.047	0.070	0.050	100.0	44.8	4.5
sync_1361	excinuclease ABC, C subunit	0.010	0.183	0.056	0.059	100.0	70.2	6.1
sync_1362	pantetheine-phosphate adenylyltransferase (EC:2.7.7.3)	0.009	0.161	0.055	0.045	100.0	96.7	10.5
sync_1363	hypothetical protein	0.007	0.224	0.033	0.041	100.0	97.4	7.8
sync_1364	D-alanyl-D-alanine carboxypeptidase (EC:3.4.16.4)	0.017	0.223	0.077	0.079	100.0	91.4	8.0
sync_1365	hypothetical protein	0.036	0.175	0.203	0.211	100.0	71.5	6.5
sync_1366	hypothetical protein	0.042	0.204	0.205	0.185	100.0	50.9	4.2
sync_1367	possible cysteine desulfurase (class-V aminotransferase family) (EC:2.8.1.7)	0.116	0.152	0.764	0.644	67.3	3.6	1.6
sync_1368	diaminopimelate epimerase (EC:5.1.1.7)	0.014	0.148	0.093	0.081	84.8	63.3	6.1
sync_1369	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1370	CD209 antigen	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1371	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1372	Phage integrase, putative	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1373	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1374	hypothetical protein	0.009	0.134	0.068	0.108	100.0	100.0	6.5
sync_1375	leucyl-tRNA synthetase (EC:6.1.1.4)	0.025	0.201	0.122	0.114	100.0	95.0	9.2
sync_1376	glucose-6-phosphate isomerase (EC:5.3.1.9)	0.010	0.237	0.043	0.042	100.0	85.4	9.3
sync_1377	possible DnaB-like helicase	0.023	0.220	0.105	0.112	99.9	74.7	10.2
sync_1378	N-acetylmuramoyl-L-alanine amidase (family 2)	0.029	0.213	0.135	0.140	100.0	85.8	8.7
sync_1379	PDZ domain	0.028	0.183	0.151	0.159	100.0	73.8	6.7
sync_1380	hypothetical protein	0.033	0.142	0.235	0.227	100.0	87.6	6.6
sync_1381	phosphoribosylglycinamide formyltransferase (EC:2.1.2.2)	0.040	0.239	0.169	0.203	100.0	85.7	9.8
sync_1382	N-acetyl-gamma-glutamyl-phosphate reductase (EC:1.2.1.38)	0.018	0.229	0.079	0.088	100.0	93.2	8.9
sync_1383	riboflavin biosynthesis protein RibA (EC:3.5.4.25)	0.017	0.240	0.070	0.071	100.0	98.4	13.0
sync_1384	peptidylprolyl cis-trans isomerase, cyclophilin-type (EC:5.2.1.8)	0.022	0.179	0.123	0.119	100.0	94.6	11.0
sync_1385	methylthioadenosine phosphorylase (EC:2.4.2.28)	0.016	0.157	0.103	0.096	100.0	100.0	11.0
sync_1386	glucokinase regulator homolog	0.023	0.275	0.085	0.073	100.0	89.3	9.1
sync_1387	hypothetical protein	0.018	0.283	0.065	0.066	100.0	100.0	11.4
sync_1388	DnaJ2 protein	0.025	0.173	0.143	0.136	100.0	83.9	8.1
sync_1389	chaperone protein DnaK	0.011	0.263	0.041	0.043	100.0	97.2	12.8
sync_1390	phosphate ABC transporter, permease protein PstC	0.009	0.249	0.037	0.036	100.0	68.9	7.6
sync_1391	phosphate ABC transporter, permease protein PtsA	0.014	0.065	0.219	0.227	87.4	7.1	1.9
sync_1392	phosphate ABC transporter, ATP-binding protein (EC:3.6.3.27)	0.014	0.322	0.044	0.052	97.2	84.9	10.7
sync_1394	ferredoxin, 2Fe-2S	0.028	0.309	0.089	0.076	100.0	100.0	15.8
sync_1395	Inositol-1-monophosphatase (EC:3.1.3.25)	0.033	0.311	0.107	0.102	100.0	100.0	15.8
sync_1396	ATP phosphoribosyltransferase regulatory subunit	0.038	0.293	0.131	0.126	100.0	100.0	23.8
sync_1397	iron-sulfur cluster-binding protein	0.041	0.265	0.154	0.148	100.0	100.0	24.2
sync_1398	chaperone protein HtpG	0.028	0.368	0.075	0.077	100.0	100.0	27.2
sync_1399	ribosomal protein L28	0.022	0.386	0.056	0.045	100.0	100.0	35.7
sync_1400	Alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergens family protein	0.063	0.440	0.143	0.146	100.0	100.0	32.6
sync_1401	glucosylglycerol-phosphate synthase (EC:2.4.1.213)	0.021	0.396	0.054	0.056	100.0	100.0	32.8
sync_1402	ABC transporter for sugars, solute-binding protein	0.234	0.147	1.591	1.512	97.7	2.3	1.9
sync_1403	ABC transporter for sugars, membrane component	0.004	0.171	0.024	0.021	76.0	12.7	2.0
sync_1404	ABC transporter component, likely for sugar transport	0.020	0.282	0.070	0.073	85.2	50.6	7.8
sync_1405	ABC transporter for sugars, ATP binding protein	0.050	0.376	0.134	0.153	100.0	100.0	17.9
sync_1406	hypothetical protein	0.036	0.198	0.181	0.188	99.4	78.5	13.2
sync_1407	hypothetical protein	0.020	0.282	0.070	0.083	100.0	100.0	15.1
sync_1408	hypothetical protein	0.024	0.170	0.143	0.134	100.0	100.0	14.4
sync_1409	Photosystem I reaction center subunit psaK (Photosystem I subunit X)	0.014	0.083	0.169	0.170	75.2	58.5	10.6
sync_1410	1-deoxy-D-xylulose-5-phosphate synthase (EC:2.2.1.7)	0.010	0.273	0.035	0.036	100.0	94.2	11.5
sync_1411	threonine dehydratase (EC:4.3.1.19)	0.011	0.233	0.045	0.046	100.0	93.1	10.4
sync_1412	segregation and condensation protein B	0.013	0.259	0.051	0.048	100.0	100.0	15.6
sync_1413	hypothetical protein	0.002	0.151	0.010	0.000	100.0	100.0	8.9

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_1414	Predicted MazG family pyrophosphatase	0.024	0.154	0.154	0.150	100.0	99.2	8.6
sync_1415	pyruvate kinase (EC:2.7.1.40)	0.005	0.191	0.029	0.031	91.5	79.3	8.8
sync_1416	macrolide ABC transporter, permease protein	0.008	0.240	0.032	0.031	100.0	98.1	11.6
sync_1417	metalloprotease, ATP-dependent, FtsH family protein (EC:3.4.24.-)	0.011	0.340	0.033	0.035	100.0	100.0	15.2
sync_1418	ATP-dependent Clp protease, proteolytic subunit ClpP (EC:3.4.21.92)	0.005	0.274	0.020	0.021	100.0	100.0	24.6
sync_1419	hypothetical protein	0.020	0.160	0.124	0.130	100.0	99.5	14.2
sync_1420	hypothetical protein	0.062	0.278	0.222	0.215	96.5	57.5	12.0
sync_1421	hypothetical protein	0.009	0.091	0.103	0.105	100.0	100.0	22.0
sync_1422	Cytochrome b6-f complex subunit VIII (Cytochrome b6f complex subunitpetN)-related protein	0.005	0.169	0.027	0.025	100.0	100.0	21.7
sync_1423	ATP-dependent Clp protease adaptor protein ClpS	0.020	0.242	0.082	0.089	100.0	100.0	25.6
sync_1424	transporter, major facilitator family protein	0.026	0.329	0.079	0.081	100.0	100.0	27.8
sync_1425	hypothetical protein	0.052	0.340	0.153	0.199	100.0	100.0	37.2
sync_1426	hypothetical protein	0.020	0.238	0.083	0.068	100.0	67.0	16.5
sync_1428	phosphoribosylaminoimidazole carboxylase, ATPase subunit (EC:4.1.1.21)	0.026	0.272	0.097	0.108	100.0	100.0	17.8
sync_1429	lactose transport system permease protein	0.015	0.263	0.057	0.062	100.0	100.0	21.5
sync_1430	3-dehydroquinate synthase (EC:4.2.3.4)	0.045	0.249	0.181	0.168	100.0	100.0	21.1
sync_1431	hypothetical protein	0.055	0.187	0.296	0.304	100.0	100.0	16.4
sync_1432	Uncharacterized membrane protein	0.010	0.257	0.038	0.030	100.0	79.3	9.6
sync_1433	possible endolysin (EC:3.2.1.17)	0.033	0.353	0.093	0.091	100.0	100.0	23.6
sync_1434	hypothetical protein	0.028	0.126	0.222	0.213	100.0	91.9	9.0
sync_1435	Predicted phosphoesterase	0.033	0.314	0.107	0.105	100.0	100.0	17.7
sync_1436	quinolinate synthetase complex, A subunit	0.035	0.369	0.095	0.104	95.3	69.4	14.5
sync_1437	predicted exonuclease involved in mRNA processing	0.056	0.297	0.189	0.181	95.9	85.2	15.9
sync_1438	possible ATP-dependent DNA ligase (EC:6.5.1.1)	0.041	0.243	0.169	0.171	42.9	34.3	5.8
sync_1439	ATP-dependent helicase, DEAD/DEAH box family protein	0.050	0.268	0.187	0.190	91.0	74.9	15.3
sync_1440	hypothetical protein	0.000	0.000	NaN	NaN	94.6	0.0	1.7
sync_1441	hypothetical protein	0.030	0.158	0.186	0.240	53.3	48.5	4.6
sync_1442	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1443	Acyltransferase family, putative	0.000	0.000	NaN	NaN	66.6	0.0	1.0
sync_1444	hypothetical protein	0.000	0.000	NaN	NaN	34.9	0.0	0.4
sync_1445	hypothetical protein	0.000	0.000	NaN	NaN	37.7	0.0	0.4
sync_1446	tetratricopeptide repeat protein	0.000	0.000	NaN	NaN	51.6	0.0	0.5
sync_1447	hypothetical protein	0.000	0.000	NaN	NaN	61.5	0.0	0.7
sync_1448	SPFH domain / Band 7 family protein	0.021	0.373	0.056	0.061	99.7	88.0	18.3
sync_1449	SPFH domain / Band 7 family protein	0.029	0.390	0.074	0.080	100.0	100.0	37.2
sync_1450	hypothetical protein	0.032	0.262	0.122	0.107	80.6	68.5	12.0
sync_1451	hypothetical protein	0.066	0.274	0.240	0.249	86.0	59.2	10.9
sync_1452	DnaJ domain-containing protein	0.053	0.234	0.226	0.226	88.1	73.1	14.5
sync_1453	Protein of unknown function (DUF558) family protein	0.030	0.234	0.130	0.129	54.8	37.1	4.3
sync_1454	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	3.5
sync_1455	Short-chain dehydrogenase/reductase	0.038	0.225	0.168	0.169	100.0	100.0	15.0
sync_1456	hypothetical protein	0.030	0.375	0.080	0.081	100.0	99.9	16.7
sync_1457	cytochrome P450 family protein (EC:1.14.-)	0.034	0.171	0.198	0.193	99.9	97.9	12.3
sync_1458	SAM (and some other nucleotide) binding motif	0.033	0.095	0.351	0.300	100.0	100.0	17.9
sync_1459	NAD binding site	0.066	0.305	0.215	0.211	100.0	100.0	24.1
sync_1460	Chalcone synthase (EC:2.3.1.74)	0.064	0.297	0.217	0.220	100.0	100.0	24.4
sync_1461	hypothetical protein	0.113	0.363	0.311	0.324	100.0	100.0	27.9
sync_1462	putative esterase/lipase/thioesterase	0.036	0.153	0.233	0.224	100.0	100.0	16.9
sync_1463	D-isomer specific 2-hydroxyacid dehydrogenase family protein (EC:1.-)	0.015	0.183	0.082	0.071	100.0	96.8	15.4
sync_1464	hypothetical protein	0.036	0.336	0.108	0.110	100.0	100.0	21.3
sync_1465	glycosyl transferase, group 2 family protein domain protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1466	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1467	endoglucanase (EC:3.2.1.4)	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1468	hypothetical protein	0.070	0.291	0.239	0.208	100.0	49.5	11.9
sync_1469	hypothetical protein	0.035	0.314	0.112	0.101	100.0	100.0	33.8
sync_1470	hypothetical protein	0.054	0.176	0.308	0.292	100.0	88.4	8.7
sync_1471	feruloyl-CoA synthetase (EC:6.2.1.34)	0.030	0.118	0.254	0.237	83.8	72.2	15.0
sync_1472	possible transcriptional regulator protein	0.058	0.020	2.916	3.238	44.6	3.8	0.9
sync_1473	Predicted calcineurin family phosphoesterase	0.074	0.359	0.206	0.219	100.0	98.9	18.9
sync_1474	hypothetical protein	0.000	0.000	NaN	NaN	38.3	0.0	0.4
sync_1475	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1476	hypothetical protein	0.000	0.000	NaN	NaN	14.4	0.0	0.1
sync_1477	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_1478	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1479	Predicted ATPase specific for cyanobacteria	0.000	0.000	NaN	NaN	100.0	0.0	4.0
sync_1480	Predicted ATPase specific for cyanobacteria	0.016	0.047	0.336	0.344	100.0	50.2	3.9
sync_1481	hypothetical protein	0.188	0.125	1.500	1.312	54.1	6.0	1.0
sync_1482	Mechanosensitive ion channel family protein	0.045	0.241	0.188	0.179	98.7	91.4	19.4
sync_1483	possible ABC transporter component, ATP binding protein	0.063	0.305	0.206	0.216	100.0	92.8	16.4
sync_1484	hypothetical protein	0.046	0.175	0.265	0.260	100.0	50.4	6.4
sync_1485	hypothetical protein	0.074	0.184	0.402	0.386	100.0	97.2	10.3
sync_1486	hypothetical protein	0.000	0.000	NaN	NaN	51.2	0.0	0.9
sync_1487	hypothetical protein	0.000	0.000	NaN	NaN	76.7	0.0	0.8
sync_1488	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1489	multicopper oxidase	0.021	0.111	0.193	0.163	92.8	70.7	5.6
sync_1490	4-methyl-5(B-hydroxyethyl)-thiazole monophosphate biosynthesis enzyme	0.004	0.000	Inf	Inf	87.1	29.3	2.9
sync_1491	indolepyruvate decarboxylase (EC:4.1.1.74)	0.000	0.000	NaN	NaN	88.3	0.0	2.1
sync_1492	hypothetical protein	0.000	0.000	NaN	NaN	21.8	0.0	0.6
sync_1493	hypothetical protein	0.000	0.000	NaN	NaN	14.9	0.0	0.1
sync_1494	RNA polymerase sigma-70 factor family protein	0.000	0.000	NaN	NaN	62.3	0.0	0.8
sync_1495	biogenesis protein	0.000	0.000	NaN	NaN	18.4	0.0	0.2
sync_1496	thioredoxin-like protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1497	manganese/zinc/iron chelate ABC transporter (MZT) family, permease protein	0.038	0.367	0.104	0.111	98.6	98.2	35.9
sync_1498	manganese/zinc/iron chelate ABC ransporter (MZT) family, ATP-binding protein	0.020	0.218	0.091	0.097	100.0	100.0	20.6
sync_1499	ABC transporter substrate-binding protein	0.107	0.784	0.137	0.173	62.2	54.4	13.0
sync_1500	possible porin	0.006	0.049	0.127	0.138	91.6	65.7	5.3
sync_1501	possible porin	0.029	0.175	0.164	0.155	95.7	71.2	8.3
sync_1502	Iron-regulated protein A precursor, putative	0.000	0.000	NaN	NaN	25.0	0.0	0.3
sync_1503	thiol oxidoreductase	0.000	0.000	NaN	NaN	17.3	0.0	0.3
sync_1504	hypothetical protein	0.000	0.000	NaN	NaN	21.3	0.0	0.2
sync_1505	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1506	hypothetical protein	0.000	0.000	NaN	NaN	80.2	0.8	1.5
sync_1507	hypothetical protein	0.000	0.000	NaN	NaN	39.4	0.0	0.4
sync_1508	possible CobW protein involved in cobalamin synthesis	0.000	0.000	NaN	NaN	2.1	0.0	0.0
sync_1509	hypothetical protein	0.000	0.000	NaN	NaN	23.9	0.0	0.2
sync_1510	cation efflux system protein czcA-1	0.000	0.000	NaN	NaN	7.4	0.0	0.1
sync_1511	hypothetical protein	0.021	0.206	0.104	0.136	89.4	34.1	3.4
sync_1512	GGDEF/EAL domain protein	0.000	0.000	NaN	NaN	29.4	0.0	0.3
sync_1513	lipoprotein, putative	0.000	0.000	NaN	NaN	85.4	0.0	0.9
sync_1514	GGDEF domain protein	0.000	0.000	NaN	NaN	25.0	0.0	0.3
sync_1515	long-chain-fatty-acid-CoA ligase (EC:6.2.1.-)	0.020	0.073	0.275	0.273	100.0	100.0	24.7
sync_1516	hydrolase CocE/NonD family protein subfamily	0.005	0.018	0.307	0.309	90.1	18.6	3.0
sync_1517	hypothetical protein	0.063	0.089	0.704	0.642	100.0	75.5	6.2
sync_1518	hypothetical protein	0.063	0.170	0.370	0.312	100.0	30.7	4.3
sync_1519	Breast cancer type 1 susceptibility protein homolog	0.107	0.105	1.018	1.051	100.0	40.0	4.4
sync_1520	hypothetical protein	0.122	0.160	0.760	0.570	100.0	18.6	3.4
sync_1521	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	3.0
sync_1522	oxidoreductase alpha (molybdopterin) subunit	0.019	0.109	0.171	0.170	90.6	36.2	3.5
sync_1523	hypothetical protein	0.023	0.137	0.169	0.168	100.0	62.0	6.3
sync_1524	hypothetical protein	0.007	0.149	0.046	0.046	100.0	36.6	4.4
sync_1525	signal peptidase I (EC:3.4.21.89)	0.018	0.107	0.164	0.178	100.0	35.1	4.1
sync_1526	2-succinyl-6-hydroxy-2, 4-cyclohexadiene-1-carboxylic acid synthase/2-oxoglutarate decarboxylase (EC:4.1.1.71)	0.037	0.103	0.362	0.361	100.0	4.4	2.7
sync_1527	naphthoate synthase (EC:4.1.3.36)	0.019	0.169	0.111	0.121	100.0	32.4	3.9
sync_1528	Protein erfK/srfK precursor	0.025	0.170	0.147	0.181	100.0	85.9	7.6
sync_1529	glycogen synthase (EC:2.4.1.21)	0.008	0.133	0.064	0.062	100.0	77.3	7.0
sync_1530	hypothetical protein	0.034	0.117	0.291	0.268	100.0	81.2	6.8
sync_1531	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D- alanine ligase (EC:6.3.2.10)	0.035	0.214	0.164	0.168	100.0	95.6	9.3
sync_1532	hypothetical protein	0.019	0.163	0.115	0.122	100.0	74.9	5.1
sync_1533	UDP-N-acetylglucosamine pyrophosphorylase, putative	0.012	0.197	0.063	0.057	99.7	98.5	12.4
sync_1534	hypothetical protein	0.000	0.000	NaN	NaN	84.7	0.6	1.7
sync_1535	hypothetical protein	0.035	0.209	0.168	0.189	100.0	100.0	11.2
sync_1536	hypothetical protein	0.050	0.226	0.220	0.241	100.0	100.0	11.4
sync_1537	3-phosphoshikimate 1-carboxyvinyltransferase (EC:2.5.1.19)	0.028	0.260	0.106	0.100	100.0	94.5	13.5
sync_1538	Predicted HTH of cAMP family transcriptional regulator, putative	0.047	0.192	0.242	0.216	100.0	100.0	9.2
sync_1539	ferritin	0.023	0.347	0.067	0.075	100.0	100.0	15.9
sync_1540	Bacterial regulatory proteins, Crp family protein	0.007	0.307	0.023	0.022	100.0	100.0	20.5

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_1541	hypothetical protein	0.122	0.230	0.530	0.530	100.0	100.0	17.8
sync_1542	possible porin	0.051	0.237	0.215	0.206	100.0	89.0	15.3
sync_1543	possible Peptidase family M20/M25/M40	0.048	0.320	0.151	0.174	100.0	85.7	9.8
sync_1544	2OG-Fe(II) oxygenase superfamily enzyme	0.043	0.292	0.147	0.144	100.0	100.0	19.7
sync_1545	ABC-type Fe3+ transport system periplasmic component	0.027	0.262	0.103	0.106	100.0	100.0	14.1
sync_1546	3-octaprenyl-4-hydroxybenzoate carboxy-lyase (EC:4.1.1.-)	0.005	0.238	0.020	0.022	100.0	97.1	8.9
sync_1547	2-phosphosulfolactate phosphatase (EC:3.1.3.71)	0.017	0.173	0.096	0.130	100.0	76.7	7.3
sync_1548	Possible nitrilase	0.008	0.155	0.055	0.053	100.0	100.0	8.5
sync_1549	Cell wall hydrolase/autolysin	0.007	0.135	0.052	0.060	100.0	55.4	4.8
sync_1550	glutamate racemase (EC:5.1.1.3)	0.007	0.067	0.107	0.094	100.0	33.6	4.6
sync_1551	solanesyl diphosphate synthase (EC:2.5.1.11)	0.005	0.191	0.028	0.025	100.0	64.0	5.8
sync_1552	HAD-superfamily hydrolase, subfamily IA, variant 3	0.017	0.111	0.151	0.146	92.9	53.6	4.2
sync_1553	acetate--CoA ligase (EC:6.2.1.1)	0.008	0.107	0.075	0.074	100.0	54.7	5.0
sync_1554	hypothetical protein	0.026	0.068	0.376	0.389	100.0	18.9	2.9
sync_1555	hypothetical protein	0.014	0.266	0.053	0.056	100.0	59.5	5.7
sync_1556	bacterioferritin comigratory protein	0.008	0.104	0.076	0.075	100.0	76.0	8.3
sync_1557	possible DNA polymerase III, epsilon subunit (EC:2.7.7.7)	0.020	0.152	0.133	0.137	100.0	95.3	8.4
sync_1558	hypothetical protein	0.022	0.146	0.151	0.151	100.0	93.4	8.4
sync_1559	hypothetical protein	0.000	0.052	0.000	0.123	90.4	50.9	3.6
sync_1560	BioY family protein	0.003	0.055	0.051	0.046	100.0	29.7	3.5
sync_1561	signal peptidase II (EC:3.4.23.36)	0.006	0.052	0.108	0.099	100.0	52.4	4.7
sync_1562	penicillin-binding protein (EC:2.4.2.-)	0.007	0.074	0.099	0.093	93.4	40.3	4.2
sync_1563	GTP-binding protein	0.008	0.111	0.071	0.076	100.0	72.2	6.4
sync_1564	Pyridoxal-dependent decarboxylase family protein (EC:4.1.1.29)	0.025	0.146	0.169	0.172	100.0	82.3	6.4
sync_1565	cytidine/deoxycytidylate deaminase family protein	0.006	0.080	0.075	0.072	100.0	100.0	10.2
sync_1566	serine:pyruvate/alanine:glyoxylate aminotransferase (EC:2.6.1.44)	0.011	0.151	0.073	0.075	100.0	97.0	9.9
sync_1567	allophycocyanin, beta subunit	0.009	0.177	0.049	0.062	100.0	97.0	12.6
sync_1568	hypothetical protein	0.076	0.105	0.727	0.961	100.0	100.0	20.0
sync_1569	glutamine synthetase, type I (EC:6.3.1.2)	0.022	0.274	0.079	0.086	100.0	90.1	13.3
sync_1570	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1571	urea transporter, putative	0.000	0.000	NaN	NaN	24.1	0.0	0.5
sync_1572	hypothetical protein	0.000	0.000	NaN	NaN	45.8	6.3	0.7
sync_1573	hypothetical protein	0.058	0.287	0.201	0.200	100.0	93.0	9.2
sync_1574	possible SAP domain	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1575	hypothetical protein	0.067	0.194	0.346	0.359	100.0	91.6	14.0
sync_1576	methyltransferase, UbiE/COQ5 family protein	0.021	0.299	0.070	0.069	100.0	100.0	14.0
sync_1577	possible Copper binding proteins, plastocyanin	0.020	0.312	0.066	0.065	100.0	63.8	9.9
sync_1578	hypothetical protein	0.010	0.051	0.205	0.191	100.0	94.0	5.2
sync_1579	hypothetical protein	0.009	0.052	0.164	0.117	100.0	85.1	5.1
sync_1580	pmgA protein	0.014	0.126	0.111	0.120	100.0	74.0	6.0
sync_1581	GUN4-like family protein	0.015	0.136	0.114	0.121	100.0	100.0	12.4
sync_1582	Predicted ATPase	0.025	0.198	0.125	0.128	100.0	97.9	10.6
sync_1583	photosystem II reaction center protein PsbW	0.006	0.200	0.029	0.033	100.0	100.0	13.7
sync_1584	hypothetical protein	0.014	0.165	0.085	0.095	100.0	99.3	8.6
sync_1585	Domain of unknown function, putative	0.016	0.138	0.114	0.116	100.0	100.0	9.7
sync_1586	Flavin reductase like domain, putative	0.012	0.171	0.069	0.067	100.0	100.0	11.4
sync_1587	hypothetical protein	0.055	0.172	0.317	0.289	100.0	98.4	10.6
sync_1588	protein-export membrane protein SecF	0.021	0.202	0.105	0.092	100.0	97.2	14.6
sync_1589	protein-export membrane protein SecD	0.023	0.251	0.092	0.097	100.0	100.0	16.8
sync_1590	pyruvate dehydrogenase E1 beta subunit (EC:1.2.4.1)	0.005	0.192	0.026	0.025	100.0	100.0	12.3
sync_1591	hypothetical protein	0.005	0.172	0.029	0.028	100.0	100.0	9.3
sync_1592	Uncharacterized membrane protein	0.073	0.240	0.305	0.329	100.0	98.1	9.6
sync_1593	4-diphosphocytidyl-2C-methyl-D-erythritol kinase (EC:2.7.1.148)	0.030	0.208	0.142	0.143	100.0	94.0	11.4
sync_1594	dimethyladenosine transferase (EC:2.1.1.-)	0.031	0.255	0.123	0.133	100.0	100.0	15.0
sync_1595	Possible phosphoribulokinase/uridine kinase family protein (EC:2.7.1.48)	0.037	0.139	0.268	0.278	100.0	100.0	13.8
sync_1596	Uncharacterised protein family protein	0.052	0.199	0.260	0.281	100.0	94.9	11.8
sync_1597	hypothetical protein	0.040	0.171	0.234	0.238	100.0	100.0	13.1
sync_1598	pentapeptide repeat protein	0.005	0.154	0.032	0.037	100.0	100.0	20.1
sync_1599	hypothetical protein	0.046	0.304	0.152	0.154	100.0	100.0	17.8
sync_1600	hypothetical protein	0.020	0.077	0.262	0.281	100.0	99.3	15.7
sync_1601	hypothetical protein	0.042	0.192	0.219	0.226	100.0	99.4	18.6
sync_1602	hypothetical protein	0.036	0.137	0.265	0.245	100.0	100.0	11.8
sync_1603	hypothetical protein	0.089	0.000	Inf	Inf	39.0	15.9	1.3
sync_1604	GAF domain protein	0.017	0.258	0.065	0.084	84.4	61.0	9.5

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_1605	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1606	UmuD protein (EC:3.4.21.-)	0.014	0.137	0.099	0.131	100.0	100.0	16.4
sync_1607	UmuC protein (EC:3.4.21.-)	0.016	0.125	0.131	0.133	100.0	100.0	19.0
sync_1608	FAD-dependent monooxygenase, putative	0.031	0.053	0.588	0.588	83.5	9.2	2.3
sync_1609	Auxin Efflux Carrier superfamily protein	0.014	0.130	0.108	0.108	100.0	100.0	17.3
sync_1610	hypothetical protein	0.022	0.238	0.092	0.085	100.0	100.0	17.7
sync_1611	hypothetical protein	0.021	0.223	0.094	0.092	100.0	100.0	22.5
sync_1612	hypothetical protein	0.000	0.000	NaN	NaN	13.0	0.0	0.1
sync_1613	DNA primase (EC:2.7.7.-)	0.010	0.218	0.046	0.047	100.0	100.0	13.9
sync_1614	Integral membrane protein, DUF6	0.016	0.130	0.123	0.110	100.0	98.8	9.5
sync_1615	hypothetical protein	0.047	0.226	0.206	0.195	100.0	100.0	14.0
sync_1616	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1617	hypothetical protein	0.000	0.000	NaN	NaN	72.6	0.0	1.4
sync_1618	hypothetical protein	0.021	0.022	0.938	0.997	100.0	100.0	10.2
sync_1619	hypothetical protein	0.000	0.000	NaN	NaN	50.4	0.0	0.5
sync_1620	hypothetical protein	0.058	0.090	0.642	0.821	100.0	100.0	8.6
sync_1621	hypothetical protein	0.120	0.165	0.725	0.680	100.0	32.7	2.8
sync_1622	hypothetical protein	0.055	0.108	0.513	0.477	100.0	100.0	9.2
sync_1623	possible Kelch motif	0.045	0.197	0.228	0.214	100.0	100.0	11.6
sync_1624	hypothetical protein	0.000	0.000	NaN	NaN	88.9	0.0	0.9
sync_1625	hypothetical protein	0.000	0.000	NaN	NaN	48.8	0.0	1.0
sync_1626	hypothetical protein	0.075	0.038	1.944	2.208	98.9	54.8	3.9
sync_1627	Holliday junction DNA helicase RuvA	0.015	0.234	0.064	0.068	100.0	100.0	13.1
sync_1628	ribosomal protein S15	0.009	0.134	0.064	0.075	100.0	100.0	18.5
sync_1629	hypothetical protein	0.020	0.199	0.098	0.082	100.0	100.0	14.1
sync_1630	DNA polymerase III, alpha subunit (EC:2.7.7.7)	0.014	0.282	0.049	0.049	100.0	96.4	13.0
sync_1631	glutamyl-tRNA(Gln) amidotransferase, A subunit (EC:6.3.5.-)	0.007	0.168	0.042	0.037	100.0	96.7	9.8
sync_1632	hypothetical protein	0.004	0.153	0.025	0.022	100.0	100.0	15.2
sync_1633	RNA methyltransferase, TrmH family, group 3	0.032	0.174	0.184	0.182	97.8	75.3	6.7
sync_1634	putative RNase3 domain	0.014	0.152	0.091	0.096	100.0	100.0	9.1
sync_1635	STAS domain protein	0.023	0.186	0.124	0.124	100.0	100.0	8.8
sync_1636	carbamoyl-phosphate synthase, small subunit (EC:6.3.5.5)	0.014	0.170	0.083	0.082	100.0	82.5	8.0
sync_1637	anthranilate phosphoribosyltransferase (EC:2.4.2.18)	0.026	0.239	0.110	0.110	100.0	99.6	10.8
sync_1638	heavy metal ABC transporter (HMT) family, permease/ATP-binding protein	0.017	0.258	0.064	0.071	100.0	99.0	9.9
sync_1639	hypothetical protein	0.013	0.178	0.073	0.084	100.0	100.0	11.3
sync_1640	methionine-S-sulfoxide reductase (EC:1.8.4.6)	0.020	0.238	0.086	0.085	100.0	100.0	13.5
sync_1641	hypothetical protein	0.022	0.164	0.137	0.135	100.0	100.0	17.2
sync_1642	hypothetical protein	0.038	0.257	0.150	0.145	100.0	100.0	22.1
sync_1644	hypothetical protein	0.042	0.204	0.205	0.182	100.0	100.0	15.0
sync_1645	hypothetical protein	0.000	0.000	NaN	NaN	28.1	0.0	0.4
sync_1646	hypothetical protein	0.045	0.338	0.134	0.159	84.9	47.6	4.9
sync_1647	hypothetical protein	0.129	0.074	1.736	1.067	100.0	33.3	3.6
sync_1648	hypothetical protein	0.103	0.089	1.159	1.367	100.0	100.0	11.0
sync_1649	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1650	peptidase M16B family, nonpeptidase homolog	0.040	0.245	0.163	0.156	100.0	100.0	17.8
sync_1651	peptidase, M16B family protein (EC:3.4.24.-)	0.023	0.207	0.109	0.112	100.0	100.0	14.8
sync_1652	hypothetical protein	0.000	0.000	NaN	NaN	57.8	0.0	0.6
sync_1653	hypothetical protein	0.000	0.000	NaN	NaN	11.7	0.0	0.1
sync_1654	hypothetical protein	0.000	0.000	NaN	NaN	70.1	0.0	1.0
sync_1655	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1656	phycocyanobilin:ferredoxin oxidoreductase (EC:1.3.7.5)	0.010	0.139	0.073	0.072	100.0	98.9	9.0
sync_1657	possible ABC transporter component	0.012	0.160	0.074	0.080	100.0	80.3	6.6
sync_1658	ABC-transporter, membrane spanning component	0.003	0.117	0.028	0.030	100.0	83.6	7.4
sync_1659	ABC transporter, ATP-binding component	0.014	0.130	0.107	0.103	100.0	100.0	11.5
sync_1660	hypothetical protein	0.000	0.142	0.000	0.000	100.0	100.0	15.6
sync_1661	glycosyl transferase, group 2 family protein (EC:2.4.1.-)	0.012	0.146	0.085	0.090	100.0	100.0	15.6
sync_1662	conserved hypothetical protein TIGR00046	0.049	0.179	0.274	0.279	98.8	90.1	10.4
sync_1663	hypothetical protein	0.037	0.119	0.310	0.312	86.6	53.1	5.8
sync_1664	hypothetical protein	0.014	0.146	0.099	0.090	100.0	100.0	19.6
sync_1665	putative type III sigma factor	0.047	0.164	0.283	0.283	100.0	76.0	11.6
sync_1666	hypothetical protein	0.077	0.079	0.978	0.951	100.0	100.0	20.5
sync_1667	Arylesterase	0.006	0.082	0.078	0.073	100.0	100.0	15.2
sync_1668	ABC-type phosphate/phosphonate transport system permease component	0.011	0.072	0.151	0.177	100.0	96.1	10.4
sync_1669	phosphonates transport	0.005	0.051	0.101	0.098	100.0	82.0	10.5

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_1670	ABC-type phosphate/phosphonate transport system periplasmic component	0.026	0.111	0.233	0.240	100.0	99.5	7.2
sync_1671	aspartate aminotransferase (EC:2.6.1.1)	0.008	0.090	0.092	0.097	100.0	47.1	5.5
sync_1672	Lactoylglutathione lyase family enzyme, putative	0.013	0.078	0.164	0.209	100.0	27.1	3.9
sync_1673	Uracil-DNA glycosylase (EC:2.7.7.7)	0.034	0.089	0.379	0.353	100.0	36.8	4.6
sync_1674	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (EC:1.17.4.3)	0.007	0.193	0.034	0.036	100.0	100.0	9.8
sync_1675	C-terminal processing peptidase (EC:3.4.21.102)	0.010	0.158	0.061	0.079	100.0	100.0	12.0
sync_1676	hypothetical protein	0.033	0.115	0.282	0.234	100.0	94.8	9.4
sync_1677	hypothetical protein	0.022	0.157	0.140	0.127	100.0	100.0	14.3
sync_1678	hypothetical protein	0.000	0.176	0.000	0.015	100.0	100.0	14.0
sync_1679	transcription-repair coupling factor	0.010	0.221	0.046	0.043	100.0	97.8	14.1
sync_1680	hypothetical protein	0.022	0.234	0.096	0.121	100.0	100.0	15.1
sync_1681	hypothetical protein	0.033	0.168	0.198	0.251	81.9	26.9	2.6
sync_1682	hypothetical protein	0.000	0.000	NaN	NaN	97.6	0.0	1.4
sync_1683	ribulose-phosphate 3-epimerase (EC:5.1.3.1)	0.005	0.177	0.028	0.027	94.1	60.0	9.0
sync_1684	ribulose-phosphate 3-epimerase (EC:5.1.3.1)	0.006	0.051	0.123	0.119	100.0	98.0	13.0
sync_1685	fructose-1,6-bisphosphatase, class II (EC:3.1.3.11)	0.005	0.120	0.040	0.038	100.0	87.2	9.2
sync_1686	glutamyl-tRNA reductase (EC:1.2.1.-)	0.004	0.173	0.025	0.030	100.0	97.3	11.3
sync_1687	glucose-1-phosphate adenylyltransferase (EC:2.7.7.27)	0.009	0.200	0.046	0.042	100.0	99.4	10.3
sync_1688	6-phosphogluconate dehydrogenase, decarboxylating (EC:1.1.1.44)	0.010	0.231	0.042	0.040	100.0	98.9	13.5
sync_1689	6-phosphogluconolactonase (EC:3.1.1.31)	0.020	0.213	0.095	0.094	100.0	100.0	13.8
sync_1690	hypothetical protein	0.025	0.218	0.114	0.115	100.0	100.0	15.4
sync_1691	hypothetical protein	0.027	0.138	0.197	0.205	100.0	89.6	11.8
sync_1692	dihydroxy-acid dehydratase (EC:4.2.1.9)	0.009	0.234	0.039	0.045	100.0	98.1	12.1
sync_1693	Uncharacterized membrane protein	0.005	0.151	0.032	0.032	100.0	100.0	16.1
sync_1694	uracil phosphoribosyltransferase (EC:2.4.2.9)	0.018	0.163	0.110	0.112	100.0	100.0	15.2
sync_1695	Secreted pentapeptide repeats protein	0.019	0.217	0.088	0.076	100.0	100.0	8.9
sync_1696	cobalamin biosynthesis protein CobW	0.018	0.222	0.081	0.085	100.0	95.5	10.6
sync_1697	hypothetical protein	0.016	0.158	0.100	0.090	100.0	97.8	9.1
sync_1698	hypothetical protein	0.037	0.160	0.232	0.289	100.0	100.0	10.2
sync_1699	hypothetical protein	0.025	0.163	0.152	0.149	100.0	92.7	11.0
sync_1700	phosphoribosylformylglycinamidine synthase, PurS protein (EC:6.3.5.3)	0.008	0.332	0.023	0.029	100.0	100.0	16.4
sync_1701	phosphoribosylformylglycinamidine synthase I (EC:6.3.5.3)	0.025	0.276	0.089	0.091	100.0	87.6	9.9
sync_1702	Uncharacterized protein family protein	0.033	0.187	0.176	0.181	100.0	93.8	14.4
sync_1703	fructose-bisphosphate aldolase, class II, Calvin cycle subtype (EC:4.1.2.13)	0.007	0.189	0.038	0.037	100.0	100.0	17.5
sync_1704	Fructose-bisphosphate aldolase class-I (EC:4.1.2.13)	0.008	0.146	0.057	0.058	100.0	100.0	17.6
sync_1705	oxidoreductase, NAD-binding	0.023	0.188	0.122	0.106	100.0	100.0	14.1
sync_1706	hypothetical protein	0.022	0.177	0.122	0.158	100.0	100.0	17.8
sync_1707	acetyl-CoA carboxylase, carboxyl transferase, beta subunit (EC:6.4.1.2)	0.009	0.143	0.062	0.062	100.0	100.0	12.7
sync_1708	hypothetical protein	0.041	0.081	0.511	0.607	100.0	97.0	6.0
sync_1709	prepilin peptidase (EC:3.4.23.43)	0.026	0.193	0.133	0.142	100.0	99.9	10.6
sync_1710	hypothetical protein	0.016	0.144	0.110	0.123	100.0	100.0	9.8
sync_1711	phosphoribulokinase (EC:2.7.1.19)	0.003	0.127	0.027	0.026	100.0	100.0	13.9
sync_1712	3-isopropylmalate dehydrogenase (EC:1.1.1.85)	0.016	0.202	0.078	0.069	100.0	90.4	12.0
sync_1713	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase (EC:2.3.1.-)	0.010	0.162	0.062	0.061	100.0	84.9	7.9
sync_1714	glutamate 5-kinase (EC:2.7.2.11)	0.009	0.151	0.062	0.059	94.1	64.8	6.0
sync_1715	phosphatase, HAD superfamily, subfamily IIIA (EC:3.1.3.-)	0.009	0.121	0.076	0.078	100.0	80.2	5.9
sync_1716	hypothetical protein	0.003	0.148	0.021	0.021	100.0	79.9	6.2
sync_1717	Predicted endonuclease (EC:3.1.-)	0.027	0.137	0.197	0.195	100.0	50.1	4.5
sync_1718	hypothetical protein	0.032	0.159	0.202	0.192	97.7	62.9	5.4
sync_1719	Uncharacterized conserved protein with a signal peptide	0.010	0.124	0.079	0.096	100.0	65.6	6.0
sync_1720	hypothetical protein	0.016	0.118	0.132	0.128	100.0	78.9	8.4
sync_1721	CheY-like domain containing protein	0.015	0.165	0.092	0.097	100.0	92.8	8.9
sync_1722	transcriptional regulator, Fur family protein	0.007	0.199	0.036	0.035	100.0	90.5	6.7
sync_1723	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (EC:5.3.1.16)	0.018	0.193	0.092	0.090	100.0	89.5	7.7
sync_1724	Possible nucleotide sugar epimerase (EC:4.2.1.46)	0.021	0.132	0.156	0.163	100.0	80.3	7.0
sync_1725	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (EC:2.7.8.5)	0.019	0.134	0.138	0.205	93.8	38.3	3.8
sync_1726	CBS domain protein	0.009	0.156	0.057	0.060	100.0	100.0	10.1
sync_1727	hypothetical protein	0.024	0.179	0.133	0.131	100.0	100.0	14.4
sync_1728	ATP synthase (EC:1.3.99.-)	0.034	0.158	0.213	0.219	100.0	100.0	14.1
sync_1729	hypothetical protein	0.024	0.156	0.152	0.142	100.0	100.0	9.8
sync_1730	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_1731	Proline-rich region	0.034	0.207	0.164	0.159	100.0	97.0	20.7
sync_1732	Serine/threonine specific protein phosphatase	0.058	0.216	0.267	0.263	100.0	100.0	22.4
sync_1733	hypothetical protein	0.052	0.092	0.563	0.603	100.0	100.0	18.8
sync_1734	possible ligand gated channel (GIC family)	0.059	0.361	0.162	0.182	100.0	100.0	20.9
sync_1735	possible Helix-turn-helix protein, copG family-related protein	0.030	0.415	0.071	0.081	100.0	100.0	33.1
sync_1736	Protein containing glutaredoxin domain and PD1-like DNA-binding domain	0.074	0.317	0.232	0.250	100.0	100.0	33.7
sync_1737	hypothetical protein	0.076	0.317	0.239	0.245	100.0	100.0	36.9
sync_1738	hypothetical protein	0.052	0.247	0.212	0.187	100.0	100.0	17.4
sync_1739	hypothetical protein	0.050	0.389	0.128	0.119	100.0	100.0	40.8
sync_1740	hypothetical protein	0.004	0.260	0.015	0.015	100.0	100.0	24.2
sync_1741	hypothetical protein	0.000	0.000	NaN	NaN	74.3	0.0	1.9
sync_1742	Cytochrome c, class IC:Cytochrome c, class I	0.054	0.195	0.278	0.257	100.0	100.0	16.0
sync_1743	putative lipoprotein	0.057	0.306	0.187	0.191	100.0	100.0	35.2
sync_1744	hypothetical protein	0.055	0.165	0.334	0.345	100.0	100.0	37.7
sync_1745	possible tRNA synthetases class I	0.054	0.250	0.218	0.238	100.0	81.3	22.1
sync_1746	hypothetical protein	0.000	0.000	NaN	NaN	55.8	0.0	0.7
sync_1747	ABC transporter, substrate binding protein, possibly Mn	0.028	0.260	0.107	0.113	94.9	88.4	16.9
sync_1748	ABC transporter component, possibly Mn transport	0.018	0.185	0.097	0.094	100.0	99.4	19.8
sync_1749	ABC transporter, ATP binding domain, possibly Mn transport	0.042	0.227	0.187	0.170	100.0	100.0	25.1
sync_1750	hypothetical protein	0.034	0.215	0.158	0.157	100.0	100.0	22.6
sync_1751	hypothetical protein	0.032	0.207	0.153	0.153	100.0	100.0	25.0
sync_1752	putative permease	0.014	0.214	0.066	0.069	100.0	100.0	26.3
sync_1753	Cytochrome c, class IC:Cytochrome c, class I	0.025	0.281	0.090	0.094	100.0	100.0	30.3
sync_1754	hypothetical protein	0.000	0.000	NaN	0.414	66.2	19.3	3.1
sync_1755	hypothetical protein	0.049	0.165	0.297	0.296	100.0	87.8	16.2
sync_1756	ABC transporter component, possibly Mn transport	0.030	0.354	0.086	0.090	100.0	100.0	42.8
sync_1757	iron chelate ABC transporter (FeCT) family, ATP-binding protein	0.020	0.305	0.064	0.061	100.0	100.0	42.8
sync_1758	cation ABC transporter, cation-binding protein	0.045	0.293	0.155	0.156	100.0	100.0	29.1
sync_1759	hypothetical protein	0.074	0.332	0.222	0.212	100.0	100.0	23.7
sync_1760	hypothetical protein	0.000	0.000	NaN	NaN	100.0	5.0	2.7
sync_1761	hypothetical protein	0.150	0.074	2.013	1.452	100.0	93.0	10.8
sync_1762	hypothetical protein	0.015	0.255	0.057	0.068	100.0	100.0	27.3
sync_1763	hypothetical protein	0.086	0.284	0.302	0.222	100.0	100.0	15.3
sync_1764	hypothetical protein	0.071	0.287	0.248	0.239	100.0	88.7	22.1
sync_1765	possible Protein of unknown function DUF131	0.040	0.169	0.235	0.247	100.0	100.0	23.4
sync_1766	hypothetical protein	0.041	0.209	0.196	0.186	100.0	100.0	14.5
sync_1767	Predicted membrane protein	0.013	0.182	0.071	0.066	100.0	100.0	15.5
sync_1768	hypothetical protein	0.045	0.167	0.272	0.274	100.0	100.0	17.3
sync_1769	hypothetical protein	0.050	0.140	0.361	0.384	100.0	97.8	15.2
sync_1770	glycosyl transferase, group 2 family protein (EC:2.4.1.-)	0.053	0.248	0.212	0.226	100.0	100.0	24.0
sync_1771	Copper/zinc superoxide dismutase	0.023	0.245	0.095	0.099	100.0	100.0	23.8
sync_1772	hypothetical protein	0.051	0.223	0.231	0.246	100.0	100.0	28.4
sync_1773	hypothetical protein	0.021	0.178	0.120	0.105	100.0	100.0	25.7
sync_1774	Uncharacterized membrane protein	0.017	0.231	0.073	0.075	100.0	100.0	15.3
sync_1775	Steroid 5-alpha reductase, C-terminal domain, putative	0.067	0.316	0.212	0.200	86.2	55.2	6.7
sync_1776	hypothetical protein	0.074	0.285	0.259	0.228	100.0	100.0	18.5
sync_1777	hypothetical protein	0.066	0.321	0.205	0.217	100.0	100.0	36.0
sync_1778	hypothetical protein	0.215	0.231	0.934	1.070	100.0	100.0	47.7
sync_1779	hypothetical protein	0.056	0.286	0.194	0.214	100.0	100.0	35.9
sync_1780	AbrB family transciptional regulator	0.034	0.335	0.103	0.107	100.0	100.0	52.3
sync_1781	hypothetical protein	0.009	0.347	0.026	0.038	100.0	100.0	31.6
sync_1782	hypothetical protein	0.148	0.046	3.235	4.079	100.0	100.0	35.3
sync_1783	hypothetical protein	0.097	0.380	0.256	0.244	100.0	100.0	59.0
sync_1784	hypothetical protein	0.097	0.042	2.329	1.390	100.0	93.2	6.7
sync_1785	hypothetical protein	0.036	0.188	0.192	0.197	100.0	74.5	6.6
sync_1786	hypothetical protein	0.069	0.377	0.184	0.175	100.0	100.0	31.4
sync_1787	hypothetical protein	0.016	0.359	0.046	0.051	100.0	97.1	17.1
sync_1788	hypothetical protein	0.075	0.384	0.195	0.259	100.0	76.5	6.4
sync_1789	hypothetical protein	0.053	0.304	0.173	0.177	100.0	100.0	31.3
sync_1790	hypothetical protein	0.005	0.317	0.015	0.008	100.0	100.0	28.5
sync_1791	hypothetical protein	0.054	0.031	1.738	1.863	100.0	100.0	18.9
sync_1792	hypothetical protein	0.049	0.331	0.149	0.132	100.0	100.0	40.1
sync_1793	hypothetical protein	0.048	0.386	0.126	0.130	100.0	100.0	28.5
sync_1794	hypothetical protein	0.014	0.270	0.052	0.050	100.0	100.0	25.8
sync_1795	hypothetical protein	0.068	0.342	0.200	0.187	100.0	100.0	37.0

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_1796	hypothetical protein	0.116	0.015	7.558	2.105	100.0	100.0	13.0
sync_1797	hypothetical protein	0.104	0.121	0.858	0.631	100.0	33.3	4.0
sync_1798	transporter, stomatin/podocin/band 7/nephrosis.2/SPFH (Stomatin) family protein	0.013	0.271	0.049	0.056	100.0	100.0	26.5
sync_1799	hypothetical protein	0.056	0.128	0.440	0.459	100.0	100.0	20.0
sync_1800	hypothetical protein	0.054	0.292	0.186	0.215	100.0	100.0	15.9
sync_1801	hypothetical protein	0.026	0.141	0.183	0.173	87.8	40.6	4.2
sync_1802	oxidoreductase yneD , putative	0.050	0.168	0.297	0.392	100.0	78.9	16.9
sync_1803	carotenoid binding protein	0.020	0.281	0.071	0.073	100.0	100.0	29.3
sync_1804	possible beta-carotene ketolase (EC:1.13.-)	0.036	0.224	0.163	0.154	100.0	90.2	19.1
sync_1805	hypothetical protein	0.015	0.282	0.052	0.062	100.0	97.9	20.4
sync_1806	hypothetical protein	0.066	0.346	0.191	0.204	100.0	100.0	28.5
sync_1807	hypothetical protein	0.064	0.245	0.261	0.264	100.0	100.0	38.1
sync_1808	hypothetical protein	0.039	0.225	0.174	0.168	100.0	100.0	24.0
sync_1809	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1810	hypothetical protein	0.028	0.228	0.121	0.125	100.0	100.0	34.2
sync_1811	hypothetical protein	0.051	0.282	0.181	0.176	100.0	100.0	39.6
sync_1812	hypothetical protein	0.031	0.200	0.157	0.144	100.0	100.0	40.3
sync_1813	Acetyltransferase	0.030	0.215	0.138	0.145	100.0	100.0	26.3
sync_1814	hypothetical protein	0.078	0.215	0.363	0.407	100.0	100.0	38.2
sync_1815	hypothetical protein	0.036	0.259	0.138	0.149	100.0	100.0	26.8
sync_1816	predicted membrane protein	0.013	0.089	0.147	0.135	56.7	12.8	1.8
sync_1817	possible transcriptional regulator	0.028	0.162	0.171	0.163	100.0	77.6	9.5
sync_1818	possible high light inducible protein-related protein	0.017	0.362	0.047	0.052	100.0	100.0	23.0
sync_1819	DnaJ domain protein	0.075	0.262	0.288	0.293	100.0	100.0	23.7
sync_1820	hypothetical protein	0.046	0.283	0.161	0.167	100.0	100.0	25.9
sync_1821	hypothetical protein	0.021	0.166	0.127	0.183	100.0	100.0	18.0
sync_1822	hypothetical protein	0.138	0.130	1.061	1.207	100.0	90.1	8.1
sync_1823	hypothetical protein	0.076	0.261	0.293	0.342	100.0	100.0	22.0
sync_1824	pyridoxamine 5'-phosphate oxidase family protein	0.075	0.263	0.284	0.296	100.0	100.0	18.5
sync_1825	hypothetical protein	0.050	0.313	0.161	0.171	100.0	100.0	26.2
sync_1826	hypothetical protein	0.039	0.279	0.139	0.151	100.0	100.0	23.4
sync_1827	hypothetical protein	0.070	0.286	0.244	0.261	100.0	100.0	17.5
sync_1828	hypothetical protein	0.110	0.040	2.790	2.097	100.0	73.6	8.6
sync_1829	hypothetical protein	0.025	0.168	0.151	0.139	100.0	100.0	24.4
sync_1830	Predicted enzyme of the cupin superfamily protein	0.041	0.175	0.235	0.222	100.0	100.0	20.9
sync_1831	cell death suppressor protein Lls1 homolog (EC:1.-)	0.022	0.208	0.104	0.101	100.0	100.0	21.6
sync_1832	Protein of unknown function (DUF1499) superfamily protein	0.038	0.174	0.220	0.221	100.0	100.0	21.0
sync_1833	hypothetical protein	0.043	0.268	0.160	0.155	100.0	100.0	13.5
sync_1834	hypothetical protein	0.019	0.215	0.089	0.096	100.0	100.0	17.4
sync_1835	aldehyde dehydrogenase family protein (EC:1.2.1.16)	0.040	0.218	0.182	0.186	100.0	100.0	27.0
sync_1836	NAD dependent epimerase/dehydratase	0.053	0.234	0.227	0.229	100.0	100.0	37.4
sync_1837	hypothetical protein	0.029	0.179	0.163	0.154	100.0	100.0	19.3
sync_1838	hypothetical protein	0.045	0.207	0.217	0.222	100.0	100.0	41.5
sync_1839	hypothetical protein	0.000	0.000	NaN	NaN	92.9	0.0	1.2
sync_1840	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1841	putative lipoprotein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1842	tryptophan-rich conserved hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1843	hypothetical protein	0.029	0.395	0.074	0.066	100.0	96.8	17.6
sync_1844	carotenoid isomerase, putative	0.025	0.216	0.116	0.115	100.0	100.0	16.9
sync_1845	Possible outer membrane protein	0.031	0.214	0.143	0.147	100.0	100.0	15.4
sync_1846	Rhomboid family protein	0.017	0.223	0.077	0.084	100.0	83.2	13.6
sync_1847	hypothetical protein	0.006	0.068	0.083	0.069	100.0	80.3	5.6
sync_1848	hypothetical protein	0.011	0.204	0.056	0.057	100.0	58.8	5.4
sync_1849	hypothetical protein	0.036	0.124	0.288	0.367	100.0	100.0	13.3
sync_1850	endonuclease III (EC:4.2.99.18)	0.014	0.167	0.081	0.093	100.0	93.3	8.7
sync_1851	putative high light inducible protein	0.006	0.192	0.033	0.029	100.0	91.6	6.1
sync_1852	hypothetical protein	0.015	0.164	0.093	0.086	100.0	100.0	10.7
sync_1853	hypothetical protein	0.016	0.107	0.152	0.174	91.4	47.9	5.4
sync_1854	possible pfkB family carbohydrate kinase	0.027	0.230	0.119	0.136	98.7	55.1	5.5
sync_1855	hypothetical protein	0.000	0.000	NaN	NaN	13.6	0.0	0.4
sync_1856	photosystem q(b) protein	0.058	0.495	0.117	0.119	100.0	100.0	19.6
sync_1857	Ser/Thr protein phosphatase family protein	0.040	0.194	0.205	0.231	92.1	80.2	11.2
sync_1858	photosystem q(b) protein	0.027	0.261	0.105	0.095	100.0	100.0	14.7
sync_1859	hypothetical protein	0.102	0.069	1.482	1.582	100.0	100.0	18.5
sync_1860	Membrane associated GTPase	0.027	0.202	0.135	0.138	100.0	99.9	11.1

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_1861	cation efflux system protein	0.016	0.183	0.086	0.082	100.0	83.2	8.5
sync_1862	hypothetical protein	0.025	0.142	0.175	0.167	100.0	76.1	9.7
sync_1863	tryptophanyl-tRNA synthetase (EC:6.1.1.2)	0.011	0.162	0.067	0.068	100.0	100.0	7.5
sync_1864	possible endolysin (EC:3.2.1.17)	0.033	0.245	0.134	0.132	100.0	100.0	10.7
sync_1865	hypothetical protein	0.042	0.187	0.224	0.215	100.0	100.0	8.3
sync_1866	threonyl-tRNA synthetase (EC:6.1.1.3)	0.009	0.174	0.052	0.052	100.0	80.6	8.1
sync_1867	hypothetical protein	0.051	0.154	0.330	0.363	100.0	72.2	7.5
sync_1868	glucokinase (EC:2.7.1.2)	0.048	0.225	0.212	0.196	100.0	100.0	12.4
sync_1869	homoserine kinase (EC:2.7.1.39)	0.010	0.216	0.046	0.051	100.0	100.0	21.0
sync_1870	NAD(P)H-quinone oxidoreductase NdhD subunit (EC:1.6.5.3)	0.013	0.264	0.049	0.051	100.0	100.0	18.2
sync_1871	Phosphofructokinase:Peptidase family M3 (EC:3.4.24.70)	0.037	0.276	0.134	0.134	100.0	100.0	20.6
sync_1872	possible Arginine repressor, C-terminal domain	0.072	0.301	0.240	0.234	100.0	100.0	29.0
sync_1873	hypothetical protein	0.029	0.347	0.084	0.090	100.0	100.0	24.2
sync_1874	hypothetical protein	0.045	0.313	0.143	0.140	100.0	100.0	27.7
sync_1875	hypothetical protein	0.022	0.357	0.063	0.060	100.0	100.0	35.5
sync_1876	lipase family protein	0.050	0.325	0.153	0.159	100.0	100.0	41.0
sync_1877	dihydroneopterin aldolase (EC:4.1.2.25)	0.071	0.335	0.213	0.197	100.0	100.0	47.3
sync_1878	gamma-glutamyl phosphate reductase (EC:1.2.1.41)	0.050	0.321	0.155	0.153	100.0	100.0	28.9
sync_1879	ROK family sugar kinase (EC:2.7.1.2)	0.036	0.275	0.129	0.123	100.0	100.0	13.6
sync_1880	hypothetical protein	0.036	0.196	0.184	0.186	100.0	74.0	7.6
sync_1881	hypothetical protein	0.023	0.155	0.146	0.152	100.0	100.0	9.6
sync_1882	hypothetical protein	0.015	0.106	0.137	0.143	100.0	97.2	9.6
sync_1883	hypothetical protein	0.036	0.190	0.190	0.225	100.0	100.0	11.9
sync_1884	X-Pro dipeptidyl-peptidase (S15 family)	0.027	0.141	0.193	0.199	100.0	83.0	7.2
sync_1885	hypothetical protein	0.016	0.188	0.088	0.084	100.0	33.7	3.7
sync_1886	1,4-alpha-glucan branching enzyme (EC:2.4.1.18)	0.009	0.199	0.044	0.045	100.0	92.3	9.1
sync_1887	uroporphyrinogen decarboxylase (EC:4.1.1.37)	0.008	0.200	0.041	0.042	100.0	100.0	14.7
sync_1888	NAD dependent epimerase/dehydratase	0.015	0.232	0.066	0.068	100.0	100.0	15.1
sync_1889	plastocyanin	0.010	0.324	0.031	0.029	100.0	100.0	17.4
sync_1890	possible cytochrome C6 (soluble cytochrome F) (cytochrome c553)	0.067	0.298	0.225	0.221	100.0	100.0	13.1
sync_1891	hypothetical protein	0.041	0.173	0.236	0.207	100.0	87.3	6.2
sync_1892	hypothetical protein	0.096	0.333	0.289	0.301	100.0	100.0	13.9
sync_1893	hypothetical protein	0.034	0.354	0.096	0.112	100.0	100.0	24.5
sync_1894	hypothetical protein	0.114	0.120	0.949	0.849	80.6	80.6	11.0
sync_1895	hypothetical protein	0.044	0.320	0.138	0.138	100.0	87.3	12.5
sync_1896	hypothetical protein	0.110	0.512	0.216	0.237	75.8	60.1	9.8
sync_1897	ATP-dependent Clp protease, Hsp 100, ATP-binding subunit ClpB	0.028	0.343	0.083	0.083	100.0	100.0	34.6
sync_1898	hypothetical protein	0.026	0.183	0.140	0.127	100.0	95.1	19.0
sync_1899	histidine biosynthesis bifunctional protein HisIE (EC:3.5.4.19,EC:3.6.1.31)	0.020	0.216	0.091	0.089	100.0	100.0	29.2
sync_1900	hypothetical protein	0.046	0.334	0.137	0.131	100.0	100.0	29.6
sync_1901	hypothetical protein	0.036	0.274	0.132	0.133	100.0	100.0	22.9
sync_1902	Possible type II alternative RNA polymerase sigma factor	0.007	0.255	0.028	0.031	100.0	100.0	40.0
sync_1903	hypothetical protein	0.039	0.297	0.133	0.149	100.0	100.0	43.6
sync_1904	hypothetical protein	0.035	0.282	0.125	0.120	100.0	100.0	39.3
sync_1905	hypothetical protein	0.033	0.190	0.172	0.177	100.0	100.0	28.7
sync_1906	hypothetical protein	0.038	0.178	0.212	0.226	100.0	100.0	31.9
sync_1907	hypothetical protein	0.066	0.263	0.249	0.261	100.0	100.0	26.8
sync_1908	Cytochrome c, class IC:Cytochrome c, class I	0.030	0.200	0.152	0.158	100.0	100.0	27.7
sync_1909	uncharacterized ycii family conserved protein	0.033	0.287	0.113	0.115	100.0	100.0	39.5
sync_1910	tryptophan synthase, alpha subunit (EC:4.2.1.20)	0.028	0.316	0.090	0.089	100.0	100.0	23.8
sync_1911	hypothetical protein	0.018	0.131	0.136	0.139	100.0	100.0	22.4
sync_1912	possible inorganic carbon transport protein	0.006	0.179	0.031	0.039	100.0	100.0	21.4
sync_1914	hypothetical protein	0.021	0.244	0.087	0.091	99.0	78.3	9.7
sync_1915	dehydrogenase subunit-like protein (EC:1.1.99.-)	0.015	0.205	0.074	0.073	100.0	100.0	20.3
sync_1916	hypothetical protein	0.020	0.160	0.125	0.111	100.0	100.0	24.4
sync_1917	hypothetical protein	0.038	0.194	0.196	0.200	100.0	100.0	19.4
sync_1918	Glucose 1-dehydrogenase (EC:1.1.1.47)	0.019	0.252	0.074	0.071	100.0	98.8	29.8
sync_1919	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1920	gluconolactonase precursor (EC:3.1.1.17)	0.031	0.183	0.170	0.171	100.0	99.8	14.1
sync_1921	hypothetical protein	0.000	0.000	NaN	0.151	32.2	12.5	1.4
sync_1922	dihydroorotase, homodimeric type (EC:3.5.2.3)	0.073	0.281	0.258	0.253	87.5	36.1	4.6
sync_1923	sulfate transporter, sulfate permease (SulP) family protein	0.018	0.278	0.065	0.062	100.0	100.0	15.0
sync_1924	K+-dependent Na+/Ca+ exchanger related-protein	0.019	0.269	0.069	0.077	100.0	100.0	27.8
sync_1925	hypothetical protein	0.014	0.189	0.074	0.070	100.0	100.0	27.0
sync_1926	hypothetical protein	0.021	0.229	0.093	0.111	100.0	100.0	20.6

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_1927	possible cytochrome c oxidase subunit II (EC:1.9.3.1)	0.000	0.000	NaN	NaN	65.5	0.0	0.8
sync_1928	cytochrome c oxidase subunit I (EC:1.9.3.1)	0.020	0.299	0.068	0.040	63.2	9.6	1.8
sync_1929	Cytochrome c oxidase subunit III	0.026	0.242	0.106	0.121	100.0	100.0	18.8
sync_1930	dehydrogenase subunit-like protein (EC:1.1.99.-)	0.015	0.230	0.066	0.068	100.0	100.0	19.1
sync_1931	Glutathione reductase (EC:1.8.1.7)	0.023	0.222	0.102	0.102	100.0	100.0	11.0
sync_1932	hypothetical protein	0.013	0.165	0.081	0.100	100.0	100.0	9.8
sync_1933	chromosomal replication initiator protein DnaA	0.010	0.206	0.051	0.055	98.9	82.4	8.9
sync_1934	peptidase, S1C (protease Do) family protein (EC:3.4.21.-)	0.009	0.185	0.049	0.045	100.0	84.9	7.8
sync_1935	hypothetical protein	0.023	0.139	0.166	0.203	100.0	72.1	7.5
sync_1936	glycosyl transferase, group 2 family protein	0.038	0.131	0.291	0.315	93.9	56.1	5.8
sync_1937	possible acetyltransferase	0.011	0.142	0.080	0.078	100.0	96.2	11.5
sync_1938	esterase	0.022	0.125	0.174	0.145	100.0	89.9	7.9
sync_1939	ABC transporter, permease/ATP-binding protein	0.014	0.182	0.078	0.077	100.0	83.1	9.1
sync_1940	ATP phosphoribosyltransferase (EC:2.4.2.17)	0.008	0.162	0.052	0.043	100.0	80.6	9.1
sync_1941	Metallo-beta-lactamase superfamily hydrolase (EC:3.1.2.6)	0.033	0.115	0.285	0.242	100.0	100.0	10.5
sync_1942	ABC transporter, ATP binding component, possibly iron transporter	0.023	0.172	0.134	0.122	100.0	72.6	10.0
sync_1943	putative endoribonuclease L-PSP	0.019	0.288	0.065	0.069	100.0	100.0	11.0
sync_1944	hypothetical protein	0.015	0.166	0.088	0.068	100.0	100.0	18.4
sync_1945	cbbX protein homolog	0.043	0.327	0.133	0.144	100.0	100.0	22.8
sync_1946	Possible pterin-4 alpha-carbinolamine dehydratase-like protein (EC:4.2.1.96)	0.049	0.413	0.119	0.122	100.0	100.0	30.1
sync_1947	hypothetical protein	0.275	0.182	1.513	1.513	100.0	11.3	2.7
sync_1948	hypothetical protein	0.071	0.149	0.474	0.560	66.3	27.5	3.1
sync_1949	hypothetical protein	0.155	0.333	0.466	0.776	100.0	14.0	2.7
sync_1950	flavodoxin	0.030	0.354	0.085	0.090	100.0	100.0	44.1
sync_1951	Thioredoxin reductase (EC:1.8.1.9)	0.029	0.310	0.094	0.091	100.0	84.3	10.8
sync_1952	hypothetical protein	0.050	0.322	0.155	0.165	100.0	100.0	30.9
sync_1953	ferredoxin	0.036	0.333	0.107	0.085	100.0	100.0	37.0
sync_1954	hypothetical protein	0.189	0.247	0.766	0.864	100.0	97.4	16.2
sync_1955	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	3.5
sync_1956	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_1957	CO2 hydration protein	0.021	0.273	0.077	0.076	100.0	99.8	23.2
sync_1958	NADH-ubiquinone/plastoquinone oxidoreductase, subunit 4 family protein	0.018	0.280	0.063	0.066	100.0	99.2	21.9
sync_1959	NAD(P)H dehydrogenase, subunit NdhF3 family protein	0.010	0.267	0.039	0.045	100.0	100.0	19.0
sync_1960	possible carbon dioxide concentrating mechanism protein CcmK	0.025	0.216	0.115	0.116	100.0	89.4	11.1
sync_1961	hypothetical protein	0.075	0.144	0.522	0.597	100.0	100.0	19.7
sync_1962	Carboxysome peptide A	0.004	0.220	0.016	0.015	100.0	100.0	19.9
sync_1963	Carboxysome peptide A	0.008	0.164	0.049	0.049	100.0	100.0	16.2
sync_1964	carboxysome shell protein CsoS3	0.019	0.235	0.080	0.084	99.9	82.8	12.9
sync_1965	hypothetical protein	0.015	0.183	0.085	0.080	100.0	90.9	9.0
sync_1966	ribulose bisphosphate carboxylase, small subunit (EC:4.1.1.39)	0.000	0.000	NaN	NaN	100.0	14.0	3.5
sync_1967	ribulose bisphosphate carboxylase, large subunit (EC:4.1.1.39)	0.001	0.132	0.010	0.019	100.0	74.0	9.5
sync_1968	Carbon dioxide concentrating mechanism protein ccmK	0.000	0.058	0.000	0.000	100.0	100.0	12.5
sync_1969	Ham1 family protein	0.012	0.066	0.182	0.168	100.0	79.1	7.7
sync_1970	hypothetical protein	0.002	0.167	0.013	0.016	100.0	98.8	9.4
sync_1971	hypothetical protein	0.006	0.072	0.089	0.079	100.0	61.5	6.3
sync_1972	lipoprotein, putative	0.000	0.000	NaN	NaN	85.7	0.0	2.0
sync_1973	light-independent protochlorophyllide reductase, N subunit (EC:1.18.-)	0.003	0.103	0.033	0.031	100.0	96.5	11.4
sync_1974	light-independent protochlorophyllide reductase, B subunit (EC:1.18.-)	0.012	0.153	0.077	0.083	100.0	96.6	11.0
sync_1975	light-independent protochlorophyllide reductase, iron-sulfur ATP-binding protein (EC:1.18.-)	0.006	0.170	0.035	0.045	100.0	93.3	10.9
sync_1976	light-dependent protochlorophyllide reductase (EC:1.3.1.33)	0.022	0.208	0.104	0.098	100.0	100.0	11.7
sync_1977	hypothetical protein	0.032	0.185	0.174	0.157	100.0	100.0	15.8
sync_1978	photosystem I reaction center subunit XII	0.052	0.134	0.391	0.360	100.0	100.0	15.8
sync_1979	hypothetical protein	0.032	0.175	0.182	0.172	100.0	100.0	9.8
sync_1980	hypothetical protein	0.032	0.142	0.224	0.217	100.0	87.4	9.4
sync_1981	hypothetical protein	0.013	0.118	0.113	0.117	100.0	94.8	9.1
sync_1982	biotin/lipoate A/B protein ligase family protein	0.043	0.141	0.306	0.301	100.0	46.7	5.8
sync_1983	peptidase, M50B family protein (EC:3.4.24.-)	0.011	0.098	0.109	0.123	99.7	71.0	5.9
sync_1984	N-(5'phosphoribosyl)anthranilate isomerase (EC:5.3.1.24)	0.015	0.088	0.172	0.275	100.0	62.7	5.1
sync_1985	hypothetical protein	0.018	0.087	0.212	0.239	98.0	36.1	3.5
sync_1986	GTP cyclohydrolase I (EC:3.5.4.16)	0.009	0.205	0.043	0.053	100.0	89.7	6.4
sync_1987	Short-chain dehydrogenase/reductase family enzyme (EC:1.-)	0.014	0.143	0.101	0.124	100.0	95.3	11.6
sync_1988	acetyl-CoA carboxylase, carboxyl transferase, alpha subunit (EC:6.4.1.2)	0.011	0.199	0.057	0.064	100.0	100.0	10.9
sync_1989	Predicted dehydrogenase	0.005	0.185	0.030	0.031	100.0	75.6	8.0

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_1990	hypothetical protein	0.009	0.270	0.034	0.040	100.0	88.1	8.1
sync_1991	creatininase (EC:3.5.2.10)	0.016	0.141	0.114	0.138	100.0	81.3	6.2
sync_1992	hypothetical protein	0.000	0.000	NaN	NaN	100.0	34.5	3.4
sync_1993	Ribosomal protein S1	0.007	0.149	0.045	0.045	100.0	90.0	8.1
sync_1994	hypothetical protein	0.020	0.189	0.106	0.109	100.0	85.9	8.4
sync_1995	conserved hypothetical protein TIGR00726	0.020	0.143	0.141	0.148	97.3	54.4	5.4
sync_1996	uncharacterized FAD-dependent dehydrogenase	0.016	0.134	0.116	0.120	100.0	73.1	7.4
sync_1997	hypothetical protein	0.006	0.081	0.071	0.057	100.0	39.3	4.2
sync_1998	Thermonuclease homolog	0.020	0.151	0.129	0.136	100.0	79.3	4.9
sync_1999	acetolactate synthase, large subunit, biosynthetic type (EC:2.2.1.6)	0.010	0.170	0.061	0.060	99.5	86.5	8.1
sync_2000	ferrochelatase (EC:4.99.1.1)	0.007	0.121	0.058	0.053	100.0	62.7	6.4
sync_2001	phage integrase family protein	0.016	0.126	0.126	0.130	100.0	98.7	9.5
sync_2002	possible methyltransferase (EC:2.1.1.-)	0.024	0.145	0.163	0.153	100.0	90.5	6.9
sync_2003	cob(I)alamin adenosyltransferase (EC:2.5.1.17)	0.005	0.174	0.030	0.040	100.0	40.4	5.6
sync_2004	uridylate kinase (EC:2.7.4.-)	0.004	0.134	0.029	0.028	100.0	100.0	9.3
sync_2005	ribosome recycling factor	0.011	0.255	0.044	0.040	100.0	95.6	8.1
sync_2006	geranylgeranyl reductase family protein	0.013	0.133	0.096	0.097	97.9	71.8	7.4
sync_2007	deoxyribodipyrimidine photolyase family protein (EC:4.1.99.3)	0.024	0.188	0.127	0.135	100.0	95.3	11.4
sync_2008	transaldolase (EC:2.2.1.2)	0.008	0.203	0.038	0.038	100.0	96.3	11.8
sync_2009	Cell division protein FtsI/penicillin-binding protein 2 (EC:2.4.1.129)	0.008	0.116	0.071	0.066	100.0	72.1	6.2
sync_2010	hypothetical protein	0.028	0.152	0.181	0.168	100.0	57.2	5.7
sync_2011	CAAX amino terminal protease family protein	0.019	0.154	0.124	0.133	100.0	74.3	6.5
sync_2012	possible alpha-ribazole-5-P phosphatase	0.012	0.157	0.074	0.075	100.0	68.6	6.8
sync_2013	dihydroorotase, putative	0.027	0.158	0.168	0.167	100.0	72.1	6.9
sync_2014	Signal peptidase I (EC:3.4.21.89)	0.014	0.156	0.088	0.089	100.0	87.1	7.0
sync_2015	Uncharacterized membrane protein	0.005	0.152	0.030	0.037	100.0	67.0	5.8
sync_2016	arsenate reductase BH3485	0.051	0.339	0.149	0.139	100.0	89.0	7.9
sync_2017	iron-sulfur cluster-binding protein	0.006	0.198	0.033	0.032	100.0	100.0	9.5
sync_2018	inorganic pyrophosphatase (EC:3.6.1.1)	0.004	0.148	0.025	0.024	100.0	72.7	6.1
sync_2019	hypothetical protein	0.020	0.049	0.408	0.415	100.0	100.0	7.9
sync_2020	possible Helix-turn-helix domain of resolvase	0.020	0.198	0.101	0.096	100.0	100.0	9.3
sync_2021	prolyl-tRNA synthetase (EC:6.1.1.15)	0.017	0.142	0.118	0.123	100.0	97.8	10.1
sync_2022	possible photosystem II Psb27 protein	0.008	0.168	0.046	0.054	100.0	100.0	14.6
sync_2023	adenylosuccinate synthetase (EC:6.3.4.4)	0.005	0.161	0.032	0.031	100.0	95.3	9.9
sync_2024	Possible carbohydrate kinase (EC:2.7.1.4)	0.012	0.135	0.088	0.087	100.0	93.9	11.7
sync_2025	precorrin-6x reductase (EC:1.3.1.54)	0.034	0.159	0.212	0.221	100.0	99.5	6.8
sync_2026	Single-stranded DNA-binding protein	0.013	0.110	0.118	0.112	100.0	97.7	9.1
sync_2027	hypothetical protein	0.024	0.152	0.155	0.142	100.0	83.1	5.9
sync_2028	hypothetical protein	0.010	0.118	0.089	0.090	100.0	71.6	5.7
sync_2029	acetylglutamate kinase (EC:2.7.2.8)	0.009	0.178	0.048	0.046	100.0	89.2	6.8
sync_2030	Uncharacterized membrane protein	0.024	0.181	0.132	0.134	100.0	97.3	12.2
sync_2031	primosomal protein N'	0.022	0.176	0.122	0.127	100.0	100.0	10.5
sync_2032	RNA polymerase sigma factor RpoD	0.022	0.323	0.069	0.070	100.0	100.0	15.9
sync_2033	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2034	hypothetical protein	0.232	0.357	0.650	0.691	76.5	4.2	1.6
sync_2035	hypothetical protein	0.039	0.293	0.133	0.164	100.0	100.0	17.0
sync_2036	porphobilinogen deaminase (EC:2.5.1.61)	0.013	0.244	0.054	0.058	100.0	100.0	18.3
sync_2037	ErfK/YbiS/YcfS/YnhG family protein	0.024	0.218	0.111	0.128	100.0	100.0	20.1
sync_2038	hypothetical protein	0.035	0.249	0.141	0.156	100.0	100.0	17.5
sync_2039	inorganic pyrophosphatase (EC:3.6.1.1)	0.005	0.174	0.026	0.045	100.0	100.0	19.0
sync_2040	Carboxypeptidase Taq (EC:3.4.17.19)	0.037	0.273	0.136	0.139	100.0	100.0	23.1
sync_2041	Uncharacterized membrane protein	0.041	0.186	0.223	0.211	100.0	100.0	17.1
sync_2042	hypothetical protein	0.024	0.149	0.161	0.140	100.0	100.0	9.9
sync_2043	hypothetical protein	0.047	0.181	0.262	0.277	100.0	100.0	16.2
sync_2044	putative pterin-4a-carbinolamine dehydratase	0.040	0.224	0.176	0.179	100.0	100.0	12.9
sync_2045	putative cobalamin biosynthesis protein CobW	0.017	0.227	0.073	0.076	100.0	94.2	14.0
sync_2046	hypothetical protein	0.009	0.081	0.115	0.113	81.4	32.1	3.1
sync_2047	hypothetical protein	0.000	0.000	NaN	NaN	48.3	0.0	1.4
sync_2048	ABC-type Fe3+ transport system permease component	0.009	0.227	0.040	0.041	100.0	100.0	14.2
sync_2049	hypothetical protein	0.034	0.123	0.272	0.256	100.0	64.3	8.5
sync_2050	hypothetical protein	0.028	0.232	0.120	0.113	100.0	95.8	12.5
sync_2051	conserved hypothetical protein TIGR00299	0.033	0.182	0.181	0.184	100.0	100.0	12.4
sync_2052	hypothetical protein	0.009	0.181	0.050	0.048	100.0	100.0	10.8
sync_2053	hypothetical protein	0.013	0.149	0.086	0.079	100.0	100.0	12.4

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_2054	hypothetical protein	0.072	0.148	0.486	0.511	100.0	94.8	9.7
sync_2055	hypothetical protein	0.031	0.144	0.215	0.216	100.0	100.0	10.1
sync_2056	exodeoxyribonuclease III (EC:3.1.11.2)	0.024	0.181	0.132	0.134	100.0	100.0	20.0
sync_2057	glycolate oxidase subunit (EC:1.1.3.15)	0.022	0.245	0.091	0.089	100.0	97.3	17.3
sync_2058	glutamate-1-semialdehyde-2,1-aminomutase (EC:5.4.3.8)	0.013	0.254	0.052	0.052	100.0	98.9	20.7
sync_2059	hypothetical protein	0.068	0.314	0.217	0.227	100.0	100.0	27.4
sync_2060	hypothetical protein	0.024	0.131	0.181	0.198	86.4	55.3	6.8
sync_2061	hypothetical protein	0.016	0.397	0.040	0.040	100.0	100.0	57.8
sync_2062	hypothetical protein	0.049	0.348	0.141	0.169	100.0	100.0	36.8
sync_2063	hypothetical protein	0.036	0.224	0.162	0.189	99.7	80.4	21.5
sync_2064	hypothetical protein	0.064	0.337	0.190	0.189	100.0	100.0	48.0
sync_2065	hypothetical protein	0.019	0.376	0.052	0.054	100.0	100.0	40.4
sync_2066	hypothetical protein	0.102	0.219	0.467	0.419	100.0	100.0	38.0
sync_2067	hypothetical protein	0.000	0.000	NaN	0.000	43.0	7.9	1.6
sync_2068	hypothetical protein	0.076	0.312	0.244	0.237	100.0	100.0	33.7
sync_2069	hypothetical protein	0.038	0.289	0.133	0.150	100.0	100.0	21.8
sync_2070	hypothetical protein	0.096	0.104	0.923	1.466	100.0	87.0	10.6
sync_2071	hypothetical protein	0.053	0.294	0.180	0.178	100.0	100.0	23.4
sync_2072	hypothetical protein	0.111	0.402	0.277	0.312	100.0	100.0	42.5
sync_2073	Predicted SH3-like domain containing protein	0.078	0.161	0.483	0.452	100.0	100.0	17.7
sync_2074	putative high light inducible protein	0.018	0.136	0.130	0.136	100.0	100.0	25.5
sync_2075	Lhc-like protein Lhl4 (EC:4.99.1.1)	0.069	0.212	0.326	0.385	100.0	100.0	24.5
sync_2076	hypothetical protein	0.173	0.126	1.370	1.215	98.0	82.1	23.4
sync_2077	hypothetical protein	0.052	0.404	0.130	0.138	100.0	100.0	41.6
sync_2078	hypothetical protein	0.061	0.270	0.225	0.223	100.0	100.0	38.6
sync_2079	hypothetical protein	0.182	0.048	3.818	1.364	100.0	20.5	2.8
sync_2080	aldehyde dehydrogenase family protein	0.000	0.000	NaN	NaN	59.8	0.0	0.7
sync_2081	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2082	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2084	hypothetical protein	0.013	0.205	0.063	0.061	81.2	68.8	12.3
sync_2085	hypothetical protein	0.034	0.215	0.157	0.166	100.0	100.0	19.6
sync_2086	conserved hypothetical protein TIGR00096	0.029	0.264	0.112	0.120	100.0	100.0	19.1
sync_2087	CysQ protein homolog	0.016	0.158	0.099	0.097	100.0	100.0	13.9
sync_2088	polyribonucleotide nucleotidyltransferase (EC:2.7.7.8)	0.003	0.178	0.019	0.021	100.0	99.5	12.7
sync_2089	ribosomal protein S14	0.009	0.134	0.067	0.065	100.0	100.0	10.9
sync_2090	membrane-associated zinc metalloprotease, putative	0.008	0.100	0.084	0.080	100.0	77.5	7.1
sync_2091	seryl-tRNA synthetase (EC:6.1.1.11)	0.013	0.114	0.111	0.119	100.0	90.0	8.5
sync_2092	hypothetical protein	0.009	0.144	0.060	0.081	100.0	100.0	6.6
sync_2093	ATPase, AAA family protein	0.007	0.170	0.038	0.037	100.0	66.4	6.8
sync_2094	hypothetical protein	0.030	0.196	0.155	0.149	96.6	64.9	6.0
sync_2095	Inner membrane protein oxaA	0.023	0.177	0.130	0.152	100.0	26.6	3.8
sync_2096	hypothetical protein	0.011	0.188	0.057	0.056	100.0	71.9	8.1
sync_2097	Ribonuclease P protein component (RNaseP protein) (RNaseP protein) (Protein C5) (EC:3.1.26.5)	0.017	0.203	0.084	0.114	100.0	88.2	9.8
sync_2098	ribosomal protein L34	0.000	0.077	0.000	0.000	100.0	37.0	4.3
sync_2099	hypothetical protein	0.004	0.193	0.021	0.027	100.0	66.7	6.1
sync_2100	chorismate mutase (EC:5.4.99.5)	0.015	0.143	0.101	0.100	100.0	100.0	7.9
sync_2101	signal peptide peptidase SppA (protease IV) (EC:3.4.21.-)	0.011	0.163	0.065	0.099	100.0	28.3	4.6
sync_2102	Permease of the drug/metabolite transporter, DMT superfamily protein	0.005	0.092	0.060	0.050	100.0	63.2	4.7
sync_2103	glycosyl transferase, group 1 family protein (EC:2.4.1.-)	0.030	0.184	0.160	0.150	100.0	60.2	5.0
sync_2104	dolichyl-phosphate-mannose-protein mannosyltransferase	0.018	0.179	0.099	0.106	84.9	52.8	4.2
sync_2105	Uncharacterized Zn ribbon-containing conserved protein	0.014	0.157	0.088	0.080	95.7	69.8	5.7
sync_2106	pspA/IM30 family protein	0.014	0.212	0.067	0.069	100.0	71.4	6.2
sync_2107	thioredoxin	0.013	0.130	0.098	0.111	100.0	100.0	8.5
sync_2108	Aminotransferases class-I (EC:2.6.1.17)	0.023	0.162	0.142	0.153	100.0	84.3	7.2
sync_2109	biotin--acetyl-CoA-carboxylase ligase (EC:6.3.4.15)	0.038	0.177	0.213	0.225	100.0	90.4	10.7
sync_2110	peptidase, M23B family protein (EC:3.4.24.-)	0.017	0.146	0.117	0.102	100.0	78.4	8.1
sync_2111	DNA-binding response regulator	0.023	0.199	0.117	0.124	100.0	100.0	10.8
sync_2112	ABC transporter, ATP-binding protein	0.020	0.218	0.091	0.090	100.0	97.4	11.7
sync_2113	NADH-ubiquinone/plastoquinone oxidoreductase, B subunit (EC:1.6.5.-)	0.007	0.209	0.032	0.038	100.0	98.5	12.4
sync_2114	DNA topoisomerase I (EC:5.99.1.2)	0.013	0.206	0.062	0.062	100.0	97.4	14.3
sync_2115	hypothetical protein	0.062	0.025	2.474	2.474	100.0	100.0	18.4
sync_2116	hypothetical protein	0.037	0.104	0.360	0.356	100.0	57.8	5.2
sync_2117	hypothetical protein	0.009	0.166	0.055	0.067	100.0	100.0	13.2
sync_2118	conserved hypothetical protein TIGR00303	0.032	0.163	0.196	0.180	100.0	100.0	13.1

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_2119	Periplasmic binding protein-like II superfamily protein	0.042	0.169	0.247	0.253	100.0	100.0	12.9
sync_2120	Oxidoreductase, Aldo/keto reductase family; 4Fe-4S ferredoxin, iron-sulfur binding domain	0.023	0.141	0.163	0.151	100.0	98.9	10.7
sync_2121	hypothetical protein	0.020	0.140	0.141	0.138	100.0	62.2	7.3
sync_2122	riboflavin synthase, alpha subunit (EC:2.5.1.9)	0.017	0.181	0.093	0.096	100.0	84.5	11.0
sync_2123	hypothetical protein	0.025	0.166	0.152	0.133	100.0	100.0	14.1
sync_2124	AbrB family transciptional regulator	0.018	0.135	0.136	0.146	100.0	100.0	20.3
sync_2125	cytochrome c oxidase, subunit III	0.007	0.181	0.038	0.045	100.0	100.0	14.1
sync_2126	cytochrome c oxidase subunit I (EC:1.9.3.1)	0.005	0.259	0.021	0.022	100.0	100.0	17.6
sync_2127	cytochrome c oxidase, subunit II (EC:1.9.3.1)	0.017	0.239	0.070	0.073	100.0	100.0	26.8
sync_2128	cytochrome oxidase assembly protein	0.023	0.187	0.125	0.138	100.0	100.0	19.4
sync_2129	protoheme IX farnesyltransferase	0.005	0.142	0.036	0.035	100.0	95.0	11.8
sync_2130	drug exporter-1 ABC transporter (DrugE1) family, ATP-binding protein	0.018	0.164	0.108	0.104	100.0	100.0	13.7
sync_2131	ABC-type multidrug transport system permease component	0.016	0.161	0.097	0.096	100.0	100.0	13.5
sync_2132	hypothetical protein	0.036	0.090	0.397	0.417	100.0	84.2	6.2
sync_2133	N-acetylmannosamine-6-phosphate 2-epimerase (EC:5.1.3.9)	0.026	0.176	0.145	0.160	100.0	100.0	10.7
sync_2134	hypothetical protein	0.018	0.196	0.094	0.083	100.0	100.0	11.0
sync_2135	GroEL2 protein (Chaperonin cpn60-2)	0.010	0.168	0.062	0.057	100.0	95.8	8.5
sync_2136	hypothetical protein	0.015	0.227	0.067	0.094	100.0	100.0	9.2
sync_2137	3-oxoacyl-(acyl-carrier-protein) reductase (EC:1.1.1.100)	0.013	0.151	0.088	0.083	100.0	97.5	9.0
sync_2138	potassium transporter, voltage-gated ion channel (VIC) family protein	0.008	0.074	0.103	0.101	100.0	33.7	4.1
sync_2139	hypothetical protein	0.010	0.073	0.143	0.163	100.0	42.4	4.7
sync_2140	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (EC:2.7.7.60)	0.020	0.101	0.201	0.180	100.0	42.0	4.3
sync_2141	hypothetical protein	0.022	0.094	0.239	0.230	100.0	76.4	8.0
sync_2142	4-hydroxybenzoate polyprenyl transferase	0.023	0.124	0.184	0.166	100.0	60.5	8.3
sync_2143	Exopolyphosphatase (EC:3.6.1.11)	0.011	0.153	0.070	0.081	100.0	91.7	7.5
sync_2144	hypothetical protein	0.043	0.159	0.271	0.283	100.0	96.2	9.7
sync_2146	precorrin-4 C11-methyltransferase (EC:2.1.1.133)	0.015	0.207	0.074	0.081	100.0	100.0	14.8
sync_2147	prolipoprotein diacylglyceryl transferase (EC:2.4.99.-)	0.015	0.202	0.073	0.070	100.0	70.4	5.7
sync_2148	apocytochrome F	0.003	0.128	0.023	0.017	100.0	100.0	12.5
sync_2149	cytochrome b6-f complex, iron-sulfur subunit (EC:1.10.99.1)	0.003	0.143	0.018	0.017	100.0	100.0	12.6
sync_2150	hypothetical protein	0.024	0.159	0.151	0.161	100.0	100.0	10.9
sync_2151	Sec-independent protein translocase TatC	0.008	0.175	0.046	0.060	100.0	99.7	11.3
sync_2152	fibronectin-binding protein	0.018	0.169	0.109	0.106	93.7	72.0	7.7
sync_2153	guanylate kinase, putative	0.019	0.182	0.105	0.071	100.0	68.4	5.9
sync_2154	photosystem I reaction centre subunit IX-related protein	0.008	0.156	0.049	0.049	100.0	50.0	4.5
sync_2155	photosystem I reaction center subunit III	0.008	0.100	0.078	0.072	100.0	85.4	9.5
sync_2156	O-sialoglycoprotein endopeptidase (EC:3.4.24.57)	0.016	0.116	0.135	0.157	100.0	58.5	5.5
sync_2157	possible high light inducible protein-related protein	0.012	0.079	0.154	0.333	100.0	61.5	5.1
sync_2158	twitching mobility protein PilT	0.019	0.197	0.096	0.094	100.0	63.8	7.1
sync_2159	UDP-N-acetylglucosamine 2-epimerase (EC:5.1.3.14)	0.010	0.153	0.067	0.064	100.0	82.0	8.3
sync_2160	Protein of unknown function (DUF1643) family protein	0.041	0.169	0.239	0.208	90.4	66.5	7.5
sync_2161	hypothetical protein	0.002	0.076	0.025	0.025	100.0	100.0	9.7
sync_2162	transporter, monovalent cation:proton antiporter-2 (CPA2) family protein	0.005	0.109	0.046	0.041	100.0	100.0	9.8
sync_2163	glutamyl-tRNA synthetase (EC:6.1.1.17)	0.028	0.221	0.127	0.124	100.0	100.0	12.3
sync_2165	hypothetical protein	0.070	0.178	0.395	0.313	100.0	23.0	2.8
sync_2166	hypothetical protein	0.002	0.101	0.019	0.020	100.0	96.5	11.7
sync_2168	ribosomal protein L19	0.013	0.171	0.076	0.070	100.0	100.0	7.1
sync_2169	hypothetical protein	0.018	0.211	0.086	0.084	100.0	100.0	11.8
sync_2170	methionine aminopeptidase, type I (EC:3.4.11.18)	0.006	0.171	0.035	0.034	100.0	92.5	8.2
sync_2171	Short-chain dehydrogenase/reductase family enzyme	0.039	0.187	0.206	0.202	100.0	100.0	15.8
sync_2172	BioD-like N-terminal domain of phosphotransacetylase (EC:2.3.1.8)	0.007	0.164	0.042	0.055	100.0	100.0	14.2
sync_2173	cyanobacteria-specific protein	0.011	0.272	0.039	0.040	100.0	98.8	15.3
sync_2174	MAPEG family protein	0.021	0.237	0.090	0.086	100.0	100.0	13.2
sync_2175	hypothetical protein	0.020	0.017	1.179	1.300	100.0	47.0	2.9
sync_2176	hypothetical protein	0.025	0.179	0.138	0.174	100.0	100.0	17.6
sync_2177	hypothetical protein	0.092	0.103	0.897	0.957	100.0	100.0	15.3
sync_2178	hypothetical protein	0.114	0.062	1.826	1.490	100.0	100.0	15.2
sync_2179	hypothetical protein	0.000	0.000	NaN	NaN	97.3	0.6	2.0
sync_2180	possible Carboxylesterase	0.010	0.275	0.038	0.052	100.0	99.6	11.1
sync_2181	hypothetical protein	0.000	0.000	NaN	NaN	47.4	8.3	1.2
sync_2182	hypothetical protein	0.000	0.000	NaN	NaN	14.2	0.0	0.2
sync_2183	hypothetical protein	0.032	0.301	0.106	0.102	100.0	100.0	12.2
sync_2184	possible high light inducible protein-related protein	0.079	0.290	0.274	0.287	100.0	100.0	13.1

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_2185	Acetyltransferase, GNAT family protein (EC:2.3.1.-)	0.050	0.186	0.267	0.285	100.0	99.1	11.0
sync_2186	hypothetical protein	0.039	0.030	1.314	0.678	100.0	100.0	6.5
sync_2187	hypothetical protein	0.042	0.117	0.355	0.389	100.0	98.5	11.3
sync_2188	hypothetical protein	0.037	0.208	0.178	0.182	100.0	100.0	15.6
sync_2189	hypothetical protein	0.038	0.103	0.364	0.393	100.0	100.0	8.1
sync_2190	Glutamyl/Glutaminyl-tRNA synthetase (EC:6.1.1.17)	0.037	0.175	0.209	0.216	100.0	45.1	5.7
sync_2191	DNA-binding protein HU	0.002	0.115	0.015	0.000	100.0	100.0	13.7
sync_2192	Metallo-beta-lactamase superfamily hydrolase	0.028	0.123	0.226	0.205	100.0	100.0	13.0
sync_2193	Glycogen debranching enzyme (EC:3.2.1.-)	0.020	0.206	0.097	0.091	100.0	100.0	13.7
sync_2195	putative GPH family sugar transporter	0.015	0.176	0.084	0.088	100.0	99.4	13.4
sync_2196	possible transporter, membrane component	0.007	0.166	0.045	0.038	100.0	100.0	13.6
sync_2197	hypothetical protein	0.044	0.203	0.215	0.213	100.0	100.0	11.3
sync_2198	hypothetical protein	0.017	0.202	0.082	0.088	100.0	79.8	12.3
sync_2199	conserved hypothetical protein TIGR00023	0.025	0.228	0.110	0.111	100.0	100.0	19.9
sync_2200	orotidine 5'-phosphate decarboxylase (EC:4.1.1.23)	0.086	0.280	0.307	0.310	100.0	24.5	5.5
sync_2201	regucalcin family protein	0.048	0.233	0.206	0.230	100.0	100.0	18.3
sync_2202	hypothetical protein	0.015	0.242	0.064	0.082	100.0	100.0	7.7
sync_2203	tyrosyl-tRNA synthetase (EC:6.1.1.1)	0.022	0.196	0.111	0.120	100.0	89.3	8.5
sync_2204	hypothetical protein	0.003	0.157	0.016	0.016	100.0	79.4	12.5
sync_2205	possible DNA-binding response regulator	0.013	0.113	0.116	0.113	100.0	85.3	8.6
sync_2206	hypothetical protein	0.032	0.188	0.173	0.189	100.0	99.8	9.4
sync_2207	cyanobacteria-specific protein containing UvrC-like endonuclease domain	0.025	0.150	0.167	0.144	100.0	86.3	8.8
sync_2208	peptidase, M17 (leucyl aminopeptidase) family protein (EC:3.4.11.-)	0.022	0.225	0.098	0.102	100.0	90.1	9.6
sync_2209	hypothetical protein	0.010	0.142	0.071	0.086	100.0	100.0	8.5
sync_2210	peptide methionine sulfoxide reductase MsrA (EC:1.8.4.6)	0.031	0.178	0.174	0.169	100.0	100.0	9.5
sync_2211	lipid-A-disaccharide synthase (EC:2.4.1.182)	0.015	0.191	0.079	0.082	100.0	84.4	8.0
sync_2212	acyl-[acyl-carrier-protein]--UDP-N- acetylglucosamine O-acyltransferase (EC:2.3.1.129)	0.009	0.164	0.057	0.052	100.0	95.2	10.0
sync_2213	beta-hydroxyacyl-(acyl-carrier-protein) dehydratase FabZ (EC:4.2.1.-)	0.004	0.130	0.035	0.034	100.0	45.4	5.0
sync_2214	UDP-3-0-acyl N-acetylglucosamine deacetylase (EC:3.5.1.-)	0.022	0.225	0.097	0.108	100.0	100.0	8.7
sync_2215	membrane protein, OMP85 family protein	0.028	0.371	0.076	0.070	90.0	52.1	6.1
sync_2216	phosphoribosylaminoimidazole-succinocarboxamide synthase (EC:6.3.2.6)	0.022	0.186	0.121	0.158	100.0	64.5	7.5
sync_2217	hypothetical protein	0.021	0.141	0.152	0.148	100.0	97.6	8.1
sync_2218	phosphoribosylamine--glycine ligase (EC:6.3.4.13)	0.027	0.166	0.164	0.173	100.0	99.4	12.2
sync_2219	sensory box histidine kinase	0.012	0.207	0.057	0.057	100.0	100.0	15.6
sync_2220	circadian clock protein KaiC	0.006	0.282	0.022	0.023	100.0	100.0	20.5
sync_2221	circadian clock protein KaiB	0.022	0.250	0.087	0.112	100.0	100.0	18.3
sync_2222	circadian clock protein KaiA	0.025	0.240	0.104	0.098	100.0	100.0	22.9
sync_2223	ribosomal protein L21	0.018	0.155	0.118	0.132	100.0	100.0	26.1
sync_2224	ribosomal protein L27	0.005	0.207	0.026	0.032	100.0	100.0	26.4
sync_2225	hypothetical protein	0.043	0.000	Inf	0.444	45.5	15.9	1.4
sync_2226	hypothetical protein	0.003	0.095	0.029	0.023	100.0	100.0	13.9
sync_2227	Lhc-like protein Lhl4 (EC:4.99.1.1)	0.000	0.000	NaN	NaN	29.9	0.0	0.7
sync_2228	CpeY protein	0.044	0.220	0.200	0.177	100.0	90.1	20.2
sync_2229	transcriptional regulator, araC family domain protein	0.061	0.301	0.204	0.230	100.0	99.4	42.8
sync_2230	hypothetical protein	0.039	0.326	0.120	0.115	100.0	100.0	39.2
sync_2231	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2232	hypothetical protein	0.035	0.280	0.125	0.117	100.0	97.9	33.5
sync_2233	tRNA pseudouridine synthase B (EC:4.2.1.70)	0.050	0.234	0.214	0.206	100.0	100.0	36.8
sync_2234	conserved hypothetical protein TIGR01033	0.039	0.215	0.180	0.180	98.6	97.4	29.7
sync_2235	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2236	hypothetical protein	0.030	0.371	0.081	0.076	100.0	100.0	42.0
sync_2237	cupin domain protein	0.094	0.331	0.284	0.303	96.5	82.2	10.8
sync_2238	hypothetical protein	0.045	0.333	0.135	0.132	100.0	100.0	31.6
sync_2239	hypothetical protein	0.000	0.000	NaN	NaN	18.8	0.0	0.4
sync_2240	hypothetical protein	0.000	0.000	NaN	NaN	92.8	0.0	2.2
sync_2241	probable acetyltransferase	0.110	0.326	0.336	0.343	95.3	91.0	16.6
sync_2242	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2243	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	2.6
sync_2244	hypothetical protein	0.000	0.000	NaN	NaN	98.6	0.9	1.8
sync_2245	Hydantoinase/oxoprolinase:Hydantoinase B/oxoprolinase (EC:3.5.2.9)	0.026	0.141	0.181	0.179	99.9	92.4	9.8
sync_2246	glucosamine-6-phosphate isomerase (EC:3.5.99.6)	0.029	0.121	0.243	0.224	100.0	94.9	8.3
sync_2247	hypothetical protein	0.014	0.000	Inf	Inf	100.0	100.0	9.9

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_2248	SpoIID/LytB domain protein	0.019	0.133	0.144	0.147	100.0	82.3	6.9
sync_2249	Ribonuclease Z (RNase Z) (tRNA 3 endonuclease) (EC:3.1.26.11)	0.007	0.143	0.048	0.054	100.0	99.8	8.6
sync_2250	cytochrome c-550	0.011	0.094	0.117	0.130	100.0	99.5	8.6
sync_2251	Phycobilisome linker polypeptide	0.020	0.119	0.166	0.166	100.0	95.3	10.9
sync_2252	hypothetical protein	0.015	0.107	0.143	0.165	100.0	58.8	5.2
sync_2253	hypothetical protein	0.035	0.098	0.356	0.349	100.0	97.4	8.1
sync_2254	Ferredoxin	0.022	0.152	0.147	0.149	100.0	100.0	16.9
sync_2255	ribosomal protein L11 methyltransferase (EC:2.1.1.-)	0.020	0.172	0.117	0.108	100.0	100.0	12.7
sync_2256	D-3-phosphoglycerate dehydrogenase (EC:1.1.1.95)	0.004	0.144	0.024	0.029	100.0	100.0	15.7
sync_2257	hypothetical protein	0.024	0.173	0.139	0.133	100.0	83.0	7.5
sync_2258	S4 domain protein	0.017	0.147	0.115	0.107	100.0	92.8	12.5
sync_2260	UDP-N-acetylmuramoylalanine--D-glutamate ligase (EC:6.3.2.9)	0.038	0.224	0.170	0.154	98.8	97.7	19.3
sync_2261	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	1.1
sync_2262	lipoprotein, putative	0.054	0.153	0.355	0.373	100.0	70.4	6.9
sync_2263	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2264	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2265	hypothetical protein	0.027	0.296	0.092	0.091	100.0	100.0	20.0
sync_2266	hypothetical protein	0.026	0.196	0.133	0.127	100.0	100.0	21.3
sync_2267	hypothetical protein	0.039	0.188	0.207	0.186	100.0	100.0	21.2
sync_2268	hypothetical protein	0.045	0.216	0.210	0.191	100.0	100.0	18.0
sync_2269	putative DNA-directed RNA polymerase, omega subunit	0.006	0.231	0.026	0.026	100.0	100.0	14.0
sync_2270	possible Ribosomal protein L36	0.019	0.246	0.078	0.071	100.0	100.0	25.8
sync_2271	DnaK family protein	0.018	0.175	0.102	0.109	100.0	83.6	9.1
sync_2272	hypothetical protein	0.034	0.232	0.148	0.164	100.0	97.3	11.0
sync_2273	hypothetical protein	0.028	0.012	2.259	2.198	100.0	100.0	17.0
sync_2274	ferredoxin-thioredoxin reductase variable chain (EC:1.18.-)	0.006	0.163	0.037	0.048	100.0	100.0	26.8
sync_2275	pyrR bifunctional protein (EC:2.4.2.9)	0.014	0.196	0.072	0.091	100.0	100.0	14.5
sync_2276	2,3-bisphosphoglycerate-independent phosphoglycerate mutase (EC:5.4.2.1)	0.024	0.268	0.091	0.095	100.0	98.8	18.9
sync_2277	preprotein translocase, SecG subunit	0.048	0.243	0.197	0.188	100.0	100.0	15.1
sync_2278	hypothetical protein	0.172	0.111	1.552	1.552	100.0	15.1	2.6
sync_2279	ammonium transporter alr0992	0.007	0.177	0.037	0.037	100.0	89.6	6.3
sync_2280	ammonium transporter alr0992	0.020	0.131	0.149	0.149	100.0	75.4	7.2
sync_2281	large-conductive mechanosensitive channel	0.043	0.073	0.587	0.723	100.0	100.0	12.6
sync_2282	GroEL chaperonin	0.006	0.299	0.021	0.024	100.0	100.0	16.2
sync_2283	chaperonin, 10 kDa	0.006	0.265	0.022	0.027	100.0	100.0	10.3
sync_2284	ATP synthase F1, beta subunit (EC:3.6.3.14)	0.005	0.250	0.019	0.018	100.0	93.2	12.5
sync_2285	ATP synthase F1, epsilon subunit (EC:3.6.3.14)	0.012	0.121	0.101	0.127	94.6	28.4	3.4
sync_2286	Ycf65 homolog	0.008	0.192	0.041	0.024	100.0	46.2	4.4
sync_2287	Putative fructose-6-phosphate aldolase (EC:4.1.2.-)	0.000	0.000	NaN	NaN	9.3	0.0	0.1
sync_2288	Predicted HAD superfamily phosphatase (EC:4.2.1.19)	0.038	0.143	0.264	0.324	55.9	39.1	5.6
sync_2289	calcium/proton antiporter	0.013	0.088	0.142	0.167	100.0	71.0	5.9
sync_2290	peptidase, M24B family protein (EC:3.4.11.9)	0.006	0.177	0.032	0.029	100.0	89.1	7.1
sync_2291	Membrane associated SBC domains	0.009	0.163	0.053	0.047	100.0	97.7	9.8
sync_2292	GTP-binding protein	0.017	0.138	0.127	0.131	100.0	100.0	12.4
sync_2293	nicotinate (nicotinamide) nucleotide adenylyltransferase (EC:2.7.7.18)	0.028	0.159	0.177	0.192	100.0	96.7	10.2
sync_2294	NAD+ synthetase (EC:6.3.1.5)	0.017	0.111	0.154	0.147	100.0	91.7	10.5
sync_2295	alanine dehydrogenase (EC:1.4.1.1)	0.008	0.175	0.046	0.042	100.0	100.0	20.3
sync_2296	hypothetical protein	0.021	0.139	0.152	0.143	78.5	31.4	6.1
sync_2297	hypothetical protein	0.043	0.186	0.229	0.252	100.0	100.0	29.7
sync_2298	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	2.3
sync_2299	photosystem II PsbY protein-related protein	0.000	0.000	NaN	NaN	84.1	0.0	1.6
sync_2300	hypothetical protein	0.012	0.149	0.079	0.082	100.0	100.0	27.3
sync_2301	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2302	hypothetical protein	0.034	0.043	0.778	0.971	100.0	100.0	9.9
sync_2303	hypothetical protein	0.123	0.135	0.911	0.737	100.0	100.0	19.5
sync_2304	CAB/ELIP/HLIP family protein	0.024	0.257	0.093	0.092	100.0	100.0	52.3
sync_2305	hypothetical protein	0.047	0.082	0.571	0.613	100.0	100.0	39.6
sync_2306	Chlorophyll a/b binding light harvesting protein PcbD	0.015	0.117	0.124	0.099	100.0	100.0	31.0
sync_2307	SEC-C motif domain protein	0.032	0.127	0.252	0.251	100.0	99.5	19.5
sync_2308	Uncharacterized membrane protein	0.024	0.134	0.180	0.188	100.0	100.0	12.5
sync_2309	hypothetical protein	0.042	0.102	0.414	0.398	100.0	100.0	15.1
sync_2310	ferredoxin, 2Fe-2S	0.007	0.123	0.060	0.070	100.0	73.7	6.0
sync_2311	hypothetical protein	0.025	0.113	0.220	0.232	100.0	83.1	6.6
sync_2312	ATP synthase F1, gamma subunit (EC:3.6.3.14)	0.012	0.222	0.056	0.069	100.0	86.4	9.6

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_2313	ATP synthase F1, alpha subunit (EC:3.6.3.14)	0.005	0.186	0.027	0.023	100.0	85.0	8.7
sync_2314	ATP synthase F1, delta subunit (EC:3.6.3.14)	0.026	0.177	0.144	0.136	100.0	100.0	10.1
sync_2315	ATP synthase B/B'''' CF(0) (EC:3.6.3.14)	0.027	0.159	0.167	0.153	100.0	100.0	8.5
sync_2316	ATP synthase chain b'''' (EC:3.6.3.14)	0.021	0.108	0.190	0.174	100.0	93.3	8.0
sync_2317	ATP synthase F0, C subunit (EC:3.6.3.14)	0.004	0.068	0.058	0.059	100.0	100.0	14.2
sync_2318	ATP synthase F0, A subunit (EC:3.6.3.14)	0.003	0.100	0.032	0.030	100.0	100.0	13.8
sync_2319	possible ATP synthase protein 1 (EC:3.6.3.14)	0.010	0.054	0.184	0.174	100.0	90.6	13.4
sync_2320	possible methyltransferase	0.041	0.253	0.163	0.162	100.0	98.3	17.4
sync_2321	anchor polypeptide LCM	0.013	0.191	0.067	0.068	100.0	93.2	12.4
sync_2322	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	3.0
sync_2323	allophycocyanin, alpha subunit	0.006	0.172	0.036	0.033	100.0	91.8	6.9
sync_2324	allophycocyanin, beta subunit	0.010	0.135	0.077	0.077	100.0	99.4	7.1
sync_2325	phycobilisome 7.8 kDa linker polypeptide, allophycocyanin-associated, core	0.003	0.045	0.056	0.057	100.0	89.1	5.1
sync_2326	putative cell division protein FtsW	0.009	0.173	0.051	0.060	100.0	85.9	7.4
sync_2327	Cytochrome c biogenesis protein (EC:1.8.1.8)	0.007	0.143	0.046	0.047	100.0	82.1	6.0
sync_2328	ResB-like protein required for cytochrome c biosynthesis	0.007	0.133	0.054	0.053	100.0	84.3	6.4
sync_2329	GTP cyclohydrolase I family enzyme	0.008	0.087	0.094	0.090	100.0	94.4	6.8
sync_2330	hypothetical protein	0.018	0.120	0.152	0.240	100.0	100.0	7.9
sync_2331	nitrogen regulatory protein P-II	0.008	0.140	0.059	0.087	100.0	98.2	9.1
sync_2332	hemolysin A	0.020	0.150	0.131	0.160	100.0	100.0	9.1
sync_2333	hypothetical protein	0.006	0.087	0.074	0.135	100.0	81.8	6.5
sync_2334	adenylosuccinate lyase (EC:4.3.2.2)	0.010	0.132	0.076	0.073	100.0	74.5	6.4
sync_2335	hypothetical protein	0.000	0.029	0.000	0.000	100.0	100.0	9.2
sync_2336	fumarate hydratase, class II (EC:4.2.1.2)	0.009	0.129	0.072	0.076	100.0	81.3	6.6
sync_2337	Superfamily II RNA helicase (EC:3.6.1.-)	0.005	0.128	0.038	0.041	100.0	87.0	7.6
sync_2338	8-amino-7-oxononanoate synthase (EC:2.3.1.47)	0.010	0.077	0.130	0.123	94.6	43.7	3.9
sync_2339	Alpha/beta hydrolase family protein	0.030	0.059	0.501	0.455	100.0	46.7	4.5
sync_2340	SAM dependent methyltransferase	0.019	0.084	0.227	0.263	100.0	38.0	3.8
sync_2341	Dethiobiotin synthetase (EC:6.3.3.3)	0.028	0.149	0.187	0.190	100.0	61.3	4.9
sync_2342	hypothetical protein	0.010	0.079	0.125	0.130	100.0	30.1	4.0
sync_2343	adenosylmethionine-8-amino-7-oxononanoate transaminase (EC:2.6.1.62)	0.024	0.155	0.157	0.144	95.8	22.7	3.3
sync_2344	hypothetical protein	0.000	0.000	NaN	NaN	79.2	0.0	2.2
sync_2345	DnaJ domain protein	0.006	0.171	0.035	0.044	100.0	46.0	4.9
sync_2346	methyltransferase GidB (EC:2.1.-)	0.021	0.140	0.149	0.160	100.0	68.9	5.6
sync_2347	oxidoreductase, aldo/keto reductase family protein	0.020	0.169	0.118	0.127	100.0	76.0	6.2
sync_2348	3Fe-4S ferredoxin	0.016	0.191	0.084	0.086	100.0	100.0	12.8
sync_2349	hypothetical protein	0.005	0.168	0.032	0.057	100.0	87.8	9.4
sync_2350	hypothetical protein	0.029	0.153	0.191	0.208	100.0	100.0	10.0
sync_2351	HEAT-like repeat containing protein	0.009	0.148	0.062	0.069	100.0	100.0	14.1
sync_2352	hypothetical protein	0.010	0.119	0.083	0.082	100.0	100.0	9.8
sync_2353	Sodium:solute symporter family, possible glucose transporter	0.009	0.115	0.076	0.073	100.0	80.1	7.2
sync_2354	radical SAM enzyme, Cfr family protein	0.009	0.209	0.045	0.043	100.0	100.0	16.7
sync_2355	possible high light inducible protein	0.003	0.173	0.017	0.017	100.0	100.0	23.1
sync_2356	DNA-directed RNA polymerase, beta'' subunit (EC:2.7.7.6)	0.007	0.201	0.033	0.036	100.0	97.8	11.3
sync_2357	DNA-directed RNA polymerase, gamma subunit (EC:2.7.7.6)	0.005	0.213	0.025	0.021	100.0	97.8	12.3
sync_2358	DNA-directed RNA polymerase, beta subunit (EC:2.7.7.6)	0.003	0.213	0.013	0.012	100.0	89.9	8.2
sync_2359	hypothetical protein	0.041	0.014	2.844	1.232	100.0	80.6	5.4
sync_2360	hydrolase, TatD family protein	0.012	0.160	0.074	0.069	100.0	55.7	5.1
sync_2361	ribosomal protein S20	0.004	0.144	0.025	0.022	100.0	85.7	7.0
sync_2362	histidinol dehydrogenase (EC:1.1.1.23)	0.005	0.136	0.038	0.043	99.5	76.2	6.3
sync_2363	ribose 5-phosphate isomerase A (EC:5.3.1.6)	0.011	0.127	0.084	0.077	100.0	95.7	9.4
sync_2364	Serine proteases, trypsin family protein (EC:3.4.21.-)	0.007	0.206	0.036	0.040	99.4	56.2	6.5
sync_2365	hypothetical protein	0.040	0.198	0.199	0.192	100.0	76.2	8.1
sync_2366	possible ABC transporter, solute binding protein	0.017	0.188	0.088	0.094	100.0	81.4	7.2
sync_2367	hypothetical protein	0.011	0.181	0.062	0.078	100.0	92.1	6.2
sync_2368	arylsulfatase regulator (EC:1.-)	0.022	0.218	0.101	0.094	100.0	96.9	8.9
sync_2369	lipoprotein, putative	0.019	0.288	0.065	0.070	100.0	55.3	5.2
sync_2370	DUF150	0.018	0.116	0.154	0.193	100.0	92.4	9.6
sync_2371	transcription termination factor NusA	0.003	0.098	0.030	0.037	100.0	61.8	5.8
sync_2372	hypothetical protein	0.012	0.019	0.635	0.632	100.0	100.0	12.9
sync_2373	translation initiation factor IF-2	0.003	0.126	0.025	0.027	100.0	68.4	6.0
sync_2374	hypothetical protein	0.028	0.080	0.354	0.444	92.0	60.4	4.8
sync_2375	hypothetical protein	0.014	0.187	0.074	0.099	100.0	90.0	7.2
sync_2377	Orotate phosphoribosyltransferase (OPRT) (OPRTase), putative	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2378	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_2379	Keratin, ultra high-sulfur matrix protein B	0.000	0.090	0.000	0.000	100.0	58.5	5.1
sync_2380	Domain of unknown function (DUF389) family protein	0.015	0.158	0.097	0.101	82.1	53.2	4.5
sync_2381	hypothetical protein	0.000	0.000	NaN	NaN	46.0	0.0	1.4
sync_2382	2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase(2,5-ddol dehydrogenase) (EC:1.1.1.-)	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2383	hypothetical protein	0.079	0.056	1.427	0.803	50.0	21.5	1.8
sync_2384	photosystem q(b) protein	0.053	0.568	0.093	0.092	100.0	96.7	25.3
sync_2385	hypothetical protein	0.016	0.120	0.135	0.141	100.0	86.9	7.6
sync_2386	putative lipoprotein	0.066	0.207	0.320	0.325	100.0	100.0	9.4
sync_2387	hypothetical protein	0.027	0.134	0.201	0.181	100.0	85.4	9.3
sync_2388	hypothetical protein	0.007	0.085	0.085	0.142	100.0	86.2	8.7
sync_2389	aspartoacylase (EC:3.5.1.15)	0.033	0.182	0.180	0.170	100.0	87.2	6.5
sync_2390	Predicted glutathione S-transferase (EC:2.5.1.18)	0.019	0.104	0.188	0.199	100.0	79.4	7.6
sync_2391	hypothetical protein	0.023	0.102	0.222	0.159	100.0	70.1	7.8
sync_2392	ABC transporter, ATP binding domain	0.013	0.152	0.084	0.077	100.0	98.2	9.6
sync_2393	CP12 domain, putative	0.045	0.144	0.311	0.293	100.0	100.0	17.3
sync_2394	GTP-binding protein, GTP1/OBG family protein	0.005	0.153	0.035	0.034	100.0	98.2	10.9
sync_2395	ferric iron ABC transporter (FeT) family, ATP-binding protein	0.020	0.180	0.110	0.103	100.0	100.0	11.8
sync_2396	MutS2 family protein	0.012	0.161	0.073	0.078	100.0	94.9	10.6
sync_2397	possible ring-cleaving dioxygenase	0.006	0.150	0.041	0.046	100.0	94.5	8.3
sync_2398	delta-aminolevulinic acid dehydratase (EC:4.2.1.24)	0.008	0.151	0.053	0.055	100.0	97.2	9.6
sync_2399	DnaJ3 protein	0.026	0.298	0.088	0.087	100.0	100.0	11.7
sync_2400	hypothetical protein	0.065	0.225	0.290	0.322	100.0	100.0	14.5
sync_2401	sulfate transporter, sulfate permease (SulP) family protein	0.023	0.310	0.074	0.076	100.0	100.0	16.3
sync_2402	hypothetical protein	0.000	0.000	NaN	NaN	27.1	0.0	0.6
sync_2403	hypothetical protein	0.068	0.257	0.262	0.267	100.0	98.9	12.7
sync_2404	hypothetical protein	0.010	0.408	0.025	0.033	100.0	100.0	26.4
sync_2405	tonB-dependent receptor protein (EC:1.3.99.-)	0.064	0.337	0.190	0.184	100.0	100.0	32.8
sync_2406	hypothetical protein	0.134	0.127	1.050	0.947	100.0	80.2	10.7
sync_2407	hypothetical protein	0.014	0.256	0.056	0.042	100.0	100.0	28.8
sync_2408	hypothetical protein	0.038	0.288	0.133	0.147	100.0	100.0	29.9
sync_2409	ParB nuclease domain protein	0.058	0.295	0.197	0.207	100.0	100.0	30.4
sync_2410	hypothetical protein	0.165	0.281	0.587	0.575	100.0	87.8	22.1
sync_2411	hypothetical protein	0.000	0.000	NaN	NaN	79.5	0.0	1.0
sync_2412	dienelactone hydrolase	0.000	0.000	NaN	NaN	68.7	0.0	1.5
sync_2413	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	1.0
sync_2414	Phage integrase family domain protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2415	bacteriophage protein homolog	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2416	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2417	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2418	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2419	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2420	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2421	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2422	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2423	hypothetical protein	0.006	0.008	0.654	0.619	100.0	81.1	7.2
sync_2424	cytochrome P450 (EC:1.14.-)	0.000	0.000	NaN	NaN	32.7	0.0	0.3
sync_2425	hypothetical protein	0.000	0.000	NaN	NaN	18.5	0.0	0.2
sync_2426	bacterial metallothionein-related protein	0.000	0.000	NaN	NaN	100.0	0.0	1.8
sync_2427	hypothetical protein	0.000	0.000	NaN	NaN	10.3	0.0	0.2
sync_2428	hypothetical protein	0.184	0.120	1.533	1.287	100.0	98.3	7.6
sync_2429	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2430	hypothetical protein	0.000	0.000	NaN	NaN	83.5	0.0	1.0
sync_2431	hypothetical protein	0.000	0.000	NaN	NaN	18.0	0.0	0.2
sync_2432	hypothetical protein	0.000	0.000	NaN	NaN	41.7	0.0	0.4
sync_2433	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2434	hypothetical protein	0.000	0.000	NaN	NaN	77.6	0.0	1.1
sync_2435	hypothetical protein	0.068	0.050	1.356	1.356	100.0	39.5	3.6
sync_2436	hypothetical protein	0.000	0.000	NaN	NaN	60.2	1.3	1.8
sync_2437	Domain of unknown function DUF81	0.000	0.000	NaN	NaN	68.1	0.0	1.0
sync_2438	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	2.0
sync_2439	possible high light inducible protein-related protein	0.000	0.000	NaN	NaN	9.1	0.0	0.1
sync_2440	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2441	hypothetical protein	0.008	0.070	0.109	0.116	100.0	18.4	2.6
sync_2442	transporter, ZIP family, putative	0.019	0.072	0.266	0.385	100.0	47.8	5.7
sync_2443	transporter, ZIP family protein	0.026	0.088	0.297	0.299	100.0	41.1	5.8

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_2444	hypothetical protein	0.058	0.181	0.320	0.385	100.0	89.0	6.0
sync_2445	hypothetical protein	0.000	0.000	NaN	Inf	39.0	7.8	1.2
sync_2446	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2447	O-linked GlcNAc transferase, putative	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2448	probable hemagglutinin-related	0.000	0.000	NaN	NaN	18.3	0.0	0.2
sync_2450	Putative RNA methylase family UPF0020 family	0.017	0.217	0.079	0.089	100.0	97.5	18.5
sync_2451	hypothetical protein	0.022	0.200	0.111	0.115	100.0	100.0	15.8
sync_2452	hypothetical protein	0.023	0.158	0.144	0.172	100.0	100.0	12.9
sync_2453	hypothetical protein	0.039	0.183	0.212	0.187	100.0	98.3	10.8
sync_2454	chromosome segregation protein SMC	0.014	0.177	0.081	0.075	100.0	80.3	8.3
sync_2455	PRC-barrel domain protein	0.018	0.159	0.114	0.121	100.0	91.4	8.1
sync_2456	methionine-R-sulfoxide reductase (EC:1.8.4.-)	0.026	0.168	0.153	0.140	100.0	100.0	10.8
sync_2457	hypothetical protein	0.013	0.125	0.104	0.105	100.0	89.6	8.8
sync_2459	acetyl-CoA carboxylase, biotin carboxylase (EC:6.4.1.2)	0.004	0.188	0.021	0.017	100.0	99.6	9.6
sync_2460	hypothetical protein	0.014	0.145	0.095	0.081	100.0	100.0	12.6
sync_2461	photosystem II reaction centre X protein (PsbX)	0.000	0.129	0.000	0.000	100.0	100.0	13.3
sync_2462	hypothetical protein	0.017	0.220	0.079	0.079	86.8	71.5	8.4
sync_2463	lipoprotein, putative	0.021	0.205	0.102	0.100	100.0	82.4	8.9
sync_2464	lipoprotein, putative	0.012	0.109	0.114	0.135	100.0	84.4	9.0
sync_2465	possible high light inducible protein	0.035	0.113	0.312	0.295	100.0	78.7	6.4
sync_2466	peroxisomal fatty acyl CoA ABC transporter (P-FAT) family, ATP-binding protein	0.009	0.271	0.034	0.034	100.0	89.9	9.7
sync_2467	histidine triad family protein	0.008	0.148	0.057	0.064	100.0	96.2	8.1
sync_2468	Mg chelatase homolog	0.025	0.184	0.138	0.122	100.0	64.2	6.1
sync_2469	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	1.5
sync_2470	ribosomal protein S21	0.000	0.046	0.000	0.000	100.0	72.5	6.9
sync_2471	hypothetical protein	0.026	0.199	0.128	0.147	100.0	56.3	5.8
sync_2472	hypothetical protein	0.008	0.187	0.040	0.040	100.0	100.0	8.2
sync_2473	peptide deformylase (EC:3.5.1.88)	0.009	0.185	0.047	0.060	100.0	93.8	6.2
sync_2474	peptidase, S9C (acylaminoacyl-peptidase) family protein (EC:3.4.-)	0.015	0.107	0.144	0.127	99.1	65.1	5.4
sync_2475	L-fuculose phosphate aldolase (EC:4.1.2.17)	0.026	0.237	0.111	0.111	100.0	58.5	6.8
sync_2476	Initiation factor 2 subunit family protein (EC:5.3.1.23)	0.012	0.094	0.132	0.148	100.0	91.9	7.6
sync_2477	hypothetical protein	0.033	0.036	0.927	1.849	100.0	64.0	4.2
sync_2478	hypothetical protein	0.019	0.192	0.099	0.099	100.0	90.8	8.1
sync_2479	hypothetical protein	0.016	0.042	0.380	0.376	100.0	29.6	3.9
sync_2480	cysteine desulfurase, SufS family protein	0.018	0.169	0.106	0.104	100.0	85.2	9.2
sync_2481	FeS assembly protein SufD	0.034	0.267	0.128	0.132	100.0	90.3	9.3
sync_2482	FeS assembly ATPase SufC	0.012	0.219	0.057	0.056	100.0	98.6	10.6
sync_2483	FeS assembly protein SufB	0.006	0.243	0.026	0.029	97.8	89.1	10.0
sync_2484	Ferredoxin thioredoxin reductase, catalytic beta chain (EC:1.18.-)	0.004	0.196	0.022	0.021	100.0	100.0	11.0
sync_2485	iron-sulfur cluster biosynthesis transcriptional regulator SufR	0.007	0.144	0.048	0.050	100.0	62.2	4.9
sync_2486	hypothetical protein	0.022	0.077	0.285	0.425	100.0	53.5	3.9
sync_2487	hypothetical protein	0.012	0.124	0.098	0.100	100.0	92.3	10.8
sync_2488	phycobilisome rod-core linker polypeptide cpcG1	0.017	0.206	0.081	0.085	100.0	98.2	10.2
sync_2489	hypothetical protein	0.012	0.211	0.058	0.059	100.0	100.0	13.8
sync_2490	Predicted membrane protein-related protein	0.013	0.153	0.084	0.070	100.0	94.4	13.6
sync_2491	phospholipase D domain protein	0.054	0.289	0.187	0.198	96.6	89.3	19.6
sync_2492	Predicted permease	0.000	0.000	NaN	NaN	70.0	0.0	1.2
sync_2493	hypothetical protein	0.000	0.000	NaN	NaN	61.6	0.0	0.7
sync_2494	hypothetical protein	0.087	0.276	0.316	0.306	87.0	76.3	11.4
sync_2495	hypothetical protein	0.000	0.000	NaN	NaN	6.1	0.0	0.1
sync_2496	hypothetical protein	0.034	0.199	0.172	0.202	100.0	98.8	12.1
sync_2497	hypothetical protein	0.065	0.238	0.271	0.275	98.0	98.0	34.0
sync_2498	hypothetical protein	0.110	0.300	0.366	0.272	100.0	34.2	4.4
sync_2499	hypothetical protein	0.050	0.276	0.182	0.202	100.0	100.0	20.1
sync_2500	phosphoglucomutase/phosphomannomutase family protein	0.021	0.194	0.107	0.118	100.0	100.0	29.4
sync_2501	lipoprotein, putative	0.055	0.347	0.160	0.156	100.0	100.0	27.5
sync_2502	RND multidrug efflux transporter all3143	0.014	0.273	0.053	0.053	100.0	100.0	27.1
sync_2503	efflux transporter, RND family, MFP subunit subfamily	0.031	0.257	0.120	0.127	100.0	100.0	15.9
sync_2504	ATPase	0.020	0.160	0.127	0.143	100.0	98.1	12.8
sync_2505	possible Protein phosphatase 2C	0.017	0.102	0.169	0.158	100.0	100.0	13.1
sync_2506	hypothetical protein	0.046	0.000	Inf	Inf	100.0	17.2	3.7
sync_2507	possible 4'-phosphopantetheinyl transferase family protein	0.047	0.156	0.302	0.281	100.0	100.0	13.0
sync_2508	bacterioferritin comigratory protein (EC:1.11.1.15)	0.012	0.118	0.098	0.096	100.0	100.0	15.3
sync_2509	transcriptional activator, putative, Baf family protein	0.063	0.268	0.236	0.232	100.0	100.0	13.7

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_2510	Phosphoadenosine phosphosulfate reductase (EC:1.8.4.8)	0.010	0.150	0.069	0.077	100.0	52.1	6.7
sync_2511	NADH dehydrogenase, FAD-containing subunit (EC:1.6.99.3)	0.023	0.114	0.200	0.208	100.0	88.9	8.2
sync_2512	GTP-binding protein HflX homolog	0.008	0.139	0.057	0.061	100.0	73.8	8.2
sync_2513	hypothetical protein	0.047	0.000	Inf	Inf	100.0	19.1	2.4
sync_2514	possible sodium transporter, Trk family protein	0.006	0.163	0.034	0.036	100.0	96.9	9.7
sync_2515	K+ transporter, Trk family protein	0.005	0.195	0.027	0.028	100.0	100.0	14.5
sync_2516	Predicted molecular chaperone	0.031	0.173	0.178	0.168	100.0	92.5	10.4
sync_2517	hypothetical protein	0.016	0.045	0.362	0.493	100.0	51.7	5.3
sync_2518	hypothetical protein	0.061	0.090	0.671	0.715	100.0	98.6	7.4
sync_2519	Predicted membrane protein	0.007	0.141	0.048	0.044	100.0	100.0	8.9
sync_2520	hypothetical protein	0.005	0.045	0.102	0.102	100.0	43.4	4.2
sync_2521	hypothetical protein	0.024	0.106	0.223	0.233	100.0	81.9	8.3
sync_2522	putative drug resistance ATPase-1 (Drug RA1) ABC transporter family, ATP-binding protein	0.013	0.132	0.097	0.094	100.0	82.5	8.6
sync_2523	peptidase, S1C (protease Do) family protein (EC:3.4.21.-)	0.022	0.184	0.121	0.126	100.0	99.5	10.0
sync_2524	Uncharacterized membrane protein	0.022	0.246	0.090	0.080	100.0	95.1	7.2
sync_2525	hypothetical protein	0.002	0.168	0.015	0.021	100.0	100.0	15.7
sync_2526	hypothetical protein	0.004	0.231	0.019	0.056	100.0	100.0	12.4
sync_2527	ATP-dependent helicase HrpB	0.032	0.194	0.167	0.170	100.0	97.7	13.6
sync_2528	Lhc-like protein Lhl4 (EC:4.99.1.1)	0.012	0.239	0.050	0.052	100.0	100.0	21.1
sync_2529	hypothetical protein	0.042	0.173	0.242	0.266	100.0	100.0	17.7
sync_2530	exodeoxyribonuclease VII, large subunit (EC:3.1.11.6)	0.049	0.300	0.163	0.166	100.0	100.0	21.3
sync_2531	Exodeoxyribonuclease VII small subunit (EC:3.1.11.6)	0.043	0.352	0.123	0.143	100.0	100.0	24.9
sync_2532	Uncharacterized membrane protein	0.056	0.309	0.183	0.201	100.0	100.0	27.9
sync_2533	possible Zinc finger, C3HC4 type (RING finger)	0.056	0.280	0.199	0.190	100.0	100.0	29.8
sync_2534	YihY family protein	0.036	0.321	0.111	0.108	100.0	100.0	24.9
sync_2540	inositol monophosphate family protein (EC:3.1.3.25)	0.030	0.252	0.119	0.120	100.0	98.0	8.4
sync_2541	possible RND family outer membrane efflux protein	0.056	0.288	0.194	0.220	98.3	52.5	5.9
sync_2542	possible Fe-S oxidoreductase	0.025	0.246	0.103	0.106	100.0	99.6	11.9
sync_2543	undecaprenol kinase, putative	0.011	0.151	0.072	0.077	100.0	91.7	10.7
sync_2544	hypothetical protein	0.027	0.157	0.171	0.139	96.3	52.6	4.8
sync_2545	photosystem II 12 kDa extrinsic protein PsbU	0.002	0.099	0.025	0.022	100.0	100.0	10.0
sync_2546	L-aspartate oxidase (EC:1.4.3.16)	0.015	0.171	0.089	0.097	100.0	100.0	10.3
sync_2547	VKORC1/thioredoxin domain protein	0.013	0.158	0.080	0.087	100.0	98.7	9.2
sync_2548	MiaB-like tRNA modifying enzyme YliG, TIGR01125	0.014	0.233	0.059	0.059	100.0	100.0	11.4
sync_2549	multidrug efflux MFS family transporter	0.015	0.156	0.099	0.093	100.0	100.0	10.3
sync_2550	hypothetical protein	0.017	0.086	0.200	0.194	100.0	100.0	11.1
sync_2551	radical SAM domain/ B12 binding domain protein (EC:2.1.1.-)	0.005	0.185	0.029	0.028	100.0	79.7	8.2
sync_2552	hypothetical protein	0.023	0.191	0.123	0.128	100.0	100.0	8.9
sync_2553	hypothetical protein	0.028	0.205	0.137	0.149	100.0	58.0	6.0
sync_2554	hypothetical protein	0.032	0.113	0.285	0.247	99.3	66.2	4.7
sync_2555	riboflavin biosynthesis protein RibD C-terminal domain protein	0.041	0.181	0.227	0.213	100.0	37.9	4.5
sync_2556	6-pyruvoyl-tetrahydropterin synthase homolog (EC:4.2.3.12)	0.007	0.095	0.070	0.068	97.9	77.0	6.9
sync_2557	shikimate kinase (EC:2.7.1.71)	0.008	0.095	0.082	0.079	100.0	66.5	6.2
sync_2558	hypothetical protein	0.009	0.183	0.049	0.050	100.0	89.6	7.0
sync_2559	hypothetical protein	0.027	0.096	0.277	0.263	100.0	82.3	9.3
sync_2560	Glutathione S-transferase zeta class (EC:2.5.1.18)	0.012	0.103	0.120	0.122	100.0	72.5	7.2
sync_2561	Protein of unknown function (DUF751) family protein	0.009	0.045	0.188	0.165	100.0	82.2	5.8
sync_2562	ribosome-binding factor A	0.004	0.119	0.034	0.030	100.0	84.7	6.1
sync_2563	Possible beta-N-acetylglucosaminidase (EC:3.2.1.52)	0.026	0.099	0.259	0.240	97.2	70.1	7.5
sync_2564	Glycosyl transferase, family 2:TPR repeat	0.009	0.113	0.078	0.079	100.0	86.1	6.3
sync_2565	uroporphyrinogen-III synthase (EC:4.2.1.75)	0.016	0.164	0.095	0.091	100.0	100.0	9.6
sync_2566	hypothetical protein	0.051	0.155	0.329	0.345	100.0	94.7	8.0
sync_2567	Oligoketide cyclase/lipid transport protein homolog	0.032	0.191	0.169	0.182	100.0	100.0	8.0
sync_2568	zeta-carotene desaturase (EC:1.14.99.30)	0.006	0.193	0.034	0.037	99.4	65.2	5.8
sync_2569	iron-sulfur cluster assembly accessory protein	0.005	0.159	0.033	0.032	100.0	89.6	6.3
sync_2570	TPR-repeat protein, specific for cyanobacteria	0.005	0.143	0.035	0.031	100.0	86.3	7.5
sync_2571	hypothetical protein	0.027	0.125	0.215	0.228	100.0	82.1	7.8
sync_2572	hypothetical protein	0.029	0.172	0.171	0.141	100.0	86.6	8.1
sync_2573	conserved hypothetical protein TIGR01777	0.025	0.202	0.125	0.123	100.0	98.2	12.1
sync_2574	transporter, monovalent cation:proton antiporter-2 (CPA2) family protein	0.025	0.141	0.178	0.167	100.0	78.6	7.8
sync_2575	hypothetical protein	0.026	0.160	0.160	0.141	100.0	100.0	13.1
sync_2576	possible heat shock protein DnaJ	0.020	0.162	0.126	0.146	100.0	100.0	8.6
sync_2577	cysteine synthase A (EC:2.5.1.47)	0.011	0.145	0.073	0.069	100.0	87.1	6.5
sync_2578	hypothetical protein	0.000	0.000	NaN	NaN	89.6	0.0	0.9

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_2579	possible porin	0.039	0.273	0.142	0.146	72.7	50.2	9.1
sync_2580	hypothetical protein	0.000	0.000	NaN	0.547	33.3	24.2	2.4
sync_2581	possible porin	0.060	0.414	0.145	0.126	78.0	42.0	7.8
sync_2582	hypothetical protein	0.000	0.000	NaN	NaN	59.0	0.0	0.8
sync_2583	hypothetical protein	0.000	0.000	NaN	NaN	35.6	0.0	0.4
sync_2584	possible Pectate lyase, putative	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2585	hypothetical protein	0.000	0.000	NaN	NaN	1.3	0.0	0.0
sync_2586	photosystem II D2 protein (photosystem q(a) protein)	0.004	0.168	0.023	0.025	100.0	99.6	13.2
sync_2587	hypothetical protein	0.044	0.335	0.130	0.143	100.0	100.0	28.8
sync_2588	possible ABC transporter, ATP-binding component	0.063	0.297	0.211	0.202	100.0	100.0	24.3
sync_2590	hypothetical protein	0.147	0.150	0.980	1.004	100.0	100.0	15.6
sync_2591	DNA-binding response regulator RpaA	0.010	0.312	0.031	0.031	100.0	100.0	17.2
sync_2592	DNA polymerase III, delta' subunit (EC:2.7.7.7)	0.028	0.220	0.127	0.118	100.0	100.0	13.2
sync_2593	thymidylate kinase (EC:2.7.4.9)	0.021	0.110	0.188	0.193	100.0	97.5	10.4
sync_2594	copper-translocating P-type ATPase (EC:3.6.3.4)	0.009	0.158	0.059	0.060	100.0	96.3	11.6
sync_2595	photosystem I assembly protein Ycf3	0.004	0.190	0.021	0.015	100.0	86.0	11.5
sync_2596	DNA repair protein RadA	0.008	0.141	0.058	0.061	96.5	40.5	4.1
sync_2597	DNA-binding response regulator	0.008	0.095	0.088	0.089	100.0	81.1	7.6
sync_2598	fatty acid/phospholipid synthesis protein PlsX	0.006	0.153	0.042	0.044	100.0	80.4	7.0
sync_2599	3-oxoacyl-(acyl-carrier-protein) synthase III (EC:2.3.1.41)	0.010	0.147	0.071	0.069	100.0	93.0	9.0
sync_2600	malonyl CoA-acyl carrier protein transacylase (EC:2.3.1.39)	0.016	0.110	0.141	0.136	100.0	79.7	7.7
sync_2601	1-acyl-sn-glycerol-3-phosphate acyltransferase (EC:2.3.1.51)	0.025	0.144	0.175	0.173	100.0	100.0	10.6
sync_2602	hypothetical protein	0.020	0.135	0.150	0.134	100.0	96.9	7.6
sync_2603	glycoprotease family protein	0.028	0.138	0.203	0.239	100.0	81.3	6.9
sync_2604	Predicted metal-binding cluster containing protein	0.021	0.082	0.257	0.293	100.0	87.8	7.5
sync_2605	polyA polymerase family protein	0.047	0.215	0.220	0.231	100.0	99.0	10.1
sync_2606	putative RNA-binding protein	0.007	0.244	0.030	0.029	100.0	100.0	9.6
sync_2607	phytoene synthase	0.017	0.282	0.061	0.067	100.0	100.0	7.9
sync_2608	phytoene desaturase (EC:1.14.99.-)	0.013	0.351	0.036	0.035	100.0	100.0	21.5
sync_2609	hypothetical protein	0.024	0.212	0.113	0.133	100.0	100.0	14.6
sync_2610	hypothetical protein	0.018	0.350	0.050	0.052	100.0	100.0	19.0
sync_2611	transcriptional regulator, LysR family protein	0.008	0.179	0.046	0.048	100.0	95.5	11.3
sync_2612	Conserved NnrU/NnuR ortholog membrane enzyme	0.015	0.186	0.078	0.072	100.0	97.5	13.3
sync_2613	NADH dehydrogenase I chain 5 (or L) (EC:1.6.5.-)	0.006	0.170	0.034	0.036	100.0	99.7	10.7
sync_2614	NADH dehydrogenase I chain 4 (or M) (EC:1.6.5.-)	0.014	0.210	0.067	0.067	100.0	96.6	10.1
sync_2615	hypothetical protein	0.015	0.084	0.176	0.182	100.0	96.7	13.1
sync_2616	segregation and condensation protein A	0.008	0.154	0.055	0.069	100.0	77.8	7.3
sync_2617	nucleotidyl transferase family protein	0.009	0.212	0.042	0.038	100.0	99.0	9.6
sync_2618	methylenetetrahydrofolate reductase family protein	0.018	0.178	0.103	0.097	100.0	100.0	11.2
sync_2619	transcriptional regulator, LuxR family protein	0.009	0.170	0.054	0.053	100.0	100.0	16.5
sync_2620	Adenylate cyclase	0.038	0.201	0.187	0.194	100.0	100.0	10.3
sync_2621	predicted inorganic polyphosphate (EC:2.7.1.23)	0.012	0.211	0.058	0.065	100.0	93.3	11.2
sync_2622	NADH dehydrogenase I chain 4L or K (EC:1.6.5.-)	0.025	0.193	0.130	0.123	100.0	100.0	11.3
sync_2623	NADH-ubiquinone/plastoquinone oxidoreductase, subunit 6	0.011	0.217	0.051	0.048	100.0	100.0	11.1
sync_2624	NADH-plastoquinone oxidoreductase, I subunit (EC:1.6.5.3)	0.009	0.208	0.044	0.046	100.0	100.0	14.0
sync_2625	NADH dehydrogenase I chain 1 (or H) (EC:1.6.5.-)	0.005	0.251	0.020	0.021	100.0	100.0	14.9
sync_2626	Citrate synthase (EC:2.3.3.1)	0.011	0.236	0.046	0.047	100.0	100.0	13.5
sync_2627	phosphohistidine phosphatase SixA (EC:3.1.3.-)	0.037	0.102	0.358	0.334	100.0	86.4	7.8
sync_2628	hypothetical protein	0.033	0.183	0.179	0.176	100.0	90.5	10.8
sync_2629	hypothetical protein	0.027	0.126	0.212	0.204	100.0	78.2	5.8
sync_2630	heat-inducible transcription repressor HrcA	0.010	0.174	0.059	0.066	100.0	71.9	7.3
sync_2631	receptor family ligand binding protein	0.031	0.189	0.164	0.168	100.0	97.3	9.6
sync_2632	hypothetical protein	0.030	0.308	0.098	0.091	100.0	96.5	11.5
sync_2633	tryptophan synthase, beta subunit (EC:4.2.1.20)	0.006	0.141	0.039	0.033	100.0	100.0	9.6
sync_2634	hypothetical protein	0.012	0.139	0.088	0.087	100.0	100.0	10.8
sync_2635	Translation initiation factor SUI1	0.014	0.100	0.144	0.194	100.0	67.3	6.1
sync_2636	adenylylsulfate kinase (EC:2.7.1.25)	0.016	0.121	0.129	0.120	100.0	74.7	6.8
sync_2637	hypothetical protein	0.013	0.129	0.101	0.096	100.0	84.7	7.9
sync_2638	phosphoribosylaminoimidazole carboxylase, catalytic subunit (EC:4.1.1.21)	0.021	0.106	0.198	0.167	100.0	88.8	7.1
sync_2639	N-acetylglucosamine-6-phosphate deacetylase (EC:3.5.1.25)	0.022	0.108	0.205	0.203	100.0	99.6	8.9
sync_2640	magnesium protoporphyrin O-methyltransferase (EC:2.1.1.11)	0.038	0.292	0.132	0.132	100.0	96.9	17.0
sync_2641	Pseudouridine synthase (EC:4.2.1.70)	0.038	0.145	0.264	0.261	99.9	91.7	9.9
sync_2642	DNA-binding response regulator	0.008	0.229	0.035	0.038	100.0	97.8	11.8
sync_2643	hypothetical protein	0.052	0.237	0.221	0.227	100.0	100.0	14.0
sync_2644	NifS-like aminotransferase class-V (EC:2.8.1.7)	0.027	0.194	0.138	0.140	100.0	98.2	9.2

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_2645	hypothetical protein	0.021	0.122	0.172	0.171	100.0	49.5	4.2
sync_2646	hypothetical protein	0.007	0.097	0.073	0.062	100.0	100.0	7.0
sync_2647	S-adenosyl-methyltransferase MraW (EC:2.1.1.-)	0.015	0.138	0.109	0.117	100.0	91.8	7.9
sync_2648	NADH dehydrogenase I chain 7 (or D) (EC:1.6.5.-)	0.004	0.152	0.024	0.032	100.0	84.4	7.6
sync_2649	hypothetical protein	0.200	0.000	Inf	Inf	100.0	0.5	3.0
sync_2650	Predicted thioesterase	0.000	0.000	NaN	NaN	100.0	8.1	3.8
sync_2651	putative inner membrane protein	0.000	0.000	NaN	NaN	100.0	0.0	2.3
sync_2652	TM2 domain family protein	0.000	0.000	NaN	NaN	100.0	0.0	2.0
sync_2653	putative O-succinylbenzoate--CoA ligase	0.000	0.000	NaN	NaN	87.8	0.0	1.9
sync_2654	O-succinylbenzoate-CoA synthase (EC:4.2.1.-)	0.024	0.140	0.173	0.145	92.2	19.1	2.6
sync_2655	1,4-dihydroxy-2-naphthoate phytyltransferase	0.006	0.062	0.100	0.098	71.6	24.7	2.9
sync_2656	isochorismate synthase (EC:5.4.4.2)	0.000	0.047	0.000	0.000	64.7	7.3	1.7
sync_2657	glutathione synthase (EC:6.3.2.3)	0.029	0.373	0.077	0.075	83.8	14.8	2.7
sync_2658	glutaredoxin 3	0.010	0.037	0.260	0.262	100.0	39.0	3.8
sync_2660	hypothetical protein	0.000	0.100	0.000	0.000	100.0	55.2	4.6
sync_2661	Hypothetical UPF0054 protein	0.031	0.094	0.333	0.300	100.0	65.4	5.5
sync_2662	diacylglycerol kinase (EC:2.7.1.107)	0.015	0.165	0.093	0.091	100.0	57.6	5.3
sync_2663	para-aminobenzoate synthase component II (EC:6.3.5.8)	0.009	0.167	0.054	0.068	100.0	68.7	6.3
sync_2664	hydrolase, metallo-beta-lactamase family protein	0.016	0.075	0.211	0.204	100.0	78.5	7.3
sync_2665	L-asparaginase (EC:3.5.1.1)	0.049	0.171	0.285	0.288	99.0	71.7	6.9
sync_2666	histidinol-phosphate aminotransferase (EC:2.6.1.9)	0.033	0.180	0.185	0.177	100.0	91.9	8.2
sync_2667	arginyl-tRNA synthetase (EC:6.1.1.19)	0.018	0.180	0.098	0.097	100.0	100.0	11.8
sync_2668	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2669	Iron-containing alcohol dehydrogenase (EC:1.1.1.1)	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2670	Antibiotic biosynthesis monooxygenase	0.000	0.000	NaN	NaN	58.6	0.0	0.6
sync_2671	sensory box/GGDEF family protein	0.000	0.000	NaN	NaN	14.6	0.0	0.1
sync_2672	hypothetical protein	0.000	0.000	NaN	NaN	41.1	0.0	0.4
sync_2673	putative peroxidase	0.000	0.000	NaN	NaN	15.1	0.0	0.2
sync_2674	hypothetical protein	0.000	0.000	NaN	NaN	35.4	0.0	0.5
sync_2675	hypothetical protein	0.000	0.000	NaN	NaN	82.3	16.4	2.1
sync_2676	hypothetical protein	0.000	0.000	NaN	NaN	56.0	0.0	0.6
sync_2677	hypothetical protein	0.000	0.000	NaN	NaN	37.9	0.0	0.6
sync_2678	Carboxyvinyl-carboxyphosphonate phosphorylmutase(Carboxyphosphonoenolpyruvate phosphonomutase) (EC:2.7.8.23)	0.007	0.112	0.060	0.068	54.1	30.6	3.6
sync_2679	enantiomer-selective amidase (EC:6.3.5.-)	0.000	0.000	NaN	NaN	45.6	0.0	0.5
sync_2680	hypothetical protein	0.000	0.000	NaN	NaN	18.4	0.0	0.3
sync_2681	vanadium-dependent bromoperoxidase 2	0.000	0.000	NaN	NaN	16.0	0.0	0.3
sync_2682	hypothetical protein	0.000	0.000	NaN	NaN	45.5	0.0	0.5
sync_2683	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2684	hypothetical protein	0.000	0.000	NaN	NaN	13.0	0.0	0.1
sync_2685	mechanosensitive ion channel family protein	0.000	0.000	NaN	NaN	76.8	0.0	1.4
sync_2686	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2687	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	1.3
sync_2688	hypothetical protein	0.000	0.000	NaN	NaN	40.4	0.0	0.4
sync_2689	hypothetical protein	0.011	0.061	0.183	0.189	87.2	5.5	1.9
sync_2690	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	1.7
sync_2691	hypothetical protein	0.000	0.000	NaN	NaN	95.8	0.0	1.4
sync_2692	nicotinate-nucleotide pyrophosphorylase (EC:2.4.2.19)	0.034	0.206	0.167	0.184	97.5	64.9	9.1
sync_2693	hypothetical protein	0.011	0.029	0.378	0.374	97.1	5.0	2.5
sync_2694	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	1.0
sync_2695	hypothetical protein	0.000	0.000	NaN	NaN	6.1	0.0	0.1
sync_2696	conserved hypothetical protein TIGR00149	0.000	0.000	NaN	NaN	47.4	0.0	0.9
sync_2697	hypothetical protein	0.000	0.000	NaN	NaN	62.7	0.0	1.1
sync_2698	hypothetical protein	0.021	0.176	0.118	0.099	95.4	3.0	2.4
sync_2699	tRNA modification GTPase TrmE	0.024	0.178	0.138	0.126	100.0	97.6	9.4
sync_2700	hypothetical protein	0.011	0.120	0.093	0.114	92.4	85.7	7.7
sync_2701	Uncharacterized protein family protein	0.023	0.187	0.125	0.114	93.5	67.7	6.2
sync_2702	RelA/SpoT family protein	0.009	0.163	0.052	0.051	100.0	77.9	7.5
sync_2703	ABC transporter, ATP binding component, possibly for oligopeptides	0.009	0.105	0.089	0.086	100.0	90.0	9.6
sync_2704	hypothetical protein	0.000	0.000	NaN	0.000	89.1	11.5	2.9
sync_2705	cellulase (EC:3.2.1.4)	0.046	0.156	0.295	0.256	87.6	73.0	9.7
sync_2706	23S RNA-specific pseudouridylate synthase (EC:4.2.1.70)	0.018	0.252	0.070	0.066	100.0	100.0	12.7
sync_2707	GTPase family protein	0.009	0.169	0.056	0.056	100.0	96.4	10.0
sync_2708	universal stress family protein	0.023	0.113	0.201	0.200	100.0	100.0	11.0
sync_2709	phosphoglycerate kinase (EC:2.7.2.3)	0.006	0.159	0.035	0.034	100.0	96.0	8.9

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_2710	possible Ribosomal L29e protein family protein	0.031	0.178	0.176	0.180	100.0	96.5	8.5
sync_2711	lipoprotein, putative	0.023	0.075	0.303	2.878	100.0	75.2	5.4
sync_2712	3-hydroxyisobutyrate dehydrogenase (EC:1.1.1.31)	0.024	0.094	0.252	0.316	100.0	44.9	4.4
sync_2713	Uncharacterized secreted protein	0.013	0.149	0.090	0.085	98.1	66.4	5.5
sync_2714	undecaprenyldiphospho-muramoylpentapeptide beta-N-acetylglucosaminyltransferase (EC:2.4.1.227)	0.013	0.104	0.123	0.141	100.0	68.9	5.1
sync_2715	L-threonine-O-3-phosphate decarboxylase (EC:4.1.1.81)	0.034	0.105	0.323	0.310	100.0	70.3	5.7
sync_2716	hypothetical protein	0.028	0.142	0.195	0.192	66.0	49.6	4.4
sync_2717	possible general (type II) secretion pathway protein D precursor	0.000	0.000	NaN	0.133	28.8	26.6	2.9
sync_2718	hypothetical protein	0.048	0.155	0.312	0.321	81.8	57.4	5.1
sync_2719	Fimbrial assembly protein (PilN) family protein	0.000	0.000	NaN	0.164	48.6	25.7	1.8
sync_2720	hypothetical protein	0.049	0.177	0.276	0.286	100.0	64.3	5.7
sync_2721	dihydroorotate oxidase (EC:1.3.3.1)	0.011	0.150	0.070	0.063	100.0	75.7	7.4
sync_2722	ribonuclease H (EC:3.1.26.4)	0.013	0.079	0.171	0.179	100.0	46.9	4.9
sync_2723	hypothetical protein	0.029	0.168	0.176	0.164	100.0	60.9	5.8
sync_2724	ribosomal protein L7/L12	0.002	0.059	0.038	0.038	100.0	54.5	5.3
sync_2725	ribosomal protein L10	0.011	0.074	0.145	0.139	100.0	98.1	7.3
sync_2726	ribosomal protein L1	0.004	0.115	0.033	0.031	100.0	84.8	7.3
sync_2727	ribosomal protein L11	0.002	0.029	0.073	0.060	100.0	100.0	7.8
sync_2728	transcription termination/antitermination factor NusG	0.004	0.145	0.029	0.028	100.0	93.6	6.9
sync_2729	preprotein translocase, SecE subunit	0.007	0.058	0.122	0.121	100.0	83.8	6.2
sync_2730	ATPases with chaperone activity, ATP-binding subunit	0.008	0.177	0.047	0.051	100.0	80.8	7.3
sync_2731	lactoylglutathione lyase (EC:4.4.1.5)	0.022	0.124	0.175	0.167	100.0	100.0	9.9
sync_2732	phosphopyruvate hydratase (EC:4.2.1.11)	0.006	0.172	0.034	0.037	100.0	92.0	10.7
sync_2733	ABC1 domain protein	0.015	0.236	0.065	0.076	100.0	99.2	16.1
sync_2734	hypothetical protein	0.038	0.207	0.186	0.182	100.0	100.0	15.3
sync_2735	Proline-rich region	0.074	0.216	0.342	0.337	100.0	100.0	15.5
sync_2736	arginine biosynthesis bifunctional protein ArgJ (EC:2.3.1.35,EC:2.3.1.1)	0.040	0.283	0.142	0.141	100.0	100.0	12.7
sync_2737	dephospho-CoA kinase (EC:2.7.1.24)	0.063	0.200	0.313	0.269	72.7	17.4	2.2
sync_2738	hypothetical protein	0.000	0.000	NaN	NaN	55.1	0.0	0.7
sync_2739	glutamyl-tRNA(Gln) amidotransferase, B subunit (EC:6.3.5.-)	0.017	0.356	0.048	0.051	100.0	96.9	22.7
sync_2740	glycine oxidase ThiO (EC:1.4.3.19)	0.058	0.356	0.163	0.164	100.0	85.9	12.4
sync_2741	nucleoside diphosphate kinase (EC:2.7.4.6)	0.016	0.307	0.053	0.053	100.0	92.8	10.9
sync_2742	arginine decarboxylase (EC:4.1.1.19)	0.022	0.327	0.068	0.068	100.0	100.0	13.7
sync_2743	hypothetical protein	0.016	0.130	0.122	0.204	100.0	62.4	4.9
sync_2744	Cyclic nucleotide-binding domain	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2745	adenylate cyclase, putative	0.000	0.000	NaN	NaN	46.3	0.0	0.7
sync_2746	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2747	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2748	hypothetical protein	0.000	0.000	NaN	NaN	12.1	0.0	0.2
sync_2749	hypothetical protein	0.000	0.000	NaN	NaN	50.1	0.0	0.8
sync_2750	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2751	hypothetical protein	0.000	0.000	NaN	NaN	93.9	0.0	0.9
sync_2752	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	1.0
sync_2753	hypothetical protein	0.000	0.000	NaN	NaN	61.2	0.0	0.6
sync_2754	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2755	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2756	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2757	transcriptional regulator	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2758	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2759	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2760	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2761	hypothetical protein	0.000	0.000	NaN	NaN	43.3	0.0	0.4
sync_2762	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2763	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2764	transcriptional regulator	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2765	transcriptional regulator, Crp/Fnr family protein	0.000	0.000	NaN	NaN	61.9	0.0	1.2
sync_2766	RND multidrug efflux transporter	0.031	0.239	0.130	0.134	63.9	8.1	1.7
sync_2767	efflux transporter, RND family, MFP subunit subfamily	0.000	0.000	NaN	0.538	60.0	4.9	1.5
sync_2768	alanyl-tRNA synthetase (EC:6.1.1.7)	0.023	0.237	0.097	0.100	100.0	91.1	12.1
sync_2769	hypothetical protein	0.022	0.072	0.306	0.371	100.0	61.9	5.2
sync_2770	helicase, Snf2 family protein	0.005	0.126	0.043	0.046	100.0	67.8	6.1
sync_2771	uncharacterized swim-like zn-finger-containing conserved protein	0.006	0.094	0.068	0.051	93.1	47.8	4.8
sync_2772	hypothetical protein	0.035	0.097	0.361	0.372	100.0	99.8	7.4
sync_2773	Mn2+ and Fe2+ transporter, NRAMP family protein	0.009	0.118	0.078	0.090	100.0	62.3	5.3

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_2774	hypothetical protein	0.108	0.176	0.613	0.542	100.0	15.8	4.1
sync_2775	small heat shock protein (HSP20) family protein	0.013	0.094	0.143	0.148	100.0	86.9	6.4
sync_2776	possible Tripartite transporter component (TRAP-T family), substrate binding protein	0.009	0.074	0.120	0.114	100.0	85.9	6.4
sync_2777	possible Tripartite transporter component (TRAP-T family)	0.036	0.065	0.556	0.442	100.0	48.6	4.9
sync_2778	possible Tripartite transporter component (TRAP-T family)	0.014	0.130	0.109	0.105	100.0	81.1	7.1
sync_2779	flavoprotein (EC:1.-)	0.008	0.154	0.054	0.056	100.0	93.3	9.5
sync_2780	Flavodoxin:Flavin reductase-like domain (EC:1.-)	0.010	0.118	0.085	0.086	100.0	93.7	7.8
sync_2781	rubredoxin	0.005	0.089	0.051	0.049	96.2	54.9	4.7
sync_2782	putative transcriptional regulator	0.000	0.000	NaN	NaN	2.0	0.0	0.1
sync_2783	possible reductase	0.037	0.211	0.177	0.169	100.0	100.0	11.3
sync_2784	hypothetical protein	0.038	0.143	0.269	0.305	87.9	62.5	6.2
sync_2785	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	4.0
sync_2786	hypothetical protein	0.033	0.268	0.125	0.118	100.0	100.0	10.9
sync_2787	hypothetical protein	0.000	0.145	0.000	0.073	98.2	29.6	3.4
sync_2788	glycine dehydrogenase (EC:1.4.4.2)	0.013	0.167	0.076	0.075	100.0	63.1	5.3
sync_2789	glycine cleavage system H protein	0.013	0.114	0.113	0.097	100.0	73.1	7.5
sync_2790	aluminium resistance protein	0.013	0.165	0.078	0.083	100.0	93.7	8.7
sync_2791	fatty acid desaturase (EC:1.14.19.1)	0.003	0.137	0.019	0.018	100.0	39.1	4.3
sync_2792	hypothetical protein	0.024	0.156	0.154	0.170	100.0	88.6	6.2
sync_2793	fatty acid desaturase (EC:1.14.19.1)	0.009	0.194	0.046	0.050	100.0	87.7	8.4
sync_2794	ribosomal protein L9	0.001	0.130	0.008	0.007	100.0	99.8	7.1
sync_2795	replicative DNA helicase (EC:3.6.1.-)	0.017	0.183	0.091	0.093	100.0	88.4	11.2
sync_2796	glucose-inhibited division protein A	0.022	0.305	0.073	0.073	100.0	100.0	19.9
sync_2797	possible DNA polymerase III beta subunit, C-te	0.066	0.234	0.283	0.308	100.0	100.0	17.0
sync_2798	lysostaphin	0.038	0.304	0.125	0.120	100.0	63.8	10.9
sync_2799	hypothetical protein	0.122	0.154	0.791	0.881	100.0	100.0	9.1
sync_2800	hypothetical protein	0.003	0.055	0.049	0.049	100.0	100.0	13.1
sync_2801	hypothetical protein	0.040	0.241	0.166	0.180	100.0	100.0	30.4
sync_2802	hypothetical protein	0.063	0.173	0.366	0.365	100.0	97.5	10.0
sync_2803	hypothetical protein	0.005	0.230	0.022	0.027	100.0	88.3	11.8
sync_2804	possible Viral (Superfamily 1) RNA helicase	0.039	0.178	0.220	0.248	100.0	86.9	7.2
sync_2805	hypothetical protein	0.009	0.143	0.060	0.056	100.0	75.2	8.0
sync_2806	DNA ligase, NAD-dependent (EC:6.5.1.2)	0.026	0.200	0.129	0.134	100.0	100.0	13.4
sync_2807	Domain of unknown function (DUF389) family protein	0.011	0.166	0.065	0.062	100.0	100.0	16.3
sync_2808	hypothetical protein	0.051	0.216	0.238	0.239	100.0	100.0	12.5
sync_2809	hypothetical protein	0.011	0.178	0.063	0.069	100.0	84.7	20.5
sync_2810	hypothetical protein	0.036	0.029	1.248	1.576	100.0	81.1	9.7
sync_2811	hypothetical protein	0.039	0.239	0.162	0.192	100.0	82.6	18.7
sync_2812	mechanosensitive ion channel family protein, putative	0.029	0.238	0.121	0.131	75.0	61.2	16.2
sync_2813	hypothetical protein	0.042	0.268	0.157	0.150	100.0	100.0	37.4
sync_2814	hypothetical protein	0.030	0.231	0.130	0.132	100.0	100.0	41.7
sync_2815	ABC-type transport protein slr1901	0.034	0.252	0.133	0.149	100.0	100.0	41.9
sync_2816	hypothetical protein	0.112	0.291	0.383	0.409	100.0	100.0	30.1
sync_2817	hypothetical protein	0.193	0.442	0.437	0.469	100.0	100.0	49.0
sync_2818	mechanosensitive ion channel family protein	0.074	0.253	0.294	0.297	99.8	99.2	54.1
sync_2819	ABC-type phosphate/phosphonate transport system, periplasmic component, putative	0.000	0.000	NaN	NaN	28.8	0.0	0.3
sync_2820	GGDEF/EAL domain protein	0.000	0.000	NaN	NaN	50.4	0.0	0.7
sync_2821	valyl-tRNA synthetase	0.000	0.000	NaN	NaN	100.0	0.0	1.0
sync_2822	valyl-tRNA synthetase (EC:6.1.1.9)	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2823	valyl-tRNA synthetase	0.136	0.189	0.723	0.721	100.0	87.5	11.9
sync_2824	transporter, small conductance mechanosensitive ion channel (MscS) family protein	0.054	0.419	0.129	0.130	100.0	100.0	42.7
sync_2825	valyl-tRNA synthetase (EC:6.1.1.9)	0.023	0.254	0.090	0.093	100.0	100.0	20.5
sync_2826	hypothetical protein	0.033	0.089	0.366	0.338	100.0	96.4	12.2
sync_2827	possible Trypsin	0.023	0.184	0.124	0.121	95.7	89.6	14.6
sync_2828	possible Protein phosphatase 2A regulatory B s	0.029	0.179	0.165	0.163	100.0	96.6	12.0
sync_2829	hypothetical protein	0.009	0.109	0.081	0.071	100.0	100.0	12.5
sync_2830	oxidoreductase, short chain dehydrogenase/reductase family protein	0.019	0.119	0.159	0.208	100.0	100.0	15.4
sync_2831	transcriptional regulator, Fur family protein	0.017	0.201	0.086	0.089	100.0	100.0	14.6
sync_2832	hypothetical protein	0.027	0.199	0.135	0.116	100.0	100.0	15.4
sync_2833	hypothetical protein	0.075	0.037	2.046	3.161	100.0	100.0	14.9
sync_2834	hypothetical protein	0.051	0.042	1.235	0.951	80.5	56.9	5.2
sync_2835	hypothetical protein	0.014	0.226	0.060	0.057	100.0	100.0	14.0
sync_2836	hypothetical protein	0.059	0.200	0.295	0.284	100.0	84.2	9.0
sync_2837	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_2838	hypothetical protein	0.000	0.000	NaN	NaN	40.7	0.0	0.4
sync_2839	possible high light inducible protein	0.051	0.201	0.254	0.291	100.0	77.9	5.1
sync_2840	hypothetical protein	0.027	0.127	0.211	0.231	100.0	100.0	8.3
sync_2841	ferredoxin	0.008	0.012	0.633	0.322	100.0	39.3	4.4
sync_2842	hypothetical protein	0.000	0.067	0.000	0.000	100.0	60.7	5.9
sync_2843	hypothetical protein	0.018	0.218	0.082	0.098	100.0	100.0	14.6
sync_2845	hypothetical protein	0.008	0.033	0.227	0.208	100.0	96.3	8.6
sync_2846	MazG family protein	0.010	0.107	0.098	0.100	100.0	87.1	7.5
sync_2847	Class I peptide chain release factor	0.000	0.000	NaN	NaN	99.1	0.0	1.6
sync_2848	hypothetical protein	0.000	0.000	NaN	NaN	37.0	0.0	0.5
sync_2849	hypothetical protein	0.056	0.048	1.153	1.674	99.5	56.7	4.2
sync_2850	spermidine synthase (EC:2.5.1.16)	0.037	0.189	0.198	0.204	100.0	70.6	6.6
sync_2851	agmatinase (EC:3.5.3.11)	0.016	0.133	0.119	0.116	100.0	85.8	7.6
sync_2852	hypothetical protein	0.028	0.037	0.756	0.727	100.0	82.3	5.9
sync_2853	glycine cleavage system T protein (EC:2.1.2.10)	0.015	0.144	0.104	0.099	100.0	82.8	7.3
sync_2854	aspartyl-tRNA synthetase (EC:6.1.1.12)	0.008	0.118	0.065	0.059	100.0	68.1	6.2
sync_2855	RNA polymerase sigma-70 factor family protein	0.005	0.126	0.042	0.041	100.0	85.0	6.6
sync_2856	dps protein	0.021	0.227	0.091	0.096	100.0	47.8	4.5
sync_2857	Short-chain dehydrogenase/reductase family enzyme (EC:1.-)	0.020	0.204	0.098	0.091	100.0	84.2	8.4
sync_2858	CTP synthase (EC:6.3.4.2)	0.012	0.187	0.064	0.063	100.0	84.9	8.5
sync_2859	possible organic radical activating enzyme	0.038	0.195	0.193	0.198	100.0	76.7	6.6
sync_2860	hypothetical protein	0.031	0.205	0.153	0.155	82.5	61.6	4.1
sync_2861	ABC transporter, multi drug efflux family protein	0.024	0.290	0.082	0.084	100.0	71.6	6.7
sync_2862	possible transporter component	0.027	0.202	0.135	0.110	81.1	11.0	2.4
sync_2863	Ecotin precursor	0.028	0.191	0.145	0.155	100.0	94.3	9.3
sync_2864	exsB protein	0.012	0.120	0.102	0.098	100.0	96.5	15.2
sync_2865	possible p-aminobenzoate synthetase (EC:6.3.5.8)	0.035	0.166	0.213	0.204	100.0	98.8	13.1
sync_2866	Aminotransferases class-IV	0.036	0.136	0.263	0.246	100.0	100.0	13.0
sync_2867	amino acid permease family protein	0.023	0.162	0.140	0.144	100.0	100.0	17.4
sync_2868	ABC transporter, ATP-binding protein	0.018	0.239	0.076	0.072	100.0	58.8	6.3
sync_2869	ABC transporter ATP-binding protein	0.007	0.178	0.038	0.055	89.4	72.4	11.4
sync_2870	ABC transporter permease protein	0.019	0.205	0.094	0.089	98.0	82.0	11.6
sync_2871	high-affinity branched-chain amino acid transport protein LivH	0.012	0.202	0.061	0.054	100.0	100.0	16.1
sync_2872	urea ABC transporter, periplasmic urea-binding protein	0.021	0.279	0.075	0.075	100.0	99.9	18.1
sync_2873	urease accessory protein UreG	0.025	0.268	0.095	0.091	100.0	100.0	35.6
sync_2874	urease accessory protein UreF	0.041	0.149	0.276	0.268	100.0	89.4	15.0
sync_2875	urease accessory protein E	0.014	0.112	0.127	0.152	100.0	100.0	12.8
sync_2876	urease accessory protein UreD	0.046	0.157	0.294	0.296	100.0	100.0	19.6
sync_2877	urease, gamma subunit (EC:3.5.1.5)	0.012	0.245	0.049	0.042	100.0	100.0	29.0
sync_2878	urease, beta subunit (EC:3.5.1.5)	0.017	0.150	0.114	0.121	100.0	100.0	35.5
sync_2879	urease, alpha subunit (EC:3.5.1.5)	0.016	0.227	0.070	0.075	100.0	100.0	21.2
sync_2880	glutamine amidotransferase, class I	0.038	0.145	0.263	0.263	100.0	100.0	21.2
sync_2881	hypothetical protein	0.022	0.183	0.122	0.121	100.0	100.0	26.9
sync_2882	Bacterial domain of unknown function (DUF403) superfamily protein	0.013	0.102	0.130	0.129	100.0	100.0	28.1
sync_2883	possible transglutaminase-like enzyme	0.026	0.146	0.180	0.179	100.0	100.0	23.2
sync_2884	Predicted redox protein	0.025	0.201	0.125	0.126	100.0	100.0	32.9
sync_2885	molybdenum cofactor biosynthesis protein MoaA	0.022	0.094	0.230	0.230	100.0	100.0	23.6
sync_2886	molybdopterin-guanine dinucleotide biosynthesis protein A	0.019	0.078	0.243	0.241	100.0	100.0	25.4
sync_2887	nitrate permease NapA	0.008	0.196	0.040	0.040	100.0	100.0	26.9
sync_2888	nitrate reductase (EC:1.7.99.4)	0.017	0.154	0.112	0.113	100.0	100.0	23.9
sync_2889	nitrate reductase associated protein	0.020	0.132	0.151	0.151	100.0	100.0	23.1
sync_2890	hypothetical protein	0.016	0.097	0.168	0.169	100.0	100.0	25.2
sync_2891	molybdenum cofactor biosynthesis protein C	0.017	0.086	0.193	0.179	100.0	100.0	26.0
sync_2892	molybdopterin biosynthesis protein MoeA	0.022	0.090	0.249	0.243	100.0	100.0	24.4
sync_2893	molybdopterin converting factor, subunit 2	0.027	0.118	0.228	0.229	100.0	100.0	21.5
sync_2894	molybdopterin biosynthesis	0.054	0.227	0.237	0.237	100.0	100.0	23.9
sync_2895	molybdenum cofactor biosynthesis protein B1	0.022	0.172	0.130	0.127	100.0	100.0	24.7
sync_2896	uroporphyrin-III C-methyltransferase (EC:2.1.1.107)	0.014	0.209	0.068	0.068	100.0	100.0	16.0
sync_2897	hypothetical protein	0.050	0.168	0.297	0.317	100.0	69.4	7.2
sync_2898	Ferredoxin--nitrite reductase (EC:1.7.7.1)	0.024	0.253	0.097	0.088	100.0	99.4	11.7
sync_2899	Formate and nitrite transporters	0.007	0.159	0.042	0.037	100.0	100.0	21.0
sync_2900	possible Homeobox domain	0.061	0.220	0.276	0.273	100.0	98.6	19.1
sync_2901	cyanate hydratase (EC:4.2.1.104)	0.018	0.268	0.069	0.075	100.0	100.0	21.9
sync_2902	hypothetical protein	0.052	0.000	Inf	Inf	100.0	98.3	7.9
sync_2903	hypothetical protein	0.000	0.000	NaN	NaN	90.9	0.0	1.6

Table A5.1. dN/dS ratios of CC9311 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9311	% of gene covered 1X	% of gene covered 5X	avg. read depth
sync_2904	glycosyl transferase family protein	0.021	0.174	0.123	0.121	100.0	98.0	14.8
sync_2905	transporter, drug:H+ antiporter-1 (12 spanner) (DHA1) family protein	0.012	0.106	0.111	0.104	100.0	92.0	8.6
sync_2906	polyphosphate kinase (EC:2.7.4.1)	0.013	0.187	0.068	0.065	93.7	87.3	10.9
sync_2907	RNA polymerase sigma factor	0.013	0.177	0.076	0.062	100.0	93.3	8.8
sync_2908	phosphatidate cytidylyltransferase, putative	0.027	0.138	0.196	0.196	98.5	39.6	4.3
sync_2909	phospho-2-dehydro-3-deoxyheptonate aldolase (EC:2.5.1.54)	0.008	0.125	0.066	0.058	100.0	74.1	7.7
sync_2910	aconitate hydratase 2 (EC:4.2.1.3)	0.006	0.125	0.048	0.060	100.0	92.4	8.2
sync_2911	chloride transporter, chloride channel (ClC) family protein	0.012	0.134	0.086	0.084	100.0	93.0	11.4
sync_2912	hypothetical protein	0.182	0.071	2.545	1.622	61.1	12.5	2.3
sync_2913	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2914	peptidase, S8A (subtilisin) family protein	0.000	0.000	NaN	NaN	11.6	0.0	0.1
sync_2915	SAM radical enzyme	0.014	0.199	0.071	0.068	100.0	98.6	15.7
sync_2916	O-Antigen Polymerase family protein	0.033	0.196	0.166	0.161	99.5	94.3	12.5
sync_2917	formyltetrahydrofolate deformylase (EC:3.5.1.10)	0.019	0.168	0.111	0.097	100.0	98.6	12.9
sync_2918	lipoprotein, putative	0.001	0.070	0.014	0.013	100.0	100.0	7.9
sync_2919	NAD binding site:D-amino acid oxidase (EC:1.-)	0.029	0.114	0.251	0.247	94.0	90.2	10.0
sync_2920	phosphate ABC transporter, phosphate-binding protein	0.014	0.162	0.088	0.089	100.0	97.2	10.6
sync_2921	putative acyltransferase	0.048	0.374	0.129	0.126	53.0	38.2	3.6
sync_2922	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2923	chaperone protein dnaK2 (Heat shock protein 70-2) (Heat shock 70 kDaprotein 2) (HSP70-2)	0.005	0.202	0.025	0.030	100.0	98.2	12.9
sync_2924	shikimate 5-dehydrogenase (EC:1.1.1.25)	0.050	0.180	0.276	0.271	100.0	99.7	13.2
sync_2925	Uncharacterized membrane protein	0.002	0.114	0.019	0.026	100.0	93.8	10.2
sync_2926	ribosomal protein S6	0.001	0.190	0.007	0.013	100.0	100.0	14.5
sync_2927	hypothetical protein	0.038	0.199	0.191	0.168	100.0	100.0	13.7
sync_2928	argininosuccinate synthase (EC:6.3.4.5)	0.008	0.170	0.050	0.053	100.0	100.0	18.2
sync_2929	possible NADH-Ubiquinone/plastoquinone (complex I)	0.035	0.096	0.370	0.309	100.0	85.1	13.1
sync_2930	hypothetical protein	0.005	0.126	0.042	0.038	100.0	100.0	13.2
sync_2931	phospho-N-acetylmuramoyl-pentapeptide- transferase (EC:2.7.8.13)	0.037	0.285	0.129	0.120	100.0	98.1	19.0
sync_2932	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
sync_2933	hypothetical protein	0.072	0.213	0.336	0.374	100.0	100.0	35.7
sync_2934	Cytochrome b6, putative	0.018	0.261	0.067	0.066	100.0	100.0	28.0
sync_2935	phosphoribosylglycinamide formyltransferase 2 (EC:2.1.2.-)	0.029	0.181	0.162	0.155	100.0	100.0	27.1
sync_2936	Sucrose phosphate synthase (EC:2.4.1.14)	0.018	0.197	0.092	0.101	100.0	100.0	26.9
sync_2937	hypothetical protein	0.096	0.106	0.908	0.711	100.0	100.0	30.6
sync_2938	Esterase/lipase/thioesterase family active site	0.060	0.294	0.205	0.221	100.0	100.0	26.5
sync_2939	excinuclease ABC, A subunit	0.011	0.267	0.042	0.043	100.0	100.0	21.3
sync_2940	DNA repair protein RecN	0.020	0.166	0.123	0.127	100.0	99.8	10.9
sync_2941	ABC1 domain protein	0.008	0.208	0.038	0.036	100.0	83.4	6.5
sync_2942	Uncharacterized membrane protein	0.014	0.151	0.095	0.093	100.0	91.8	6.7
sync_2943	threonine synthase (EC:4.2.3.1)	0.003	0.148	0.019	0.019	96.6	62.0	5.0
sync_2944	hypothetical protein	0.050	0.043	1.167	1.109	100.0	45.0	3.4

Table A5.2. dN/dS ratios calculated from the tiling of metagenome sequences to the *Synechococcus* CC9902 genome (see Chapter 3). dN/dS relative to the CC9902 genome is also provided (dN/dS CC9902). NaN indicates that there were no polymorphisms found in the regions aligned with at least 5-fold read coverage Inf indicates that only non-synonymous polymorphisms were found.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_0001	DNA polymerase III subunit beta (EC:2.7.7.7)	0.010	0.194	0.052	0.056	100.0	51.5	5.6
Syncc9902_0002	hypothetical protein	0.020	0.196	0.103	0.105	100.0	86.3	7.3
Syncc9902_0003	phosphoribosylformylglycinamidine synthase subunit II (EC:6.3.5.3)	0.017	0.174	0.100	0.100	100.0	92.4	10.4
Syncc9902_0004	amidophosphoribosyltransferase	0.012	0.230	0.052	0.054	100.0	100.0	17.4
Syncc9902_0005	DNA gyrase/topoisomerase IV, subunit A	0.020	0.236	0.083	0.086	100.0	100.0	21.1
Syncc9902_0006	TPR repeat	0.026	0.205	0.129	0.134	100.0	100.0	21.3
Syncc9902_0007	4Fe-4S cluster binding	0.027	0.201	0.136	0.135	100.0	100.0	16.3
Syncc9902_0008	hypothetical protein	0.037	0.209	0.178	0.192	100.0	100.0	14.3
Syncc9902_0009	hypothetical protein	0.007	0.204	0.036	0.033	100.0	100.0	13.5
Syncc9902_0010	transcription antitermination protein NusB	0.022	0.144	0.153	0.163	100.0	92.0	7.6
Syncc9902_0011	Cell division transporter substrate-binding protein FtsY	0.033	0.177	0.185	0.187	100.0	94.1	9.7
Syncc9902_0012	serine phosphatase	0.015	0.198	0.075	0.074	100.0	100.0	16.2
Syncc9902_0013	argininosuccinate lyase	0.017	0.200	0.086	0.086	100.0	100.0	21.2
Syncc9902_0014	RNA-binding region RNP-1	0.020	0.167	0.117	0.121	100.0	77.2	10.9
Syncc9902_0015	hypothetical protein	0.029	0.226	0.130	0.130	100.0	100.0	19.2
Syncc9902_0016	Methionine sulfoxide reductase B (EC:1.8.4.6)	0.034	0.180	0.189	0.170	100.0	100.0	20.0
Syncc9902_0017	hypothetical protein	0.031	0.246	0.128	0.129	100.0	97.5	15.8
Syncc9902_0018	general secretion pathway protein E	0.015	0.088	0.172	0.175	95.2	39.3	3.5
Syncc9902_0019	putative heat shock protein GrpE	0.015	0.148	0.103	0.095	100.0	48.9	4.1
Syncc9902_0020	Heat shock protein DnaJ	0.003	0.135	0.024	0.024	99.8	12.6	3.0
Syncc9902_0021	hypothetical protein	0.000	0.000	NaN	NaN	100.0	1.7	3.5
Syncc9902_0022	GTPase EngC	0.027	0.165	0.165	0.176	100.0	93.0	7.0
Syncc9902_0023	hypothetical protein	0.007	0.135	0.055	0.055	100.0	89.5	8.5
Syncc9902_0024	UDP-N-acetylenolpyruvoylglucosamine reductase (EC:1.1.1.158)	0.030	0.153	0.197	0.187	100.0	70.5	6.0
Syncc9902_0025	UDP-N-acetylmuramate--alanine ligase	0.015	0.160	0.091	0.100	100.0	74.2	6.8
Syncc9902_0026	Glyceraldehyde-3-phosphate dehydrogenase, type I (EC:1.2.1.12)	0.009	0.199	0.043	0.047	100.0	87.5	7.1
Syncc9902_0027	Thiamine-monophosphate kinase (EC:2.7.4.16)	0.038	0.147	0.258	0.223	100.0	40.7	4.3
Syncc9902_0028	putative cyclophilin-type peptidyl-prolyl cis-trans isomerase	0.012	0.125	0.096	0.104	100.0	55.7	5.2
Syncc9902_0029	elongation factor P	0.001	0.090	0.010	0.010	100.0	92.7	7.1
Syncc9902_0030	Acetyl-CoA biotin carboxyl carrier	0.000	0.142	0.000	0.000	95.6	32.3	3.2
Syncc9902_0031	4-hydroxythreonine-4-phosphate dehydrogenase (EC:1.1.1.262)	0.022	0.124	0.178	0.173	96.8	65.9	6.3
Syncc9902_0032	hypothetical protein	0.036	0.186	0.194	0.193	100.0	100.0	13.6
Syncc9902_0033	HNH nuclease	0.020	0.264	0.077	0.076	100.0	100.0	13.2
Syncc9902_0034	DEAD/DEAH box helicase-like	0.016	0.213	0.075	0.085	100.0	95.2	9.6
Syncc9902_0035	hypothetical protein	0.029	0.288	0.101	0.115	100.0	94.4	14.6
Syncc9902_0036	hypothetical protein	0.024	0.220	0.111	0.110	100.0	100.0	15.1
Syncc9902_0037	hypothetical protein	0.040	0.233	0.173	0.154	100.0	100.0	22.1
Syncc9902_0038	hypothetical protein	0.027	0.246	0.110	0.106	100.0	97.2	9.2
Syncc9902_0039	hypothetical protein	0.028	0.394	0.072	0.072	100.0	100.0	14.4
Syncc9902_0040	soluble hydrogenase small subunit	0.012	0.201	0.059	0.058	100.0	100.0	13.3
Syncc9902_0041	cobalt-precorrin-6A synthase	0.036	0.233	0.155	0.148	100.0	97.5	12.6
Syncc9902_0042	bifunctional GMP synthase/glutamine amidotransferase protein	0.007	0.205	0.036	0.037	99.9	93.8	7.8
Syncc9902_0043	Methylase involved in ubiquinone/menaquinone biosynthesis-like	0.000	0.000	NaN	NaN	36.1	0.0	0.4
Syncc9902_0044	hypothetical protein	0.037	0.287	0.131	0.129	100.0	85.7	10.3
Syncc9902_0045	hypothetical protein	0.011	0.199	0.057	0.057	100.0	99.8	13.4
Syncc9902_0046	hypothetical protein	0.042	0.322	0.130	0.137	100.0	100.0	14.0
Syncc9902_0047	Peptidoglycan glycosyltransferase (EC:2.4.1.129)	0.033	0.234	0.140	0.140	100.0	92.2	11.3
Syncc9902_0048	SqdX	0.019	0.232	0.082	0.084	100.0	100.0	13.0
Syncc9902_0049	sulfolipid (UDP-sulfoquinovose) biosynthesis protein	0.013	0.311	0.042	0.044	100.0	100.0	15.4
Syncc9902_0050	high light-inducible protein-like	0.005	0.110	0.047	0.047	100.0	100.0	10.3
Syncc9902_0051	thiazole synthase	0.021	0.248	0.084	0.088	100.0	100.0	14.9
Syncc9902_0052	hypothetical protein	0.034	0.285	0.121	0.121	100.0	100.0	12.9
Syncc9902_0053	TPR repeat	0.019	0.244	0.079	0.089	100.0	97.9	12.1
Syncc9902_0054	50S ribosomal protein L20	0.011	0.179	0.064	0.077	100.0	69.5	6.7
Syncc9902_0055	50S ribosomal protein L35	0.007	0.050	0.140	0.074	100.0	100.0	9.5
Syncc9902_0056	possible amidase enhancer	0.034	0.197	0.171	0.193	100.0	98.9	11.9
Syncc9902_0057	putative glycosyltransferase	0.024	0.276	0.088	0.092	100.0	90.1	8.5
Syncc9902_0058	DNA-directed DNA polymerase (EC:2.7.7.7)	0.041	0.275	0.149	0.149	100.0	100.0	17.1

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_0059	hypothetical protein	0.053	0.338	0.158	0.162	100.0	100.0	20.0
Syncc9902_0060	Peptidase, metallopeptidase	0.054	0.192	0.281	0.293	100.0	49.5	4.6
Syncc9902_0061	ATP-dependent protease ATP-binding subunit	0.015	0.335	0.044	0.043	100.0	100.0	19.0
Syncc9902_0062	Peptidase S14, ClpP (EC:3.4.21.92)	0.004	0.217	0.021	0.024	100.0	100.0	16.3
Syncc9902_0063	trigger factor	0.024	0.261	0.090	0.091	100.0	100.0	16.8
Syncc9902_0064	Aspartate-semialdehyde dehydrogenase, USG-1 related (EC:1.2.1.11)	0.024	0.263	0.092	0.108	100.0	100.0	17.1
Syncc9902_0065	dihydrodipicolinate synthase	0.024	0.258	0.094	0.103	100.0	100.0	13.8
Syncc9902_0066	hypothetical protein	0.006	0.238	0.026	0.031	100.0	84.7	10.4
Syncc9902_0067	hypothetical protein	0.099	0.336	0.293	0.288	94.8	31.4	4.6
Syncc9902_0068	aspartate kinase (EC:2.7.2.4)	0.016	0.261	0.061	0.062	100.0	97.2	14.5
Syncc9902_0069	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0070	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0071	aspartate kinase	0.000	0.000	NaN	NaN	45.6	0.0	1.0
Syncc9902_0072	hypothetical protein	0.037	0.321	0.116	0.111	100.0	95.3	11.5
Syncc9902_0073	putative precorrin-8X methylmutase CobH	0.095	0.317	0.298	0.269	83.0	32.9	4.3
Syncc9902_0074	possible MesJ-like	0.350	0.467	0.750	0.860	76.7	7.5	2.0
Syncc9902_0075	hypothetical protein	0.025	0.249	0.102	0.100	100.0	97.2	8.2
Syncc9902_0076	hypothetical protein	0.024	0.201	0.122	0.143	100.0	92.0	13.9
Syncc9902_0077	excinuclease ABC subunit B	0.016	0.330	0.050	0.050	100.0	93.6	12.6
Syncc9902_0078	hypothetical protein	0.000	0.000	NaN	NaN	37.3	0.0	0.4
Syncc9902_0079	hypothetical protein	0.000	0.000	NaN	NaN	81.1	0.0	1.1
Syncc9902_0080	hypothetical protein	0.000	0.000	NaN	NaN	60.2	0.0	1.3
Syncc9902_0081	DNA mismatch repair protein	0.046	0.376	0.123	0.131	100.0	89.0	10.1
Syncc9902_0082	photosystem II reaction center protein PsbZ	0.000	0.000	NaN	NaN	27.5	0.0	0.6
Syncc9902_0083	glycosyltransferase	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0084	N-acylneuraminate-9-phosphate synthase (EC:2.5.1.57)	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0085	dTDP-4-dehydrorhamnose reductase-like	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0086	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0087	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0088	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0089	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0090	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0091	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0092	Glutamate-1-semialdehyde aminotransferase-like	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0093	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0094	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0095	dehydrogenase and related proteins-like	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0096	spore coat polysaccharide biosynthesis protein spsC	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0097	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0098	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0099	Capsular polysaccharide biosynthesis protein-like	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0100	HAD-superfamily hydrolase subfamily IA	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0101	capsular polysaccharide biosynthesis protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0102	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0103	riboflavin synthase subunit beta (EC:2.5.1.9)	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0104	TPR repeat	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0105	Glycosyltransferase-like	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0106	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0107	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0108	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0109	probable glycosyltransferase	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0110	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0111	probable glycosyltransferase	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0112	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0113	O-linked N-acetylglucosamine transferase SPINDLY family-like	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0114	TPR repeat	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0115	TPR repeat	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0116	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0117	cyclic nucleotide-binding domain (cNMP-BD) protein	0.000	0.000	NaN	NaN	2.7	0.0	0.0
Syncc9902_0118	hemolysin secretion protein-like	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0119	Glycosyltransferase-like	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0120	TPR repeat	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0121	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0122	Glycosyltransferase-like	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0123	translocase	0.033	0.589	0.056	0.061	92.2	81.4	10.6
Syncc9902_0124	hypothetical protein	0.000	0.000	NaN	NaN	77.8	0.0	1.1

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_0125	hypothetical protein	0.043	0.273	0.157	0.164	71.7	23.0	2.4
Syncc9902_0126	hypothetical protein	0.000	0.000	NaN	NaN	66.0	0.0	1.0
Syncc9902_0127	hypothetical protein	0.000	0.000	NaN	NaN	67.7	0.0	0.7
Syncc9902_0128	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_0129	Serine O-acetyltransferase (EC:2.3.1.30)	0.014	0.328	0.044	0.042	97.9	58.3	6.7
Syncc9902_0130	transcriptional regulator, GntR family	0.004	0.271	0.014	0.019	100.0	100.0	13.4
Syncc9902_0131	translation initiation factor IF-3	0.015	0.353	0.044	0.049	100.0	100.0	14.8
Syncc9902_0132	tRNA isopentenyltransferase (EC:2.5.1.8)	0.036	0.212	0.171	0.171	100.0	100.0	11.6
Syncc9902_0133	DNA topoisomerase IV subunit B (EC:5.99.1.3)	0.016	0.384	0.042	0.044	100.0	100.0	15.1
Syncc9902_0134	hypothetical protein	0.043	0.361	0.120	0.129	100.0	100.0	13.1
Syncc9902_0135	integral membrane protein possibly involved in chromosome condensation	0.085	0.357	0.238	0.245	100.0	100.0	13.3
Syncc9902_0136	putative integral membrane protein	0.117	0.310	0.377	0.431	100.0	100.0	16.0
Syncc9902_0137	glutathione peroxidase	0.049	0.387	0.126	0.123	100.0	100.0	15.9
Syncc9902_0138	Divalent cation transporter	0.017	0.322	0.052	0.058	100.0	100.0	19.4
Syncc9902_0139	type II alternative RNA polymerase sigma factor, sigma-70 family	0.013	0.460	0.028	0.031	100.0	100.0	18.2
Syncc9902_0140	hypothetical protein	0.047	0.371	0.127	0.131	100.0	97.3	8.5
Syncc9902_0141	fatty acid desaturase	0.036	0.473	0.075	0.069	100.0	80.1	9.3
Syncc9902_0142	hypothetical protein	0.121	0.200	0.606	0.400	100.0	6.0	3.0
Syncc9902_0143	Mutator MutT	0.062	0.285	0.217	0.210	100.0	100.0	11.3
Syncc9902_0144	putative fructokinase	0.056	0.296	0.190	0.200	100.0	100.0	11.0
Syncc9902_0145	hypothetical protein	0.060	0.246	0.242	0.245	100.0	100.0	10.3
Syncc9902_0146	S-adenosyl-L-homocysteine hydrolase (EC:3.3.1.1)	0.013	0.283	0.046	0.046	100.0	100.0	15.6
Syncc9902_0147	putative alkaline phosphatase-like protein	0.019	0.234	0.080	0.074	100.0	100.0	10.4
Syncc9902_0148	single-strand DNA-binding protein	0.006	0.217	0.028	0.015	100.0	100.0	8.7
Syncc9902_0149	Cell shape determining protein MreB/Mrl	0.005	0.242	0.022	0.021	100.0	83.3	7.1
Syncc9902_0150	putative rod shape-determining protein	0.013	0.178	0.072	0.072	100.0	90.5	6.4
Syncc9902_0151	hypothetical protein	0.026	0.165	0.155	0.174	100.0	68.0	5.8
Syncc9902_0152	putative sugar-binding protein	0.028	0.218	0.129	0.129	100.0	90.1	9.0
Syncc9902_0153	two component transcriptional regulator, winged helix family	0.009	0.317	0.027	0.024	100.0	98.3	7.6
Syncc9902_0154	lysyl-tRNA synthetase	0.008	0.197	0.043	0.041	100.0	66.1	5.6
Syncc9902_0155	hypothetical protein	0.002	0.167	0.013	0.012	100.0	81.1	5.5
Syncc9902_0156	hypothetical protein	0.030	0.195	0.155	0.144	100.0	95.8	6.5
Syncc9902_0157	hypothetical protein	0.007	0.208	0.032	0.037	100.0	72.4	5.2
Syncc9902_0158	hypothetical protein	0.052	0.411	0.125	0.114	79.4	17.7	2.2
Syncc9902_0159	hypothetical protein	0.037	0.224	0.164	0.164	100.0	72.7	5.4
Syncc9902_0160	serine/threonine protein kinase	0.058	0.327	0.176	0.177	100.0	93.8	9.9
Syncc9902_0161	SsrA-binding protein	0.011	0.259	0.042	0.043	100.0	84.3	7.7
Syncc9902_0162	Holliday junction DNA helicase RuvB	0.015	0.273	0.056	0.062	100.0	100.0	10.0
Syncc9902_0163	TPR repeat	0.019	0.236	0.079	0.087	100.0	97.5	8.4
Syncc9902_0164	Peptidase M20D, amidohydrolase (EC:3.5.1.32)	0.023	0.252	0.091	0.095	100.0	86.5	10.6
Syncc9902_0165	hypothetical protein	0.070	0.262	0.268	0.271	100.0	100.0	17.0
Syncc9902_0166	hypothetical protein	0.058	0.326	0.179	0.197	100.0	100.0	23.3
Syncc9902_0167	thiamine biosynthesis protein ThiC	0.018	0.471	0.039	0.041	100.0	97.6	15.4
Syncc9902_0168	transketolase	0.010	0.307	0.034	0.035	98.5	92.7	9.7
Syncc9902_0169	3-oxoacyl-(acyl-carrier-protein) synthase II	0.015	0.212	0.071	0.072	100.0	100.0	10.8
Syncc9902_0170	acyl carrier protein	0.000	0.137	0.000	0.000	100.0	100.0	9.8
Syncc9902_0171	photosystem I subunit VII	0.028	0.312	0.088	0.092	100.0	100.0	10.9
Syncc9902_0172	D-fructose-6-phosphate amidotransferase	0.027	0.278	0.099	0.097	100.0	90.3	10.5
Syncc9902_0173	NAD dependent epimerase/dehydratase	0.000	0.000	NaN	NaN	65.8	0.0	0.8
Syncc9902_0174	GDP-mannose 4,6-dehydratase	0.000	0.000	NaN	NaN	87.8	0.0	1.4
Syncc9902_0175	Mannose-1-phosphate guanylyltransferase/mannose-6-phosphate isomerase (EC:2.7.7.22)	0.011	0.235	0.046	0.045	100.0	97.1	9.2
Syncc9902_0176	16S rRNA-processing protein	0.027	0.177	0.155	0.162	100.0	64.3	5.3
Syncc9902_0177	hypothetical protein	0.000	0.000	NaN	NaN	100.0	1.6	3.3
Syncc9902_0178	possible potassium channel, VIC family	0.028	0.257	0.109	0.107	100.0	74.4	6.8
Syncc9902_0179	Ribonuclease III (EC:3.1.26.3)	0.060	0.232	0.258	0.302	100.0	95.6	8.1
Syncc9902_0180	hypothetical protein	0.015	0.189	0.082	0.079	100.0	100.0	9.6
Syncc9902_0181	hypothetical protein	0.031	0.267	0.118	0.117	100.0	70.8	6.4
Syncc9902_0182	Glycogen/starch/alpha-glucan phosphorylase (EC:2.4.1.1)	0.014	0.226	0.061	0.057	100.0	69.9	6.4
Syncc9902_0183	putative Na+/H+ antiporter, CPA2 family	0.014	0.239	0.057	0.067	100.0	67.0	5.4
Syncc9902_0184	hypothetical protein	0.022	0.213	0.105	0.107	100.0	100.0	8.1
Syncc9902_0185	hypothetical protein	0.017	0.218	0.077	0.092	100.0	67.8	5.7
Syncc9902_0186	hypothetical protein	0.012	0.178	0.067	0.063	100.0	85.8	6.7
Syncc9902_0187	putative glycolate oxidase subunit GlcE	0.056	0.278	0.200	0.199	62.6	20.5	2.8
Syncc9902_0188	hypothetical protein	0.037	0.178	0.211	0.213	100.0	94.1	8.8
Syncc9902_0189	putative glycolate oxidase subunit (Fe-S) protein	0.022	0.205	0.109	0.107	100.0	57.0	6.3
Syncc9902_0190	hypothetical protein	0.032	0.111	0.287	0.270	97.8	26.8	3.1

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_0191	isocitrate dehydrogenase	0.017	0.261	0.065	0.067	100.0	96.5	9.7
Syncc9902_0192	putative glycosyl transferase family protein	0.023	0.280	0.082	0.086	100.0	98.7	12.7
Syncc9902_0193	hypothetical protein	0.048	0.231	0.208	0.212	100.0	90.6	8.5
Syncc9902_0194	Methylthioadenosine nucleosidase (EC:3.2.2.16)	0.064	0.277	0.230	0.235	100.0	100.0	16.0
Syncc9902_0195	hypothetical protein	0.066	0.326	0.204	0.213	100.0	100.0	19.8
Syncc9902_0196	Heme oxygenase (decyclizing) (EC:1.14.99.3)	0.015	0.346	0.043	0.043	100.0	99.0	16.3
Syncc9902_0197	hypothetical protein	0.040	0.209	0.191	0.194	100.0	83.9	8.0
Syncc9902_0198	ATPase	0.015	0.185	0.082	0.081	100.0	50.9	4.9
Syncc9902_0199	hypothetical protein	0.021	0.056	0.382	0.426	69.2	14.7	2.2
Syncc9902_0200	putative glycosyl transferase, group 1	0.000	0.014	0.000	0.000	71.0	16.2	2.1
Syncc9902_0201	glycosyl transferase, group 1	0.000	0.053	0.000	0.000	57.6	10.0	1.8
Syncc9902_0202	glycosyl transferase, group 1	0.000	0.033	0.000	0.000	86.3	4.1	1.6
Syncc9902_0203	hypothetical protein	0.007	0.091	0.082	0.083	99.2	29.0	3.9
Syncc9902_0204	hypothetical protein	0.008	0.062	0.131	0.129	95.2	14.7	2.6
Syncc9902_0205	hypothetical protein	0.041	0.237	0.172	0.162	100.0	58.6	5.3
Syncc9902_0206	2-dehydro-3-deoxyphosphooctonate aldolase	0.038	0.401	0.095	0.094	100.0	86.5	9.9
Syncc9902_0207	ATP-dependent DNA helicase PcrA	0.018	0.386	0.047	0.047	100.0	100.0	13.9
Syncc9902_0208	hypothetical protein	0.032	0.395	0.082	0.089	100.0	100.0	17.9
Syncc9902_0209	3-deoxy-manno-octulosonate cytidylyltransferase (EC:2.7.7.38)	0.031	0.277	0.112	0.110	100.0	99.7	11.3
Syncc9902_0210	Phosphatase kdsC	0.056	0.280	0.200	0.183	100.0	100.0	14.6
Syncc9902_0211	KpsF/GutQ (EC:5.3.1.13)	0.028	0.278	0.101	0.095	100.0	100.0	14.8
Syncc9902_0212	possible polyA polymerase	0.059	0.245	0.240	0.253	100.0	100.0	12.9
Syncc9902_0213	hypothetical protein	0.047	0.193	0.243	0.272	100.0	80.4	8.6
Syncc9902_0214	Selenide,water dikinase (EC:2.7.9.3)	0.064	0.277	0.230	0.251	100.0	100.0	13.3
Syncc9902_0215	UDP-glucose 4-epimerase	0.041	0.330	0.125	0.132	100.0	99.6	13.3
Syncc9902_0216	hypothetical protein	0.020	0.064	0.319	0.355	100.0	19.3	3.5
Syncc9902_0217	Type I secretion system ATPase, HlyB	0.022	0.318	0.069	0.073	95.4	47.4	4.6
Syncc9902_0218	possible similar to leukotoxin secretion protein	0.026	0.200	0.131	0.143	100.0	95.6	9.5
Syncc9902_0219	putative glycosyltransferase	0.049	0.353	0.140	0.157	91.4	7.4	1.9
Syncc9902_0220	Histidyl-tRNA synthetase, class IIa (EC:6.1.1.21)	0.035	0.268	0.129	0.129	100.0	79.6	7.6
Syncc9902_0221	NAD dependent epimerase/dehydratase family	0.004	0.173	0.025	0.028	100.0	77.0	6.5
Syncc9902_0222	UDP-glucose 6-dehydrogenase (EC:1.1.1.22)	0.021	0.149	0.139	0.150	93.4	26.6	4.5
Syncc9902_0223	putative nucleotide sugar epimerase	0.066	0.468	0.142	0.143	100.0	100.0	27.7
Syncc9902_0224	photosystem II reaction center protein PsbJ	0.019	0.400	0.047	0.047	100.0	100.0	18.7
Syncc9902_0225	photosystem II reaction center L	0.027	0.362	0.074	0.074	100.0	100.0	25.6
Syncc9902_0226	cytochrome b559 subunit beta	0.014	0.346	0.039	0.049	100.0	100.0	43.1
Syncc9902_0227	cytochrome b559 subunit alpha	0.013	0.302	0.043	0.046	100.0	100.0	40.8
Syncc9902_0228	photosystem II stability/assembly factor-like	0.035	0.408	0.085	0.092	100.0	100.0	25.5
Syncc9902_0229	probable rubredoxin	0.034	0.236	0.142	0.129	95.7	81.0	12.0
Syncc9902_0230	NADH dehydrogenase alpha subunit	0.009	0.328	0.027	0.027	100.0	100.0	22.7
Syncc9902_0231	NADH dehydrogenase beta subunit (EC:1.6.5.3)	0.010	0.355	0.029	0.028	100.0	99.2	11.6
Syncc9902_0232	NADH dehydrogenase subunit J	0.024	0.251	0.098	0.092	100.0	86.1	9.1
Syncc9902_0233	hypothetical protein	0.020	0.286	0.070	0.073	100.0	98.2	12.0
Syncc9902_0234	ATPase	0.030	0.342	0.087	0.085	100.0	100.0	11.9
Syncc9902_0235	possible ABC transporter	0.060	0.337	0.177	0.174	100.0	100.0	19.8
Syncc9902_0236	Magnesium chelatase ATPase subunit D (EC:6.6.1.1)	0.018	0.224	0.081	0.083	100.0	94.1	11.0
Syncc9902_0237	7, 8-Dihydro-6-hydroxymethylpterin-pyrophosphokinase, HPPK	0.051	0.166	0.309	0.286	100.0	99.2	7.0
Syncc9902_0238	NUDIX hydrolase	0.020	0.210	0.096	0.095	100.0	86.3	7.2
Syncc9902_0239	Deoxyribodipyrimidine photolyase (EC:4.1.99.3)	0.018	0.216	0.082	0.091	100.0	85.4	8.8
Syncc9902_0240	hypothetical protein	0.016	0.009	1.656	1.656	100.0	32.4	3.5
Syncc9902_0241	putative pleiotropic regulatory protein	0.017	0.175	0.098	0.102	100.0	91.8	8.2
Syncc9902_0242	hypothetical protein	0.026	0.153	0.171	0.168	94.8	78.5	7.4
Syncc9902_0243	hypothetical protein	0.041	0.234	0.175	0.173	100.0	100.0	12.4
Syncc9902_0244	hypothetical protein	0.024	0.245	0.100	0.112	100.0	97.3	8.9
Syncc9902_0245	probable deoxyribodipyrimidine photolyase	0.038	0.247	0.154	0.155	100.0	100.0	13.9
Syncc9902_0246	enoyl-(acyl carrier protein) reductase	0.010	0.212	0.046	0.042	100.0	97.2	10.7
Syncc9902_0247	imidazoleglycerol-phosphate dehydratase (EC:4.2.1.19)	0.013	0.286	0.045	0.050	100.0	100.0	13.1
Syncc9902_0248	lignostilbene-alpha,beta-dioxygenase and related enzyme-like	0.023	0.280	0.082	0.084	100.0	93.5	11.1
Syncc9902_0249	hypothetical protein	0.046	0.195	0.237	0.227	100.0	89.9	9.8
Syncc9902_0250	hypothetical protein	0.052	0.300	0.172	0.179	100.0	99.6	13.1
Syncc9902_0251	putative glycine betaine transporter, BCCT family	0.012	0.306	0.038	0.038	100.0	100.0	14.4
Syncc9902_0252	putative sarcosine oxidase	0.015	0.191	0.078	0.070	100.0	89.0	8.8
Syncc9902_0253	Glutamate dehydrogenase/leucine dehydrogenase-like	0.009	0.197	0.045	0.044	100.0	80.3	7.4
Syncc9902_0256	Ham1-like protein	0.022	0.119	0.184	0.198	100.0	80.2	6.3
Syncc9902_0257	phosphoglucomutase/phosphomannomutase family protein	0.005	0.091	0.052	0.054	97.9	7.3	2.2
Syncc9902_0258	hypothetical protein	0.019	0.152	0.128	0.128	96.2	51.7	4.5

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_0259	hypothetical protein	0.039	0.195	0.199	0.183	100.0	48.5	5.5
Syncc9902_0260	Orotate phosphoribosyl transferase (EC:2.4.2.10)	0.027	0.170	0.161	0.160	100.0	70.1	5.5
Syncc9902_0261	hypothetical protein	0.021	0.167	0.127	0.126	100.0	31.0	4.4
Syncc9902_0262	hypothetical protein	0.011	0.174	0.066	0.063	100.0	81.5	8.1
Syncc9902_0263	probable oxidoreductase	0.002	0.171	0.012	0.012	100.0	87.1	8.7
Syncc9902_0264	putative phosphorylase kinase	0.020	0.164	0.124	0.128	100.0	62.1	5.4
Syncc9902_0265	Glycosyl transferase WecB/TagA/CpsF	0.019	0.167	0.114	0.118	100.0	70.5	7.4
Syncc9902_0266	photosystem II reaction center protein PsbK precursor	0.009	0.135	0.069	0.063	100.0	55.2	6.2
Syncc9902_0267	Queuine tRNA-ribosyltransferase (EC:2.4.2.29)	0.009	0.161	0.055	0.052	100.0	75.4	6.5
Syncc9902_0268	Cobalamin-5-phosphate synthase CobS	0.027	0.169	0.157	0.139	100.0	61.6	5.4
Syncc9902_0269	histidine kinase	0.021	0.172	0.121	0.125	100.0	82.5	6.7
Syncc9902_0270	hypothetical protein	0.005	0.258	0.019	0.019	100.0	98.9	10.5
Syncc9902_0271	probable esterase	0.040	0.167	0.242	0.231	100.0	96.7	7.8
Syncc9902_0272	AICAR transformylase/IMP cyclohydrolase PurH (EC:2.1.2.3)	0.011	0.161	0.070	0.069	100.0	100.0	9.9
Syncc9902_0273	hypothetical protein	0.004	0.141	0.028	0.027	100.0	75.6	7.0
Syncc9902_0274	hypothetical protein	0.021	0.127	0.166	0.174	100.0	97.5	9.4
Syncc9902_0275	4-hydroxy-3-methylbut-2-enyl diphosphate reductase (EC:1.17.1.2)	0.008	0.198	0.040	0.040	100.0	100.0	11.3
Syncc9902_0276	Ammonium transporter	0.025	0.244	0.104	0.103	100.0	96.3	11.7
Syncc9902_0277	Sugar fermentation stimulation protein	0.031	0.170	0.185	0.187	100.0	100.0	10.1
Syncc9902_0278	Virulence factor MVIN-like	0.024	0.193	0.125	0.117	100.0	95.2	9.4
Syncc9902_0279	citrate synthase (EC:2.3.3.1)	0.011	0.230	0.049	0.048	100.0	100.0	14.2
Syncc9902_0280	NADH dehydrogenase subunit H	0.006	0.248	0.025	0.028	100.0	100.0	18.8
Syncc9902_0281	NADH dehydrogenase subunit I	0.010	0.245	0.041	0.041	100.0	100.0	19.9
Syncc9902_0282	NADH dehydrogenase subunit J	0.009	0.200	0.043	0.043	100.0	100.0	18.3
Syncc9902_0283	NADH dehydrogenase kappa subunit	0.007	0.151	0.046	0.046	100.0	100.0	15.0
Syncc9902_0284	inorganic polyphosphate/ATP-NAD kinase (EC:2.7.1.23)	0.015	0.226	0.067	0.064	100.0	100.0	14.9
Syncc9902_0285	hypothetical protein	0.052	0.281	0.184	0.190	100.0	100.0	17.7
Syncc9902_0286	hypothetical protein	0.066	0.221	0.298	0.331	100.0	100.0	20.0
Syncc9902_0287	transcriptional regulator, LuxR family	0.008	0.172	0.047	0.048	100.0	100.0	20.6
Syncc9902_0288	hypothetical protein	0.028	0.290	0.095	0.094	100.0	100.0	15.9
Syncc9902_0289	putative sugar-phosphate nucleotidyl transferase	0.008	0.328	0.025	0.026	100.0	100.0	13.8
Syncc9902_0290	putative segregation and condensation protein A	0.016	0.196	0.082	0.076	100.0	96.8	10.1
Syncc9902_0291	hypothetical protein	0.020	0.196	0.104	0.104	100.0	95.1	9.0
Syncc9902_0292	NAD(P)H-quinone oxidoreductase subunit 4 (EC:1.6.99.5)	0.007	0.176	0.039	0.041	100.0	90.8	11.6
Syncc9902_0293	NADH dehydrogenase subunit L (EC:1.6.99.5)	0.010	0.232	0.043	0.042	100.0	91.1	9.8
Syncc9902_0294	hypothetical protein	0.022	0.197	0.110	0.106	100.0	93.9	8.4
Syncc9902_0295	putative Rubisco transcriptional regulator	0.010	0.216	0.046	0.046	100.0	100.0	10.7
Syncc9902_0296	hypothetical protein	0.014	0.191	0.072	0.072	100.0	99.2	8.1
Syncc9902_0297	NADH dehydrogenase I subunit M	0.019	0.211	0.089	0.109	100.0	100.0	9.5
Syncc9902_0298	Carotene 7,8-desaturase (EC:1.14.99.30)	0.009	0.265	0.033	0.033	100.0	98.7	10.0
Syncc9902_0299	phytoene synthase	0.009	0.285	0.032	0.036	100.0	87.9	8.4
Syncc9902_0300	RNA-binding region RNP-1	0.005	0.255	0.020	0.019	100.0	84.5	10.0
Syncc9902_0301	poly A polymerase family	0.044	0.193	0.230	0.245	100.0	77.3	7.3
Syncc9902_0301a	hypothetical protein	0.023	0.132	0.176	0.184	100.0	87.3	9.6
Syncc9902_0302	hypothetical protein	0.049	0.213	0.228	0.241	100.0	100.0	10.6
Syncc9902_0303	hypothetical protein	0.018	0.153	0.118	0.141	100.0	80.7	6.9
Syncc9902_0304	1-acyl-sn-glycerol-3-phosphate acyltransferase	0.022	0.199	0.111	0.115	100.0	91.4	9.3
Syncc9902_0305	Malonyl CoA-acyl carrier protein transacylase	0.030	0.204	0.148	0.140	100.0	100.0	12.9
Syncc9902_0306	3-oxoacyl-(acyl carrier protein) synthase (EC:2.3.1.41)	0.022	0.172	0.129	0.125	100.0	73.6	7.7
Syncc9902_0307	fatty acid/phospholipid synthesis protein	0.012	0.151	0.080	0.085	100.0	95.0	8.6
Syncc9902_0308	two component transcriptional regulator, winged helix family	0.002	0.113	0.016	0.016	100.0	70.0	6.2
Syncc9902_0309	DNA repair protein RadA	0.010	0.215	0.047	0.047	100.0	70.1	6.0
Syncc9902_0310	photosystem I assembly protein Ycf3	0.003	0.180	0.018	0.017	100.0	100.0	13.9
Syncc9902_0311	Copper-translocating P-type ATPase	0.021	0.205	0.103	0.102	100.0	99.7	12.0
Syncc9902_0312	Thymidylate kinase (EC:2.7.4.9)	0.034	0.240	0.142	0.149	100.0	100.0	15.7
Syncc9902_0313	DNA polymerase III subunit delta	0.040	0.298	0.136	0.137	100.0	96.9	19.2
Syncc9902_0314	two component transcriptional regulator, winged helix family	0.003	0.365	0.008	0.011	100.0	100.0	22.9
Syncc9902_0315	ATPase	0.059	0.309	0.193	0.198	100.0	100.0	26.3
Syncc9902_0317	Photosystem II reaction centre protein PsbD/D2	0.002	0.169	0.010	0.010	100.0	100.0	17.9
Syncc9902_0318	hypothetical protein	0.058	0.285	0.204	0.203	100.0	100.0	16.9
Syncc9902_0319	hypothetical protein	0.066	0.296	0.224	0.223	100.0	100.0	16.6
Syncc9902_0320	hypothetical protein	0.057	0.254	0.223	0.228	100.0	100.0	20.7
Syncc9902_0321	hypothetical protein	0.059	0.227	0.262	0.274	100.0	100.0	16.2
Syncc9902_0322	hypothetical protein	0.039	0.199	0.194	0.198	100.0	100.0	13.8
Syncc9902_0323	possible porin	0.037	0.203	0.184	0.193	99.8	51.0	11.3
Syncc9902_0324	possible porin	0.052	0.216	0.240	0.235	94.9	47.0	8.3
Syncc9902_0325	Cysteine synthase K/M/A	0.011	0.191	0.058	0.068	100.0	47.7	4.1

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_0326	Heat shock protein DnaJ-like	0.017	0.152	0.110	0.103	100.0	36.9	4.3
Syncc9902_0327	hypothetical protein	0.028	0.153	0.184	0.182	100.0	72.0	6.4
Syncc9902_0328	hypothetical protein	0.014	0.118	0.121	0.107	100.0	63.5	4.9
Syncc9902_0329	putative Na+/H+ antiporter, CPA1 family	0.000	0.000	NaN	NaN	87.7	0.0	1.7
Syncc9902_0330	hypothetical protein	0.041	0.186	0.222	0.202	100.0	48.7	4.7
Syncc9902_0331	hypothetical protein	0.032	0.198	0.160	0.160	100.0	65.9	6.9
Syncc9902_0332	hypothetical protein	0.032	0.145	0.218	0.210	100.0	77.5	7.2
Syncc9902_0333	hypothetical protein	0.000	0.087	0.000	0.000	100.0	7.4	3.1
Syncc9902_0334	HesB/YadR/YfhF	0.017	0.145	0.114	0.108	100.0	85.0	5.5
Syncc9902_0335	Carotene 7,8-desaturase (EC:1.14.99.30)	0.007	0.210	0.035	0.036	100.0	70.6	6.9
Syncc9902_0336	hypothetical protein	0.007	0.131	0.053	0.070	100.0	60.1	6.4
Syncc9902_0337	hypothetical protein	0.079	0.211	0.376	0.362	100.0	42.3	4.2
Syncc9902_0338	Uroporphyrinogen-III synthase (EC:4.2.1.75)	0.016	0.148	0.107	0.095	100.0	89.4	6.6
Syncc9902_0339	TPR repeat	0.022	0.150	0.148	0.154	100.0	72.7	5.3
Syncc9902_0340	putative beta-glucosidase	0.033	0.198	0.167	0.167	99.9	68.3	6.6
Syncc9902_0341	Ribosome-binding factor A	0.001	0.179	0.007	0.007	100.0	87.7	6.1
Syncc9902_0342	hypothetical protein	0.008	0.142	0.059	0.056	100.0	79.7	6.3
Syncc9902_0343	putative glutathione S-transferase	0.021	0.205	0.103	0.104	100.0	73.4	5.8
Syncc9902_0344	hypothetical protein	0.027	0.137	0.199	0.198	100.0	82.7	7.7
Syncc9902_0345	hypothetical protein	0.004	0.152	0.024	0.038	100.0	93.6	6.2
Syncc9902_0346	shikimate kinase (EC:2.7.1.71)	0.030	0.145	0.207	0.203	98.1	74.7	6.3
Syncc9902_0347	6-pyruvoyltetrahydropterin synthase (EC:4.2.3.12)	0.013	0.176	0.072	0.069	100.0	74.2	7.5
Syncc9902_0348	putative riboflavin-specific deaminase	0.004	0.014	0.302	0.302	100.0	15.6	3.3
Syncc9902_0349	hypothetical protein	0.029	0.112	0.261	0.266	100.0	57.6	4.2
Syncc9902_0350	hypothetical protein	0.027	0.289	0.094	0.109	99.5	89.1	13.1
Syncc9902_0351	hypothetical protein	0.036	0.226	0.159	0.165	100.0	100.0	13.8
Syncc9902_0351a	cytochrome b6f complex subunit PetL	0.022	0.169	0.127	0.121	100.0	100.0	14.9
Syncc9902_0352	putative multidrug efflux MFS transporter	0.032	0.287	0.110	0.116	100.0	100.0	17.3
Syncc9902_0353	hypothetical protein	0.020	0.248	0.081	0.080	100.0	100.0	17.3
Syncc9902_0354	Thioredoxin domain 2	0.024	0.227	0.107	0.099	100.0	97.2	10.4
Syncc9902_0355	L-aspartate oxidase (EC:1.4.3.16)	0.014	0.179	0.077	0.075	100.0	100.0	12.9
Syncc9902_0356	photosystem II complex extrinsic protein precursor PsuB	0.013	0.188	0.067	0.060	100.0	100.0	17.2
Syncc9902_0357	hypothetical protein	0.038	0.327	0.115	0.122	100.0	100.0	13.1
Syncc9902_0358	undecaprenyl pyrophosphate phosphatase (EC:2.7.1.66)	0.035	0.321	0.108	0.108	100.0	100.0	14.7
Syncc9902_0359	possible Fe-S oxidoreductase	0.028	0.263	0.108	0.109	100.0	89.4	10.0
Syncc9902_0360	putative RND family outer membrane efflux protein	0.045	0.298	0.150	0.148	100.0	76.5	6.5
Syncc9902_0363	serum resistance locus BrkB-like	0.070	0.391	0.179	0.178	100.0	100.0	18.8
Syncc9902_0364	hypothetical protein	0.074	0.310	0.237	0.236	100.0	100.0	13.0
Syncc9902_0365	hypothetical protein	0.044	0.388	0.113	0.124	100.0	100.0	12.3
Syncc9902_0366	exodeoxyribonuclease VII small subunit	0.049	0.321	0.153	0.151	100.0	100.0	12.5
Syncc9902_0367	Exonuclease VII, large subunit (EC:3.1.11.6)	0.036	0.267	0.134	0.145	100.0	94.9	13.0
Syncc9902_0368	possible high light inducible protein	0.023	0.301	0.076	0.079	100.0	100.0	20.9
Syncc9902_0369	ATP-dependent helicase HrpB	0.033	0.218	0.152	0.159	100.0	95.9	11.7
Syncc9902_0370	hypothetical protein	0.006	0.218	0.028	0.028	100.0	100.0	12.2
Syncc9902_0371	hypothetical protein	0.021	0.101	0.213	0.207	100.0	100.0	9.4
Syncc9902_0372	PDZ/DHR/GLGF	0.021	0.178	0.118	0.117	100.0	86.1	6.7
Syncc9902_0373	ATPase	0.014	0.172	0.081	0.084	100.0	71.8	6.5
Syncc9902_0374	hypothetical protein	0.005	0.027	0.179	0.171	100.0	43.4	3.4
Syncc9902_0375	hypothetical protein	0.000	0.000	NaN	NaN	97.0	0.0	2.6
Syncc9902_0376	hypothetical protein	0.032	0.373	0.086	0.083	100.0	28.0	3.9
Syncc9902_0377	hypothetical protein	0.011	0.129	0.087	0.087	79.3	28.6	2.6
Syncc9902_0378	conserved hypothetical protein, transport associated	0.035	0.139	0.253	0.233	88.5	44.9	4.2
Syncc9902_0379	putative potassium channel, VIC family	0.011	0.159	0.067	0.061	100.0	71.2	6.0
Syncc9902_0380	possible sodium transporter, trk family	0.006	0.165	0.035	0.036	98.1	41.9	5.0
Syncc9902_0381	putative sodium/sulfate transporter, DASS family	0.007	0.177	0.038	0.044	95.6	27.9	3.5
Syncc9902_0382	putative GTP-binding protein, transporter associated	0.009	0.171	0.050	0.049	98.9	56.5	4.9
Syncc9902_0383	putative NADH dehydrogenase, transport associated	0.023	0.171	0.136	0.136	100.0	79.2	5.8
Syncc9902_0384	phosphoadenosine phosphosulfate reductase (EC:1.8.4.8)	0.025	0.198	0.125	0.127	100.0	77.4	9.6
Syncc9902_0385	putative transcriptional acitvator, Baf	0.046	0.177	0.260	0.253	100.0	85.1	6.5
Syncc9902_0386	putative bacterioferritin comigratory (BCP) protein	0.014	0.261	0.055	0.059	100.0	100.0	10.7
Syncc9902_0387	possible protein phosphatase 2C	0.038	0.208	0.183	0.181	100.0	98.6	13.9
Syncc9902_0388	ATPase	0.029	0.244	0.119	0.118	96.6	87.3	12.5
Syncc9902_0389	Secretion protein HlyD	0.019	0.237	0.078	0.077	100.0	99.4	9.7
Syncc9902_0390	Hydrophobe/amphiphile efflux-1 HAE1	0.013	0.304	0.041	0.041	100.0	100.0	14.8
Syncc9902_0391	hypothetical protein	0.083	0.291	0.284	0.284	100.0	100.0	19.2
Syncc9902_0392	phosphoglucomutase	0.032	0.285	0.111	0.114	100.0	100.0	18.5
Syncc9902_0393	hypothetical protein	0.050	0.283	0.177	0.181	100.0	100.0	18.5

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_0394	hypothetical protein	0.110	0.358	0.308	0.319	100.0	100.0	18.5
Syncc9902_0395	hypothetical protein	0.000	0.000	NaN	NaN	0.1	0.0	0.0
Syncc9902_0396	Phospholipase D/Transphosphatidylase	0.060	0.274	0.219	0.236	100.0	98.4	11.7
Syncc9902_0397	hypothetical protein	0.029	0.245	0.119	0.103	100.0	100.0	11.1
Syncc9902_0398	hypothetical protein	0.030	0.319	0.093	0.097	100.0	100.0	17.4
Syncc9902_0399	phycobilisome rod-core linker polypeptide cpcG	0.016	0.235	0.067	0.069	100.0	100.0	16.6
Syncc9902_0400	hypothetical protein	0.027	0.275	0.100	0.102	100.0	100.0	9.9
Syncc9902_0401	hypothetical protein	0.036	0.279	0.131	0.130	100.0	81.9	7.8
Syncc9902_0402	putative transcriptional regulator	0.024	0.209	0.113	0.119	100.0	97.5	11.6
Syncc9902_0403	ferredoxin thioredoxin reductase, catalytic beta chain	0.017	0.257	0.065	0.062	100.0	100.0	8.6
Syncc9902_0404	FeS assembly protein SufB	0.011	0.311	0.036	0.038	100.0	96.3	9.9
Syncc9902_0405	FeS assembly ATPase SufC	0.006	0.190	0.033	0.034	100.0	64.9	6.2
Syncc9902_0406	ABC transporter, membrane component	0.019	0.162	0.120	0.123	100.0	63.8	5.1
Syncc9902_0407	Cysteine desulphurases, SufS	0.019	0.191	0.099	0.106	100.0	51.8	4.7
Syncc9902_0408	putative peptidase	0.034	0.164	0.207	0.211	99.9	50.0	5.1
Syncc9902_0409	peptide deformylase (EC:3.5.1.88)	0.012	0.145	0.082	0.080	100.0	38.3	3.8
Syncc9902_0410	hypothetical protein	0.007	0.179	0.040	0.048	100.0	100.0	7.1
Syncc9902_0411	hypothetical protein	0.029	0.234	0.124	0.125	98.0	60.7	6.4
Syncc9902_0412	Mg chelatase-related protein	0.040	0.195	0.207	0.216	100.0	76.7	6.6
Syncc9902_0413	HIT (histidine triad) family protein	0.012	0.114	0.106	0.098	100.0	90.4	7.6
Syncc9902_0414	ATPase	0.016	0.261	0.062	0.065	100.0	92.6	9.8
Syncc9902_0415	possible high light inducible protein	0.012	0.212	0.056	0.054	100.0	39.5	4.2
Syncc9902_0416	hypothetical protein	0.023	0.196	0.120	0.139	100.0	87.1	7.7
Syncc9902_0417	hypothetical protein	0.033	0.198	0.166	0.166	100.0	100.0	10.2
Syncc9902_0418	hypothetical protein	0.039	0.248	0.159	0.165	100.0	99.6	9.5
Syncc9902_0419	photosystem II PsbX protein	0.000	0.075	0.000	0.000	100.0	100.0	7.0
Syncc9902_0420	YGGT family, conserved hypothetical integral membrane protein	0.020	0.211	0.096	0.093	100.0	100.0	15.6
Syncc9902_0421	acetyl-CoA carboxylase	0.009	0.188	0.045	0.047	100.0	99.8	11.1
Syncc9902_0422	hypothetical protein	0.017	0.208	0.083	0.081	100.0	99.3	8.9
Syncc9902_0423	Methionine sulfoxide reductase B (EC:1.8.4.6)	0.016	0.263	0.061	0.061	100.0	91.5	8.6
Syncc9902_0424	hypothetical protein	0.025	0.201	0.124	0.119	100.0	85.1	9.0
Syncc9902_0425	Chromosome segregation protein SMC	0.011	0.183	0.062	0.062	100.0	84.6	7.4
Syncc9902_0426	hypothetical protein	0.044	0.165	0.269	0.278	100.0	75.0	5.8
Syncc9902_0427	hypothetical protein	0.009	0.143	0.062	0.061	100.0	74.2	9.3
Syncc9902_0428	hypothetical protein	0.011	0.222	0.050	0.042	100.0	64.9	6.9
Syncc9902_0429	putative RNA methylase	0.027	0.237	0.112	0.106	100.0	95.3	9.4
Syncc9902_0430	Site-specific recombinase XerD-like	0.000	0.000	NaN	NaN	98.7	0.0	1.4
Syncc9902_0431	hypothetical protein	0.070	0.371	0.188	0.193	100.0	100.0	13.7
Syncc9902_0432	hypothetical protein	0.000	0.000	NaN	NaN	85.2	1.4	1.9
Syncc9902_0433	hypothetical protein	0.000	0.000	NaN	NaN	36.3	0.0	0.4
Syncc9902_0436	Band 7 protein	0.021	0.441	0.048	0.048	100.0	100.0	16.0
Syncc9902_0437	hypothetical protein	0.059	0.145	0.406	0.382	100.0	100.0	11.5
Syncc9902_0438	hypothetical protein	0.047	0.281	0.166	0.174	100.0	100.0	16.5
Syncc9902_0439	plastoquinol terminal oxidase	0.033	0.229	0.144	0.143	100.0	100.0	18.0
Syncc9902_0440	hypothetical protein	0.091	0.319	0.286	0.297	100.0	100.0	46.8
Syncc9902_0441	hypothetical protein	0.069	0.111	0.619	0.657	100.0	100.0	27.1
Syncc9902_0442	hypothetical protein	0.116	0.463	0.250	0.252	100.0	100.0	28.8
Syncc9902_0443	hypothetical protein	0.052	0.189	0.277	0.287	100.0	100.0	15.7
Syncc9902_0444	hypothetical protein	0.041	0.276	0.148	0.147	100.0	100.0	25.4
Syncc9902_0445	hypothetical protein	0.071	0.198	0.358	0.345	100.0	100.0	13.8
Syncc9902_0446	hypothetical protein	0.089	0.090	0.980	0.957	100.0	100.0	16.9
Syncc9902_0447	nuclease (SNase-like)	0.052	0.239	0.219	0.227	100.0	100.0	24.9
Syncc9902_0448	DEAD/DEAH box helicase-like	0.059	0.406	0.144	0.144	100.0	100.0	58.3
Syncc9902_0449	hypothetical protein	0.094	0.400	0.235	0.243	100.0	100.0	44.0
Syncc9902_0450	hypothetical protein	0.226	0.207	1.093	1.059	100.0	100.0	60.2
Syncc9902_0451	hypothetical protein	0.124	0.480	0.259	0.246	100.0	100.0	66.7
Syncc9902_0452	hypothetical protein	0.069	0.505	0.136	0.136	100.0	100.0	41.7
Syncc9902_0453	hypothetical protein	0.090	0.440	0.205	0.205	100.0	100.0	21.2
Syncc9902_0454	hypothetical protein	0.069	0.438	0.157	0.167	100.0	97.9	12.9
Syncc9902_0455	putative sulfotransferase protein	0.099	0.393	0.251	0.261	100.0	100.0	46.8
Syncc9902_0456	hypothetical protein	0.074	0.574	0.128	0.137	100.0	100.0	52.8
Syncc9902_0457	hypothetical protein	0.064	0.259	0.249	0.246	100.0	100.0	11.2
Syncc9902_0458	N-acylneuraminate-9-phosphate synthase (EC:2.5.1.57)	0.046	0.434	0.106	0.102	100.0	100.0	20.4
Syncc9902_0459	hypothetical protein	0.056	0.313	0.180	0.178	100.0	100.0	15.7
Syncc9902_0460	hypothetical protein	0.086	0.287	0.299	0.310	100.0	98.9	14.7
Syncc9902_0461	hypothetical protein	0.043	0.294	0.148	0.148	100.0	100.0	13.6
Syncc9902_0462	putative acylneuraminate cytidylyltransferase	0.042	0.232	0.182	0.187	100.0	100.0	10.4

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_0463	possible polysaccharide export protein	0.047	0.272	0.173	0.180	100.0	99.6	11.8
Syncc9902_0464	capsule polysaccharide protein KpsS-like	0.030	0.246	0.120	0.122	99.7	81.6	10.1
Syncc9902_0465	putative 3-oxoacyl-(acyl-carrier-protein) synthase I	0.038	0.280	0.137	0.138	100.0	97.1	8.7
Syncc9902_0466	possible acyl carrier protein	0.018	0.238	0.077	0.081	100.0	100.0	9.2
Syncc9902_0467	hypothetical protein	0.033	0.314	0.104	0.118	100.0	70.2	7.6
Syncc9902_0468	polysaccharide export outer membrane protein	0.063	0.267	0.238	0.256	100.0	83.4	7.6
Syncc9902_0469	adenylosuccinate lyase	0.017	0.251	0.066	0.062	100.0	99.0	9.0
Syncc9902_0470	nitrogen regulatory protein P-II	0.005	0.289	0.019	0.019	100.0	100.0	11.7
Syncc9902_0471	hypothetical protein	0.024	0.313	0.075	0.069	100.0	100.0	6.8
Syncc9902_0472	putative C-type cytochrome biogenesis protein Ccs1	0.026	0.288	0.089	0.093	100.0	75.0	6.1
Syncc9902_0473	putative c-type cytochrome biogenesis protein CcdA	0.007	0.216	0.032	0.031	100.0	39.4	4.4
Syncc9902_0474	cell division protein FtsW	0.005	0.183	0.028	0.026	100.0	61.5	4.9
Syncc9902_0475	hypothetical protein	0.015	0.133	0.113	0.114	100.0	77.6	5.5
Syncc9902_0476	Hemolysin A	0.014	0.130	0.109	0.114	100.0	59.1	4.7
Syncc9902_0477	linker polypeptide, allophycocyanin-associated	0.000	0.076	0.000	0.000	100.0	100.0	11.0
Syncc9902_0478	Allophycocyanin, beta subunit	0.002	0.129	0.015	0.014	100.0	100.0	9.2
Syncc9902_0479	allophycocyanin alpha chain	0.001	0.089	0.012	0.012	100.0	91.2	12.2
Syncc9902_0480	anchor polypeptide LCM	0.016	0.236	0.070	0.071	100.0	84.5	8.3
Syncc9902_0481	hypothetical protein	0.030	0.239	0.125	0.126	100.0	84.1	7.9
Syncc9902_0482	possible ATP synthase protein 1	0.014	0.070	0.202	0.168	100.0	100.0	11.1
Syncc9902_0483	ATP synthase subunit A	0.004	0.133	0.034	0.034	100.0	99.9	9.2
Syncc9902_0484	ATP synthase subunit C	0.000	0.087	0.000	0.000	100.0	73.2	4.9
Syncc9902_0485	ATP synthase subunit B	0.022	0.150	0.147	0.146	100.0	45.2	3.7
Syncc9902_0486	ATP synthase subunit B	0.008	0.245	0.034	0.039	94.0	40.4	5.6
Syncc9902_0487	H+-transporting two-sector ATPase, delta (OSCP) subunit	0.004	0.172	0.023	0.022	100.0	65.4	6.3
Syncc9902_0488	ATP synthase subunit A (EC:3.6.3.15)	0.003	0.151	0.022	0.018	100.0	63.6	5.3
Syncc9902_0489	ATP synthase subunit C (EC:3.6.3.15)	0.006	0.264	0.022	0.023	96.5	16.6	2.9
Syncc9902_0490	hypothetical protein	0.000	0.000	NaN	NaN	84.1	0.0	2.0
Syncc9902_0491	Ferredoxin (2Fe-2S)	0.029	0.200	0.143	0.143	100.0	5.3	2.6
Syncc9902_0492	hypothetical protein	0.012	0.082	0.144	0.142	100.0	63.5	4.6
Syncc9902_0493	hypothetical protein	0.010	0.151	0.067	0.073	100.0	39.4	4.0
Syncc9902_0494	hypothetical protein	0.089	0.276	0.321	0.300	78.8	50.1	3.4
Syncc9902_0495	NAD+ synthase (EC:6.3.5.1)	0.028	0.139	0.201	0.190	100.0	76.3	6.1
Syncc9902_0496	nicotinic acid mononucleotide adenyltransferase	0.033	0.218	0.150	0.149	100.0	85.9	8.2
Syncc9902_0497	possible GTPase	0.020	0.191	0.106	0.108	100.0	91.8	7.5
Syncc9902_0498	hypothetical protein	0.072	0.253	0.285	0.257	100.0	52.2	4.5
Syncc9902_0499	hypothetical protein	0.010	0.197	0.052	0.056	100.0	88.5	6.9
Syncc9902_0500	putative aminopeptidase P	0.015	0.186	0.078	0.080	100.0	62.4	4.8
Syncc9902_0501	HAD-superfamily subfamily IA	0.036	0.195	0.186	0.213	100.0	77.5	6.8
Syncc9902_0502	hypothetical protein	0.010	0.079	0.127	0.117	100.0	59.1	4.6
Syncc9902_0503	ATP synthase subunit epsilon	0.006	0.162	0.036	0.036	100.0	77.4	10.9
Syncc9902_0504	ATP synthase subunit B (EC:3.6.3.15)	0.009	0.272	0.033	0.034	100.0	100.0	11.6
Syncc9902_0505	co-chaperonin GroES	0.007	0.237	0.031	0.031	100.0	100.0	12.4
Syncc9902_0506	chaperonin GroEL	0.008	0.310	0.027	0.027	100.0	100.0	15.0
Syncc9902_0507	hypothetical protein	0.055	0.269	0.206	0.207	100.0	62.9	6.4
Syncc9902_0508	hypothetical protein	0.045	0.239	0.187	0.189	100.0	96.0	8.3
Syncc9902_0509	hypothetical protein	0.077	0.351	0.219	0.218	100.0	100.0	21.9
Syncc9902_0510	hypothetical protein	0.085	0.429	0.199	0.200	100.0	94.9	14.6
Syncc9902_0511	hypothetical protein	0.058	0.297	0.196	0.195	100.0	85.6	14.2
Syncc9902_0512	protein-export membrane protein	0.020	0.342	0.057	0.064	100.0	100.0	24.8
Syncc9902_0513	phosphoglyceromutase (EC:5.4.2.1)	0.029	0.279	0.105	0.104	100.0	100.0	23.6
Syncc9902_0514	pyrimidine regulatory protein PyrR	0.022	0.328	0.067	0.066	100.0	100.0	16.8
Syncc9902_0515	putative ferredoxin-thioredoxin reductase, variable chain	0.046	0.286	0.160	0.198	100.0	100.0	22.9
Syncc9902_0516	hypothetical protein	0.042	0.192	0.218	0.210	100.0	96.8	12.5
Syncc9902_0517	putative DnaK-type molecular chaperone (HSP70 family)	0.033	0.218	0.153	0.148	97.5	95.8	12.2
Syncc9902_0518	hypothetical protein	0.026	0.243	0.107	0.143	100.0	19.3	4.1
Syncc9902_0519	DNA-directed RNA polymerase omega subunit	0.019	0.236	0.079	0.078	100.0	90.4	9.0
Syncc9902_0520	hypothetical protein	0.053	0.241	0.218	0.229	100.0	100.0	9.1
Syncc9902_0521	hypothetical protein	0.015	0.243	0.061	0.068	100.0	100.0	14.1
Syncc9902_0522	hypothetical protein	0.066	0.236	0.280	0.282	100.0	92.6	9.8
Syncc9902_0523	hypothetical protein	0.027	0.197	0.136	0.140	100.0	90.0	11.6
Syncc9902_0524	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase	0.052	0.229	0.229	0.231	100.0	100.0	10.8
Syncc9902_0525	RNA-binding S4	0.034	0.246	0.137	0.135	100.0	84.0	8.0
Syncc9902_0526	hypothetical protein	0.019	0.180	0.106	0.107	100.0	95.8	8.8
Syncc9902_0527	D-3-phosphoglycerate dehydrogenase	0.007	0.210	0.033	0.036	100.0	82.2	7.7
Syncc9902_0528	Ribosomal protein L11 methyltransferase	0.025	0.252	0.100	0.107	100.0	100.0	10.8
Syncc9902_0529	Ferredoxin (2Fe-2S)	0.003	0.154	0.021	0.020	100.0	92.0	7.6

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_0530	hypothetical protein	0.028	0.303	0.092	0.105	100.0	83.9	9.0
Syncc9902_0531	cytochrome c-550	0.006	0.189	0.032	0.037	100.0	88.1	12.2
Syncc9902_0532	Ribonuclease Z (EC:3.1.26.11)	0.013	0.183	0.071	0.071	100.0	100.0	10.0
Syncc9902_0533	stage II sporulation protein D-like	0.040	0.192	0.210	0.209	100.0	100.0	12.7
Syncc9902_0534	putative N-acetylglucosamine-6-phosphate isomerase	0.044	0.212	0.209	0.202	100.0	100.0	12.7
Syncc9902_0535	hypothetical protein	0.053	0.212	0.248	0.234	100.0	100.0	19.8
Syncc9902_0536	hypothetical protein	0.008	0.194	0.040	0.040	100.0	100.0	15.2
Syncc9902_0537	hypothetical protein	0.016	0.400	0.039	0.039	100.0	100.0	15.4
Syncc9902_0538	hypothetical protein	0.007	0.352	0.020	0.020	100.0	100.0	14.1
Syncc9902_0539	hypothetical protein	0.027	0.265	0.103	0.106	100.0	100.0	15.4
Syncc9902_0540	hypothetical protein	0.091	0.216	0.421	0.416	100.0	93.1	6.3
Syncc9902_0541	hypothetical protein	0.000	0.150	0.000	0.000	75.8	13.5	2.7
Syncc9902_0542	Protein of unknown function DUF28	0.028	0.242	0.114	0.120	100.0	100.0	14.5
Syncc9902_0543	tRNA pseudouridine synthase B (EC:4.2.1.70)	0.043	0.218	0.195	0.194	100.0	99.6	11.9
Syncc9902_0544	hypothetical protein	0.024	0.219	0.110	0.104	100.0	100.0	10.6
Syncc9902_0545	50S ribosomal protein L27	0.002	0.202	0.009	0.009	100.0	100.0	10.4
Syncc9902_0546	Ribosomal protein L21	0.010	0.201	0.049	0.048	100.0	100.0	15.5
Syncc9902_0547	possible circadian clock protein KaiA	0.028	0.263	0.108	0.108	100.0	100.0	14.6
Syncc9902_0548	circadian clock protein KaiB	0.033	0.231	0.144	0.142	100.0	100.0	8.5
Syncc9902_0549	circadian clock protein KaiC	0.004	0.267	0.013	0.014	100.0	99.7	10.6
Syncc9902_0550	multi-sensor signal transduction histidine kinase	0.011	0.243	0.044	0.041	97.9	50.7	6.0
Syncc9902_0551	Phosphoribosylglycinamide synthetase (EC:6.3.4.13)	0.016	0.171	0.096	0.091	100.0	46.3	4.8
Syncc9902_0552	hypothetical protein	0.034	0.161	0.210	0.223	100.0	69.0	6.4
Syncc9902_0553	phosphoribosylaminoimidazole-succinocarboxamide synthase (EC:6.3.2.6)	0.024	0.177	0.135	0.151	100.0	87.1	6.9
Syncc9902_0554	chloroplast outer envelope membrane protein-like	0.026	0.264	0.097	0.096	100.0	90.1	7.4
Syncc9902_0555	UDP-3-0-acyl N-acetylglucosamine deacetylase	0.027	0.159	0.170	0.165	100.0	59.1	4.6
Syncc9902_0556	(3R)-hydroxymyristoyl ACP dehydratase	0.011	0.192	0.060	0.065	100.0	97.8	6.9
Syncc9902_0557	UDP-N-acetylglucosamine acyltransferase (EC:2.3.1.129)	0.021	0.194	0.111	0.106	100.0	90.6	8.0
Syncc9902_0558	lipid-A-disaccharide synthase (EC:2.4.1.182)	0.021	0.160	0.130	0.130	100.0	80.8	7.4
Syncc9902_0559	Methionine sulfoxide reductase A (EC:1.8.4.6)	0.029	0.199	0.144	0.144	100.0	80.4	9.0
Syncc9902_0560	hypothetical protein	0.004	0.175	0.024	0.046	100.0	73.4	6.0
Syncc9902_0561	leucyl aminopeptidase (EC:3.4.11.1)	0.020	0.251	0.080	0.080	91.2	85.0	7.3
Syncc9902_0562	hypothetical protein	0.030	0.216	0.138	0.128	100.0	63.1	5.7
Syncc9902_0563	hypothetical protein	0.007	0.338	0.021	0.020	100.0	11.7	3.2
Syncc9902_0564	hypothetical protein	0.041	0.116	0.350	0.445	100.0	53.0	4.4
Syncc9902_0565	transcriptional regulator, LuxR family	0.028	0.208	0.136	0.133	100.0	45.7	4.4
Syncc9902_0566	hypothetical protein	0.013	0.259	0.052	0.050	100.0	100.0	10.3
Syncc9902_0567	tyrosyl-tRNA synthetase (EC:6.1.1.1)	0.013	0.205	0.065	0.066	100.0	90.9	9.4
Syncc9902_0568	regucalcin family protein	0.020	0.135	0.149	0.152	94.1	10.8	2.1
Syncc9902_0569	hypothetical protein	0.000	0.000	NaN	NaN	71.0	0.0	1.0
Syncc9902_0570	Orotidine 5'-phosphate decarboxylase (EC:4.1.1.23)	0.049	0.376	0.129	0.137	100.0	78.3	9.3
Syncc9902_0571	Outer membrane autotransporter barrel	0.033	0.183	0.182	0.201	84.1	50.6	4.8
Syncc9902_0572	Protein of unknown function DUF205	0.030	0.228	0.132	0.128	100.0	83.4	7.2
Syncc9902_0573	hypothetical protein	0.022	0.207	0.104	0.106	100.0	75.6	6.6
Syncc9902_0574	hypothetical protein	0.033	0.212	0.154	0.159	100.0	61.6	7.6
Syncc9902_0575	Protein of unknown function DUF140	0.015	0.211	0.073	0.071	100.0	69.5	6.7
Syncc9902_0576	putative GPH family sugar transporter	0.021	0.197	0.107	0.115	99.8	82.0	9.2
Syncc9902_0577	Alpha amylase, catalytic subdomain	0.024	0.260	0.092	0.094	100.0	99.0	10.3
Syncc9902_0578	hypothetical protein	0.031	0.164	0.186	0.158	100.0	96.8	11.0
Syncc9902_0579	Histone-like DNA-binding protein	0.006	0.097	0.057	0.057	100.0	90.2	6.9
Syncc9902_0580	glutamyl-tRNA synthetase-like	0.060	0.251	0.239	0.255	100.0	77.4	7.8
Syncc9902_0581	hypothetical protein	0.030	0.197	0.153	0.157	100.0	61.8	4.6
Syncc9902_0582	hypothetical protein	0.069	0.303	0.227	0.217	100.0	100.0	9.6
Syncc9902_0583	hypothetical protein	0.029	0.222	0.128	0.127	100.0	22.2	3.6
Syncc9902_0584	succinate dehydrogenase	0.028	0.270	0.105	0.105	100.0	97.5	12.6
Syncc9902_0585	succinate dehydrogenase (EC:1.3.99.1)	0.028	0.298	0.093	0.090	100.0	100.0	21.8
Syncc9902_0586	succinate dehydrogenase cytochrome b-556 subunit	0.047	0.229	0.205	0.242	100.0	100.0	18.9
Syncc9902_0587	hypothetical protein	0.055	0.296	0.185	0.177	100.0	100.0	31.3
Syncc9902_0588	hypothetical protein	0.066	0.316	0.209	0.207	100.0	100.0	26.6
Syncc9902_0589	hypothetical protein	0.027	0.339	0.081	0.081	100.0	100.0	29.0
Syncc9902_0590	hypothetical protein	0.099	0.287	0.344	0.319	100.0	100.0	21.8
Syncc9902_0591	hypothetical protein	0.093	0.206	0.450	0.455	100.0	89.3	12.4
Syncc9902_0592	translation initiation factor IF-2	0.017	0.207	0.080	0.080	100.0	92.0	12.9
Syncc9902_0593	hypothetical protein	0.008	0.077	0.109	0.109	100.0	100.0	10.5
Syncc9902_0594	transcription elongation factor NusA	0.012	0.264	0.047	0.052	100.0	95.5	9.6
Syncc9902_0595	hypothetical protein	0.029	0.198	0.147	0.151	100.0	100.0	15.0
Syncc9902_0596	hypothetical protein	0.053	0.339	0.158	0.142	100.0	76.3	7.5

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_0597	putative arylsulfatase regulatory protein	0.024	0.228	0.106	0.119	100.0	55.9	5.8
Syncc9902_0598	hypothetical protein	0.010	0.190	0.054	0.058	100.0	91.3	5.6
Syncc9902_0599	extracellular solute-binding protein, family 3	0.024	0.235	0.101	0.097	100.0	89.7	11.3
Syncc9902_0600	PDZ/DHR/GLGF	0.011	0.211	0.053	0.053	100.0	68.5	6.7
Syncc9902_0601	Ribose 5-phosphate isomerase (EC:5.3.1.6)	0.021	0.270	0.078	0.075	100.0	100.0	10.0
Syncc9902_0602	Histidinol dehydrogenase (EC:1.1.1.23)	0.024	0.193	0.123	0.128	100.0	99.3	12.0
Syncc9902_0603	30S ribosomal protein S20	0.017	0.230	0.074	0.073	100.0	100.0	11.4
Syncc9902_0604	TatD-related deoxyribonuclease	0.033	0.246	0.133	0.133	100.0	100.0	11.9
Syncc9902_0605	DNA-directed RNA polymerase beta subunit (EC:2.7.7.6)	0.004	0.243	0.017	0.020	100.0	96.5	13.6
Syncc9902_0606	DNA-directed RNA polymerase gamma chain (EC:2.7.7.6)	0.007	0.236	0.030	0.029	100.0	100.0	11.5
Syncc9902_0607	DNA-directed RNA polymerase beta' subunit (EC:2.7.7.6)	0.011	0.292	0.037	0.037	100.0	100.0	20.5
Syncc9902_0608	putative high light inducible protein	0.015	0.272	0.055	0.065	100.0	100.0	19.0
Syncc9902_0609	hypothetical protein	0.028	0.323	0.085	0.084	100.0	100.0	16.3
Syncc9902_0610	sodium:solute symporter family, possibly glucose transporter	0.017	0.291	0.060	0.060	100.0	100.0	19.9
Syncc9902_0611	hypothetical protein	0.028	0.179	0.157	0.156	100.0	100.0	13.0
Syncc9902_0612	PBS lyase HEAT-like repeat	0.032	0.342	0.093	0.096	100.0	100.0	18.6
Syncc9902_0613	hypothetical protein	0.028	0.118	0.237	0.232	100.0	100.0	10.5
Syncc9902_0614	hypothetical protein	0.022	0.345	0.065	0.065	100.0	100.0	12.7
Syncc9902_0615	possible 3Fe-4S ferredoxin	0.022	0.304	0.071	0.074	100.0	100.0	10.0
Syncc9902_0616	hypothetical protein	0.022	0.196	0.114	0.113	100.0	87.7	9.8
Syncc9902_0617	Glucose inhibited division protein	0.027	0.150	0.179	0.176	100.0	74.9	6.6
Syncc9902_0618	Heat shock protein DnaJ-like	0.026	0.100	0.261	0.235	100.0	99.1	8.2
Syncc9902_0619	hypothetical protein	0.065	0.381	0.169	0.169	100.0	14.1	3.8
Syncc9902_0620	putative diaminopelargonic acid synthase	0.008	0.125	0.065	0.063	98.9	71.1	6.1
Syncc9902_0621	Dethiobiotin synthase (EC:6.3.3.3)	0.019	0.125	0.155	0.171	100.0	68.9	6.5
Syncc9902_0622	biotin biosynthesis protein BioC	0.054	0.114	0.469	0.530	100.0	31.5	4.1
Syncc9902_0623	hypothetical protein	0.043	0.175	0.244	0.239	100.0	35.4	4.3
Syncc9902_0624	8-amino-7-oxononanoate synthase (EC:2.3.1.47)	0.026	0.142	0.184	0.190	100.0	71.6	5.6
Syncc9902_0625	sodium/bile acid cotransporter family	0.019	0.153	0.123	0.134	100.0	80.1	6.5
Syncc9902_0626	hypothetical protein	0.020	0.230	0.088	0.104	100.0	100.0	12.4
Syncc9902_0627	DEAD/DEAH box helicase-like	0.012	0.191	0.065	0.066	100.0	100.0	12.7
Syncc9902_0628	fumarate hydratase	0.041	0.289	0.142	0.140	100.0	100.0	23.9
Syncc9902_0629	hypothetical protein	0.080	0.329	0.243	0.242	100.0	100.0	16.5
Syncc9902_0630	hypothetical protein	0.056	0.273	0.206	0.207	100.0	80.2	8.4
Syncc9902_0631	possible ABC transporter involved in polysaccharide efflux	0.052	0.508	0.102	0.105	100.0	100.0	22.3
Syncc9902_0632	possible ABC transporter involved in polysaccharide efflux	0.062	0.363	0.171	0.180	100.0	100.0	11.8
Syncc9902_0633	Lipopolysaccharide biosynthesis protein-like	0.102	0.541	0.189	0.194	100.0	94.5	26.3
Syncc9902_0634	dTDP-glucose 4,6-dehydratase	0.065	0.471	0.138	0.140	100.0	100.0	27.2
Syncc9902_0635	dTDP-4-dehydrorhamnose reductase (EC:1.1.1.133)	0.066	0.406	0.162	0.170	100.0	100.0	25.1
Syncc9902_0636	dTDP-4-dehydrorhamnose 3,5-epimerase related (EC:5.1.3.13)	0.096	0.453	0.211	0.215	100.0	100.0	32.3
Syncc9902_0637	Glucose-1-phosphate thymidylyltransferase, long form	0.047	0.365	0.128	0.125	100.0	100.0	27.5
Syncc9902_0638	possible glycosyltransferase	0.069	0.289	0.238	0.234	100.0	98.2	14.3
Syncc9902_0639	hypothetical protein	0.046	0.151	0.303	0.328	100.0	38.5	5.4
Syncc9902_0640	Lipopolysaccharide biosynthesis protein-like	0.047	0.170	0.278	0.294	100.0	90.3	10.8
Syncc9902_0641	hypothetical protein	0.034	0.174	0.196	0.196	100.0	69.9	6.2
Syncc9902_0642	Dihydrouridine synthase TIM-barrel protein nifR3	0.048	0.316	0.153	0.150	100.0	99.8	17.2
Syncc9902_0643	hypothetical protein	0.035	0.287	0.123	0.112	100.0	100.0	16.1
Syncc9902_0644	GTP-binding protein EngA	0.008	0.291	0.027	0.024	100.0	100.0	19.3
Syncc9902_0645	possible cobalt transport protein	0.034	0.280	0.122	0.122	100.0	100.0	20.8
Syncc9902_0646	hypothetical protein	0.007	0.261	0.026	0.033	100.0	100.0	24.2
Syncc9902_0647	hypothetical protein	0.034	0.237	0.145	0.129	100.0	100.0	12.2
Syncc9902_0648	hypothetical protein	0.015	0.249	0.061	0.054	100.0	100.0	13.7
Syncc9902_0649	Delta 1-pyrroline-5-carboxylate reductase (EC:1.5.1.2)	0.031	0.199	0.156	0.154	100.0	100.0	10.8
Syncc9902_0650	Precorrin-2 C20-methyltransferase	0.029	0.184	0.155	0.176	100.0	95.5	10.6
Syncc9902_0651	methionine salvage pathway enzyme E-2/E-2' -like	0.023	0.180	0.126	0.140	100.0	98.8	7.4
Syncc9902_0652	hypothetical protein	0.080	0.215	0.372	0.398	100.0	89.4	8.8
Syncc9902_0653	hypothetical protein	0.041	0.178	0.227	0.238	100.0	80.5	7.4
Syncc9902_0654	probable glycosyltransferase	0.023	0.202	0.113	0.108	100.0	100.0	11.6
Syncc9902_0655	putative multidrug efflux transporter, MFS family	0.019	0.209	0.089	0.087	100.0	98.6	10.3
Syncc9902_0656	Recombination protein O, RecO	0.011	0.188	0.056	0.061	100.0	100.0	8.2
Syncc9902_0657	Deoxyribose-phosphate aldolase (EC:4.1.2.4)	0.016	0.143	0.112	0.106	100.0	94.0	8.9
Syncc9902_0658	putative sigma 54 modulation protein/ribosomal protein S30EA	0.011	0.132	0.081	0.093	100.0	86.6	5.8
Syncc9902_0659	Lipoate-protein ligase B	0.034	0.153	0.220	0.216	100.0	100.0	9.1
Syncc9902_0660	putative long-chain-fatty-acid--CoA ligase	0.017	0.162	0.105	0.110	100.0	83.4	9.3
Syncc9902_0661	hypothetical protein	0.013	0.150	0.090	0.089	100.0	100.0	9.8
Syncc9902_0662	dihydrolipoamide acetyltransferase	0.009	0.150	0.063	0.062	100.0	87.4	8.0

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_0663	queuosine biosynthesis protein	0.032	0.202	0.160	0.163	100.0	87.4	8.0
Syncc9902_0664	Cysteine synthase K/M/A	0.007	0.204	0.034	0.036	100.0	99.2	8.5
Syncc9902_0665	possible cystathionine gamma-synthase	0.036	0.160	0.224	0.211	96.3	83.5	7.0
Syncc9902_0666	Cystathionine gamma-synthase (EC:2.5.1.48)	0.019	0.169	0.114	0.114	93.7	22.9	3.4
Syncc9902_0667	Photosynthetic II protein PsbC	0.005	0.166	0.030	0.032	100.0	98.9	9.9
Syncc9902_0668	Photosystem II reaction centre protein PsbD/D2	0.002	0.082	0.025	0.025	100.0	98.7	12.4
Syncc9902_0669	photosystem I assembly protein Ycf4	0.011	0.261	0.042	0.037	100.0	96.5	10.0
Syncc9902_0670	probable peptidyl-prolyl cis-trans isomerase, cyclophilin type	0.023	0.251	0.090	0.096	100.0	54.0	4.9
Syncc9902_0671	acetolactate synthase III small subunit	0.010	0.218	0.048	0.043	100.0	93.8	8.5
Syncc9902_0672	alpha/beta hydrolase superfamily protein	0.029	0.163	0.179	0.197	100.0	62.5	5.5
Syncc9902_0673	tRNA (guanine-N(2)-)-methyltransferase (EC:2.1.1.32)	0.030	0.130	0.230	0.226	93.8	52.6	4.8
Syncc9902_0674	hypothetical protein	0.026	0.121	0.219	0.207	100.0	100.0	6.7
Syncc9902_0675	putative chaperon-like protein for quinone binding in photosystem II	0.006	0.202	0.029	0.040	100.0	98.3	8.7
Syncc9902_0676	hypothetical protein	0.043	0.400	0.109	0.000	100.0	5.1	3.5
Syncc9902_0677	methyltransferase-like	0.025	0.219	0.113	0.114	100.0	48.7	5.5
Syncc9902_0678	Pseudouridine synthase, Rsu	0.055	0.222	0.249	0.247	100.0	94.4	9.8
Syncc9902_0679	Translation initiation factor IF-1	0.002	0.198	0.009	0.009	100.0	100.0	9.0
Syncc9902_0680	Thioredoxin reductase	0.015	0.247	0.059	0.061	100.0	88.1	8.8
Syncc9902_0681	hypothetical protein	0.053	0.130	0.407	0.405	100.0	42.6	4.2
Syncc9902_0682	hypothetical protein	0.021	0.194	0.109	0.111	100.0	77.9	7.9
Syncc9902_0683	hypothetical protein	0.013	0.149	0.086	0.085	100.0	89.0	5.5
Syncc9902_0684	putative nicotinamide nucleotide transhydrogenase, subunit alpha 1 (A1)	0.010	0.165	0.064	0.066	100.0	74.4	6.6
Syncc9902_0685	putative nicotinamide nucleotide transhydrogenase, subunit alpha 2 (A2)	0.003	0.223	0.016	0.024	100.0	87.9	7.1
Syncc9902_0686	putative nicotinamide nucleotide transhydrogenase, subunit beta	0.003	0.184	0.017	0.021	100.0	51.6	5.4
Syncc9902_0687	hypothetical protein	0.021	0.144	0.145	0.146	100.0	99.6	8.7
Syncc9902_0688	putative ferredoxin like protein	0.021	0.168	0.126	0.142	100.0	95.0	7.2
Syncc9902_0689	1-deoxy-D-xylulose 5-phosphate reductoisomerase (EC:1.1.1.267)	0.011	0.163	0.070	0.076	100.0	93.0	8.2
Syncc9902_0690	putative sodium-dependent transporter, NSS family	0.010	0.151	0.065	0.060	100.0	81.0	7.6
Syncc9902_0691	hypothetical protein	0.033	0.137	0.240	0.243	100.0	87.2	5.7
Syncc9902_0692	Cysteinyl-tRNA synthetase, class Ia (EC:6.1.1.16)	0.026	0.239	0.111	0.111	100.0	95.8	10.2
Syncc9902_0693	DNA polymerase I (EC:2.7.7.7)	0.022	0.222	0.098	0.098	100.0	96.2	9.7
Syncc9902_0694	Secretion protein HlyD	0.036	0.224	0.162	0.167	100.0	100.0	13.6
Syncc9902_0695	hypothetical protein	0.010	0.210	0.046	0.050	100.0	100.0	16.6
Syncc9902_0696	hypothetical protein	0.048	0.242	0.196	0.203	100.0	93.6	9.1
Syncc9902_0697	hypothetical protein	0.053	0.252	0.213	0.226	100.0	100.0	12.8
Syncc9902_0698	hypothetical protein	0.046	0.218	0.213	0.199	100.0	100.0	10.0
Syncc9902_0699	possible multidrug efflux transporter, MFS family	0.029	0.271	0.105	0.096	100.0	100.0	12.9
Syncc9902_0700	hypothetical protein	0.043	0.179	0.243	0.253	100.0	100.0	9.2
Syncc9902_0701	putative ABC transporter, oligopeptides	0.011	0.211	0.052	0.053	100.0	100.0	10.7
Syncc9902_0702	ABC transporter, substrate binding protein, possibly oligopeptides	0.027	0.179	0.149	0.146	100.0	99.0	10.1
Syncc9902_0703	hypothetical protein	0.025	0.158	0.157	0.137	100.0	100.0	13.7
Syncc9902_0704	Homoserine dehydrogenase (EC:1.1.1.3)	0.017	0.200	0.086	0.093	100.0	95.3	12.5
Syncc9902_0705	possible SufE protein	0.035	0.260	0.136	0.137	100.0	100.0	16.6
Syncc9902_0706	hypothetical protein	0.045	0.198	0.225	0.221	100.0	87.3	7.8
Syncc9902_0707	hypothetical protein	0.043	0.134	0.320	0.319	97.6	93.3	8.6
Syncc9902_0708	hypothetical protein	0.003	0.152	0.017	0.016	100.0	94.6	14.5
Syncc9902_0709	hypothetical protein	0.098	0.171	0.572	0.594	100.0	17.0	3.7
Syncc9902_0710	hypothetical protein	0.000	0.143	0.000	0.000	100.0	3.9	3.7
Syncc9902_0711	hypothetical protein	0.022	0.095	0.231	0.231	100.0	46.2	4.5
Syncc9902_0712	Holliday junction resolvase RuvC	0.005	0.160	0.028	0.026	100.0	67.5	5.3
Syncc9902_0713	Magnesium chelatase ATPase subunit I (EC:6.6.1.1)	0.007	0.109	0.064	0.064	100.0	24.6	3.5
Syncc9902_0714	hypothetical protein	0.016	0.143	0.111	0.112	97.7	28.1	3.4
Syncc9902_0715	possible beta-lactamase family protein	0.009	0.146	0.065	0.065	97.9	29.0	3.7
Syncc9902_0716	probable tRNA/rRNA methyltransferase	0.000	0.025	0.000	0.000	95.3	7.0	2.5
Syncc9902_0717	cytochrome cM	0.015	0.118	0.127	0.111	100.0	95.1	7.0
Syncc9902_0717a	cytochrome b6f complex subunit PetG	0.041	0.050	0.811	0.865	100.0	84.6	5.1
Syncc9902_0718	hypothetical protein	0.000	0.119	0.000	0.000	100.0	27.2	3.5
Syncc9902_0719	Imidazole glycerol phosphate synthase, glutamine amidotransferase subunit	0.035	0.135	0.257	0.290	100.0	43.8	5.5
Syncc9902_0720	Thioredoxin	0.000	0.143	0.000	0.000	100.0	38.0	4.0
Syncc9902_0721	inositol-5-monophosphate dehydrogenase	0.000	0.177	0.000	0.000	90.2	19.2	2.9
Syncc9902_0722	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	2.1
Syncc9902_0723	DNA gyrase subunit A	0.009	0.153	0.057	0.062	93.1	26.6	3.4
Syncc9902_0724	Lycopene cyclase, beta and epsilon	0.044	0.299	0.147	0.148	100.0	87.8	9.7
Syncc9902_0725	Outer membrane autotransporter barrel	0.037	0.157	0.237	0.250	84.2	25.2	3.1

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_0726	hypothetical protein	0.023	0.226	0.103	0.104	100.0	100.0	9.6
Syncc9902_0727	2-isopropylmalate synthase	0.015	0.244	0.060	0.058	100.0	74.6	7.1
Syncc9902_0728	putative permease protein of sugar ABC transporter	0.010	0.188	0.056	0.074	100.0	84.5	7.0
Syncc9902_0729	hypothetical protein	0.036	0.172	0.211	0.235	100.0	85.8	7.0
Syncc9902_0730	hypothetical protein	0.021	0.188	0.111	0.116	100.0	85.7	6.5
Syncc9902_0731	hypothetical protein	0.024	0.101	0.238	0.244	100.0	40.7	4.7
Syncc9902_0732	hypothetical protein	0.030	0.143	0.212	0.206	100.0	100.0	7.3
Syncc9902_0733	Protein of unknown function DUF77	0.017	0.167	0.099	0.098	100.0	19.5	3.8
Syncc9902_0734	hypothetical protein	0.026	0.096	0.267	0.266	100.0	38.4	4.6
Syncc9902_0735	putative pseudouridylate synthase	0.026	0.119	0.221	0.255	100.0	57.2	5.5
Syncc9902_0736	metal dependent phosphohydrolase	0.011	0.141	0.078	0.081	99.8	53.1	5.3
Syncc9902_0737	Methylenetetrahydrofolate dehydrogenase (NADP+) (EC:1.5.1.5)	0.006	0.117	0.055	0.070	100.0	47.9	4.8
Syncc9902_0738	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	3.0
Syncc9902_0739	geranylgeranyl pyrophosphate synthase	0.002	0.095	0.017	0.015	100.0	65.3	4.8
Syncc9902_0740	hypothetical protein	0.024	0.147	0.165	0.151	100.0	66.1	5.6
Syncc9902_0741	exodeoxyribonuclease V	0.007	0.116	0.056	0.058	100.0	67.0	5.5
Syncc9902_0742	possible cytochrome c biogenesis protein trans	0.021	0.129	0.163	0.142	100.0	46.0	4.3
Syncc9902_0743	putative phosphoglycerate mutase family protein	0.029	0.151	0.191	0.174	99.8	55.8	5.1
Syncc9902_0744	hypothetical protein	0.029	0.134	0.215	0.222	100.0	90.1	7.6
Syncc9902_0745	possible acylphosphatase	0.010	0.155	0.061	0.045	100.0	69.8	4.7
Syncc9902_0746	Cobyrinic acid a,c-diamide synthase CbiA	0.008	0.106	0.080	0.101	98.3	35.7	3.8
Syncc9902_0747	possible glucose 6-phosphate dehydrogenase effector OpcA	0.012	0.104	0.120	0.122	98.5	15.2	2.8
Syncc9902_0748	Glucose-6-phosphate dehydrogenase (EC:1.1.1.49)	0.007	0.194	0.036	0.038	100.0	64.4	5.7
Syncc9902_0749	ferredoxin--NADP reductase (FNR)	0.024	0.200	0.119	0.128	100.0	96.1	10.9
Syncc9902_0750	hypothetical protein	0.019	0.223	0.085	0.086	100.0	93.1	9.3
Syncc9902_0751	adaptive-response sensory kinase	0.009	0.191	0.047	0.044	100.0	100.0	10.4
Syncc9902_0752	hypothetical protein	0.016	0.223	0.070	0.072	100.0	99.5	9.0
Syncc9902_0753	hypothetical protein	0.011	0.223	0.047	0.053	98.4	59.7	6.6
Syncc9902_0754	Peptidase M, neutral zinc metallopeptidase, zinc-binding site	0.029	0.243	0.121	0.124	100.0	100.0	15.4
Syncc9902_0755	hypothetical protein	0.053	0.177	0.299	0.278	100.0	100.0	12.8
Syncc9902_0756	hypothetical protein	0.061	0.084	0.726	0.636	70.4	54.2	3.4
Syncc9902_0757	hypothetical protein	0.030	0.141	0.211	0.211	100.0	63.2	5.1
Syncc9902_0758	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	2.9
Syncc9902_0759	hypothetical protein	0.043	0.210	0.204	0.190	100.0	94.4	10.9
Syncc9902_0760	hypothetical protein	0.053	0.215	0.246	0.248	100.0	100.0	13.1
Syncc9902_0761	hypothetical protein	0.021	0.206	0.101	0.099	100.0	95.6	9.2
Syncc9902_0762	hypothetical protein	0.015	0.226	0.068	0.066	100.0	100.0	15.7
Syncc9902_0763	hypothetical protein	0.047	0.168	0.276	0.342	100.0	100.0	13.6
Syncc9902_0764	hypothetical protein	0.019	0.187	0.103	0.095	100.0	96.5	10.0
Syncc9902_0765	hypothetical protein	0.011	0.235	0.046	0.046	100.0	90.3	6.9
Syncc9902_0766	Ionotropic glutamate receptor	0.039	0.229	0.171	0.170	100.0	89.5	11.7
Syncc9902_0767	hypothetical protein	0.023	0.134	0.175	0.189	100.0	100.0	8.1
Syncc9902_0768	hypothetical protein	0.032	0.181	0.176	0.172	100.0	96.4	8.4
Syncc9902_0769	Phosphopantetheinyl transferase-like	0.046	0.219	0.210	0.211	100.0	98.5	12.2
Syncc9902_0770	hypothetical protein	0.012	0.193	0.063	0.076	100.0	100.0	17.4
Syncc9902_0771	hypothetical protein	0.095	0.420	0.225	0.323	80.6	40.3	3.6
Syncc9902_0772	hypothetical protein	0.031	0.312	0.099	0.096	100.0	99.8	8.9
Syncc9902_0773	hypothetical protein	0.030	0.238	0.127	0.128	100.0	90.7	7.9
Syncc9902_0774	CBS	0.011	0.187	0.058	0.064	100.0	89.9	7.3
Syncc9902_0775	CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase	0.027	0.172	0.154	0.157	100.0	100.0	9.6
Syncc9902_0776	possible nucleotide sugar epimerase	0.027	0.240	0.113	0.112	100.0	65.7	8.1
Syncc9902_0777	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase	0.013	0.162	0.082	0.085	100.0	86.5	8.2
Syncc9902_0778	putative ferric uptake regulator, FUR family	0.007	0.158	0.047	0.064	100.0	92.8	8.6
Syncc9902_0779	hypothetical protein	0.033	0.177	0.189	0.198	100.0	52.1	4.7
Syncc9902_0780	hypothetical protein	0.021	0.213	0.096	0.096	100.0	60.3	5.0
Syncc9902_0781	hypothetical protein	0.015	0.112	0.134	0.129	100.0	81.8	5.9
Syncc9902_0782	hypothetical protein	0.019	0.165	0.117	0.111	100.0	49.2	4.6
Syncc9902_0783	Holliday junction resolvase YqgF	0.016	0.165	0.097	0.099	94.6	47.1	3.6
Syncc9902_0784	hypothetical protein	0.003	0.152	0.021	0.021	99.5	78.0	5.1
Syncc9902_0785	HAD-superfamily phosphatase subfamily IIIA	0.037	0.216	0.169	0.151	100.0	91.2	8.3
Syncc9902_0786	gamma-glutamyl kinase (EC:2.7.2.11)	0.013	0.151	0.085	0.087	100.0	61.9	5.3
Syncc9902_0787	UDP-3-O-(3-hydroxymyristoyl) glucosamine N-acyltransferase, LpxD	0.008	0.139	0.060	0.060	100.0	69.0	7.4
Syncc9902_0788	3-isopropylmalate dehydrogenase (EC:1.1.1.85)	0.014	0.173	0.081	0.088	100.0	91.9	9.0
Syncc9902_0789	Phosphoribulokinase (EC:2.7.1.19)	0.005	0.194	0.026	0.025	100.0	100.0	10.0
Syncc9902_0790	hypothetical protein	0.046	0.171	0.267	0.284	100.0	96.9	7.4

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_0793	acetyl-CoA carboxylase beta subunit	0.008	0.206	0.038	0.043	100.0	100.0	10.2
Syncc9902_0794	hypothetical protein	0.019	0.222	0.087	0.091	100.0	100.0	9.3
Syncc9902_0795	putative oxidoreductase	0.027	0.197	0.136	0.135	100.0	98.2	9.8
Syncc9902_0796	Fructose-bisphosphate aldolase (EC:4.1.2.13)	0.023	0.192	0.120	0.116	100.0	100.0	17.7
Syncc9902_0797	fructose-bisphosphate aldolase (EC:4.1.2.13)	0.004	0.199	0.019	0.019	100.0	99.3	13.1
Syncc9902_0798	Gamma-glutamyltransferase (EC:2.3.2.2)	0.044	0.209	0.209	0.219	100.0	91.1	10.3
Syncc9902_0799	hypothetical protein	0.041	0.187	0.220	0.215	100.0	98.8	10.1
Syncc9902_0800	Phosphoribosylformylglycinamidine synthase I	0.024	0.233	0.104	0.095	100.0	100.0	12.2
Syncc9902_0801	Phosphoribosylformylglycinamidine synthetase PurS	0.023	0.183	0.126	0.142	100.0	100.0	12.0
Syncc9902_0802	hypothetical protein	0.059	0.180	0.326	0.377	100.0	100.0	9.0
Syncc9902_0803	small mechanosensitive ion channel, MscS family	0.031	0.285	0.109	0.103	100.0	100.0	10.3
Syncc9902_0804	hypothetical protein	0.084	0.311	0.270	0.266	100.0	100.0	22.1
Syncc9902_0805	nuclease (SNase-like)	0.077	0.329	0.234	0.255	100.0	100.0	14.9
Syncc9902_0806	hypothetical protein	0.025	0.344	0.074	0.075	100.0	100.0	35.0
Syncc9902_0807	hypothetical protein	0.094	0.344	0.273	0.293	100.0	100.0	50.1
Syncc9902_0808	hypothetical protein	0.028	0.382	0.074	0.076	100.0	100.0	51.6
Syncc9902_0809	hypothetical protein	0.103	0.135	0.763	0.869	100.0	100.0	11.4
Syncc9902_0810	hypothetical protein	0.074	0.199	0.370	0.326	100.0	100.0	23.4
Syncc9902_0811	hypothetical protein	0.070	0.222	0.316	0.315	100.0	100.0	19.1
Syncc9902_0812	hypothetical protein	0.085	0.287	0.297	0.300	100.0	100.0	13.4
Syncc9902_0813	hypothetical protein	0.094	0.349	0.270	0.257	100.0	100.0	31.8
Syncc9902_0814	hypothetical protein	0.076	0.424	0.179	0.181	100.0	100.0	44.5
Syncc9902_0815	two-component sensor histidine kinase	0.000	0.111	0.000	0.000	100.0	13.0	3.7
Syncc9902_0816	hypothetical protein	0.036	0.271	0.134	0.126	100.0	82.7	12.6
Syncc9902_0817	hypothetical protein	0.019	0.200	0.095	0.097	100.0	100.0	22.7
Syncc9902_0818	hypothetical protein	0.034	0.312	0.109	0.120	100.0	100.0	20.3
Syncc9902_0819	hypothetical protein	0.029	0.259	0.113	0.114	100.0	100.0	19.5
Syncc9902_0820	ADP-heptose:LPS heptosyltransferase-like	0.000	0.000	NaN	NaN	69.7	0.0	1.7
Syncc9902_0821	adenine phosphoribosyltransferase (EC:2.4.2.7)	0.046	0.280	0.164	0.164	100.0	96.0	14.7
Syncc9902_0822	hypothetical protein	0.018	0.296	0.060	0.060	100.0	100.0	13.7
Syncc9902_0823	possible 2-octaprenyl-6-methoxyphenol 4-monoxygenase; UbiH	0.046	0.268	0.173	0.166	100.0	100.0	13.4
Syncc9902_0824	possible high light inducible protein	0.031	0.301	0.105	0.093	100.0	100.0	16.7
Syncc9902_0825	hypothetical protein	0.029	0.199	0.147	0.148	100.0	100.0	10.8
Syncc9902_0826	dihydrodipicolinate reductase (EC:1.3.1.26)	0.026	0.296	0.086	0.083	100.0	100.0	17.5
Syncc9902_0827	cobaltochelatase (EC:6.6.1.2)	0.016	0.296	0.056	0.054	100.0	99.7	17.2
Syncc9902_0828	hypothetical protein	0.029	0.111	0.261	0.288	85.8	31.1	3.5
Syncc9902_0829	Dihydropteroate synthase (EC:2.5.1.15)	0.051	0.386	0.131	0.135	100.0	100.0	29.1
Syncc9902_0830	triosephosphate isomerase (EC:5.3.1.1)	0.016	0.276	0.057	0.055	100.0	100.0	29.3
Syncc9902_0831	hypothetical protein	0.043	0.183	0.234	0.230	100.0	100.0	20.3
Syncc9902_0832	ATPase	0.027	0.283	0.096	0.103	100.0	100.0	26.2
Syncc9902_0833	hypothetical protein	0.032	0.392	0.082	0.082	100.0	100.0	22.5
Syncc9902_0834	Sodium:alanine symporter	0.021	0.250	0.085	0.090	100.0	100.0	10.7
Syncc9902_0835	hypothetical protein	0.021	0.231	0.090	0.093	100.0	100.0	9.8
Syncc9902_0836	Carbamoyl-phosphate synthase, large subunit, glutamine-dependent	0.016	0.206	0.076	0.076	100.0	96.2	11.8
Syncc9902_0837	hypothetical protein	0.001	0.248	0.004	0.004	100.0	70.6	5.7
Syncc9902_0838	Iojap-related protein	0.036	0.184	0.198	0.234	100.0	85.8	6.1
Syncc9902_0839	hypothetical protein	0.025	0.247	0.099	0.095	100.0	100.0	10.4
Syncc9902_0840	hypothetical protein	0.010	0.224	0.043	0.047	100.0	98.8	12.2
Syncc9902_0841	probable oxidoreductase	0.029	0.235	0.125	0.134	100.0	100.0	13.4
Syncc9902_0842	hypothetical protein	0.045	0.200	0.226	0.186	100.0	100.0	16.5
Syncc9902_0843	hypothetical protein	0.020	0.233	0.085	0.090	100.0	90.4	12.3
Syncc9902_0844	hypothetical protein	0.054	0.204	0.263	0.269	100.0	100.0	13.3
Syncc9902_0845	putative sulfate transporter	0.015	0.220	0.070	0.077	100.0	100.0	14.1
Syncc9902_0846	hypothetical protein	0.065	0.223	0.292	0.290	100.0	84.6	5.9
Syncc9902_0847	extracellular solute-binding protein, family 3	0.025	0.236	0.107	0.110	100.0	100.0	13.6
Syncc9902_0848	Amino acid ABC transporter, permease protein, 3-TM region, His/Glu/Gln/Arg/opine	0.024	0.273	0.088	0.095	100.0	100.0	18.4
Syncc9902_0849	Amino acid ABC transporter, permease protein, 3-TM region, His/Glu/Gln/Arg/opine	0.020	0.219	0.089	0.098	100.0	100.0	18.0
Syncc9902_0850	ATPase	0.025	0.306	0.083	0.082	100.0	100.0	19.5
Syncc9902_0851	hypothetical protein	0.051	0.247	0.207	0.172	100.0	100.0	13.6
Syncc9902_0852	possible AEC transporter family	0.041	0.199	0.207	0.205	100.0	100.0	12.6
Syncc9902_0853	possible alkylated DNA repair protein	0.056	0.232	0.240	0.247	100.0	100.0	17.7
Syncc9902_0854	hypothetical protein	0.062	0.202	0.309	0.307	100.0	100.0	13.6
Syncc9902_0855	Homoserine O-succinyltransferase (EC:2.3.1.46)	0.010	0.220	0.047	0.050	100.0	100.0	12.4
Syncc9902_0856	O-acetylhomoserine/O-acetylserine sulfhydrylase (EC:2.5.1.49)	0.009	0.215	0.043	0.041	100.0	96.0	16.7
Syncc9902_0857	hypothetical protein	0.015	0.205	0.075	0.077	100.0	100.0	15.2

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_0858	hypothetical protein	0.040	0.144	0.274	0.275	100.0	96.2	11.5
Syncc9902_0859	hypothetical protein	0.029	0.205	0.141	0.140	100.0	100.0	11.4
Syncc9902_0860	hypothetical protein	0.042	0.230	0.182	0.185	100.0	97.2	10.7
Syncc9902_0861	hypothetical protein	0.018	0.212	0.083	0.082	100.0	72.7	6.2
Syncc9902_0862	putative glutathione S-transferase	0.016	0.186	0.088	0.092	100.0	78.3	6.5
Syncc9902_0863	hypothetical protein	0.023	0.183	0.128	0.137	100.0	74.3	6.9
Syncc9902_0864	cell death suppressor protein Lls1-like	0.016	0.199	0.083	0.083	99.6	75.6	8.3
Syncc9902_0865	Glucosylglycerol 3-phosphatase (EC:3.1.3.69)	0.027	0.197	0.137	0.141	100.0	97.5	10.7
Syncc9902_0866	possible L-asparaginase II	0.026	0.174	0.147	0.149	100.0	100.0	11.1
Syncc9902_0867	hypothetical protein	0.049	0.148	0.327	0.353	100.0	96.3	10.3
Syncc9902_0868	hypothetical protein	0.019	0.191	0.098	0.102	100.0	100.0	15.3
Syncc9902_0869	putative ferric uptake regulator, FUR family	0.038	0.222	0.171	0.184	100.0	91.7	9.7
Syncc9902_0870	hypothetical protein	0.032	0.237	0.135	0.123	100.0	100.0	22.6
Syncc9902_0871	hypothetical protein	0.075	0.205	0.365	0.358	100.0	100.0	14.6
Syncc9902_0872	hypothetical protein	0.063	0.228	0.275	0.295	100.0	100.0	10.7
Syncc9902_0873	hypothetical protein	0.026	0.147	0.174	0.176	100.0	94.8	8.8
Syncc9902_0874	hypothetical protein	0.055	0.227	0.243	0.234	100.0	100.0	10.2
Syncc9902_0875	hypothetical protein	0.013	0.231	0.057	0.071	100.0	100.0	9.0
Syncc9902_0876	hypothetical protein	0.019	0.271	0.069	0.052	100.0	100.0	23.0
Syncc9902_0877	YjgF-like protein	0.021	0.212	0.097	0.112	100.0	100.0	17.3
Syncc9902_0878	ATPase	0.022	0.222	0.099	0.102	100.0	100.0	12.8
Syncc9902_0879	Hydroxyacylglutathione hydrolase (EC:3.1.2.6)	0.041	0.213	0.191	0.180	100.0	100.0	13.5
Syncc9902_0880	ATP phosphoribosyltransferase (EC:2.4.2.17)	0.019	0.207	0.092	0.100	100.0	98.5	7.8
Syncc9902_0881	ATPase	0.012	0.236	0.053	0.055	100.0	98.7	12.9
Syncc9902_0882	possible acetyltransferase	0.011	0.147	0.075	0.088	100.0	100.0	10.6
Syncc9902_0883	hypothetical protein	0.045	0.167	0.269	0.278	100.0	100.0	9.3
Syncc9902_0884	hypothetical protein	0.042	0.168	0.250	0.248	100.0	89.4	6.7
Syncc9902_0885	PDZ/DHR/GLGF	0.010	0.184	0.052	0.049	100.0	100.0	11.3
Syncc9902_0886	chromosomal replication initiation protein	0.007	0.180	0.038	0.039	100.0	78.0	8.6
Syncc9902_0887	hypothetical protein	0.020	0.119	0.166	0.182	100.0	65.3	5.4
Syncc9902_0888	hypothetical protein	0.032	0.193	0.166	0.160	100.0	86.8	7.0
Syncc9902_0889	probable glutathione reductase (NADPH)	0.028	0.208	0.132	0.137	100.0	98.5	8.5
Syncc9902_0890	K+-dependent Na+/Ca+ exchanger related-protein	0.016	0.268	0.058	0.057	100.0	100.0	11.7
Syncc9902_0891	putative sulfate transporter	0.010	0.239	0.041	0.042	100.0	88.2	9.3
Syncc9902_0892	dihydroorotase (EC:3.5.2.3)	0.024	0.208	0.115	0.126	100.0	99.2	10.0
Syncc9902_0893	hypothetical protein	0.004	0.182	0.025	0.016	100.0	100.0	8.0
Syncc9902_0894	possible inorganic carbon transport protein	0.011	0.171	0.066	0.078	100.0	100.0	14.1
Syncc9902_0895	hypothetical protein	0.014	0.184	0.075	0.074	100.0	100.0	8.6
Syncc9902_0896	tryptophan synthase subunit alpha (EC:4.2.1.20)	0.020	0.262	0.078	0.077	100.0	77.0	9.7
Syncc9902_0897	hypothetical protein	0.044	0.275	0.159	0.164	100.0	100.0	9.2
Syncc9902_0898	hypothetical protein	0.035	0.142	0.243	0.225	100.0	100.0	15.4
Syncc9902_0899	cytochrome c6 (soluble cytochrome f) (cytochrome c553)	0.035	0.214	0.163	0.192	100.0	96.2	8.9
Syncc9902_0900	hypothetical protein	0.037	0.179	0.206	0.258	100.0	84.6	8.7
Syncc9902_0901	hypothetical protein	0.032	0.240	0.133	0.133	100.0	100.0	23.1
Syncc9902_0902	hypothetical protein	0.078	0.214	0.363	0.343	100.0	100.0	14.0
Syncc9902_0903	hypothetical protein	0.047	0.301	0.155	0.151	100.0	100.0	17.9
Syncc9902_0904	hypothetical protein	0.053	0.273	0.195	0.227	100.0	100.0	12.9
Syncc9902_0905	possible type II alternative RNA polymerase sigma factor	0.010	0.183	0.054	0.055	100.0	96.3	9.6
Syncc9902_0906	hypothetical protein	0.011	0.248	0.046	0.034	100.0	100.0	12.2
Syncc9902_0907	Peptidase, metallopeptidase	0.050	0.195	0.255	0.249	100.0	95.2	11.6
Syncc9902_0908	hypothetical protein	0.049	0.220	0.223	0.201	99.3	92.6	13.7
Syncc9902_0909	histidine biosynthesis bifunctional protein HisIE	0.031	0.253	0.123	0.125	100.0	100.0	12.6
Syncc9902_0910	hypothetical protein	0.023	0.224	0.104	0.119	100.0	99.1	7.6
Syncc9902_0911	ATPase	0.011	0.226	0.049	0.049	100.0	88.4	9.6
Syncc9902_0912	hypothetical protein	0.037	0.283	0.132	0.131	100.0	100.0	8.8
Syncc9902_0913	hypothetical protein	0.030	0.231	0.128	0.136	100.0	68.2	5.8
Syncc9902_0914	hypothetical protein	0.047	0.304	0.154	0.170	100.0	100.0	10.0
Syncc9902_0915	possible cytochrome c6	0.042	0.257	0.162	0.156	100.0	100.0	9.2
Syncc9902_0916	plastocyanin precursor	0.001	0.146	0.009	0.038	100.0	95.0	6.7
Syncc9902_0917	hypothetical protein	0.030	0.188	0.159	0.154	100.0	88.7	7.0
Syncc9902_0918	uroporphyrinogen decarboxylase (EC:4.1.1.37)	0.009	0.182	0.048	0.043	100.0	58.9	6.2
Syncc9902_0919	glycogen branching enzyme	0.011	0.132	0.081	0.085	99.3	34.9	4.1
Syncc9902_0920	hypothetical protein	0.014	0.148	0.096	0.088	100.0	88.2	6.2
Syncc9902_0921	acyl esterase	0.035	0.170	0.208	0.209	97.0	30.6	3.4
Syncc9902_0922	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	2.4
Syncc9902_0923	hypothetical protein	0.006	0.000	Inf	Inf	100.0	20.3	3.4
Syncc9902_0924	hypothetical protein	0.018	0.154	0.117	0.132	100.0	93.2	7.2

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_0925	hypothetical protein	0.035	0.177	0.199	0.205	94.8	37.2	4.0
Syncc9902_0926	ROK family sugar kinase	0.020	0.170	0.120	0.117	100.0	70.0	5.5
Syncc9902_0927	gamma-glutamyl phosphate reductase (EC:1.2.1.41)	0.022	0.192	0.117	0.113	100.0	97.5	11.1
Syncc9902_0928	Dihydroneopterin aldolase family	0.029	0.198	0.146	0.140	100.0	94.4	10.2
Syncc9902_0929	lipase family protein	0.052	0.193	0.268	0.258	100.0	100.0	13.2
Syncc9902_0930	Oligopeptidase A (EC:3.4.24.70)	0.030	0.235	0.130	0.125	100.0	100.0	17.7
Syncc9902_0931	NAD(P)H-quinone oxidoreductase subunit 4 (EC:1.6.99.5)	0.012	0.320	0.036	0.037	100.0	100.0	23.0
Syncc9902_0932	Homoserine kinase (EC:2.7.1.39)	0.012	0.254	0.046	0.046	100.0	100.0	20.8
Syncc9902_0933	Glucokinase (EC:2.7.1.2)	0.046	0.211	0.220	0.234	100.0	100.0	14.6
Syncc9902_0934	hypothetical protein	0.063	0.265	0.239	0.238	100.0	100.0	14.6
Syncc9902_0935	Threonyl-tRNA synthetase, class IIa	0.022	0.170	0.129	0.128	100.0	56.2	6.4
Syncc9902_0936	hypothetical protein	0.015	0.151	0.101	0.101	100.0	92.2	6.8
Syncc9902_0937	possible muramidase, COG4678	0.058	0.353	0.165	0.176	100.0	100.0	15.5
Syncc9902_0938	tryptophanyl-tRNA synthetase (EC:6.1.1.2)	0.024	0.288	0.082	0.083	100.0	100.0	19.3
Syncc9902_0939	hypothetical protein	0.056	0.277	0.203	0.198	100.0	99.8	18.1
Syncc9902_0940	hypothetical protein	0.042	0.260	0.163	0.183	100.0	100.0	22.6
Syncc9902_0941	Cation efflux protein	0.032	0.273	0.116	0.111	100.0	100.0	16.1
Syncc9902_0942	hypothetical protein	0.033	0.208	0.157	0.156	100.0	100.0	18.7
Syncc9902_0943	Photosystem II reaction centre protein PsbA/D1	0.027	0.261	0.105	0.106	100.0	100.0	21.7
Syncc9902_0944	hypothetical protein	0.026	0.392	0.066	0.078	100.0	100.0	26.5
Syncc9902_0945	hypothetical protein	0.059	0.265	0.223	0.221	100.0	100.0	26.4
Syncc9902_0946	putative redox protein	0.029	0.263	0.110	0.120	100.0	100.0	25.6
Syncc9902_0947	hypothetical protein	0.047	0.201	0.235	0.233	100.0	100.0	14.4
Syncc9902_0948	hypothetical protein	0.047	0.266	0.177	0.205	100.0	100.0	21.7
Syncc9902_0949	possible transcriptional regulator	0.057	0.361	0.157	0.158	100.0	100.0	28.3
Syncc9902_0950	hypothetical protein	0.068	0.316	0.214	0.221	100.0	100.0	22.1
Syncc9902_0951	hypothetical protein	0.028	0.355	0.080	0.075	100.0	100.0	41.8
Syncc9902_0952	hypothetical protein	0.033	0.166	0.201	0.203	100.0	100.0	26.4
Syncc9902_0953	hypothetical protein	0.051	0.306	0.167	0.175	100.0	100.0	37.9
Syncc9902_0954	hypothetical protein	0.041	0.337	0.122	0.135	100.0	100.0	33.8
Syncc9902_0955	hypothetical protein	0.041	0.179	0.231	0.225	100.0	100.0	18.5
Syncc9902_0956	hypothetical protein	0.036	0.230	0.155	0.151	100.0	100.0	15.8
Syncc9902_0957	membrane protein	0.027	0.318	0.084	0.085	100.0	100.0	13.2
Syncc9902_0958	esterase/lipase/thioesterase family active site	0.026	0.107	0.241	0.242	100.0	100.0	10.4
Syncc9902_0959	hypothetical protein	0.009	0.173	0.052	0.050	100.0	100.0	14.5
Syncc9902_0960	hypothetical protein	0.067	0.172	0.393	0.357	100.0	100.0	21.0
Syncc9902_0961	cytochrome P450 enzyme	0.040	0.294	0.137	0.138	100.0	100.0	19.8
Syncc9902_0962	hypothetical protein	0.039	0.234	0.167	0.172	100.0	100.0	17.7
Syncc9902_0963	Ribosomal protein S21	0.003	0.258	0.010	0.010	100.0	100.0	26.2
Syncc9902_0964	hypothetical protein	0.072	0.326	0.221	0.202	100.0	100.0	29.5
Syncc9902_0965	hypothetical protein	0.054	0.370	0.147	0.141	100.0	100.0	21.4
Syncc9902_0966	hypothetical protein	0.034	0.368	0.093	0.092	100.0	100.0	24.4
Syncc9902_0967	hypothetical protein	0.075	0.281	0.268	0.282	100.0	100.0	14.4
Syncc9902_0968	hypothetical protein	0.096	0.288	0.335	0.367	100.0	100.0	25.5
Syncc9902_0969	hypothetical protein	0.034	0.331	0.102	0.104	100.0	100.0	29.0
Syncc9902_0970	hypothetical protein	0.057	0.349	0.164	0.174	100.0	100.0	24.1
Syncc9902_0971	hypothetical protein	0.051	0.389	0.131	0.120	100.0	100.0	40.5
Syncc9902_0972	possible beta-carotene ketolase	0.087	0.373	0.234	0.233	100.0	100.0	29.7
Syncc9902_0973	carotenoid binding protein	0.019	0.523	0.035	0.035	100.0	100.0	50.0
Syncc9902_0974	hypothetical protein	0.026	0.153	0.173	0.173	100.0	100.0	17.6
Syncc9902_0975	short-chain dehydrogenase/reductase (SDR) superfamily	0.064	0.350	0.184	0.181	100.0	100.0	37.0
Syncc9902_0976	Heat shock protein DnaJ-like	0.040	0.173	0.233	0.233	100.0	100.0	22.2
Syncc9902_0977	hypothetical protein	0.073	0.353	0.208	0.203	100.0	100.0	41.1
Syncc9902_0978	hypothetical protein	0.045	0.397	0.112	0.111	100.0	100.0	72.6
Syncc9902_0979	transcriptional regulator, LuxR family	0.045	0.405	0.110	0.109	100.0	100.0	41.6
Syncc9902_0980	RNA-binding region RNP-1	0.020	0.293	0.069	0.078	100.0	100.0	54.1
Syncc9902_0981	hypothetical protein	0.069	0.396	0.175	0.169	100.0	100.0	29.5
Syncc9902_0982	putative superoxide dismutase	0.089	0.385	0.230	0.226	100.0	100.0	32.9
Syncc9902_0983	hypothetical protein	0.016	0.395	0.040	0.044	100.0	100.0	42.4
Syncc9902_0984	hypothetical protein	0.054	0.208	0.259	0.257	100.0	100.0	16.2
Syncc9902_0985	hypothetical protein	0.041	0.297	0.138	0.135	100.0	100.0	25.4
Syncc9902_0986	hypothetical protein	0.061	0.282	0.215	0.215	100.0	100.0	20.3
Syncc9902_0987	hypothetical protein	0.025	0.261	0.094	0.094	100.0	100.0	20.4
Syncc9902_0988	hypothetical protein	0.064	0.275	0.231	0.244	100.0	100.0	30.4
Syncc9902_0989	hypothetical protein	0.056	0.277	0.202	0.208	100.0	100.0	26.5
Syncc9902_0990	hypothetical protein	0.083	0.260	0.318	0.361	100.0	100.0	33.1
Syncc9902_0991	putative glycosyltransferase family 2	0.052	0.275	0.190	0.194	100.0	100.0	30.2

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_0992	hypothetical protein	0.043	0.175	0.243	0.241	100.0	100.0	9.9
Syncc9902_0993	hypothetical protein	0.077	0.087	0.890	0.897	100.0	35.5	3.9
Syncc9902_0994	hypothetical protein	0.030	0.094	0.321	0.254	100.0	41.0	4.3
Syncc9902_0995	putative hydrolase transmembrane protein	0.076	0.263	0.290	0.298	100.0	100.0	21.0
Syncc9902_0996	hypothetical protein	0.054	0.213	0.254	0.249	100.0	100.0	25.8
Syncc9902_0997	hypothetical protein	0.072	0.433	0.167	0.167	100.0	100.0	30.4
Syncc9902_0998	hypothetical protein	0.050	0.164	0.305	0.297	100.0	100.0	18.5
Syncc9902_0999	Ferredoxin (2Fe-2S)	0.011	0.201	0.053	0.053	100.0	100.0	27.7
Syncc9902_1000	possible high light inducible protein	0.023	0.355	0.065	0.065	100.0	100.0	18.6
Syncc9902_1003	possible phycobilisome rod-core linker polypeptide	0.078	0.385	0.203	0.209	100.0	100.0	47.8
Syncc9902_1004	hypothetical protein	0.087	0.439	0.199	0.203	100.0	100.0	69.4
Syncc9902_1005	chlorophyll a/b binding light harvesting protein PcbD	0.025	0.290	0.085	0.086	100.0	100.0	68.2
Syncc9902_1006	hypothetical protein	0.093	0.136	0.686	0.685	100.0	100.0	11.4
Syncc9902_1007	possible phosphoribulokinase/uridine kinase family protein	0.070	0.220	0.317	0.324	100.0	100.0	24.1
Syncc9902_1008	Ferredoxin (2Fe-2S)	0.046	0.261	0.175	0.168	100.0	100.0	16.7
Syncc9902_1009	Endonuclease III/Nth	0.040	0.292	0.139	0.129	100.0	100.0	18.6
Syncc9902_1010	ABC transporter, substrate binding protein, possibly Mn	0.051	0.265	0.192	0.196	100.0	100.0	24.0
Syncc9902_1011	ATPase	0.021	0.374	0.056	0.059	100.0	100.0	34.7
Syncc9902_1012	Phospholipid/glycerol acyltransferase	0.027	0.225	0.119	0.119	100.0	100.0	23.0
Syncc9902_1013	ABC transporter component, possibly Mn transport	0.016	0.223	0.071	0.068	100.0	100.0	21.7
Syncc9902_1014	hypothetical protein	0.038	0.350	0.109	0.110	100.0	100.0	22.3
Syncc9902_1015	hypothetical protein	0.031	0.327	0.095	0.111	100.0	100.0	12.0
Syncc9902_1016	ATPase	0.060	0.308	0.194	0.201	100.0	100.0	19.9
Syncc9902_1017	NUDIX family protein	0.040	0.268	0.149	0.151	100.0	100.0	16.4
Syncc9902_1018	hypothetical protein	0.086	0.237	0.361	0.370	100.0	96.5	14.3
Syncc9902_1019	hypothetical protein	0.068	0.250	0.273	0.282	100.0	100.0	29.0
Syncc9902_1020	hypothetical protein	0.010	0.271	0.036	0.036	100.0	100.0	34.9
Syncc9902_1021	hypothetical protein	0.044	0.273	0.161	0.167	100.0	100.0	16.7
Syncc9902_1022	hypothetical protein	0.067	0.377	0.177	0.179	100.0	100.0	24.0
Syncc9902_1023	hypothetical protein	0.048	0.311	0.153	0.155	100.0	100.0	15.8
Syncc9902_1024	hypothetical protein	0.084	0.331	0.253	0.250	100.0	100.0	31.5
Syncc9902_1025	hypothetical protein	0.047	0.386	0.121	0.121	100.0	100.0	24.9
Syncc9902_1026	hypothetical protein	0.054	0.022	2.430	1.973	100.0	100.0	12.6
Syncc9902_1027	hypothetical protein	0.058	0.166	0.350	0.371	100.0	100.0	28.4
Syncc9902_1028	hypothetical protein	0.089	0.353	0.253	0.262	100.0	100.0	30.1
Syncc9902_1029	hypothetical protein	0.076	0.298	0.255	0.255	100.0	100.0	17.1
Syncc9902_1030	possible high light inducible protein	0.045	0.385	0.117	0.112	100.0	100.0	40.2
Syncc9902_1031	hypothetical protein	0.051	0.505	0.100	0.101	100.0	100.0	39.4
Syncc9902_1032	hypothetical protein	0.087	0.206	0.421	0.424	100.0	100.0	26.6
Syncc9902_1033	hypothetical protein	0.060	0.439	0.137	0.140	100.0	100.0	52.2
Syncc9902_1034	hypothetical protein	0.076	0.333	0.228	0.225	100.0	100.0	21.1
Syncc9902_1035	hypothetical protein	0.105	0.255	0.411	0.403	100.0	100.0	29.0
Syncc9902_1036	hypothetical protein	0.094	0.387	0.243	0.251	100.0	100.0	35.1
Syncc9902_1037	hypothetical protein	0.037	0.219	0.170	0.210	100.0	100.0	13.7
Syncc9902_1038	hypothetical protein	0.069	0.291	0.239	0.236	100.0	100.0	18.7
Syncc9902_1039	DEAD/DEAH box helicase-like	0.050	0.398	0.125	0.128	100.0	100.0	26.7
Syncc9902_1040	DNA ligase	0.060	0.347	0.174	0.173	100.0	100.0	17.8
Syncc9902_1041	hypothetical protein	0.037	0.179	0.206	0.221	100.0	100.0	14.2
Syncc9902_1042	exonuclease	0.032	0.342	0.093	0.091	100.0	100.0	15.1
Syncc9902_1043	hypothetical protein	0.044	0.336	0.130	0.131	100.0	86.8	6.9
Syncc9902_1044	quinolinate synthetase	0.038	0.374	0.102	0.098	100.0	100.0	21.1
Syncc9902_1045	putative transcripton factor	0.036	0.268	0.133	0.128	100.0	100.0	15.2
Syncc9902_1046	hypothetical protein	0.056	0.265	0.211	0.213	100.0	100.0	14.4
Syncc9902_1047	possible muramidase, COG4678	0.031	0.371	0.083	0.089	100.0	100.0	16.4
Syncc9902_1048	hypothetical protein	0.009	0.243	0.039	0.043	100.0	100.0	18.6
Syncc9902_1049	hypothetical protein	0.051	0.206	0.249	0.230	100.0	100.0	13.7
Syncc9902_1050	3-dehydroquinate synthase	0.037	0.236	0.156	0.155	100.0	100.0	12.5
Syncc9902_1051	lactose transport system permease protein	0.022	0.200	0.108	0.103	100.0	100.0	13.9
Syncc9902_1052	Phosphoribosylaminoimidazole carboxylase, ATPase subunit (EC:4.1.1.21)	0.024	0.226	0.108	0.113	100.0	100.0	16.0
Syncc9902_1053	hypothetical protein	0.046	0.382	0.119	0.115	100.0	100.0	14.0
Syncc9902_1054	hypothetical protein	0.060	0.274	0.219	0.206	100.0	100.0	16.2
Syncc9902_1055	ATP-dependent Clp protease adaptor protein ClpS	0.022	0.212	0.103	0.094	100.0	100.0	14.9
Syncc9902_1056	cytochrome b6-f complex subunit VIII	0.024	0.125	0.196	0.195	100.0	100.0	17.9
Syncc9902_1057	hypothetical protein	0.078	0.199	0.392	0.419	99.9	84.4	8.5
Syncc9902_1058	hypothetical protein	0.034	0.266	0.128	0.120	100.0	90.6	14.2
Syncc9902_1059	ATP-dependent Clp protease proteolytic subunit (EC:3.4.21.92)	0.013	0.322	0.041	0.044	100.0	92.6	16.7

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_1060	hypothetical protein	0.041	0.180	0.226	0.226	100.0	93.3	8.1
Syncc9902_1061	Peptidase M41, FtsH (EC:3.6.4.6)	0.010	0.340	0.030	0.032	100.0	95.7	15.3
Syncc9902_1062	possible ABC transporter	0.007	0.230	0.030	0.031	100.0	93.7	13.6
Syncc9902_1063	pyruvate kinase (EC:2.7.1.40)	0.014	0.201	0.068	0.072	100.0	92.4	10.1
Syncc9902_1064	hypothetical protein	0.016	0.129	0.121	0.121	100.0	100.0	9.7
Syncc9902_1065	hypothetical protein	0.027	0.248	0.109	0.104	100.0	85.8	9.7
Syncc9902_1066	hypothetical protein	0.016	0.308	0.051	0.038	100.0	100.0	14.4
Syncc9902_1067	putative transcriptional regulator	0.019	0.211	0.091	0.088	100.0	100.0	10.0
Syncc9902_1068	threonine dehydratase	0.012	0.167	0.074	0.071	100.0	71.7	5.7
Syncc9902_1069	1-deoxy-D-xylulose-5-phosphate synthase	0.014	0.215	0.067	0.062	100.0	81.6	7.5
Syncc9902_1070	hypothetical protein	0.039	0.205	0.190	0.199	98.8	54.2	4.4
Syncc9902_1071	possible photosystem I reaction center subunit X	0.000	0.210	0.000	0.033	100.0	28.2	3.7
Syncc9902_1072	hypothetical protein	0.014	0.231	0.062	0.063	100.0	100.0	9.6
Syncc9902_1073	hypothetical protein	0.004	0.153	0.028	0.029	100.0	78.3	5.5
Syncc9902_1074	hypothetical protein	0.041	0.181	0.229	0.232	100.0	91.3	7.7
Syncc9902_1075	ATPase	0.032	0.197	0.164	0.173	100.0	95.1	10.6
Syncc9902_1076	ABC transporter, likely for trehalose/maltose, membrane component	0.019	0.190	0.098	0.090	100.0	100.0	9.9
Syncc9902_1077	ABC transporter possibly for trehalose/maltose, membrane component	0.025	0.253	0.098	0.107	100.0	100.0	14.8
Syncc9902_1078	ABC transporter, likely sugar solute binding protein	0.043	0.233	0.184	0.184	100.0	100.0	18.5
Syncc9902_1079	Glucosylglycerol-phosphate synthase (EC:2.4.1.213)	0.022	0.330	0.066	0.065	100.0	100.0	18.1
Syncc9902_1080	alkyl hydroperoxide reductase/thiol specific antioxidant/Mal allergens family protein	0.028	0.311	0.091	0.112	100.0	100.0	20.9
Syncc9902_1081	50S ribosomal protein L28	0.002	0.247	0.008	0.008	100.0	100.0	19.7
Syncc9902_1082	heat shock protein 90	0.014	0.298	0.048	0.047	100.0	100.0	16.4
Syncc9902_1083	hypothetical protein	0.070	0.327	0.215	0.225	100.0	100.0	15.0
Syncc9902_1084	ferredoxin	0.020	0.243	0.083	0.084	100.0	100.0	14.7
Syncc9902_1085	Histidine--tRNA ligase (EC:6.1.1.21)	0.031	0.233	0.135	0.132	100.0	100.0	8.5
Syncc9902_1086	Inositol-1(or 4)-monophosphatase (EC:3.1.3.25)	0.028	0.252	0.111	0.112	100.0	92.5	12.9
Syncc9902_1087	ferredoxin	0.014	0.295	0.049	0.059	100.0	100.0	10.7
Syncc9902_1088	hypothetical protein	0.036	0.381	0.095	0.098	100.0	100.0	19.5
Syncc9902_1089	Phosphate transport system permease protein 1	0.024	0.357	0.066	0.067	100.0	91.1	10.5
Syncc9902_1090	Phosphate transport system permease protein 2	0.034	0.330	0.103	0.104	100.0	97.3	10.1
Syncc9902_1091	Phosphate ABC transporter, permease protein PstC	0.036	0.321	0.111	0.106	100.0	100.0	13.6
Syncc9902_1092	molecular chaperone DnaK, heat shock protein HSP70	0.008	0.242	0.035	0.038	100.0	74.7	7.7
Syncc9902_1093	Heat shock protein DnaJ-like	0.022	0.235	0.093	0.096	100.0	100.0	10.7
Syncc9902_1094	hypothetical protein	0.031	0.244	0.127	0.144	100.0	100.0	10.2
Syncc9902_1095	hypothetical protein	0.016	0.161	0.098	0.098	100.0	67.1	5.7
Syncc9902_1096	Methylthioadenosine phosphorylase (EC:2.4.2.28)	0.026	0.229	0.113	0.121	100.0	83.0	9.2
Syncc9902_1097	Peptidylprolyl isomerase (EC:5.2.1.8)	0.005	0.168	0.030	0.030	100.0	100.0	10.9
Syncc9902_1098	bifunctional 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II protein (EC:3.5.4.25)	0.023	0.243	0.093	0.091	100.0	87.4	7.4
Syncc9902_1099	N-acetyl-gamma-glutamyl-phosphate reductase (EC:1.2.1.38)	0.017	0.241	0.070	0.080	100.0	90.4	7.6
Syncc9902_1100	phosphoribosylglycinamide formyltransferase	0.027	0.179	0.151	0.150	100.0	93.7	8.3
Syncc9902_1101	hypothetical protein	0.034	0.175	0.195	0.196	98.9	67.0	7.1
Syncc9902_1102	PDZ/DHR/GLGF	0.025	0.170	0.147	0.137	100.0	67.6	5.2
Syncc9902_1103	N-acetylmuramoyl-L-alanine amidase, family 2	0.026	0.188	0.141	0.117	100.0	43.7	4.7
Syncc9902_1104	possible DnaB-like helicase	0.021	0.168	0.123	0.122	100.0	88.6	8.4
Syncc9902_1105	glucose-6-phosphate isomerase (EC:5.3.1.9)	0.014	0.254	0.056	0.058	100.0	84.9	8.3
Syncc9902_1106	Leucyl-tRNA synthetase class Ia	0.025	0.206	0.123	0.127	100.0	91.8	8.6
Syncc9902_1107	Diaminopimelate epimerase (EC:5.1.1.7)	0.024	0.233	0.102	0.104	100.0	79.5	7.1
Syncc9902_1108	possible cysteine desulfurase (class-V aminotransferase family)	0.043	0.185	0.232	0.235	100.0	91.0	6.9
Syncc9902_1109	hypothetical protein	0.021	0.226	0.092	0.112	100.0	69.3	6.7
Syncc9902_1110	hypothetical protein	0.044	0.207	0.214	0.201	100.0	91.2	11.5
Syncc9902_1111	Peptidase S13, D-Ala-D-Ala carboxypeptidase C	0.035	0.236	0.150	0.153	100.0	90.4	13.7
Syncc9902_1112	hypothetical protein	0.014	0.238	0.060	0.060	100.0	96.9	11.0
Syncc9902_1113	Coenzyme A biosynthesis protein (EC:2.7.7.3)	0.014	0.183	0.076	0.073	100.0	100.0	8.2
Syncc9902_1114	hypothetical protein	0.011	0.260	0.044	0.044	100.0	100.0	11.7
Syncc9902_1115	excinuclease ABC subunit C	0.013	0.251	0.050	0.054	100.0	97.4	11.4
Syncc9902_1116	hypothetical protein	0.027	0.221	0.121	0.127	100.0	99.0	8.3
Syncc9902_1117	photolyase protein family	0.032	0.248	0.128	0.128	100.0	91.3	9.7
Syncc9902_1118	hypothetical protein	0.016	0.244	0.066	0.068	100.0	100.0	11.7
Syncc9902_1119	Phosphoesterase PHP-like	0.041	0.253	0.160	0.154	100.0	97.2	10.9
Syncc9902_1120	hypothetical protein	0.033	0.211	0.156	0.127	100.0	100.0	14.9
Syncc9902_1121	Cobaltochelatase, CobN subunit (EC:6.6.1.2)	0.038	0.244	0.155	0.146	100.0	79.9	9.2
Syncc9902_1122	branched-chain amino acid aminotransferase	0.008	0.216	0.036	0.036	100.0	83.0	6.6
Syncc9902_1123	5-methyltetrahydrofolate--homocysteine methyltransferase	0.018	0.312	0.058	0.059	100.0	100.0	13.9

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_1124	hypothetical protein	0.043	0.286	0.152	0.170	100.0	100.0	15.1
Syncc9902_1125	hypothetical protein	0.021	0.179	0.117	0.117	100.0	100.0	19.1
Syncc9902_1126	hypothetical protein	0.029	0.219	0.131	0.129	100.0	100.0	24.2
Syncc9902_1127	possible CsgA C-factor signaling protein	0.042	0.298	0.141	0.137	100.0	100.0	18.5
Syncc9902_1128	hypothetical protein	0.059	0.268	0.220	0.216	100.0	100.0	17.8
Syncc9902_1129	possible ATP adenylyltransferase	0.061	0.328	0.185	0.201	100.0	100.0	14.0
Syncc9902_1130	possible type 3 alternative RNA polymerase sigma factor	0.057	0.250	0.229	0.232	100.0	63.5	7.1
Syncc9902_1131	hypothetical protein	0.044	0.275	0.161	0.158	100.0	96.3	11.9
Syncc9902_1132	possible DnaJ domain	0.038	0.291	0.129	0.124	100.0	95.7	14.7
Syncc9902_1133	allophycocyanin alpha-B chain	0.006	0.282	0.020	0.026	100.0	100.0	16.5
Syncc9902_1134	23S rRNA methyltransferase/RumA	0.040	0.270	0.149	0.147	100.0	99.3	13.5
Syncc9902_1135	phenylalanyl-tRNA synthetase, beta subunit	0.031	0.231	0.134	0.142	100.0	100.0	11.7
Syncc9902_1136	50S ribosomal protein L33	0.000	0.059	0.000	0.000	100.0	100.0	11.5
Syncc9902_1137	30S ribosomal protein S18	0.002	0.236	0.009	0.000	100.0	100.0	15.1
Syncc9902_1138	Exoribonuclease II (EC:3.1.13.1)	0.036	0.199	0.180	0.185	100.0	93.2	9.9
Syncc9902_1139	hypothetical protein	0.051	0.256	0.201	0.193	94.5	84.2	9.3
Syncc9902_1140	methionine--tRNA ligase	0.031	0.324	0.094	0.094	100.0	100.0	14.8
Syncc9902_1141	hypothetical protein	0.082	0.338	0.244	0.250	100.0	100.0	14.6
Syncc9902_1142	hypothetical protein	0.050	0.253	0.196	0.195	100.0	100.0	19.6
Syncc9902_1143	hypothetical protein	0.043	0.331	0.130	0.127	100.0	100.0	22.0
Syncc9902_1144	hypothetical protein	0.039	0.364	0.107	0.105	100.0	100.0	30.0
Syncc9902_1145	proton extrusion protein PcxA	0.035	0.406	0.087	0.084	100.0	100.0	19.6
Syncc9902_1146	putative cobinamide kinase	0.093	0.309	0.302	0.301	100.0	96.3	17.2
Syncc9902_1147	putative tRNA/rRNA methyltransferase (SpoU family)	0.041	0.359	0.113	0.112	100.0	100.0	24.5
Syncc9902_1148	Peptidoglycan-binding LysM	0.048	0.332	0.144	0.142	100.0	99.9	17.2
Syncc9902_1149	hypothetical protein	0.041	0.346	0.119	0.134	100.0	100.0	15.9
Syncc9902_1150	thioredoxin peroxidase	0.007	0.318	0.021	0.021	100.0	100.0	29.4
Syncc9902_1151	hypothetical protein	0.037	0.309	0.121	0.120	100.0	100.0	16.5
Syncc9902_1152	Peptidase M41, FtsH (EC:3.6.4.6)	0.019	0.278	0.070	0.067	100.0	100.0	13.9
Syncc9902_1153	50S ribosomal protein L32	0.012	0.222	0.056	0.056	100.0	86.4	6.5
Syncc9902_1154	hypothetical protein	0.051	0.158	0.320	0.335	100.0	67.1	5.2
Syncc9902_1155	hypothetical protein	0.030	0.187	0.163	0.173	100.0	100.0	6.9
Syncc9902_1156	HAD-superfamily hydrolase subfamily IA, variant 3	0.045	0.188	0.237	0.236	100.0	96.2	8.2
Syncc9902_1157	RecJ exonuclease	0.024	0.220	0.108	0.113	100.0	77.3	7.8
Syncc9902_1158	hypothetical protein	0.046	0.195	0.236	0.249	100.0	89.3	9.2
Syncc9902_1159	hypothetical protein	0.016	0.191	0.082	0.084	100.0	100.0	11.5
Syncc9902_1160	hypothetical protein	0.015	0.179	0.085	0.081	100.0	100.0	8.7
Syncc9902_1161	RNA binding S1	0.045	0.308	0.147	0.150	100.0	98.2	12.0
Syncc9902_1162	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	0.017	0.323	0.053	0.062	100.0	100.0	11.9
Syncc9902_1163	hypothetical protein	0.036	0.129	0.279	0.299	100.0	84.2	6.3
Syncc9902_1164	hypothetical protein	0.018	0.302	0.059	0.063	100.0	100.0	10.3
Syncc9902_1165	Magnesium-protoporphyrin IX monomethyl ester aerobic oxidative cyclase	0.017	0.413	0.041	0.042	100.0	99.7	14.5
Syncc9902_1166	putative modulator of DNA gyrase; TldD	0.021	0.378	0.055	0.055	100.0	100.0	17.7
Syncc9902_1167	possible modulator of DNA gyrase	0.023	0.362	0.064	0.063	100.0	100.0	22.0
Syncc9902_1168	Methionyl-tRNA formyltransferase	0.047	0.238	0.197	0.197	100.0	100.0	19.3
Syncc9902_1169	hypothetical protein	0.035	0.206	0.168	0.168	100.0	100.0	17.3
Syncc9902_1170	hypothetical protein	0.033	0.225	0.145	0.156	100.0	100.0	17.8
Syncc9902_1171	TerC family protein	0.041	0.311	0.132	0.135	100.0	100.0	20.0
Syncc9902_1172	hypothetical protein	0.061	0.295	0.207	0.205	100.0	100.0	16.9
Syncc9902_1173	hypothetical protein	0.027	0.140	0.195	0.195	100.0	100.0	14.5
Syncc9902_1174	hypothetical protein	0.202	0.068	2.962	2.962	100.0	100.0	22.3
Syncc9902_1175	hypothetical protein	0.018	0.104	0.170	0.200	100.0	100.0	10.1
Syncc9902_1176	Transcription-repair coupling factor	0.017	0.333	0.051	0.053	100.0	97.6	17.9
Syncc9902_1177	hypothetical protein	0.025	0.464	0.053	0.053	100.0	100.0	18.5
Syncc9902_1178	Peptidase S41A, C-terminal protease (EC:3.4.21.102)	0.044	0.341	0.128	0.121	100.0	100.0	19.6
Syncc9902_1179	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (EC:1.17.4.3)	0.018	0.327	0.056	0.056	100.0	99.9	13.8
Syncc9902_1180	Phage SPO1 DNA polymerase-related protein	0.032	0.159	0.202	0.219	99.1	60.1	6.2
Syncc9902_1181	hypothetical protein	0.052	0.226	0.231	0.210	100.0	85.3	9.2
Syncc9902_1182	aminotransferase class-I	0.026	0.235	0.111	0.118	100.0	94.6	9.9
Syncc9902_1183	Phosphonate-binding periplasmic protein	0.035	0.248	0.143	0.144	100.0	100.0	15.7
Syncc9902_1184	ATPase	0.037	0.214	0.173	0.185	100.0	100.0	12.2
Syncc9902_1185	putative phosphonate ABC transporter	0.025	0.202	0.122	0.127	100.0	86.3	8.1
Syncc9902_1186	hypothetical protein	0.040	0.323	0.125	0.123	100.0	90.8	19.1
Syncc9902_1187	hypothetical protein	0.005	0.150	0.036	0.036	100.0	100.0	11.6
Syncc9902_1188	hypothetical protein	0.059	0.116	0.506	0.506	100.0	93.5	8.9
Syncc9902_1189	putative type III sigma factor	0.076	0.231	0.330	0.343	100.0	100.0	14.5

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_1190	hypothetical protein	0.049	0.230	0.213	0.208	100.0	99.7	14.9
Syncc9902_1191	hypothetical protein	0.056	0.200	0.279	0.257	100.0	100.0	14.9
Syncc9902_1192	hypothetical protein	0.075	0.248	0.301	0.294	100.0	100.0	15.1
Syncc9902_1193	hypothetical protein	0.021	0.199	0.104	0.101	100.0	100.0	9.4
Syncc9902_1194	ATPase	0.041	0.242	0.168	0.168	100.0	100.0	11.5
Syncc9902_1195	possible 33kD chaperonin, heat shock protein HSP33	0.026	0.286	0.092	0.085	100.0	87.0	14.3
Syncc9902_1196	hypothetical protein	0.035	0.303	0.117	0.115	100.0	100.0	13.9
Syncc9902_1197	hypothetical protein	0.065	0.288	0.224	0.242	100.0	100.0	13.6
Syncc9902_1198	peptide chain release factor 3	0.013	0.296	0.043	0.042	100.0	100.0	13.4
Syncc9902_1199	hypothetical protein	0.049	0.236	0.209	0.203	100.0	100.0	12.3
Syncc9902_1200	ribonucleotide reductase (class II)	0.019	0.312	0.061	0.066	100.0	97.5	13.4
Syncc9902_1201	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	1.9
Syncc9902_1202	hypothetical protein	0.010	0.167	0.060	0.060	100.0	48.2	5.1
Syncc9902_1203	hypothetical protein	0.070	0.322	0.216	0.219	100.0	100.0	16.3
Syncc9902_1204	hypothetical protein	0.019	0.125	0.151	0.151	100.0	100.0	11.7
Syncc9902_1205	hypothetical protein	0.048	0.159	0.302	0.283	100.0	100.0	13.5
Syncc9902_1206	hypothetical protein	0.091	0.238	0.383	0.384	100.0	100.0	27.9
Syncc9902_1207	hypothetical protein	0.057	0.203	0.282	0.277	100.0	100.0	21.0
Syncc9902_1208	hypothetical protein	0.038	0.197	0.193	0.194	100.0	100.0	15.5
Syncc9902_1209	hypothetical protein	0.014	0.283	0.050	0.050	100.0	100.0	47.5
Syncc9902_1210	hypothetical protein	0.098	0.288	0.340	0.334	100.0	100.0	29.5
Syncc9902_1211	hypothetical protein	0.042	0.212	0.198	0.197	100.0	100.0	18.9
Syncc9902_1212	hypothetical protein	0.079	0.280	0.284	0.296	100.0	100.0	31.3
Syncc9902_1213	putative short-chain dehydrogenase family protein	0.052	0.335	0.156	0.154	100.0	100.0	31.2
Syncc9902_1214	hypothetical protein	0.055	0.228	0.239	0.240	100.0	100.0	22.3
Syncc9902_1215	conserved hypothetical inner membrane protein	0.070	0.188	0.372	0.383	100.0	100.0	12.7
Syncc9902_1216	possible light-dependent protochlorophyllide oxido-reductase	0.060	0.255	0.236	0.242	100.0	100.0	15.9
Syncc9902_1217	hypothetical protein	0.036	0.232	0.156	0.157	100.0	100.0	15.8
Syncc9902_1218	putative cobalamin synthesis protein	0.027	0.237	0.112	0.114	100.0	99.2	11.7
Syncc9902_1219	hypothetical protein	0.017	0.242	0.070	0.072	100.0	100.0	12.4
Syncc9902_1220	putative uracil phosphoribosyltransferase	0.035	0.219	0.160	0.155	100.0	100.0	13.9
Syncc9902_1221	hypothetical protein	0.030	0.201	0.149	0.165	100.0	100.0	14.5
Syncc9902_1222	dihydroxy-acid dehydratase (EC:4.2.1.9)	0.017	0.255	0.066	0.066	100.0	100.0	15.2
Syncc9902_1223	hypothetical protein	0.020	0.272	0.074	0.074	100.0	100.0	10.9
Syncc9902_1224	hypothetical protein	0.025	0.316	0.080	0.082	100.0	100.0	21.8
Syncc9902_1225	6-phosphogluconolactonase	0.030	0.245	0.121	0.122	100.0	100.0	16.7
Syncc9902_1226	6-phosphogluconate dehydrogenase	0.016	0.304	0.052	0.053	100.0	100.0	20.4
Syncc9902_1227	glucose-1-phosphate adenylyltransferase (EC:2.7.7.27)	0.013	0.365	0.036	0.040	100.0	100.0	17.4
Syncc9902_1228	glutamyl-tRNA reductase	0.017	0.326	0.051	0.052	100.0	100.0	19.7
Syncc9902_1229	GlpX	0.015	0.325	0.047	0.047	100.0	100.0	26.9
Syncc9902_1230	ribulose-phosphate 3-epimerase (EC:5.1.3.1)	0.024	0.501	0.048	0.048	100.0	100.0	19.7
Syncc9902_1231	possible heme transporter	0.031	0.491	0.062	0.060	100.0	100.0	38.2
Syncc9902_1232	possible permease	0.045	0.479	0.094	0.092	100.0	100.0	40.5
Syncc9902_1233	ATPase	0.034	0.401	0.085	0.088	100.0	100.0	27.7
Syncc9902_1234	hypothetical protein	0.071	0.399	0.177	0.181	100.0	100.0	33.1
Syncc9902_1235	hypothetical protein	0.066	0.391	0.168	0.176	100.0	100.0	27.9
Syncc9902_1236	GTP-binding protein TypA	0.007	0.236	0.029	0.030	100.0	99.9	13.1
Syncc9902_1237	putative peptidase U32 family protein	0.027	0.139	0.195	0.193	100.0	59.7	5.8
Syncc9902_1238	putative carboxypeptidase	0.023	0.192	0.117	0.127	100.0	100.0	12.4
Syncc9902_1239	Geranylgeranyl reductase	0.009	0.260	0.036	0.037	100.0	100.0	13.3
Syncc9902_1240	hypothetical protein	0.028	0.102	0.277	0.319	100.0	100.0	6.6
Syncc9902_1241	Glycyl-tRNA synthetase, beta subunit (EC:6.1.1.14)	0.026	0.194	0.135	0.145	100.0	70.9	6.7
Syncc9902_1242	Sulfite reductase, ferredoxin dependent	0.011	0.221	0.052	0.057	100.0	84.2	8.0
Syncc9902_1243	D-ala-D-ala dipeptidase	0.053	0.216	0.245	0.254	100.0	95.1	8.8
Syncc9902_1244	ATP-dependent DNA helicase RecG	0.019	0.179	0.106	0.112	98.9	85.9	9.7
Syncc9902_1245	hypothetical protein	0.045	0.197	0.230	0.232	100.0	93.8	12.9
Syncc9902_1246	elongation factor Ts	0.010	0.251	0.041	0.036	100.0	99.9	11.2
Syncc9902_1247	30S ribosomal protein S2	0.003	0.227	0.014	0.014	100.0	100.0	13.8
Syncc9902_1248	possible glycosyltransferase	0.017	0.230	0.076	0.071	100.0	100.0	11.3
Syncc9902_1249	hypothetical protein	0.000	0.369	0.000	0.000	100.0	100.0	11.9
Syncc9902_1250	ATPase	0.030	0.259	0.116	0.117	100.0	100.0	13.1
Syncc9902_1251	DevC protein	0.007	0.171	0.043	0.053	100.0	82.7	7.1
Syncc9902_1252	possible ABC transporter component	0.027	0.160	0.167	0.164	100.0	74.3	6.4
Syncc9902_1253	phycocyanobilin:ferredoxin oxidoreductase	0.041	0.201	0.202	0.205	100.0	24.3	3.5
Syncc9902_1254	possible Zn-dependent peptidase	0.013	0.188	0.070	0.077	92.7	41.7	3.9
Syncc9902_1255	possible Zn-dependent peptidase	0.027	0.159	0.170	0.164	100.0	39.8	4.3
Syncc9902_1256	hypothetical protein	0.004	0.071	0.059	0.060	100.0	38.3	3.4

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_1257	BioY protein	0.022	0.183	0.121	0.106	100.0	72.6	6.5
Syncc9902_1258	Peptidase A8, signal peptidase II	0.028	0.102	0.272	0.194	100.0	71.1	5.8
Syncc9902_1259	Peptidoglycan glycosyltransferase (EC:2.4.1.129)	0.020	0.119	0.169	0.182	99.0	15.4	2.9
Syncc9902_1260	hypothetical protein	0.026	0.175	0.147	0.155	99.3	66.3	6.1
Syncc9902_1261	pyridoxal-dependent decarboxylase family protein	0.026	0.167	0.156	0.169	100.0	57.4	5.0
Syncc9902_1262	cytidine/deoxycytidylate deaminase family protein	0.020	0.204	0.098	0.080	100.0	91.4	6.5
Syncc9902_1263	serine:pyruvate/alanine:glyoxylate aminotransferase	0.011	0.210	0.053	0.063	100.0	77.6	6.2
Syncc9902_1264	phycobilisome core component-allophycocyanin beta-18 subunit	0.012	0.212	0.055	0.049	100.0	77.3	10.1
Syncc9902_1265	Glutamine synthetase type I (EC:6.3.1.2)	0.003	0.202	0.016	0.016	100.0	79.5	7.7
Syncc9902_1266	generic methyl-transferase	0.016	0.236	0.069	0.071	100.0	80.7	7.5
Syncc9902_1267	hypothetical protein	0.017	0.215	0.081	0.082	100.0	57.4	5.3
Syncc9902_1268	hypothetical protein	0.079	0.116	0.678	0.420	100.0	48.6	4.4
Syncc9902_1269	hypothetical protein	0.012	0.171	0.070	0.064	100.0	77.7	5.9
Syncc9902_1270	putative anti-sigma regulatory factor (serine/threonine protein kinase)	0.016	0.148	0.110	0.097	100.0	100.0	6.5
Syncc9902_1271	hypothetical protein	0.015	0.198	0.076	0.095	94.2	55.4	4.7
Syncc9902_1272	Rhodanese-like	0.024	0.205	0.118	0.131	100.0	99.6	10.2
Syncc9902_1273	putative photosystem II reaction center Psb28 protein	0.012	0.184	0.066	0.066	100.0	95.0	8.6
Syncc9902_1274	hypothetical protein	0.023	0.208	0.108	0.115	100.0	89.6	8.6
Syncc9902_1275	hypothetical protein	0.011	0.253	0.045	0.050	100.0	99.3	12.4
Syncc9902_1276	hypothetical protein	0.053	0.273	0.193	0.195	100.0	100.0	21.5
Syncc9902_1277	hypothetical protein	0.068	0.277	0.246	0.244	100.0	100.0	29.0
Syncc9902_1278	SecF protein	0.026	0.266	0.097	0.095	100.0	100.0	23.5
Syncc9902_1279	SecD export membrane protein	0.019	0.268	0.071	0.070	100.0	100.0	21.2
Syncc9902_1280	pyruvate dehydrogenase E1 beta subunit	0.006	0.186	0.030	0.032	100.0	100.0	13.1
Syncc9902_1281	hypothetical protein	0.026	0.168	0.156	0.157	100.0	100.0	11.9
Syncc9902_1282	4-diphosphocytidyl-2C-methyl-D-erythritol kinase (EC:2.7.1.148)	0.044	0.208	0.211	0.212	100.0	100.0	18.6
Syncc9902_1283	16S rRNA dimethylase	0.031	0.214	0.143	0.153	100.0	100.0	21.7
Syncc9902_1284	hypothetical protein	0.051	0.173	0.296	0.316	100.0	100.0	15.3
Syncc9902_1285	hypothetical protein	0.020	0.203	0.098	0.098	100.0	100.0	18.3
Syncc9902_1286	hypothetical protein	0.027	0.210	0.130	0.148	100.0	93.4	13.1
Syncc9902_1287	hypothetical protein	0.016	0.168	0.097	0.097	100.0	100.0	17.0
Syncc9902_1288	hypothetical protein	0.068	0.321	0.210	0.204	44.5	14.9	1.8
Syncc9902_1289	hypothetical protein	0.050	0.226	0.219	0.220	100.0	50.0	4.3
Syncc9902_1290	hypothetical protein	0.066	0.211	0.313	0.323	100.0	100.0	17.1
Syncc9902_1291	putative SOS mutagenesis protein UmuD	0.023	0.241	0.094	0.091	100.0	73.9	5.7
Syncc9902_1292	DNA-directed DNA polymerase (EC:2.7.7.7)	0.014	0.151	0.094	0.091	100.0	100.0	12.9
Syncc9902_1293	hypothetical protein	0.032	0.237	0.133	0.131	100.0	100.0	15.7
Syncc9902_1294	hypothetical protein	0.027	0.244	0.110	0.102	100.0	100.0	12.1
Syncc9902_1295	DNA primase	0.006	0.164	0.039	0.043	100.0	94.4	10.9
Syncc9902_1296	hypothetical protein	0.027	0.185	0.144	0.157	100.0	91.8	6.2
Syncc9902_1297	hypothetical protein	0.034	0.237	0.142	0.144	100.0	90.1	8.9
Syncc9902_1298	DNA recombination protein, RuvA	0.011	0.170	0.064	0.064	100.0	100.0	10.7
Syncc9902_1299	Ribosomal protein S15	0.014	0.232	0.059	0.081	100.0	100.0	9.2
Syncc9902_1300	hypothetical protein	0.033	0.214	0.154	0.148	100.0	100.0	8.1
Syncc9902_1301	DNA polymerase III subunit alpha (EC:2.7.7.7)	0.012	0.211	0.059	0.059	100.0	86.9	8.4
Syncc9902_1302	Glutamyl-tRNA(Gln) amidotransferase A subunit (EC:3.5.1.4)	0.007	0.148	0.045	0.054	100.0	52.3	5.0
Syncc9902_1303	hypothetical protein	0.023	0.179	0.129	0.126	100.0	44.8	4.5
Syncc9902_1304	RNA methyltransferase TrmH, group 3	0.006	0.169	0.037	0.034	92.5	16.0	3.3
Syncc9902_1305	hypothetical protein	0.012	0.110	0.106	0.106	100.0	47.2	4.4
Syncc9902_1306	anti-anti-sigma regulatory factor, SpoIIAA	0.017	0.200	0.087	0.084	100.0	80.4	5.7
Syncc9902_1307	carbamoyl-phosphate synthase small subunit	0.027	0.142	0.188	0.178	100.0	36.3	4.3
Syncc9902_1308	Anthranilate phosphoribosyl transferase (EC:2.4.2.18)	0.007	0.104	0.064	0.061	100.0	27.7	3.5
Syncc9902_1309	ATPase	0.012	0.153	0.081	0.075	100.0	88.3	7.0
Syncc9902_1310	putative regulator	0.004	0.068	0.053	0.053	100.0	100.0	6.9
Syncc9902_1311	Methionine sulfoxide reductase A (EC:1.8.4.6)	0.024	0.220	0.110	0.107	100.0	95.2	11.9
Syncc9902_1312	hypothetical protein	0.020	0.104	0.194	0.189	100.0	84.7	7.3
Syncc9902_1313	Periplasmic phosphate binding protein	0.022	0.167	0.132	0.136	100.0	66.8	5.9
Syncc9902_1314	hypothetical protein	0.030	0.113	0.266	0.266	100.0	100.0	11.3
Syncc9902_1315	Exonuclease	0.015	0.120	0.127	0.130	100.0	90.6	6.1
Syncc9902_1316	putative bacterioferritin comigratory protein	0.031	0.142	0.221	0.238	92.3	38.8	4.3
Syncc9902_1317	hypothetical protein	0.012	0.126	0.095	0.115	100.0	43.3	4.2
Syncc9902_1318	Acetate--CoA ligase	0.003	0.155	0.020	0.019	96.7	18.6	2.6
Syncc9902_1319	HAD-superfamily hydrolase subfamily IA, variant 3	0.023	0.123	0.189	0.197	88.3	27.1	3.4
Syncc9902_1320	Trans-hexaprenyltranstransferase (EC:2.5.1.30)	0.021	0.286	0.074	0.077	100.0	25.1	3.7
Syncc9902_1321	Glutamate racemase (EC:5.1.1.3)	0.026	0.143	0.185	0.118	95.4	22.4	2.9
Syncc9902_1322	N-acetylmuramoyl-L-alanine amidase (EC:3.5.1.28)	0.000	0.084	0.000	0.051	84.3	13.8	2.2

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_1323	possible nitrilase	0.004	0.182	0.025	0.024	98.1	25.8	3.2
Syncc9902_1324	2-phosphosulfolactate phosphatas (EC:3.1.3.71)	0.000	0.000	NaN	NaN	60.9	0.0	1.1
Syncc9902_1325	Carboxylyase-related protein	0.022	0.293	0.074	0.072	92.2	17.7	2.4
Syncc9902_1326	3-phosphoshikimate 1-carboxyvinyltransferase	0.021	0.165	0.125	0.107	91.6	22.2	3.6
Syncc9902_1327	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.2	2.4
Syncc9902_1328	UDP-N-acetylglucosamine pyrophosphorylase (EC:2.7.7.23)	0.012	0.148	0.082	0.078	100.0	25.1	3.4
Syncc9902_1329	UDP-N-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate-D-alanyl-D-alanyl ligase	0.029	0.159	0.182	0.200	97.9	13.6	2.6
Syncc9902_1330	hypothetical protein	0.062	0.227	0.271	0.406	99.9	4.1	2.8
Syncc9902_1331	glycogen synthase	0.010	0.168	0.058	0.058	100.0	39.8	4.8
Syncc9902_1332	hypothetical protein	0.025	0.147	0.168	0.146	69.8	7.3	1.9
Syncc9902_1333	Naphthoate synthase	0.005	0.103	0.044	0.042	98.2	12.0	2.7
Syncc9902_1334	Menaquinone biosynthesis protein (EC:2.5.1.64)	0.017	0.156	0.112	0.111	78.0	33.7	3.6
Syncc9902_1335	Peptidase S26A, signal peptidase I (EC:3.4.21.89)	0.021	0.194	0.107	0.108	90.8	54.3	4.3
Syncc9902_1336	hypothetical protein	0.011	0.207	0.051	0.042	100.0	100.0	8.2
Syncc9902_1337	hypothetical protein	0.050	0.197	0.256	0.258	100.0	97.7	8.0
Syncc9902_1338	hypothetical protein	0.032	0.184	0.176	0.172	100.0	91.5	10.5
Syncc9902_1339	Naringenin-chalcone synthase (EC:2.3.1.74)	0.054	0.222	0.243	0.246	100.0	92.4	10.3
Syncc9902_1340	hypothetical protein	0.041	0.222	0.185	0.179	100.0	100.0	12.1
Syncc9902_1341	hypothetical protein	0.031	0.198	0.156	0.161	100.0	100.0	11.1
Syncc9902_1342	hypothetical protein	0.028	0.135	0.207	0.209	100.0	84.8	7.5
Syncc9902_1343	hypothetical protein	0.037	0.402	0.093	0.095	100.0	100.0	15.8
Syncc9902_1344	hypothetical protein	0.058	0.250	0.231	0.259	100.0	92.2	12.8
Syncc9902_1345	hypothetical protein	0.024	0.312	0.077	0.084	100.0	100.0	16.0
Syncc9902_1346	zeta-carotene desaturase-like	0.036	0.232	0.156	0.152	100.0	100.0	15.1
Syncc9902_1347	hypothetical protein	0.078	0.179	0.435	0.420	100.0	87.3	11.2
Syncc9902_1348	hypothetical protein	0.040	0.278	0.144	0.136	100.0	100.0	24.2
Syncc9902_1349	hypothetical protein	0.062	0.302	0.207	0.228	100.0	100.0	19.9
Syncc9902_1350	hypothetical protein	0.033	0.128	0.256	0.255	100.0	100.0	10.6
Syncc9902_1351	Rare lipoprotein A	0.086	0.220	0.390	0.398	100.0	100.0	22.1
Syncc9902_1352	oxidoreductase	0.028	0.289	0.097	0.095	100.0	100.0	28.4
Syncc9902_1353	metal-binding protein-like	0.099	0.275	0.362	0.377	100.0	100.0	26.2
Syncc9902_1354	hypothetical protein	0.036	0.242	0.150	0.150	100.0	100.0	18.2
Syncc9902_1355	hypothetical protein	0.030	0.208	0.146	0.154	100.0	100.0	22.2
Syncc9902_1356	pterin-4-alpha-carbinolamine dehydratase	0.024	0.208	0.114	0.119	100.0	100.0	17.8
Syncc9902_1357	hypothetical protein	0.021	0.049	0.419	0.420	100.0	100.0	13.3
Syncc9902_1358	5-oxoprolinase (ATP-hydrolyzing) (EC:3.5.2.9)	0.047	0.228	0.208	0.202	100.0	94.9	12.7
Syncc9902_1359	possible permease	0.031	0.299	0.103	0.107	100.0	93.7	14.9
Syncc9902_1360	hypothetical protein	0.035	0.258	0.134	0.149	91.7	73.7	8.3
Syncc9902_1361	hypothetical protein	0.030	0.154	0.194	0.209	100.0	87.8	7.4
Syncc9902_1362	hypothetical protein	0.030	0.078	0.382	0.363	100.0	28.8	4.1
Syncc9902_1363	hypothetical protein	0.045	0.409	0.109	0.113	100.0	91.5	11.2
Syncc9902_1364	ABC transporter, substrate binding protein, possibly Mn	0.049	0.357	0.138	0.135	100.0	100.0	20.5
Syncc9902_1365	ATPase	0.059	0.406	0.146	0.133	100.0	100.0	22.9
Syncc9902_1366	ABC transporter component, possibly Mn transport	0.032	0.349	0.092	0.098	100.0	99.0	16.8
Syncc9902_1367	Glycosyltransferase-like	0.068	0.289	0.234	0.232	98.9	83.6	11.7
Syncc9902_1368	hypothetical protein	0.070	0.370	0.189	0.196	100.0	87.8	11.1
Syncc9902_1369	hypothetical protein	0.087	0.490	0.178	0.180	100.0	100.0	15.8
Syncc9902_1370	hypothetical protein	0.058	0.306	0.191	0.189	100.0	100.0	24.1
Syncc9902_1371	ribose-phosphate pyrophosphokinase (EC:2.7.6.1)	0.007	0.286	0.023	0.030	100.0	100.0	16.7
Syncc9902_1372	hypothetical protein	0.052	0.228	0.226	0.235	100.0	100.0	18.8
Syncc9902_1373	Cell envelope-related transcriptional attenuator	0.054	0.207	0.260	0.255	100.0	100.0	11.3
Syncc9902_1374	hypothetical protein	0.025	0.265	0.095	0.100	100.0	100.0	20.7
Syncc9902_1375	Glycoside hydrolase, family 77 (EC:2.4.1.25)	0.037	0.255	0.144	0.158	100.0	100.0	16.8
Syncc9902_1376	hypothetical protein	0.090	0.312	0.288	0.282	100.0	100.0	20.3
Syncc9902_1377	Pseudouridine synthase, Rsu	0.065	0.264	0.245	0.222	100.0	100.0	19.8
Syncc9902_1378	hypothetical protein	0.009	0.310	0.028	0.029	100.0	99.8	12.4
Syncc9902_1379	possible kinase	0.020	0.389	0.051	0.052	100.0	100.0	20.1
Syncc9902_1380	Arginine decarboxylase (EC:4.1.1.19)	0.026	0.138	0.189	0.182	100.0	87.9	6.3
Syncc9902_1381	hypothetical protein	0.018	0.163	0.108	0.108	100.0	80.3	6.5
Syncc9902_1382	hypothetical protein	0.020	0.111	0.183	0.186	100.0	47.3	4.9
Syncc9902_1383	Arabinose efflux permease-like	0.019	0.160	0.120	0.158	73.9	20.3	2.2
Syncc9902_1384	Putative multicopper oxidase-like	0.004	0.012	0.355	0.355	88.8	20.1	2.8
Syncc9902_1385	putative phosphatidate cytidylyltransferase	0.020	0.258	0.076	0.088	100.0	94.9	14.0
Syncc9902_1386	hypothetical protein	0.048	0.243	0.196	0.202	100.0	92.2	12.9
Syncc9902_1387	putative 2-hydroxy-6-oxohepta-2,4-dienoate hydrolase	0.041	0.266	0.153	0.151	100.0	100.0	14.0
Syncc9902_1388	putative glycerol dehydrogenase	0.039	0.293	0.134	0.145	100.0	86.1	11.1
Syncc9902_1389	ATPase	0.012	0.319	0.038	0.038	100.0	96.6	11.1

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_1390	putative ribosomal-protein-alanine acetyltransferase	0.036	0.178	0.201	0.192	100.0	67.3	6.7
Syncc9902_1391	Diaminopimelate decarboxylase	0.017	0.159	0.105	0.109	100.0	85.8	9.2
Syncc9902_1392	Protein of unknown function DUF147	0.003	0.228	0.011	0.042	100.0	23.2	3.7
Syncc9902_1393	Di-trans-poly-cis-decaprenylcistransferase (EC:2.5.1.31)	0.011	0.153	0.073	0.093	100.0	37.3	4.0
Syncc9902_1394	Biotin synthase (EC:2.8.1.6)	0.010	0.212	0.046	0.055	100.0	91.9	8.1
Syncc9902_1395	Rhodanese-like	0.049	0.247	0.197	0.186	100.0	100.0	12.1
Syncc9902_1396	glutathione S-transferase domain protein	0.044	0.233	0.187	0.190	100.0	100.0	12.3
Syncc9902_1397	hypothetical protein	0.029	0.230	0.124	0.125	100.0	96.2	12.6
Syncc9902_1398	lipoyl synthase	0.020	0.251	0.080	0.083	100.0	99.2	13.3
Syncc9902_1399	RecR protein	0.014	0.296	0.048	0.059	100.0	100.0	16.8
Syncc9902_1400	putative photosystem II oxygen-evolving complex 23K protein PsbP	0.022	0.279	0.078	0.071	100.0	100.0	15.8
Syncc9902_1401	possible Na+/H+ antiporter, CPA2 family	0.019	0.278	0.067	0.064	100.0	100.0	19.4
Syncc9902_1402	ATPase	0.028	0.230	0.121	0.126	100.0	100.0	16.3
Syncc9902_1403	hypothetical protein	0.118	0.197	0.599	0.583	100.0	100.0	24.5
Syncc9902_1404	RNA-binding region RNP-1	0.006	0.277	0.023	0.047	100.0	100.0	25.0
Syncc9902_1405	hypothetical protein	0.047	0.266	0.178	0.177	100.0	99.5	11.9
Syncc9902_1406	hypothetical protein	0.048	0.339	0.141	0.142	100.0	100.0	13.0
Syncc9902_1407	DEAD/DEAH box helicase-like	0.024	0.227	0.105	0.100	100.0	100.0	15.0
Syncc9902_1408	mannose-6-phosphate isomerase WbpW-like	0.020	0.143	0.143	0.134	100.0	100.0	12.9
Syncc9902_1409	Exodeoxyribonuclease V (EC:3.1.11.5)	0.030	0.170	0.176	0.180	100.0	93.2	8.6
Syncc9902_1410	probable exodeoxyribonuclease V, beta subunit RecB	0.026	0.185	0.138	0.143	100.0	92.4	9.4
Syncc9902_1411	Exodeoxyribonuclease V, RecC subunit	0.024	0.175	0.135	0.135	100.0	79.1	7.9
Syncc9902_1412	hypothetical protein	0.009	0.286	0.033	0.045	100.0	39.8	4.3
Syncc9902_1413	hypothetical protein	0.000	0.182	0.000	0.000	100.0	5.0	3.3
Syncc9902_1414	pyridoxal phosphate biosynthetic protein	0.014	0.156	0.092	0.098	100.0	96.3	8.0
Syncc9902_1415	Phospholipid/glycerol acyltransferase	0.017	0.159	0.110	0.123	100.0	100.0	9.1
Syncc9902_1416	hypothetical protein	0.036	0.297	0.120	0.112	100.0	83.5	6.2
Syncc9902_1417	BolA family protein	0.008	0.203	0.039	0.029	100.0	100.0	7.9
Syncc9902_1418	Glutaredoxin-related protein	0.010	0.241	0.043	0.058	100.0	95.1	9.6
Syncc9902_1419	hypothetical protein	0.013	0.341	0.039	0.045	100.0	100.0	14.6
Syncc9902_1420	two component transcriptional regulator, winged helix family	0.004	0.305	0.013	0.013	100.0	100.0	11.1
Syncc9902_1421	hypothetical protein	0.045	0.284	0.157	0.150	100.0	100.0	9.6
Syncc9902_1422	hypothetical protein	0.024	0.279	0.087	0.105	100.0	100.0	11.4
Syncc9902_1423	carotenoid isomerase	0.019	0.248	0.077	0.073	100.0	100.0	14.9
Syncc9902_1424	putative cation efflux transporter (CDF family)	0.098	0.214	0.456	0.442	100.0	63.2	6.1
Syncc9902_1425	glucose-inhibited division protein A	0.010	0.202	0.050	0.047	100.0	92.2	10.7
Syncc9902_1426	hypothetical protein	0.005	0.148	0.031	0.063	100.0	100.0	9.1
Syncc9902_1427	possible carbonic anhydrase	0.033	0.222	0.147	0.150	100.0	98.9	9.2
Syncc9902_1428	hypothetical protein	0.045	0.249	0.182	0.180	100.0	96.5	8.1
Syncc9902_1429	hypothetical protein	0.114	0.232	0.493	0.531	89.1	16.3	2.7
Syncc9902_1430	hypothetical protein	0.130	0.267	0.487	0.482	100.0	86.1	8.6
Syncc9902_1431	P-loop ATPase and inactivated derivatives-like	0.017	0.153	0.113	0.122	98.2	28.8	3.5
Syncc9902_1432	hypothetical protein	0.000	0.000	NaN	NaN	28.8	0.0	0.7
Syncc9902_1433	hypothetical protein	0.000	0.000	NaN	NaN	79.6	0.0	1.7
Syncc9902_1434	possible carbamoyl-phosphate synthase L chain	0.151	0.552	0.273	0.267	91.0	88.7	33.2
Syncc9902_1435	hypothetical protein	0.000	0.000	NaN	NaN	90.5	0.0	2.7
Syncc9902_1436	HNH nuclease	0.000	0.000	NaN	NaN	16.8	0.0	0.2
Syncc9902_1437	hypothetical protein	0.055	0.167	0.331	0.344	100.0	100.0	17.9
Syncc9902_1438	hypothetical protein	0.048	0.225	0.215	0.211	100.0	88.4	9.7
Syncc9902_1439	hypothetical protein	0.039	0.164	0.237	0.230	85.8	67.8	4.5
Syncc9902_1440	hypothetical protein	0.008	0.091	0.091	0.089	100.0	63.6	5.0
Syncc9902_1441	photosystem I reaction center subunit IV	0.019	0.254	0.074	0.074	100.0	100.0	15.9
Syncc9902_1442	metalloprotease	0.014	0.082	0.167	0.167	100.0	14.8	2.9
Syncc9902_1443	possible phage integrase family	0.069	0.289	0.238	0.235	100.0	91.8	9.8
Syncc9902_1444	hypothetical protein	0.019	0.301	0.062	0.057	100.0	100.0	20.8
Syncc9902_1445	hypothetical protein	0.038	0.082	0.460	0.502	100.0	73.5	6.1
Syncc9902_1446	hypothetical protein	0.109	0.159	0.682	0.686	100.0	100.0	11.9
Syncc9902_1447	hypothetical protein	0.063	0.237	0.266	0.273	100.0	100.0	16.9
Syncc9902_1448	hypothetical protein	0.082	0.297	0.275	0.269	100.0	100.0	18.6
Syncc9902_1449	hypothetical protein	0.072	0.091	0.791	0.802	100.0	100.0	12.7
Syncc9902_1450	hypothetical protein	0.081	0.228	0.354	0.349	100.0	100.0	17.0
Syncc9902_1451	hypothetical protein	0.059	0.197	0.297	0.316	100.0	100.0	16.6
Syncc9902_1452	hypothetical protein	0.116	0.143	0.809	0.759	100.0	100.0	14.1
Syncc9902_1453	hypothetical protein	0.089	0.274	0.326	0.320	100.0	100.0	18.3
Syncc9902_1454	hypothetical protein	0.114	0.151	0.756	0.828	100.0	100.0	19.0
Syncc9902_1455	conserved hypotheical protein	0.072	0.356	0.203	0.204	100.0	100.0	27.7
Syncc9902_1456	hypothetical protein	0.055	0.215	0.255	0.276	100.0	100.0	12.9

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_1457	hypothetical protein	0.082	0.233	0.350	0.346	100.0	100.0	17.3
Syncc9902_1458	hypothetical protein	0.062	0.271	0.230	0.222	100.0	100.0	21.1
Syncc9902_1459	putative histone deacetylase/AcuC/AphA family protein	0.026	0.270	0.096	0.095	100.0	100.0	18.5
Syncc9902_1460	hypothetical protein	0.062	0.217	0.285	0.289	100.0	100.0	9.8
Syncc9902_1463	hypothetical protein	0.004	0.248	0.015	0.015	100.0	100.0	34.9
Syncc9902_1464	membrane protein-like	0.046	0.307	0.149	0.158	100.0	100.0	21.4
Syncc9902_1465	hypothetical protein	0.030	0.329	0.092	0.092	100.0	100.0	22.2
Syncc9902_1466	hypothetical protein	0.047	0.176	0.264	0.243	100.0	100.0	10.3
Syncc9902_1467	hypothetical protein	0.081	0.298	0.272	0.262	100.0	100.0	21.1
Syncc9902_1468	hypothetical protein	0.044	0.347	0.126	0.127	100.0	100.0	36.8
Syncc9902_1469	hypothetical protein	0.096	0.131	0.730	0.803	86.9	54.0	4.3
Syncc9902_1470	possible ATLS1-like light-inducible protein	0.068	0.230	0.295	0.246	100.0	70.5	6.4
Syncc9902_1471	hypothetical protein	0.047	0.096	0.489	0.476	100.0	85.8	7.6
Syncc9902_1472	hypothetical protein	0.000	0.000	NaN	NaN	54.4	0.0	0.5
Syncc9902_1473	hypothetical protein	0.000	0.000	NaN	NaN	88.7	0.0	1.7
Syncc9902_1474	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	1.9
Syncc9902_1475	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_1476	hypothetical protein	0.000	0.000	NaN	NaN	70.5	0.0	1.3
Syncc9902_1481	Peptidase S24, LexA repressor (EC:3.4.21.88)	0.016	0.279	0.059	0.063	100.0	100.0	18.2
Syncc9902_1482	Ornithine carbamoyltransferase (EC:2.1.3.3)	0.018	0.252	0.073	0.069	100.0	100.0	17.3
Syncc9902_1483	Peptidase M41, FtsH	0.010	0.365	0.026	0.026	100.0	99.6	11.0
Syncc9902_1484	hypothetical protein	0.037	0.218	0.171	0.164	100.0	96.3	10.6
Syncc9902_1485	hypothetical protein	0.028	0.165	0.169	0.170	100.0	100.0	10.7
Syncc9902_1486	Riboflavin biosynthesis protein RibD (EC:3.5.4.26,EC:1.1.1.193)	0.041	0.216	0.191	0.191	100.0	99.6	9.1
Syncc9902_1487	hypothetical protein	0.023	0.249	0.092	0.093	100.0	90.0	9.2
Syncc9902_1488	transcriptional regulator, AraC family	0.052	0.207	0.252	0.241	97.9	56.9	5.1
Syncc9902_1489	hypothetical protein	0.012	0.198	0.061	0.064	100.0	77.1	6.8
Syncc9902_1490	hypothetical protein	0.012	0.186	0.063	0.117	100.0	77.5	5.9
Syncc9902_1491	hypothetical protein	0.051	0.231	0.220	0.205	100.0	59.7	6.0
Syncc9902_1492	putative precorrin-6y methylase	0.041	0.186	0.219	0.211	100.0	85.2	8.2
Syncc9902_1493	hypothetical protein	0.043	0.205	0.207	0.224	100.0	90.1	7.0
Syncc9902_1494	response regulator receiver domain protein (CheY-like)	0.024	0.167	0.142	0.163	100.0	78.4	7.0
Syncc9902_1495	inorganic polyphosphate/ATP-NAD kinase (EC:2.7.1.23)	0.005	0.199	0.024	0.023	100.0	85.0	8.6
Syncc9902_1496	NHL repeat	0.039	0.280	0.139	0.134	100.0	86.3	12.4
Syncc9902_1497	hypothetical protein	0.000	0.027	0.000	0.000	82.4	16.0	2.4
Syncc9902_1498	Phenylalanyl-tRNA synthetase, alpha subunit (EC:6.1.1.20)	0.007	0.204	0.036	0.035	100.0	94.8	13.7
Syncc9902_1499	Survival protein SurE (EC:3.1.3.2)	0.011	0.199	0.054	0.053	100.0	94.8	12.4
Syncc9902_1500	hypothetical protein	0.009	0.154	0.057	0.062	100.0	67.2	5.5
Syncc9902_1501	Riboflavin kinase / FAD synthetase	0.021	0.159	0.134	0.138	100.0	96.4	7.9
Syncc9902_1502	thiamine-phosphate pyrophosphorylase (EC:2.5.1.3)	0.035	0.182	0.191	0.203	100.0	89.8	9.4
Syncc9902_1503	ThiS, thiamine-biosynthesis	0.028	0.159	0.178	0.191	100.0	100.0	8.8
Syncc9902_1504	hypothetical protein	0.023	0.242	0.093	0.098	100.0	100.0	11.1
Syncc9902_1505	hypothetical protein	0.028	0.309	0.090	0.086	100.0	100.0	8.7
Syncc9902_1506	circadian phase modifier CpmA-like	0.038	0.238	0.159	0.154	100.0	100.0	10.8
Syncc9902_1507	hypothetical protein	0.025	0.124	0.200	0.204	100.0	60.0	5.5
Syncc9902_1508	tRNA (guanine-N1-)-methyltransferase (EC:2.1.1.31,EC:4.6.1.12)	0.019	0.173	0.109	0.103	100.0	98.1	9.2
Syncc9902_1509	hypothetical protein	0.029	0.178	0.160	0.152	100.0	100.0	8.0
Syncc9902_1510	hypothetical protein	0.012	0.188	0.064	0.086	100.0	44.4	5.0
Syncc9902_1511	GTP-binding protein Era	0.015	0.163	0.094	0.089	100.0	82.3	8.5
Syncc9902_1512	Alpha-acetolactate decarboxylase (EC:4.1.1.5)	0.015	0.059	0.256	0.256	100.0	9.0	3.3
Syncc9902_1513	hypothetical protein	0.006	0.164	0.037	0.054	100.0	29.3	4.3
Syncc9902_1514	hypothetical protein	0.010	0.291	0.035	0.038	100.0	100.0	11.4
Syncc9902_1515	PhoH-like phosphate starvation-inducible protein	0.015	0.217	0.071	0.077	100.0	86.5	7.2
Syncc9902_1516	30S ribosomal protein S16	0.005	0.228	0.021	0.025	100.0	100.0	10.4
Syncc9902_1517	Signal recognition particle protein	0.011	0.187	0.061	0.064	100.0	95.2	9.2
Syncc9902_1518	hypothetical protein	0.033	0.174	0.191	0.181	100.0	74.9	6.5
Syncc9902_1519	Pyruvate dehydrogenase (lipoamide) (EC:1.2.4.1)	0.007	0.236	0.029	0.028	100.0	88.7	7.4
Syncc9902_1520	type II alternative RNA polymerase sigma factor, sigma-70 family	0.007	0.234	0.028	0.034	100.0	96.8	9.8
Syncc9902_1521	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase (EC:2.1.1.61)	0.025	0.218	0.114	0.125	100.0	81.2	8.3
Syncc9902_1522	possible apolipoprotein n-acyltransferase	0.044	0.204	0.215	0.217	100.0	91.1	10.4
Syncc9902_1523	hypothetical protein	0.009	0.214	0.041	0.046	100.0	100.0	10.1
Syncc9902_1524	Peptidylprolyl isomerase (EC:5.2.1.8)	0.018	0.241	0.075	0.079	100.0	100.0	18.7
Syncc9902_1525	putative nickel-containing superoxide dismutase precursor (NISOD)	0.015	0.218	0.071	0.071	100.0	100.0	19.7
Syncc9902_1526	possible peptidase S26 family protein	0.085	0.249	0.344	0.361	100.0	95.6	18.1
Syncc9902_1527	Calcium/proton exchanger	0.008	0.067	0.124	0.124	100.0	64.2	6.3

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_1528	hypothetical protein	0.019	0.258	0.073	0.074	100.0	100.0	16.5
Syncc9902_1529	indole-3-glycerol-phosphate synthase (EC:4.1.1.48)	0.024	0.230	0.105	0.106	100.0	100.0	16.0
Syncc9902_1530	dihydrolipoamide dehydrogenase (EC:1.8.1.4)	0.006	0.220	0.028	0.031	100.0	94.5	10.7
Syncc9902_1531	tRNA/rRNA methyltransferase (SpoU)	0.044	0.196	0.224	0.211	100.0	92.0	7.1
Syncc9902_1532	hypothetical protein	0.019	0.233	0.080	0.070	100.0	100.0	10.4
Syncc9902_1533	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	0.022	0.201	0.111	0.108	100.0	93.9	8.7
Syncc9902_1534	acetylornithine aminotransferase	0.015	0.162	0.094	0.095	100.0	77.2	7.4
Syncc9902_1535	Folylpolyglutamate synthetase (EC:6.3.2.17)	0.030	0.114	0.261	0.251	100.0	77.6	5.8
Syncc9902_1536	hypothetical protein	0.027	0.105	0.259	0.249	100.0	67.9	6.1
Syncc9902_1537	hypothetical protein	0.027	0.171	0.156	0.154	99.2	48.1	4.8
Syncc9902_1538	possible cytosine deaminase	0.032	0.106	0.302	0.302	99.9	22.9	3.1
Syncc9902_1539	hypothetical protein	0.031	0.086	0.367	0.340	100.0	38.9	3.2
Syncc9902_1540	hypothetical protein	0.020	0.163	0.120	0.125	100.0	38.2	4.2
Syncc9902_1541	D-alanine--D-alanine ligase (EC:6.3.2.4)	0.027	0.172	0.155	0.157	100.0	45.3	4.5
Syncc9902_1542	hypothetical protein	0.013	0.183	0.073	0.059	100.0	100.0	9.5
Syncc9902_1543	cell division protein FtsQ	0.035	0.136	0.257	0.239	93.8	45.7	5.7
Syncc9902_1544	cell division protein FtsZ	0.017	0.231	0.073	0.069	100.0	73.1	7.5
Syncc9902_1545	Ketopantoate hydroxymethyltransferase (EC:2.1.2.11)	0.019	0.161	0.115	0.098	100.0	92.7	6.3
Syncc9902_1546	Putative oxygen-independent coproporphyrinogen III oxidase	0.031	0.125	0.245	0.264	100.0	66.9	6.0
Syncc9902_1547	Nucleotide binding protein, PINc	0.015	0.260	0.056	0.059	100.0	96.7	14.5
Syncc9902_1548	ATP-dependent Clp protease proteolytic subunit (EC:3.4.21.92)	0.006	0.289	0.021	0.020	100.0	100.0	14.7
Syncc9902_1549	ATP-dependent Clp protease proteolytic subunit (EC:3.4.21.92)	0.015	0.348	0.043	0.036	100.0	100.0	15.6
Syncc9902_1550	ketol-acid reductoisomerase (EC:1.1.1.86)	0.011	0.289	0.037	0.036	100.0	100.0	14.9
Syncc9902_1551	Cobalamin biosynthesis protein CbiB	0.031	0.186	0.168	0.164	100.0	98.5	13.5
Syncc9902_1552	Undecaprenyl-phosphate galactosephosphotransferase (EC:2.7.8.6)	0.037	0.401	0.092	0.087	100.0	100.0	28.3
Syncc9902_1553	probable glycosyltransferase	0.039	0.284	0.138	0.142	100.0	100.0	26.7
Syncc9902_1554	hypothetical protein	0.016	0.277	0.059	0.055	100.0	100.0	29.2
Syncc9902_1555	hypothetical protein	0.093	0.214	0.432	0.428	100.0	100.0	12.4
Syncc9902_1556	hypothetical protein	0.080	0.404	0.199	0.209	100.0	100.0	29.4
Syncc9902_1557	hypothetical protein	0.070	0.381	0.185	0.182	100.0	100.0	25.8
Syncc9902_1558	hypothetical protein	0.047	0.403	0.116	0.112	100.0	100.0	22.8
Syncc9902_1559	GAF	0.044	0.378	0.116	0.112	100.0	100.0	37.0
Syncc9902_1560	ComEC/Rec2-related protein	0.077	0.267	0.287	0.289	100.0	100.0	18.6
Syncc9902_1561	Glycyl-tRNA synthetase, alpha subunit (EC:6.1.1.14)	0.025	0.442	0.057	0.058	100.0	100.0	40.6
Syncc9902_1562	hypothetical protein	0.138	0.312	0.442	0.458	100.0	100.0	16.5
Syncc9902_1563	hypothetical protein	0.078	0.363	0.214	0.211	100.0	100.0	29.1
Syncc9902_1564	hypothetical protein	0.059	0.280	0.212	0.207	100.0	100.0	24.0
Syncc9902_1565	hypothetical protein	0.034	0.319	0.106	0.106	100.0	100.0	24.6
Syncc9902_1566	hypothetical protein	0.056	0.399	0.139	0.138	100.0	100.0	36.1
Syncc9902_1567	transcriptional regulator, AraC family	0.079	0.349	0.227	0.223	100.0	100.0	47.5
Syncc9902_1568	PBS lyase HEAT-like repeat	0.051	0.275	0.185	0.190	100.0	100.0	21.1
Syncc9902_1569	hypothetical protein	0.026	0.250	0.106	0.106	100.0	100.0	22.5
Syncc9902_1570	Ubiquinone/menaquinone biosynthesis methyltransferase	0.031	0.168	0.184	0.184	100.0	100.0	12.5
Syncc9902_1571	Histidine biosynthesis protein HisF	0.020	0.219	0.093	0.087	100.0	99.9	11.8
Syncc9902_1572	hypothetical protein	0.000	0.130	0.000	0.000	100.0	94.0	7.8
Syncc9902_1573	bacteriochlorophyll a synthase	0.020	0.269	0.074	0.071	100.0	100.0	14.7
Syncc9902_1574	Penicillin-binding protein 1A	0.028	0.211	0.134	0.130	100.0	96.6	10.4
Syncc9902_1575	Fmu, rRNA SAM-dependent methyltransferase	0.032	0.212	0.153	0.150	100.0	89.2	9.3
Syncc9902_1576	Methylated-DNA-(protein)-cysteine S-methyltransferase	0.030	0.000	Inf	Inf	100.0	8.7	3.5
Syncc9902_1577	tRNA (guanosine-2'-O-)-methyltransferase (EC:2.1.1.34)	0.009	0.192	0.045	0.047	99.6	70.9	5.8
Syncc9902_1578	hypothetical protein	0.031	0.202	0.155	0.189	100.0	100.0	8.6
Syncc9902_1579	putative ABC transporter, oligopeptides	0.011	0.240	0.045	0.045	100.0	98.8	9.3
Syncc9902_1580	periplasmic sensor signal transduction histidine kinase	0.010	0.096	0.101	0.112	99.8	31.0	4.0
Syncc9902_1581	hypothetical protein	0.020	0.176	0.112	0.118	100.0	88.4	7.8
Syncc9902_1582	GTP-binding protein LepA	0.022	0.440	0.050	0.048	100.0	100.0	27.2
Syncc9902_1583	putative SMR family multidrug efflux transporter	0.035	0.340	0.104	0.089	100.0	97.9	15.6
Syncc9902_1584	malate:quinone oxidoreductase (EC:1.1.99.16)	0.026	0.272	0.095	0.095	100.0	100.0	16.2
Syncc9902_1585	NifU-like protein	0.011	0.305	0.036	0.049	100.0	100.0	15.2
Syncc9902_1586	putative L-cysteine/cystine lyase	0.035	0.251	0.138	0.132	100.0	94.5	16.4
Syncc9902_1587	UDP-N-acetylmuramoylalanyl-D-glutamate--2, 6-diaminopimelate ligase	0.040	0.239	0.169	0.170	100.0	100.0	19.6
Syncc9902_1588	ribonucleotide reductase (class II)	0.046	0.214	0.217	0.244	100.0	100.0	16.4
Syncc9902_1589	Protein of unknown function DUF37	0.032	0.212	0.149	0.148	100.0	100.0	17.8
Syncc9902_1590	30S ribosomal protein S4	0.003	0.138	0.022	0.022	100.0	100.0	16.7
Syncc9902_1591	possible carbamoyl-phosphate synthase L chain	0.087	0.412	0.211	0.215	100.0	95.0	22.4
Syncc9902_1592	hypothetical protein	0.045	0.282	0.161	0.171	100.0	100.0	43.1

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_1593	possible mechanosensitive ion channel, MscS family	0.022	0.148	0.148	0.146	100.0	100.0	15.0
Syncc9902_1594	Maf-like protein	0.063	0.212	0.297	0.315	100.0	100.0	12.5
Syncc9902_1595	hypothetical protein	0.048	0.325	0.146	0.150	100.0	100.0	27.2
Syncc9902_1596	cobyric acid synthase	0.053	0.377	0.141	0.143	100.0	100.0	27.9
Syncc9902_1597	hypothetical protein	0.059	0.387	0.153	0.154	100.0	100.0	29.8
Syncc9902_1598	hypothetical protein	0.123	0.228	0.542	0.567	100.0	100.0	25.2
Syncc9902_1599	hypothetical protein	0.092	0.325	0.284	0.289	100.0	100.0	19.3
Syncc9902_1600	transcriptional regulator, LuxR family	0.052	0.311	0.166	0.166	100.0	100.0	18.5
Syncc9902_1601	hypothetical protein	0.078	0.311	0.250	0.244	100.0	100.0	20.2
Syncc9902_1602	possible chloride channel	0.065	0.270	0.239	0.241	100.0	100.0	26.2
Syncc9902_1603	ATPase	0.024	0.341	0.072	0.070	100.0	100.0	33.9
Syncc9902_1604	hypothetical protein	0.047	0.339	0.138	0.146	100.0	100.0	17.7
Syncc9902_1605	CO2 hydration	0.022	0.296	0.074	0.077	100.0	100.0	25.6
Syncc9902_1606	Proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.99.5)	0.014	0.288	0.049	0.054	100.0	100.0	22.5
Syncc9902_1607	NADH dehydrogenase subunit L (EC:1.6.99.5)	0.009	0.226	0.042	0.040	100.0	100.0	17.2
Syncc9902_1608	possible carbon dioxide concentrating mechanism protein CcmK	0.024	0.186	0.131	0.133	100.0	96.0	11.3
Syncc9902_1609	putative carboxysome peptide B	0.007	0.199	0.036	0.036	100.0	100.0	12.5
Syncc9902_1610	putative carboxysome peptide A	0.015	0.198	0.075	0.075	100.0	100.0	13.7
Syncc9902_1611	putative carboxysome shell polypeptide CsoS3	0.013	0.212	0.064	0.065	100.0	99.4	14.0
Syncc9902_1612	putative carboxysome structural peptide CsoS2	0.021	0.246	0.087	0.090	100.0	100.0	10.6
Syncc9902_1613	Ribulose-bisphosphate carboxylase (EC:4.1.1.39)	0.000	0.055	0.000	0.000	100.0	92.4	5.6
Syncc9902_1614	methionine sulfoxide reductase A (EC:4.1.1.39)	0.004	0.183	0.023	0.020	100.0	96.3	12.8
Syncc9902_1615	carboxysome shell peptide, CsoS1	0.002	0.061	0.040	0.040	100.0	59.9	4.9
Syncc9902_1616	Ham1 protein-like	0.027	0.248	0.107	0.110	100.0	70.3	7.0
Syncc9902_1617	hypothetical protein	0.009	0.215	0.041	0.040	100.0	99.6	11.0
Syncc9902_1618	hypothetical protein	0.021	0.253	0.083	0.097	100.0	86.2	8.1
Syncc9902_1619	light-independent protochlorophyllide reductase subunit N	0.009	0.214	0.043	0.051	100.0	100.0	10.7
Syncc9902_1620	light-independent protochlorophyllide reductase subunit B	0.018	0.235	0.076	0.079	100.0	84.4	7.9
Syncc9902_1621	Light-independent protochlorophyllide reductase, iron-sulfur ATP-binding protein (EC:1.18.6.1)	0.009	0.237	0.036	0.035	100.0	94.1	8.6
Syncc9902_1622	hypothetical protein	0.047	0.154	0.305	0.290	100.0	81.3	6.2
Syncc9902_1623	protochlorophyllide oxidoreductase (EC:1.3.1.33)	0.034	0.262	0.128	0.128	100.0	93.0	10.3
Syncc9902_1624	hypothetical protein	0.024	0.182	0.130	0.132	100.0	100.0	11.3
Syncc9902_1625	photosystem I reaction centre subunit XII	0.005	0.108	0.047	0.047	100.0	83.3	6.2
Syncc9902_1626	hypothetical protein	0.018	0.208	0.086	0.084	100.0	88.1	7.8
Syncc9902_1627	hypothetical protein	0.015	0.150	0.099	0.110	100.0	77.2	6.2
Syncc9902_1628	biotin/lipoate A/B protein ligase family	0.033	0.149	0.225	0.209	95.5	53.4	4.9
Syncc9902_1629	hypothetical protein	0.020	0.200	0.098	0.106	99.5	52.3	5.0
Syncc9902_1630	Phosphoribosylanthranilate isomerase (EC:5.3.1.24)	0.044	0.227	0.195	0.193	100.0	81.6	7.6
Syncc9902_1631	GTP cyclohydrolase I (EC:3.5.4.16)	0.008	0.238	0.033	0.033	100.0	69.4	5.6
Syncc9902_1632	short chain dehydrogenase	0.015	0.121	0.127	0.123	100.0	42.4	4.5
Syncc9902_1633	acetyl-CoA carboxylase alpha subunit (EC:6.4.1.2)	0.009	0.182	0.050	0.051	100.0	29.7	4.0
Syncc9902_1634	hypothetical protein	0.009	0.167	0.055	0.054	100.0	46.7	4.3
Syncc9902_1635	hypothetical protein	0.007	0.188	0.039	0.031	100.0	51.3	6.0
Syncc9902_1636	creatininase	0.020	0.233	0.084	0.091	90.8	49.4	5.0
Syncc9902_1637	RNA binding S1	0.005	0.133	0.041	0.041	100.0	61.1	5.6
Syncc9902_1638	hypothetical protein	0.032	0.219	0.148	0.148	100.0	79.0	7.3
Syncc9902_1639	Protein of unknown function DUF152	0.019	0.135	0.141	0.142	100.0	44.1	4.7
Syncc9902_1640	hypothetical protein	0.021	0.177	0.118	0.122	100.0	53.4	6.0
Syncc9902_1641	hypothetical protein	0.028	0.153	0.182	0.169	100.0	51.8	4.4
Syncc9902_1642	acetolactate synthase III large subunit	0.019	0.185	0.104	0.104	98.6	47.2	4.2
Syncc9902_1643	ferrochelatase (EC:4.99.1.1)	0.010	0.189	0.053	0.053	100.0	41.8	4.6
Syncc9902_1644	phage integrase family	0.036	0.184	0.195	0.191	100.0	73.1	6.0
Syncc9902_1645	possible methyltransferase	0.021	0.193	0.109	0.105	98.1	48.3	5.1
Syncc9902_1646	corrinoid adenosyltransferase BtuR/CobO/CobP (EC:2.5.1.17)	0.009	0.193	0.049	0.045	100.0	62.7	5.3
Syncc9902_1647	hypothetical protein	0.032	0.143	0.221	0.128	100.0	24.5	2.5
Syncc9902_1648	uridylate kinase	0.008	0.186	0.045	0.045	100.0	99.9	9.0
Syncc9902_1649	ribosome releasing factor	0.011	0.235	0.046	0.043	100.0	100.0	7.8
Syncc9902_1650	Geranylgeranyl reductase	0.040	0.312	0.129	0.128	100.0	100.0	14.4
Syncc9902_1651	Deoxyribodipyrimidine photolyase (EC:4.1.99.3)	0.045	0.293	0.154	0.157	100.0	100.0	15.4
Syncc9902_1652	light-harvesting 1 (B870) complex assembly protein PucC-like	0.070	0.252	0.276	0.289	100.0	100.0	17.4
Syncc9902_1653	transaldolase (EC:2.2.1.2)	0.013	0.303	0.043	0.043	100.0	100.0	18.1
Syncc9902_1654	Peptidoglycan glycosyltransferase (EC:2.4.1.129)	0.021	0.248	0.085	0.079	100.0	81.8	11.7
Syncc9902_1655	hypothetical protein	0.020	0.099	0.200	0.199	100.0	100.0	10.9
Syncc9902_1656	possible membrane associated protease	0.033	0.183	0.179	0.179	100.0	86.6	9.0
Syncc9902_1657	putative alpha-ribazole-5'-P phosphatase	0.021	0.137	0.151	0.146	100.0	85.5	6.3

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_1658	putative dihydroorotase	0.026	0.175	0.146	0.152	92.8	38.5	3.4
Syncc9902_1659	Peptidase S26A, signal peptidase I	0.015	0.175	0.087	0.093	100.0	54.4	5.0
Syncc9902_1660	hypothetical protein	0.019	0.160	0.120	0.124	90.7	23.2	3.1
Syncc9902_1661	hypothetical protein	0.029	0.148	0.197	0.257	100.0	26.5	3.8
Syncc9902_1662	possible ferredoxin (2Fe-2S)	0.007	0.073	0.093	0.091	100.0	54.3	5.3
Syncc9902_1663	Inorganic diphosphatase (EC:3.6.1.1)	0.015	0.065	0.231	0.232	93.5	31.2	4.0
Syncc9902_1664	hypothetical protein	0.013	0.169	0.080	0.075	100.0	44.9	4.5
Syncc9902_1665	prolyl-tRNA synthetase	0.004	0.125	0.029	0.029	92.7	6.6	2.3
Syncc9902_1666	possible photosystem II Psb27 protein	0.013	0.248	0.051	0.044	100.0	56.8	4.5
Syncc9902_1667	adenylosuccinate synthetase (EC:6.3.4.4)	0.003	0.158	0.020	0.020	100.0	65.6	5.8
Syncc9902_1668	putative carbohydrate kinase, PfkB family	0.004	0.148	0.030	0.033	100.0	72.5	5.5
Syncc9902_1669	putative divalent cation tolerance protein	0.036	0.108	0.332	0.345	100.0	100.0	7.6
Syncc9902_1670	precorrin-6X reductase	0.036	0.152	0.236	0.228	100.0	61.1	5.1
Syncc9902_1671	hypothetical protein	0.026	0.155	0.169	0.162	100.0	95.9	7.6
Syncc9902_1672	hypothetical protein	0.026	0.127	0.205	0.233	100.0	100.0	9.2
Syncc9902_1673	hypothetical protein	0.011	0.214	0.049	0.058	100.0	82.3	6.0
Syncc9902_1674	acetylglutamate kinase (EC:2.7.2.8)	0.016	0.229	0.069	0.068	100.0	98.1	9.5
Syncc9902_1675	hypothetical protein	0.029	0.278	0.103	0.104	100.0	100.0	8.4
Syncc9902_1676	hypothetical protein	0.035	0.178	0.196	0.183	100.0	97.4	7.9
Syncc9902_1677	Primosomal protein n	0.028	0.202	0.139	0.143	100.0	96.9	9.3
Syncc9902_1678	RNA polymerase sigma factor	0.014	0.267	0.053	0.053	100.0	93.2	12.8
Syncc9902_1679	hypothetical protein	0.028	0.235	0.119	0.118	100.0	100.0	14.3
Syncc9902_1680	porphobilinogen deaminase (EC:2.5.1.61)	0.019	0.268	0.073	0.071	100.0	100.0	17.4
Syncc9902_1681	hypothetical protein	0.017	0.134	0.126	0.126	100.0	61.8	5.7
Syncc9902_1682	hypothetical protein	0.040	0.212	0.187	0.187	100.0	100.0	8.9
Syncc9902_1683	hypothetical protein	0.041	0.298	0.136	0.134	100.0	100.0	16.1
Syncc9902_1684	Inorganic diphosphatase (EC:3.6.1.1)	0.005	0.186	0.029	0.033	100.0	100.0	10.5
Syncc9902_1685	Thermostable carboxypeptidase 1 (EC:3.4.17.19)	0.021	0.273	0.078	0.082	100.0	100.0	13.2
Syncc9902_1686	hypothetical protein	0.021	0.192	0.108	0.104	100.0	45.5	4.8
Syncc9902_1687	hypothetical protein	0.038	0.237	0.162	0.165	100.0	100.0	6.7
Syncc9902_1688	hypothetical protein	0.026	0.202	0.127	0.134	100.0	100.0	11.4
Syncc9902_1689	4A-hydroxytetrahydrobiopterin dehydratase (PCD)	0.049	0.263	0.187	0.175	100.0	100.0	13.4
Syncc9902_1690	hypothetical protein	0.024	0.247	0.097	0.105	100.0	100.0	13.1
Syncc9902_1691	hypothetical protein	0.047	0.277	0.168	0.140	100.0	100.0	15.3
Syncc9902_1692	putative iron ABC transporter	0.015	0.234	0.062	0.070	100.0	99.7	10.9
Syncc9902_1693	hypothetical protein	0.036	0.200	0.181	0.175	100.0	100.0	10.2
Syncc9902_1694	hypothetical protein	0.041	0.231	0.175	0.178	100.0	91.0	9.7
Syncc9902_1695	hypothetical protein	0.015	0.217	0.068	0.066	100.0	75.9	8.0
Syncc9902_1696	hypothetical protein	0.018	0.155	0.115	0.110	100.0	89.2	9.6
Syncc9902_1697	hypothetical protein	0.097	0.179	0.541	0.502	99.1	17.5	3.0
Syncc9902_1698	hypothetical protein	0.047	0.207	0.229	0.237	100.0	83.5	8.1
Syncc9902_1699	Exodeoxyribonuclease III xth	0.032	0.239	0.133	0.130	100.0	80.1	8.7
Syncc9902_1700	putative glycolate oxidase subunit GlcD	0.015	0.219	0.069	0.069	100.0	100.0	15.5
Syncc9902_1701	glutamate-1-semialdehyde aminotransferase	0.007	0.300	0.025	0.024	100.0	100.0	19.9
Syncc9902_1702	hypothetical protein	0.044	0.129	0.337	0.327	100.0	89.7	7.5
Syncc9902_1703	hypothetical protein	0.047	0.348	0.135	0.138	100.0	100.0	21.8
Syncc9902_1704	possible phosphoribosyltransferase family protein	0.062	0.278	0.224	0.207	100.0	100.0	21.7
Syncc9902_1705	hypothetical protein	0.023	0.282	0.081	0.076	100.0	100.0	12.4
Syncc9902_1706	hypothetical protein	0.078	0.378	0.206	0.201	100.0	100.0	20.7
Syncc9902_1707	possible phosphate-binding protein	0.043	0.241	0.180	0.186	100.0	100.0	11.0
Syncc9902_1708	hypothetical protein	0.012	0.273	0.044	0.042	100.0	100.0	17.0
Syncc9902_1709	hypothetical protein	0.015	0.270	0.057	0.055	100.0	100.0	12.5
Syncc9902_1710	hypothetical protein	0.012	0.214	0.055	0.050	100.0	100.0	15.0
Syncc9902_1711	hypothetical protein	0.011	0.145	0.074	0.073	99.1	65.8	6.4
Syncc9902_1712	hypothetical protein	0.059	0.143	0.410	0.409	100.0	100.0	9.4
Syncc9902_1713	short chain dehydrogenase	0.044	0.160	0.272	0.279	100.0	96.7	7.9
Syncc9902_1714	methionine aminopeptidase	0.009	0.197	0.045	0.047	100.0	97.1	10.3
Syncc9902_1715	hypothetical protein	0.008	0.167	0.051	0.047	100.0	100.0	6.9
Syncc9902_1716	50S ribosomal protein L19	0.010	0.127	0.077	0.073	100.0	96.5	6.2
Syncc9902_1717	hypothetical protein	0.003	0.176	0.014	0.014	100.0	100.0	18.8
Syncc9902_1718	hypothetical protein	0.132	0.178	0.742	0.738	100.0	100.0	16.9
Syncc9902_1719	glutamyl-tRNA synthetase	0.037	0.274	0.135	0.136	100.0	100.0	19.5
Syncc9902_1720	putative Na+/H+ antiporter, CPA1 family	0.012	0.213	0.055	0.059	100.0	100.0	15.5
Syncc9902_1721	hypothetical protein	0.011	0.170	0.063	0.101	100.0	100.0	14.7
Syncc9902_1722	hypothetical protein	0.054	0.176	0.308	0.282	100.0	90.6	9.0
Syncc9902_1723	UDP-N-acetylglucosamine 2-epimerase (EC:5.1.3.14)	0.017	0.231	0.075	0.086	100.0	93.8	11.5
Syncc9902_1724	Pilus retraction protein PilT	0.040	0.183	0.219	0.226	100.0	91.6	9.1

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_1726	hypothetical protein	0.023	0.255	0.090	0.099	100.0	100.0	9.2
Syncc9902_1727	Peptidase M22, glycoprotease (EC:3.4.24.57)	0.019	0.183	0.101	0.115	100.0	98.7	7.8
Syncc9902_1728	photosystem I reaction center subunit III	0.006	0.074	0.086	0.086	92.2	62.3	7.3
Syncc9902_1729	photosystem I reaction center subunit IX	0.008	0.103	0.076	0.076	100.0	100.0	11.4
Syncc9902_1730	Guanylate kinase (EC:2.7.4.8)	0.033	0.191	0.172	0.169	100.0	100.0	12.3
Syncc9902_1731	secreted protein MPB70 precursor-like	0.021	0.160	0.130	0.132	100.0	84.2	7.6
Syncc9902_1732	Sec-independent periplasmic protein translocase	0.011	0.192	0.060	0.058	100.0	84.3	7.6
Syncc9902_1733	hypothetical protein	0.020	0.139	0.141	0.114	100.0	100.0	8.0
Syncc9902_1734	Twin-arginine translocation pathway signal	0.008	0.097	0.085	0.095	100.0	92.9	10.0
Syncc9902_1735	apocytochrome f precursor	0.014	0.196	0.072	0.072	100.0	82.8	12.1
Syncc9902_1736	prolipoprotein diacylglyceryl transferase	0.023	0.194	0.121	0.134	100.0	100.0	10.1
Syncc9902_1737	Precorrin-4 C11-methyltransferase region (EC:2.1.1.133)	0.022	0.172	0.126	0.114	100.0	100.0	11.3
Syncc9902_1738	transcriptional regulator, XRE family	0.036	0.168	0.212	0.200	100.0	59.9	9.3
Syncc9902_1739	Guanosine-5'-triphosphate,3'-diphosphate diphosphatase (EC:3.6.1.40)	0.018	0.196	0.092	0.097	100.0	78.7	7.6
Syncc9902_1740	probable 4-hydroxybenzoate-octaprenyltransferase	0.024	0.180	0.131	0.130	100.0	80.9	6.2
Syncc9902_1741	hypothetical protein	0.043	0.223	0.191	0.209	100.0	71.4	7.1
Syncc9902_1742	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	0.021	0.168	0.122	0.145	100.0	79.7	6.8
Syncc9902_1743	hypothetical protein	0.018	0.149	0.123	0.134	100.0	64.5	5.2
Syncc9902_1744	putative potassium channel, VIC family	0.016	0.175	0.093	0.095	100.0	92.9	8.0
Syncc9902_1745	3-oxoacyl-(acyl-carrier-protein) reductase	0.008	0.221	0.035	0.035	100.0	96.7	7.9
Syncc9902_1746	hypothetical protein	0.014	0.153	0.088	0.088	100.0	100.0	11.9
Syncc9902_1747	Chaperonin Cpn60/TCP-1	0.015	0.248	0.061	0.061	100.0	100.0	14.7
Syncc9902_1748	N-acetylmannosamine-6-phosphate 2-epimerase (EC:5.1.3.9)	0.041	0.283	0.145	0.143	100.0	100.0	23.1
Syncc9902_1749	hypothetical protein	0.080	0.176	0.455	0.438	100.0	80.0	15.7
Syncc9902_1750	putative multidrug efflux ABC transporter	0.019	0.231	0.084	0.083	100.0	100.0	28.1
Syncc9902_1751	ATPase	0.018	0.217	0.082	0.082	100.0	100.0	24.6
Syncc9902_1752	Protoheme IX farnesyltransferase	0.016	0.259	0.062	0.059	100.0	100.0	22.4
Syncc9902_1753	putative cytochrome c oxidase assembly protein	0.036	0.251	0.143	0.152	100.0	100.0	29.5
Syncc9902_1754	possible cytochrome c oxidase subunit II	0.017	0.235	0.071	0.068	100.0	100.0	30.0
Syncc9902_1755	Cytochrome-c oxidase (EC:1.9.3.1)	0.010	0.237	0.040	0.040	100.0	100.0	23.3
Syncc9902_1756	possible cytochrome c oxidase subunit III	0.018	0.203	0.089	0.087	100.0	100.0	19.9
Syncc9902_1757	Transcriptional regulator AbrB	0.026	0.177	0.148	0.159	100.0	100.0	19.9
Syncc9902_1758	riboflavin synthase subunit alpha	0.024	0.224	0.109	0.099	100.0	100.0	13.5
Syncc9902_1759	hypothetical protein	0.024	0.210	0.116	0.128	100.0	100.0	14.4
Syncc9902_1760	putative oxidoreductase, aldo/keto reductase family	0.033	0.196	0.170	0.168	100.0	98.8	14.3
Syncc9902_1761	hypothetical protein	0.044	0.195	0.224	0.221	100.0	93.2	8.8
Syncc9902_1762	hypothetical protein	0.040	0.173	0.232	0.226	100.0	97.2	10.4
Syncc9902_1763	hypothetical protein	0.013	0.181	0.073	0.070	100.0	96.2	9.8
Syncc9902_1764	hypothetical protein	0.038	0.145	0.259	0.270	98.2	73.0	7.5
Syncc9902_1765	DNA topoisomerase I (EC:5.99.1.2)	0.007	0.203	0.033	0.035	100.0	98.5	13.0
Syncc9902_1766	NAD(P)H-quinone oxidoreductase subunit 2 (EC:1.6.99.5)	0.008	0.224	0.038	0.036	100.0	90.2	9.2
Syncc9902_1767	ATPase	0.016	0.248	0.066	0.080	100.0	98.5	12.7
Syncc9902_1768	response regulator receiver domain protein (CheY-like)	0.071	0.204	0.349	0.313	100.0	46.6	8.6
Syncc9902_1769	histidine kinase	0.045	0.171	0.263	0.266	100.0	98.9	13.1
Syncc9902_1770	putative PAS/PAC sensor protein	0.014	0.271	0.053	0.046	100.0	100.0	12.1
Syncc9902_1771	two component transcriptional regulator, winged helix family	0.015	0.245	0.060	0.060	100.0	83.8	13.0
Syncc9902_1772	peptidase family M23/M37	0.050	0.225	0.221	0.205	100.0	100.0	12.7
Syncc9902_1773	Biotin--acetyl-CoA-carboxylase ligase	0.044	0.225	0.197	0.187	100.0	100.0	20.0
Syncc9902_1774	aminotransferase	0.037	0.230	0.162	0.159	100.0	100.0	14.0
Syncc9902_1775	thioredoxin	0.020	0.251	0.081	0.079	100.0	100.0	19.2
Syncc9902_1776	membrane-associated 30 kD protein-like	0.019	0.270	0.069	0.069	100.0	100.0	15.1
Syncc9902_1777	hypothetical protein	0.034	0.238	0.143	0.135	100.0	100.0	15.4
Syncc9902_1778	hypothetical protein	0.035	0.210	0.165	0.164	100.0	100.0	13.3
Syncc9902_1779	putative UDP-glucose:tetrahydrobiopterin glucosyltransferase	0.032	0.166	0.195	0.185	100.0	100.0	12.0
Syncc9902_1780	SMR family transporter-like	0.013	0.167	0.080	0.080	100.0	100.0	10.3
Syncc9902_1781	Peptidase S49, SppA	0.007	0.172	0.042	0.041	100.0	96.7	7.4
Syncc9902_1782	Chorismate mutase of the AroH class	0.022	0.155	0.141	0.167	100.0	81.5	9.3
Syncc9902_1783	hypothetical protein	0.020	0.219	0.092	0.090	100.0	87.2	9.0
Syncc9902_1784	50S ribosomal protein L34	0.006	0.164	0.036	0.036	100.0	57.3	5.3
Syncc9902_1785	ribonuclease P protein component of ribozyme	0.023	0.199	0.114	0.143	100.0	67.7	6.0
Syncc9902_1786	hypothetical protein	0.001	0.167	0.007	0.007	100.0	94.6	6.5
Syncc9902_1787	putative inner membrane protein translocase component YidC	0.011	0.209	0.050	0.051	100.0	82.6	6.6
Syncc9902_1788	hypothetical protein	0.024	0.147	0.161	0.189	100.0	54.9	4.8
Syncc9902_1789	ATPase	0.015	0.172	0.085	0.087	100.0	75.0	5.5
Syncc9902_1790	hypothetical protein	0.000	0.000	NaN	NaN	100.0	0.0	2.4
Syncc9902_1791	seryl-tRNA synthetase (EC:6.1.1.11)	0.011	0.204	0.056	0.060	100.0	65.3	8.9

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_1792	Peptidase M50, putative membrane-associated zinc metallopeptidase	0.015	0.209	0.073	0.073	100.0	97.1	14.2
Syncc9902_1793	30S ribosomal protein S14	0.005	0.218	0.021	0.020	100.0	100.0	8.4
Syncc9902_1794	RNA binding S1	0.008	0.179	0.044	0.046	100.0	93.4	12.8
Syncc9902_1795	CysQ protein-like	0.017	0.221	0.078	0.082	100.0	100.0	21.8
Syncc9902_1796	hypothetical protein	0.036	0.251	0.144	0.142	100.0	100.0	19.7
Syncc9902_1797	hypothetical protein	0.061	0.230	0.264	0.212	100.0	100.0	18.8
Syncc9902_1798	hypothetical protein	0.027	0.222	0.121	0.125	100.0	100.0	12.5
Syncc9902_1799	hypothetical protein	0.029	0.378	0.077	0.076	100.0	100.0	21.0
Syncc9902_1800	hypothetical protein	0.048	0.098	0.493	0.493	100.0	100.0	8.1
Syncc9902_1801	hypothetical protein	0.190	0.165	1.156	1.181	100.0	100.0	28.2
Syncc9902_1802	hypothetical protein	0.077	0.388	0.199	0.195	100.0	100.0	46.6
Syncc9902_1803	hypothetical protein	0.070	0.351	0.200	0.211	100.0	100.0	42.9
Syncc9902_1804	hypothetical protein	0.077	0.086	0.894	0.881	100.0	100.0	15.2
Syncc9902_1805	hypothetical protein	0.064	0.331	0.194	0.203	100.0	100.0	21.4
Syncc9902_1806	hypothetical protein	0.030	0.299	0.100	0.104	100.0	100.0	23.9
Syncc9902_1807	hypothetical protein	0.023	0.330	0.069	0.066	100.0	100.0	18.7
Syncc9902_1808	hypothetical protein	0.065	0.371	0.175	0.169	100.0	100.0	16.6
Syncc9902_1809	hypothetical protein	0.072	0.219	0.331	0.333	100.0	100.0	16.5
Syncc9902_1810	Band 7 protein	0.012	0.402	0.031	0.033	100.0	100.0	28.8
Syncc9902_1811	membrane protein-like	0.085	0.320	0.267	0.265	100.0	100.0	22.8
Syncc9902_1812	salt-stress induced hydrophobic peptide	0.014	0.312	0.043	0.042	100.0	100.0	12.7
Syncc9902_1813	possible pfkB family carbohydrate kinase	0.056	0.172	0.328	0.289	100.0	91.6	10.9
Syncc9902_1814	Photosystem II reaction centre protein PsbA/D1	0.035	0.410	0.085	0.086	100.0	100.0	28.2
Syncc9902_1815	Glyceraldehyde-3-phosphate dehydrogenase, type I (EC:1.2.1.12)	0.026	0.303	0.087	0.083	100.0	100.0	15.5
Syncc9902_1816	GCN5-related N-acetyltransferase	0.050	0.301	0.166	0.160	100.0	96.9	11.4
Syncc9902_1817	Photosystem II reaction centre protein PsbA/D1	0.030	0.402	0.073	0.073	100.0	100.0	22.5
Syncc9902_1818	hypothetical protein	0.073	0.271	0.269	0.277	100.0	100.0	18.5
Syncc9902_1819	aspartoacylase (EC:3.5.1.15)	0.045	0.223	0.203	0.198	100.0	100.0	16.9
Syncc9902_1820	putative glutathione S-transferase	0.037	0.201	0.187	0.205	100.0	100.0	15.9
Syncc9902_1821	hypothetical protein	0.047	0.267	0.177	0.148	100.0	100.0	16.5
Syncc9902_1822	ATPase	0.021	0.247	0.086	0.088	100.0	100.0	20.5
Syncc9902_1823	hypothetical protein	0.035	0.278	0.125	0.128	100.0	100.0	22.0
Syncc9902_1824	Small GTP-binding protein domain	0.012	0.269	0.046	0.046	100.0	100.0	18.5
Syncc9902_1825	ATPase	0.033	0.213	0.153	0.154	100.0	100.0	12.4
Syncc9902_1826	MutS 2 protein	0.026	0.250	0.106	0.109	100.0	94.7	15.8
Syncc9902_1827	possible ring-cleaving dioxygenase	0.031	0.199	0.154	0.140	100.0	100.0	12.1
Syncc9902_1828	delta-aminolevulinic acid dehydratase (EC:4.2.1.24)	0.009	0.255	0.037	0.037	100.0	96.3	13.5
Syncc9902_1829	Heat shock protein DnaJ-like	0.017	0.260	0.066	0.065	100.0	90.4	12.7
Syncc9902_1830	hypothetical protein	0.036	0.193	0.186	0.182	100.0	100.0	9.7
Syncc9902_1831	putative sulfate transporter	0.016	0.190	0.083	0.090	100.0	97.7	10.3
Syncc9902_1832	tRNA-(MS(2)IO(6)A)-hydroxylase-like	0.030	0.227	0.131	0.150	100.0	76.3	7.8
Syncc9902_1833	3-dehydroquinate dehydratase (EC:4.2.1.10)	0.030	0.219	0.139	0.141	100.0	100.0	10.2
Syncc9902_1834	Alpha amylase, catalytic subdomain	0.030	0.178	0.170	0.171	100.0	93.2	8.6
Syncc9902_1835	putative glycerol kinase	0.014	0.157	0.091	0.085	100.0	80.4	7.5
Syncc9902_1836	putative glycerol-3-phosphate dehydrogenase	0.023	0.173	0.134	0.131	100.0	94.8	8.0
Syncc9902_1837	Alpha amylase, catalytic subdomain	0.007	0.151	0.050	0.053	100.0	87.2	8.5
Syncc9902_1838	putative aldehyde dehydrogenase	0.024	0.185	0.130	0.137	100.0	95.1	9.4
Syncc9902_1839	Peptidoglycan-binding LysM	0.031	0.187	0.166	0.168	100.0	100.0	14.0
Syncc9902_1840	ATP-dependent DNA helicase RecQ	0.013	0.212	0.060	0.057	100.0	91.9	11.8
Syncc9902_1841	hypothetical protein	0.007	0.204	0.033	0.040	100.0	91.8	9.5
Syncc9902_1842	photosystem I reaction center subunit IV	0.009	0.103	0.085	0.085	100.0	100.0	10.7
Syncc9902_1843	Formamidopyrimidine-DNA glycolase (EC:3.2.2.23)	0.027	0.236	0.113	0.111	100.0	100.0	11.7
Syncc9902_1844	Initiation factor 2B alpha/beta/delta (EC:5.3.1.23)	0.043	0.267	0.159	0.156	100.0	100.0	10.0
Syncc9902_1845	putative sugar aldolase	0.065	0.260	0.250	0.241	100.0	100.0	11.5
Syncc9902_1846	2,3-diketo-5-methylthio-1-phosphopentane phosphatase	0.088	0.319	0.276	0.259	100.0	100.0	18.6
Syncc9902_1847	putative neutral invertase-like protein	0.008	0.166	0.050	0.052	100.0	83.2	7.8
Syncc9902_1848	cytochrome b6-f complex subunit 4	0.006	0.263	0.025	0.024	100.0	87.6	7.5
Syncc9902_1849	cytochrome b6	0.003	0.135	0.019	0.018	100.0	55.0	5.6
Syncc9902_1850	Peptidase S41A, C-terminal protease (EC:3.4.21.102)	0.019	0.196	0.097	0.089	100.0	100.0	14.3
Syncc9902_1851	metal dependent phosphohydrolase	0.025	0.182	0.140	0.131	100.0	86.5	9.7
Syncc9902_1852	possible septum site-determining protein	0.048	0.165	0.292	0.290	100.0	93.5	9.1
Syncc9902_1853	Septum site-determining protein MinD	0.004	0.152	0.027	0.023	100.0	99.8	10.7
Syncc9902_1854	Septum formation topological specificity factor MinE	0.014	0.244	0.059	0.060	100.0	100.0	11.1
Syncc9902_1855	transcriptional regulator, LuxR family	0.060	0.196	0.306	0.284	100.0	100.0	11.1
Syncc9902_1856	hypothetical protein	0.077	0.236	0.326	0.308	100.0	100.0	14.1
Syncc9902_1857	hypothetical protein	0.036	0.201	0.177	0.190	100.0	97.5	7.4

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_1858	Modification methylase HemK	0.036	0.148	0.247	0.257	100.0	100.0	10.1
Syncc9902_1859	SMF protein	0.045	0.181	0.247	0.251	96.7	70.7	7.9
Syncc9902_1860	hypothetical protein	0.030	0.160	0.188	0.218	98.9	63.2	6.7
Syncc9902_1861	hypothetical protein	0.010	0.224	0.043	0.041	100.0	95.3	9.0
Syncc9902_1862	possible photosystem II reaction center M protein	0.004	0.029	0.147	0.147	100.0	100.0	7.6
Syncc9902_1863	possible ferredoxin (2Fe-2S)	0.031	0.218	0.144	0.164	100.0	97.8	8.0
Syncc9902_1864	photosystem II chlorophyll-binding protein CP47	0.008	0.206	0.039	0.040	100.0	70.0	5.7
Syncc9902_1865	photosystem II reaction center T protein	0.000	0.048	0.000	0.000	100.0	100.0	14.6
Syncc9902_1866	Protein of unknown function DUF193	0.012	0.215	0.054	0.053	100.0	100.0	8.0
Syncc9902_1867	30S ribosomal protein S1	0.005	0.176	0.026	0.030	100.0	62.7	5.4
Syncc9902_1868	hypothetical protein	0.035	0.144	0.245	0.237	100.0	91.0	9.5
Syncc9902_1869	S-adenosylmethionine synthetase (EC:2.5.1.6)	0.007	0.192	0.034	0.035	100.0	100.0	13.8
Syncc9902_1870	putative carbohydrate kinase, FGGY family	0.042	0.194	0.218	0.215	100.0	100.0	15.0
Syncc9902_1871	possible phycobilisome linker polypeptide	0.024	0.211	0.115	0.124	100.0	100.0	18.0
Syncc9902_1872	hypothetical protein	0.027	0.193	0.137	0.157	100.0	100.0	17.4
Syncc9902_1873	hypothetical protein	0.071	0.159	0.444	0.457	100.0	100.0	13.3
Syncc9902_1874	hypothetical protein	0.244	0.098	2.493	2.901	100.0	22.9	3.8
Syncc9902_1875	hypothetical protein	0.065	0.055	1.192	1.220	100.0	95.6	7.9
Syncc9902_1876	hypothetical protein	0.026	0.234	0.111	0.110	100.0	100.0	16.2
Syncc9902_1877	hypothetical protein	0.035	0.197	0.178	0.194	100.0	80.4	10.0
Syncc9902_1878	hypothetical protein	0.020	0.329	0.060	0.061	100.0	100.0	10.7
Syncc9902_1879	hypothetical protein	0.041	0.201	0.203	0.219	100.0	100.0	19.5
Syncc9902_1880	possible phycobilisome polypeptide	0.049	0.195	0.254	0.280	100.0	100.0	9.6
Syncc9902_1881	possible phycobilisome rod-core linker polypeptide (L-RC 28.5)	0.034	0.232	0.146	0.153	100.0	100.0	23.0
Syncc9902_1882	hypothetical protein	0.045	0.269	0.166	0.155	100.0	100.0	22.4
Syncc9902_1883	phycobilisome linker polypeptide	0.017	0.270	0.063	0.068	100.0	100.0	26.9
Syncc9902_1884	phycobilisome linker polypeptide	0.018	0.244	0.073	0.073	100.0	100.0	20.2
Syncc9902_1885	possible phycobilisome linker polypeptide	0.038	0.309	0.124	0.126	100.0	100.0	18.0
Syncc9902_1886	conserved hypothetical protein CpeS	0.030	0.301	0.100	0.095	100.0	87.5	10.2
Syncc9902_1887	CpeT-like	0.071	0.334	0.213	0.208	100.0	100.0	27.6
Syncc9902_1888	CpeR-like, phycoerythrin linker-proteins region	0.035	0.531	0.065	0.079	100.0	100.0	31.7
Syncc9902_1889	hypothetical protein	0.045	0.333	0.135	0.141	100.0	100.0	22.2
Syncc9902_1890	hypothetical protein	0.086	0.286	0.300	0.295	100.0	100.0	24.9
Syncc9902_1891	hypothetical protein	0.047	0.247	0.191	0.203	100.0	100.0	25.1
Syncc9902_1892	CpeY-like	0.031	0.196	0.157	0.159	100.0	100.0	32.4
Syncc9902_1893	C-phycoerythrin class II beta chain	0.010	0.211	0.049	0.053	100.0	100.0	39.1
Syncc9902_1894	C-phycoerythrin class II alpha chain	0.029	0.304	0.095	0.094	100.0	100.0	47.0
Syncc9902_1895	C-phycoerythrin class II gamma chain, linker polypeptide	0.026	0.246	0.104	0.102	100.0	100.0	25.6
Syncc9902_1896	PBS lyase HEAT-like repeat	0.042	0.210	0.198	0.199	100.0	100.0	26.0
Syncc9902_1897	putative bilin biosynthesis protein cpeZ	0.034	0.210	0.161	0.162	100.0	100.0	44.4
Syncc9902_1898	putative bilin biosynthesis protein (CpeY)	0.072	0.218	0.329	0.323	100.0	100.0	30.4
Syncc9902_1899	hypothetical protein	0.058	0.182	0.319	0.296	100.0	100.0	25.6
Syncc9902_1900	hypothetical protein	0.030	0.194	0.154	0.158	100.0	100.0	17.9
Syncc9902_1901	PBS lyase HEAT-like repeat	0.033	0.206	0.161	0.167	100.0	100.0	23.6
Syncc9902_1902	C-phycoerythrin class I alpha chain	0.018	0.242	0.074	0.074	100.0	100.0	25.8
Syncc9902_1903	C-phycoerythrin class I beta chain	0.010	0.181	0.055	0.055	100.0	100.0	40.4
Syncc9902_1904	hypothetical protein	0.004	0.211	0.019	0.009	100.0	100.0	23.4
Syncc9902_1905	hypothetical protein	0.040	0.329	0.122	0.119	100.0	100.0	39.8
Syncc9902_1906	15,16 dihydrobiliverdin:ferredoxin oxidoreductase	0.025	0.274	0.090	0.091	100.0	100.0	34.3
Syncc9902_1907	phycoerythrobilin:ferredoxin oxidoreductase	0.038	0.253	0.151	0.145	100.0	100.0	20.7
Syncc9902_1908	Phycocyanin, beta subunit	0.017	0.220	0.077	0.077	100.0	100.0	22.5
Syncc9902_1909	Phycocyanin, alpha subunit	0.009	0.229	0.037	0.045	100.0	100.0	16.9
Syncc9902_1910	hypothetical protein	0.035	0.215	0.161	0.159	100.0	100.0	13.8
Syncc9902_1911	PBS lyase HEAT-like repeat	0.056	0.264	0.212	0.214	100.0	100.0	19.5
Syncc9902_1912	phycocyanobilin lyase beta subunit	0.039	0.214	0.184	0.174	100.0	100.0	12.5
Syncc9902_1913	protein tyrosine phosphatase (EC:3.1.3.48)	0.035	0.217	0.161	0.152	100.0	100.0	15.0
Syncc9902_1914	Cytidylate kinase (EC:6.3.2.1,EC:2.7.4.14)	0.037	0.225	0.162	0.173	100.0	100.0	13.3
Syncc9902_1915	Rare lipoprotein A	0.030	0.137	0.223	0.225	100.0	99.5	9.7
Syncc9902_1916	phosphoribosylaminoimidazole synthetase (EC:6.3.3.1)	0.024	0.224	0.108	0.123	100.0	100.0	11.3
Syncc9902_1917	hypothetical protein	0.007	0.262	0.027	0.030	100.0	84.1	8.8
Syncc9902_1918	hypothetical protein	0.035	0.207	0.167	0.162	100.0	99.8	8.9
Syncc9902_1919	hypothetical protein	0.035	0.069	0.506	0.427	100.0	57.0	4.8
Syncc9902_1920	hypothetical protein	0.040	0.201	0.198	0.191	100.0	95.7	10.1
Syncc9902_1921	hypothetical protein	0.027	0.224	0.120	0.127	100.0	100.0	6.3
Syncc9902_1922	3'-5' exonuclease	0.012	0.173	0.070	0.068	100.0	100.0	12.1
Syncc9902_1923	possible Pex protein	0.017	0.223	0.074	0.080	100.0	100.0	10.0
Syncc9902_1924	hypothetical protein	0.014	0.166	0.087	0.087	100.0	53.6	5.2

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_1925	hypothetical protein	0.036	0.148	0.245	0.239	100.0	100.0	8.0
Syncc9902_1926	coproporphyrinogen III oxidase (EC:1.3.3.3)	0.008	0.126	0.065	0.062	100.0	77.9	7.0
Syncc9902_1927	MRP protein-like	0.013	0.189	0.071	0.074	100.0	82.3	6.6
Syncc9902_1928	cell division protein possibly involved in shape determination	0.011	0.131	0.083	0.084	100.0	57.3	4.6
Syncc9902_1929	hypothetical protein	0.021	0.153	0.139	0.141	97.6	37.9	4.0
Syncc9902_1930	photosystem I reaction center subunit II (PsaD)	0.011	0.189	0.056	0.062	100.0	89.8	6.6
Syncc9902_1931	Anthranilate synthase (EC:4.1.3.27)	0.013	0.214	0.060	0.065	94.7	56.7	4.8
Syncc9902_1932	Glutamate--cysteine ligase, putative	0.013	0.148	0.089	0.094	100.0	70.7	5.4
Syncc9902_1933	phosphoenolpyruvate carboxylase (EC:4.1.1.31)	0.007	0.168	0.039	0.039	97.4	89.4	8.3
Syncc9902_1934	hypothetical protein	0.020	0.108	0.186	0.226	100.0	57.0	7.2
Syncc9902_1935	RecF protein	0.011	0.173	0.061	0.065	100.0	100.0	9.0
Syncc9902_1936	S-adenosylmethionine decarboxylase proenzyme (EC:4.1.1.50)	0.009	0.100	0.092	0.092	100.0	98.3	9.0
Syncc9902_1937	hypothetical protein	0.016	0.175	0.090	0.093	100.0	87.5	7.3
Syncc9902_1938	cob(I)yrinic acid a,c-diamide adenosyltransferase	0.008	0.221	0.037	0.043	100.0	99.7	10.8
Syncc9902_1939	hypothetical protein	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_1940	Rhodanese-like	0.032	0.209	0.152	0.157	100.0	100.0	15.0
Syncc9902_1941	hypothetical protein	0.066	0.183	0.363	0.366	100.0	100.0	13.1
Syncc9902_1942	hypothetical protein	0.013	0.194	0.065	0.059	100.0	100.0	17.3
Syncc9902_1943	hypothetical protein	0.052	0.247	0.212	0.214	100.0	100.0	11.9
Syncc9902_1944	phytoene dehydrogenase related enzyme	0.028	0.182	0.154	0.145	100.0	100.0	11.5
Syncc9902_1945	prephenate dehydrogenase	0.017	0.145	0.117	0.129	100.0	100.0	10.6
Syncc9902_1946	hypothetical protein	0.011	0.139	0.081	0.087	100.0	72.5	5.5
Syncc9902_1947	hypothetical protein	0.011	0.172	0.066	0.074	100.0	84.3	5.5
Syncc9902_1948	hypothetical protein	0.044	0.222	0.200	0.192	100.0	100.0	9.9
Syncc9902_1949	recombinase A	0.005	0.129	0.038	0.038	100.0	50.4	5.0
Syncc9902_1950	hypothetical protein	0.024	0.143	0.167	0.167	98.4	57.5	4.7
Syncc9902_1951	ATPase	0.013	0.138	0.097	0.095	100.0	64.6	5.0
Syncc9902_1952	putative LdpA protein	0.026	0.106	0.241	0.219	100.0	45.6	4.3
Syncc9902_1952a	hypothetical protein	0.021	0.207	0.103	0.102	100.0	100.0	26.6
Syncc9902_1953	NADH dehydrogenase I subunit N	0.014	0.200	0.068	0.068	100.0	7.0	3.1
Syncc9902_1954	hypothetical protein	0.057	0.027	2.112	2.139	100.0	89.2	7.4
Syncc9902_1955	50S ribosomal protein L3	0.009	0.104	0.085	0.083	100.0	56.0	4.9
Syncc9902_1956	50S ribosomal protein L4	0.005	0.102	0.050	0.049	100.0	100.0	7.6
Syncc9902_1957	50S ribosomal protein L23	0.000	0.054	0.000	0.000	100.0	92.4	7.3
Syncc9902_1958	50S ribosomal protein L2	0.005	0.124	0.039	0.035	100.0	73.4	6.3
Syncc9902_1959	Ribosomal protein S19	0.000	0.026	0.000	0.000	100.0	21.7	2.8
Syncc9902_1960	Ribosomal protein L22	0.019	0.093	0.204	0.204	100.0	27.9	3.8
Syncc9902_1961	30S ribosomal protein S3	0.002	0.115	0.021	0.021	100.0	54.5	5.5
Syncc9902_1962	50S ribosomal protein L16	0.004	0.137	0.028	0.028	100.0	75.5	6.1
Syncc9902_1963	Ribosomal protein L29	0.005	0.048	0.094	0.094	100.0	97.6	8.4
Syncc9902_1964	30S ribosomal protein S17	0.005	0.077	0.062	0.088	100.0	100.0	9.9
Syncc9902_1965	50S ribosomal protein L14	0.003	0.107	0.023	0.023	100.0	100.0	11.1
Syncc9902_1966	Ribosomal protein L24	0.005	0.122	0.042	0.042	100.0	100.0	14.5
Syncc9902_1967	50S ribosomal protein L5	0.004	0.116	0.037	0.036	100.0	100.0	10.7
Syncc9902_1968	30S ribosomal protein S8	0.007	0.102	0.067	0.067	100.0	100.0	9.6
Syncc9902_1969	50S ribosomal protein L6	0.004	0.132	0.033	0.032	100.0	100.0	9.2
Syncc9902_1970	Ribosomal protein L18	0.009	0.076	0.116	0.110	100.0	100.0	10.4
Syncc9902_1971	30S ribosomal protein S5	0.005	0.171	0.030	0.029	100.0	100.0	13.7
Syncc9902_1972	50S ribosomal protein L15	0.005	0.124	0.042	0.049	100.0	99.3	7.7
Syncc9902_1973	preprotein translocase SecY	0.003	0.149	0.018	0.015	100.0	95.6	7.9
Syncc9902_1974	Adenylate kinase, subfamily	0.015	0.188	0.078	0.084	100.0	91.9	8.1
Syncc9902_1975	Ribosomal protein L36	0.004	0.056	0.072	0.072	100.0	100.0	8.1
Syncc9902_1976	30S ribosomal protein S13	0.004	0.102	0.038	0.038	100.0	100.0	9.0
Syncc9902_1977	30S ribosomal protein S11	0.003	0.139	0.023	0.022	100.0	76.6	7.8
Syncc9902_1978	DNA-directed RNA polymerase alpha subunit (EC:2.7.7.6)	0.004	0.145	0.025	0.027	100.0	83.2	6.4
Syncc9902_1979	50S ribosomal protein L17	0.007	0.188	0.038	0.043	100.0	66.1	4.9
Syncc9902_1980	tRNA pseudouridine synthase (EC:4.2.1.70)	0.026	0.180	0.145	0.136	100.0	88.4	7.6
Syncc9902_1981	50S ribosomal protein L13	0.005	0.057	0.095	0.093	100.0	48.7	4.9
Syncc9902_1982	30S ribosomal protein S9	0.002	0.179	0.009	0.009	100.0	75.1	6.6
Syncc9902_1983	50S ribosomal protein L31	0.013	0.094	0.138	0.129	100.0	100.0	7.0
Syncc9902_1984	peptide chain release factor 1	0.005	0.205	0.025	0.031	100.0	59.8	6.0
Syncc9902_1985	HNH nuclease	0.012	0.140	0.084	0.084	100.0	29.7	4.0
Syncc9902_1986	alanine racemase (EC:5.1.1.1)	0.019	0.138	0.138	0.132	94.8	68.5	5.6
Syncc9902_1987	hypothetical protein	0.024	0.228	0.107	0.133	100.0	21.8	3.5
Syncc9902_1988	hypothetical protein	0.016	0.085	0.189	0.272	97.5	27.2	3.8
Syncc9902_1989	glycosyl transferase, family 2	0.010	0.118	0.087	0.106	100.0	47.1	4.3
Syncc9902_1990	hypothetical protein	0.032	0.165	0.195	0.198	100.0	100.0	8.1

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_1991	photosystem I subunit VIII (PsaI)	0.000	0.011	0.000	0.000	100.0	100.0	7.4
Syncc9902_1992	putative photosystem I reaction center subunit XI (PsaL)	0.003	0.186	0.016	0.016	100.0	94.7	7.1
Syncc9902_1993	putative RND family multidrug efflux transporter	0.019	0.199	0.096	0.093	94.9	50.2	4.9
Syncc9902_1994	Photosystem I psaB	0.001	0.093	0.012	0.014	100.0	95.3	10.2
Syncc9902_1995	Photosystem I psaA	0.009	0.159	0.058	0.059	100.0	100.0	15.6
Syncc9902_1996	hypothetical protein	0.051	0.215	0.239	0.252	100.0	94.3	10.5
Syncc9902_1997	Precorrin-3B C17-methyltransferase region	0.057	0.283	0.203	0.209	100.0	100.0	21.1
Syncc9902_1998	putative transcriptional regulator, Crp/Fnr family	0.055	0.239	0.230	0.230	100.0	100.0	17.1
Syncc9902_1999	possible porin	0.050	0.340	0.147	0.149	100.0	100.0	33.4
Syncc9902_2000	hypothetical protein	0.096	0.314	0.305	0.295	100.0	100.0	21.3
Syncc9902_2001	putative hydroxylase	0.076	0.448	0.169	0.168	100.0	100.0	47.5
Syncc9902_2002	putative iron ABC transporter, substrate binding protein	0.062	0.451	0.137	0.135	100.0	100.0	46.4
Syncc9902_2003	ferritin	0.049	0.404	0.122	0.116	100.0	100.0	44.9
Syncc9902_2004	hypothetical protein	0.090	0.353	0.255	0.257	100.0	100.0	36.2
Syncc9902_2005	hypothetical protein	0.095	0.356	0.266	0.252	100.0	100.0	38.0
Syncc9902_2006	Thioredoxin-disulfide reductase (EC:1.8.1.9)	0.032	0.342	0.095	0.097	100.0	100.0	43.0
Syncc9902_2007	flavodoxin	0.029	0.353	0.081	0.075	100.0	100.0	47.2
Syncc9902_2008	hypothetical protein	0.168	0.124	1.356	1.270	100.0	100.0	37.1
Syncc9902_2009	hypothetical protein	0.052	0.205	0.253	0.256	100.0	100.0	16.9
Syncc9902_2010	regulatory proteins, Crp family	0.011	0.340	0.033	0.035	100.0	100.0	29.8
Syncc9902_2011	putative hydrogenase accessory protein	0.049	0.261	0.189	0.196	100.0	100.0	28.3
Syncc9902_2012	possible porin	0.074	0.248	0.299	0.300	100.0	100.0	19.5
Syncc9902_2013	hypothetical protein	0.031	0.168	0.182	0.204	100.0	100.0	10.7
Syncc9902_2014	lipoyl synthase	0.020	0.222	0.089	0.090	100.0	100.0	17.5
Syncc9902_2015	hypothetical protein	0.078	0.289	0.271	0.279	100.0	100.0	14.1
Syncc9902_2016	Glutamate synthase (NADPH) (EC:1.4.1.13)	0.016	0.198	0.080	0.083	99.1	91.2	9.2
Syncc9902_2017	hypothetical protein	0.027	0.134	0.204	0.211	99.7	47.5	5.0
Syncc9902_2018	hypothetical protein	0.030	0.145	0.206	0.213	100.0	53.1	5.2
Syncc9902_2019	30S ribosomal protein S12	0.002	0.162	0.012	0.011	100.0	69.1	5.9
Syncc9902_2020	30S ribosomal protein S7	0.000	0.154	0.000	0.000	100.0	37.4	4.2
Syncc9902_2021	elongation factor EF-2	0.009	0.188	0.047	0.048	100.0	95.0	9.4
Syncc9902_2022	elongation factor Tu (EC:2.7.7.4)	0.010	0.243	0.042	0.038	100.0	100.0	17.1
Syncc9902_2023	30S ribosomal protein S10	0.011	0.239	0.048	0.048	100.0	100.0	22.9
Syncc9902_2024	Peptidase S16, lon-like	0.016	0.221	0.074	0.072	100.0	100.0	14.5
Syncc9902_2025	probable sterol-C-methyltransferase	0.028	0.244	0.115	0.123	100.0	100.0	16.8
Syncc9902_2026	Prephenate dehydratase (EC:4.2.1.51)	0.020	0.214	0.096	0.097	100.0	100.0	15.9
Syncc9902_2027	hypothetical protein	0.023	0.164	0.140	0.144	100.0	100.0	20.1
Syncc9902_2028	Ribonuclease H (EC:3.1.26.4)	0.043	0.234	0.182	0.194	100.0	100.0	21.0
Syncc9902_2029	Ribonuclease E and G	0.026	0.233	0.110	0.107	100.0	100.0	16.9
Syncc9902_2030	Elongator protein 3/MiaB/NifB	0.023	0.246	0.093	0.095	100.0	100.0	18.9
Syncc9902_2031	aspartate aminotransferase	0.020	0.266	0.076	0.081	100.0	100.0	20.0
Syncc9902_2032	ATP-dependent Clp protease adaptor protein ClpS	0.010	0.248	0.041	0.041	100.0	100.0	26.3
Syncc9902_2033	hypothetical protein	0.026	0.117	0.219	0.215	100.0	87.5	8.2
Syncc9902_2034	TPR repeat	0.027	0.094	0.286	0.285	100.0	95.9	8.7
Syncc9902_2035	Elongator protein 3/MiaB/NifB	0.025	0.414	0.061	0.061	100.0	100.0	20.6
Syncc9902_2036	Photosystem II reaction centre protein PsbA/D1	0.038	0.470	0.080	0.082	100.0	100.0	26.9
Syncc9902_2037	hypothetical protein	0.059	0.049	1.206	1.206	79.9	5.5	1.7
Syncc9902_2038	hypothetical protein	0.037	0.234	0.157	0.154	100.0	100.0	9.6
Syncc9902_2039	hypothetical protein	0.054	0.275	0.196	0.196	100.0	100.0	10.4
Syncc9902_2040	chorismate synthase (EC:4.2.3.5)	0.008	0.131	0.058	0.062	100.0	71.8	6.8
Syncc9902_2041	hypothetical protein	0.007	0.243	0.029	0.027	100.0	75.7	8.4
Syncc9902_2042	possible aldolase	0.043	0.117	0.370	0.339	100.0	22.2	4.2
Syncc9902_2043	Peptidase M41, FtsH (EC:3.6.4.6)	0.006	0.274	0.023	0.024	100.0	89.8	8.1
Syncc9902_2044	ATP-sulfurylase (EC:2.7.7.4)	0.012	0.202	0.061	0.065	100.0	63.8	6.6
Syncc9902_2045	photosystem II manganese-stabilizing polypeptide	0.017	0.172	0.097	0.099	100.0	86.6	6.8
Syncc9902_2046	DNA/pantothenate metabolism flavoprotein	0.031	0.193	0.158	0.167	100.0	96.3	10.7
Syncc9902_2047	hypothetical protein	0.017	0.108	0.160	0.165	100.0	86.9	6.3
Syncc9902_2048	hypothetical protein	0.028	0.071	0.396	0.298	100.0	34.3	4.1
Syncc9902_2049	hypothetical protein	0.027	0.221	0.123	0.130	100.0	100.0	8.6
Syncc9902_2050	putative type 2 NADH dehydrogenase (Ndh, Ndh2B or NdbA)	0.024	0.187	0.127	0.123	100.0	99.0	11.2
Syncc9902_2051	aspartate carbamoyltransferase catalytic subunit (EC:2.1.3.2)	0.017	0.201	0.082	0.081	100.0	96.8	14.0
Syncc9902_2052	Methylpurine-DNA glycosylase (MPG)	0.046	0.244	0.189	0.192	100.0	100.0	15.9
Syncc9902_2053	Protein of unknown function DUF165	0.024	0.285	0.086	0.086	100.0	100.0	15.7
Syncc9902_2054	possible glycosyltransferase	0.031	0.146	0.210	0.227	98.7	41.1	4.0
Syncc9902_2055	hypothetical protein	0.013	0.202	0.065	0.063	100.0	100.0	10.5
Syncc9902_2056	hypothetical protein	0.019	0.251	0.077	0.074	100.0	100.0	21.8
Syncc9902_2057	Glutamyl-tRNA(Gln) amidotransferase C subunit	0.014	0.336	0.042	0.043	100.0	100.0	19.2

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_2058	beta-carotene hydroxylase	0.034	0.377	0.090	0.090	100.0	100.0	22.1
Syncc9902_2059	hypothetical protein	0.039	0.229	0.169	0.172	100.0	100.0	23.0
Syncc9902_2060	isoleucyl-tRNA synthetase	0.031	0.269	0.117	0.121	100.0	100.0	23.2
Syncc9902_2061	hypothetical protein	0.054	0.249	0.217	0.210	100.0	100.0	16.7
Syncc9902_2062	hypothetical protein	0.015	0.174	0.084	0.095	100.0	100.0	14.0
Syncc9902_2063	hypothetical protein	0.025	0.212	0.119	0.120	100.0	100.0	14.4
Syncc9902_2064	tRNA (guanine-N(7)-)-methyltransferase (EC:2.1.1.33)	0.025	0.194	0.130	0.136	100.0	86.2	8.8
Syncc9902_2065	possible bicarbonate transporter, ICT family	0.030	0.201	0.150	0.159	100.0	87.4	9.1
Syncc9902_2066	hypothetical protein	0.039	0.177	0.220	0.238	99.5	78.6	7.4
Syncc9902_2067	Phosphoglucosamine mutase	0.017	0.190	0.090	0.088	100.0	84.7	9.7
Syncc9902_2068	possible soluble lytic transglycosylase	0.045	0.150	0.297	0.284	98.9	34.3	3.7
Syncc9902_2069	hypothetical protein	0.032	0.170	0.186	0.195	100.0	84.1	6.2
Syncc9902_2070	thioredoxin-like protein TxlA	0.015	0.149	0.104	0.125	100.0	65.2	5.1
Syncc9902_2071	Thymidylate synthase complementing protein ThyX	0.011	0.217	0.050	0.048	100.0	100.0	9.2
Syncc9902_2072	deoxycytidine triphosphate deaminase (EC:3.5.4.13)	0.018	0.225	0.078	0.079	100.0	30.6	3.9
Syncc9902_2073	cob(I)alamin adenosyltransferase	0.018	0.207	0.087	0.082	96.4	64.1	5.5
Syncc9902_2074	Ribonuclease PH (EC:2.7.7.56)	0.026	0.237	0.109	0.111	100.0	92.1	8.5
Syncc9902_2075	transcriptional regulator, Crp/Fnr family	0.007	0.223	0.034	0.027	100.0	84.7	8.3
Syncc9902_2076	hypothetical protein	0.018	0.188	0.095	0.095	100.0	89.3	8.0
Syncc9902_2077	Resolvase, RNase H-like fold	0.031	0.157	0.200	0.203	100.0	95.5	8.3
Syncc9902_2078	hypothetical protein	0.002	0.171	0.011	0.010	100.0	100.0	8.3
Syncc9902_2079	peptidyl-tRNA hydrolase (EC:3.1.1.29)	0.013	0.178	0.073	0.077	100.0	99.4	9.2
Syncc9902_2080	Twin-arginine translocation protein TatA/E	0.029	0.269	0.107	0.088	100.0	100.0	13.2
Syncc9902_2081	photosystem II reaction center protein PsbH	0.005	0.110	0.042	0.043	100.0	100.0	7.7
Syncc9902_2082	photosystem II reaction centre N protein (psbN)	0.007	0.070	0.094	0.042	100.0	100.0	6.6
Syncc9902_2083	possible alpha mannosidase	0.037	0.201	0.186	0.187	100.0	93.1	10.3
Syncc9902_2084	photosystem II reaction center I protein PsbI	0.004	0.099	0.038	0.038	100.0	100.0	10.7
Syncc9902_2085	hypothetical protein	0.030	0.254	0.119	0.119	100.0	88.9	11.8
Syncc9902_2086	Uncharacterized low-complexity proteins-like	0.070	0.247	0.284	0.291	100.0	64.8	9.7
Syncc9902_2087	3-isopropylmalate dehydratase small subunit	0.034	0.190	0.181	0.168	100.0	100.0	10.8
Syncc9902_2088	isopropylmalate isomerase large subunit	0.012	0.182	0.068	0.072	100.0	100.0	12.0
Syncc9902_2089	Competence-damaged protein	0.031	0.173	0.178	0.173	100.0	94.5	8.1
Syncc9902_2090	putative glycosyltransferase	0.016	0.212	0.074	0.086	100.0	74.9	6.1
Syncc9902_2091	serine hydroxymethyltransferase (EC:2.1.2.1)	0.013	0.226	0.056	0.055	100.0	89.7	9.2
Syncc9902_2092	hypothetical protein	0.014	0.155	0.089	0.075	100.0	93.1	7.1
Syncc9902_2093	hypothetical protein	0.014	0.250	0.055	0.052	100.0	77.4	6.8
Syncc9902_2094	hypothetical protein	0.004	0.160	0.024	0.025	100.0	92.7	5.9
Syncc9902_2095	hypothetical protein	0.024	0.169	0.143	0.141	100.0	82.0	6.4
Syncc9902_2096	Rhodanese-like	0.017	0.125	0.136	0.136	100.0	100.0	7.4
Syncc9902_2097	Negative regulator of class I heat shock protein	0.011	0.200	0.057	0.056	100.0	58.9	4.7
Syncc9902_2098	tryptophan synthase subunit beta	0.007	0.151	0.048	0.047	100.0	50.8	6.2
Syncc9902_2099	putative translation initiation factor SUI1	0.017	0.152	0.114	0.100	100.0	75.7	6.0
Syncc9902_2100	Adenylylsulfate kinase (EC:2.7.1.25)	0.020	0.207	0.094	0.091	100.0	88.7	7.3
Syncc9902_2101	hypothetical protein	0.013	0.166	0.081	0.082	100.0	94.9	7.1
Syncc9902_2102	hypothetical protein	0.031	0.180	0.171	0.166	100.0	55.5	4.6
Syncc9902_2103	1-(5-Phosphoribosyl)-5-amino-4-imidazole- carboxylate (AIR) carboxylase (EC:4.1.1.21)	0.016	0.158	0.101	0.098	100.0	64.3	7.9
Syncc9902_2104	N-acetylglucosamine-6-phosphate deacetylase (EC:3.5.1.25)	0.047	0.152	0.309	0.328	100.0	82.7	7.5
Syncc9902_2105	Mg-protoporphyrin IX methyl transferase (EC:2.1.1.11)	0.034	0.223	0.153	0.157	100.0	72.1	6.4
Syncc9902_2106	probable pseudouridine synthase	0.028	0.214	0.129	0.127	100.0	93.3	12.1
Syncc9902_2107	two component transcriptional regulator, LuxR family	0.005	0.218	0.023	0.025	100.0	100.0	10.4
Syncc9902_2108	hypothetical protein	0.043	0.196	0.217	0.217	100.0	100.0	13.7
Syncc9902_2109	NifS-like aminotransferase class-V	0.032	0.177	0.180	0.184	100.0	74.8	7.0
Syncc9902_2110	hypothetical protein	0.028	0.141	0.199	0.200	100.0	95.5	7.7
Syncc9902_2111	methyltransferase	0.026	0.215	0.123	0.119	100.0	81.2	10.2
Syncc9902_2112	NADH dehydrogenase delta subunit (EC:1.6.5.3)	0.009	0.249	0.037	0.036	100.0	100.0	11.1
Syncc9902_2113	hypothetical protein	0.031	0.086	0.360	0.388	100.0	89.9	7.3
Syncc9902_2114	hypothetical protein	0.073	0.311	0.236	0.242	100.0	100.0	11.9
Syncc9902_2115	hypothetical protein	0.048	0.203	0.239	0.237	100.0	97.8	8.6
Syncc9902_2116	hypothetical protein	0.052	0.250	0.208	0.200	100.0	100.0	11.1
Syncc9902_2117	probable O-succinylbenzoic acid--CoA ligase	0.035	0.136	0.257	0.255	99.2	54.5	5.0
Syncc9902_2118	putative O-succinylbenzoate synthase	0.079	0.196	0.401	0.366	96.0	49.8	5.2
Syncc9902_2119	1,4-dihydroxy-2-naphthoate octaprenyltransferase,	0.035	0.189	0.183	0.166	100.0	70.5	6.7
Syncc9902_2120	Isochorismate synthase	0.025	0.186	0.132	0.139	100.0	83.8	9.6
Syncc9902_2121	glutathione synthetase (EC:6.3.2.3)	0.004	0.185	0.023	0.023	100.0	81.5	7.3
Syncc9902_2122	Glutaredoxin, GrxC	0.011	0.162	0.070	0.070	100.0	94.6	8.8
Syncc9902_2123	peptide chain release factor RF-2	0.006	0.109	0.056	0.056	100.0	100.0	13.8

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_2124	peptide chain release factor 2	0.008	0.159	0.050	0.048	100.0	91.2	7.6
Syncc9902_2125	hypothetical protein	0.021	0.124	0.170	0.172	100.0	100.0	11.0
Syncc9902_2126	hypothetical protein	0.059	0.187	0.315	0.303	100.0	98.7	9.8
Syncc9902_2127	possible diacylglycerol kinase	0.028	0.248	0.113	0.112	100.0	97.3	8.8
Syncc9902_2128	Glutamine amidotransferase of anthranilate synthase	0.012	0.219	0.057	0.066	100.0	100.0	10.3
Syncc9902_2129	hypothetical protein	0.019	0.197	0.094	0.087	100.0	90.0	8.1
Syncc9902_2130	hypothetical protein	0.050	0.153	0.325	0.331	100.0	75.1	5.2
Syncc9902_2131	hypothetical protein	0.032	0.182	0.176	0.181	100.0	36.6	4.2
Syncc9902_2132	Asparaginase (EC:3.5.1.1)	0.033	0.206	0.162	0.174	100.0	53.0	5.6
Syncc9902_2133	aminotransferase class-I	0.020	0.194	0.102	0.104	100.0	63.1	6.1
Syncc9902_2134	arginyl-tRNA synthetase (EC:6.1.1.19)	0.013	0.208	0.064	0.068	100.0	99.2	14.7
Syncc9902_2135	Nicotinate-nucleotide pyrophosphorylase (EC:2.4.2.19)	0.026	0.188	0.138	0.138	98.2	94.8	10.7
Syncc9902_2136	tRNA modification GTPase	0.016	0.200	0.081	0.078	100.0	91.7	9.3
Syncc9902_2137	hypothetical protein	0.033	0.195	0.170	0.181	100.0	81.8	6.7
Syncc9902_2138	hypothetical protein	0.027	0.173	0.156	0.148	100.0	79.5	5.5
Syncc9902_2139	metal dependent phosphohydrolase (EC:2.7.6.5)	0.009	0.140	0.063	0.074	100.0	60.5	6.5
Syncc9902_2140	ATPase	0.020	0.152	0.131	0.130	99.5	66.2	6.0
Syncc9902_2141	Pseudouridine synthase, RluD (EC:4.2.1.70)	0.017	0.182	0.091	0.087	100.0	88.2	8.9
Syncc9902_2142	GTP-binding	0.013	0.209	0.062	0.060	100.0	59.2	6.4
Syncc9902_2143	phosphoglycerate kinase (EC:2.7.2.3)	0.006	0.158	0.035	0.030	100.0	75.2	6.6
Syncc9902_2144	hypothetical protein	0.056	0.216	0.257	0.235	100.0	80.4	6.3
Syncc9902_2145	2-hydroxy-3-oxopropionate reductase (EC:1.1.1.60)	0.056	0.216	0.260	0.243	100.0	55.0	4.7
Syncc9902_2146	hypothetical protein	0.045	0.199	0.224	0.201	100.0	53.7	4.8
Syncc9902_2147	N-acetylglucosaminyltransferase, MurG (EC:2.4.1.227)	0.024	0.163	0.147	0.125	100.0	59.7	5.5
Syncc9902_2148	L-threonine-O-3-phosphate decarboxylase (EC:4.1.1.81)	0.061	0.315	0.192	0.181	100.0	62.3	7.7
Syncc9902_2149	hypothetical protein	0.061	0.000	Inf	Inf	89.0	7.6	2.0
Syncc9902_2150	dihydroorotate dehydrogenase	0.014	0.203	0.069	0.068	100.0	80.5	9.8
Syncc9902_2151	possible ribonuclease HI	0.019	0.165	0.113	0.122	100.0	87.8	7.1
Syncc9902_2152	hypothetical protein	0.035	0.185	0.189	0.183	100.0	55.5	4.8
Syncc9902_2153	Ribosomal protein L7/L12	0.013	0.068	0.188	0.159	100.0	84.5	5.8
Syncc9902_2154	50S ribosomal protein L10	0.003	0.057	0.048	0.047	100.0	90.1	7.5
Syncc9902_2155	50S ribosomal protein L1	0.010	0.184	0.056	0.060	100.0	94.2	9.2
Syncc9902_2156	Ribosomal protein L11	0.007	0.134	0.049	0.048	100.0	97.7	7.0
Syncc9902_2157	transcription antitermination protein NusG	0.011	0.240	0.046	0.042	100.0	100.0	11.7
Syncc9902_2158	SecE subunit of protein translocation complex	0.011	0.124	0.089	0.083	100.0	71.3	7.2
Syncc9902_2159	ATPase	0.014	0.177	0.078	0.081	98.7	86.0	9.3
Syncc9902_2160	Glyoxalase I (EC:4.4.1.5)	0.009	0.166	0.056	0.054	100.0	100.0	10.4
Syncc9902_2161	Enolase (EC:4.2.1.11)	0.009	0.233	0.037	0.033	100.0	100.0	11.6
Syncc9902_2162	possible kinase	0.016	0.244	0.067	0.068	100.0	100.0	14.7
Syncc9902_2163	hypothetical protein	0.027	0.227	0.119	0.149	100.0	100.0	13.6
Syncc9902_2164	hypothetical protein	0.058	0.194	0.302	0.293	100.0	46.0	4.8
Syncc9902_2165	phage integrase	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_2166	bifunctional ornithine acetyltransferase/N-acetylglutamate synthase protein (EC:2.3.1.1)	0.055	0.365	0.152	0.159	100.0	95.5	20.1
Syncc9902_2167	Dephospho-CoA kinase (EC:2.7.1.24)	0.044	0.173	0.255	0.264	100.0	71.5	9.3
Syncc9902_2168	Prolyl 4-hydroxylase, alpha subunit	0.052	0.247	0.210	0.194	100.0	99.5	13.1
Syncc9902_2169	aspartyl/glutamyl-tRNA amidotransferase subunit B	0.020	0.394	0.050	0.049	100.0	100.0	18.3
Syncc9902_2170	Aromatic-ring hydroxylase	0.062	0.421	0.147	0.151	100.0	91.9	14.3
Syncc9902_2171	Nucleoside-diphosphate kinase (EC:2.7.4.6)	0.008	0.286	0.028	0.025	100.0	70.8	5.7
Syncc9902_2172	arginine decarboxylase	0.014	0.206	0.069	0.072	100.0	91.9	10.3
Syncc9902_2173	alanyl-tRNA synthetase (EC:6.1.1.7)	0.027	0.255	0.105	0.102	100.0	98.0	11.9
Syncc9902_2174	hypothetical protein	0.015	0.084	0.184	0.151	100.0	65.2	6.1
Syncc9902_2175	DEAD/DEAH box helicase-like	0.008	0.193	0.041	0.043	100.0	95.4	8.7
Syncc9902_2176	hypothetical protein	0.016	0.221	0.073	0.073	100.0	100.0	13.5
Syncc9902_2177	hypothetical protein	0.064	0.189	0.341	0.328	100.0	100.0	13.7
Syncc9902_2178	putative small heat shock protein	0.033	0.303	0.107	0.116	100.0	100.0	13.8
Syncc9902_2179	conserved hypothetical putative integral membrane protein	0.017	0.224	0.075	0.076	100.0	100.0	12.9
Syncc9902_2180	flavoprotein	0.013	0.241	0.054	0.057	100.0	90.6	10.2
Syncc9902_2181	hypothetical protein	0.102	0.335	0.304	0.306	76.8	42.2	4.0
Syncc9902_2182	Capsular polysaccharide biosynthesis protein-like	0.000	0.000	NaN	NaN	0.0	0.0	0.0
Syncc9902_2183	putative flavoprotein	0.014	0.163	0.086	0.084	100.0	77.6	7.9
Syncc9902_2184	rubrerythrin	0.024	0.195	0.120	0.119	100.0	79.2	6.9
Syncc9902_2185	possible reductase	0.037	0.212	0.175	0.186	100.0	99.1	8.4
Syncc9902_2186	hypothetical protein	0.000	0.019	0.000	0.000	100.0	36.7	3.7
Syncc9902_2187	hypothetical protein	0.018	0.161	0.110	0.109	100.0	85.9	7.3
Syncc9902_2188	glycine dehydrogenase (EC:1.4.4.2)	0.022	0.214	0.102	0.098	100.0	99.2	12.9
Syncc9902_2189	Glycine cleavage H-protein	0.006	0.169	0.035	0.027	100.0	100.0	15.7
Syncc9902_2190	hypothetical protein	0.017	0.203	0.084	0.084	100.0	100.0	12.2

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_2191	Stearoyl-CoA 9-desaturase (EC:1.14.19.1)	0.015	0.303	0.051	0.053	100.0	100.0	12.7
Syncc9902_2192	Stearoyl-CoA 9-desaturase (EC:1.14.19.1)	0.013	0.234	0.054	0.051	100.0	98.9	11.8
Syncc9902_2193	50S ribosomal protein L9	0.012	0.164	0.073	0.067	100.0	100.0	15.4
Syncc9902_2194	DnaB helicase	0.012	0.286	0.043	0.047	100.0	100.0	12.7
Syncc9902_2195	hypothetical protein	0.065	0.239	0.271	0.271	93.5	45.2	4.8
Syncc9902_2196	transcriptional regulator, AraC family	0.119	0.384	0.309	0.321	100.0	100.0	28.7
Syncc9902_2197	glucose-inhibited division protein A	0.029	0.318	0.091	0.092	100.0	100.0	22.0
Syncc9902_2198	hypothetical protein	0.075	0.235	0.319	0.340	100.0	100.0	12.6
Syncc9902_2199	hypothetical protein	0.021	0.233	0.090	0.091	100.0	100.0	16.2
Syncc9902_2200	hypothetical protein	0.065	0.199	0.325	0.330	100.0	100.0	17.4
Syncc9902_2201	hypothetical protein	0.015	0.225	0.066	0.070	100.0	100.0	17.9
Syncc9902_2202	hypothetical protein	0.049	0.220	0.224	0.211	100.0	100.0	12.8
Syncc9902_2203	Protein of unknown function SprT	0.020	0.146	0.136	0.140	100.0	100.0	13.2
Syncc9902_2204	hypothetical protein	0.003	0.140	0.024	0.023	100.0	100.0	15.6
Syncc9902_2205	DNA ligase (EC:6.5.1.2)	0.029	0.211	0.138	0.134	100.0	93.3	11.7
Syncc9902_2206	hypothetical protein	0.015	0.153	0.100	0.101	100.0	48.8	5.6
Syncc9902_2207	hypothetical protein	0.044	0.205	0.215	0.216	100.0	100.0	13.0
Syncc9902_2208	hypothetical protein	0.025	0.252	0.100	0.087	100.0	100.0	19.7
Syncc9902_2209	conserved membrane protein, multidrug efflux associated	0.017	0.118	0.148	0.152	100.0	100.0	17.7
Syncc9902_2210	putative multidrug efflux ABC transporter	0.025	0.269	0.094	0.087	100.0	100.0	28.8
Syncc9902_2211	ATPase	0.024	0.259	0.093	0.101	100.0	97.1	31.0
Syncc9902_2212	small mechanosensitive ion channel, MscS family	0.003	0.043	0.070	0.068	100.0	85.0	7.3
Syncc9902_2213	valyl-tRNA synthetase	0.023	0.269	0.084	0.086	100.0	90.8	14.5
Syncc9902_2214	ABC-type sugar transport system periplasmic component-like	0.021	0.148	0.144	0.150	100.0	69.4	5.5
Syncc9902_2215	Small-conductance mechanosensitive channel-like	0.023	0.220	0.104	0.101	97.7	66.8	5.4
Syncc9902_2216	hypothetical protein	0.028	0.216	0.131	0.132	100.0	98.0	11.0
Syncc9902_2217	possible protein phosphatase 2A regulatory B S	0.019	0.158	0.121	0.119	100.0	92.6	8.0
Syncc9902_2218	hypothetical protein	0.012	0.121	0.097	0.091	100.0	100.0	9.1
Syncc9902_2219	short-chain dehydrogenase/reductase (SDR) superfamily	0.028	0.202	0.141	0.143	100.0	70.5	6.7
Syncc9902_2220	putative ferric uptake regulator, FUR family	0.028	0.181	0.156	0.152	100.0	91.9	7.4
Syncc9902_2221	hypothetical protein	0.015	0.185	0.082	0.094	97.6	58.8	4.8
Syncc9902_2222	putative high light inducible protein	0.008	0.201	0.038	0.038	100.0	100.0	8.1
Syncc9902_2223	hypothetical protein	0.091	0.200	0.455	0.571	100.0	6.0	2.5
Syncc9902_2224	hypothetical protein	0.000	0.000	NaN	NaN	16.0	0.0	0.2
Syncc9902_2225	hypothetical protein	0.143	0.000	Inf	Inf	29.7	3.7	1.2
Syncc9902_2226	hypothetical protein	0.031	0.162	0.191	0.197	100.0	89.9	7.3
Syncc9902_2227	MazG	0.031	0.146	0.213	0.201	100.0	95.1	8.6
Syncc9902_2228	hypothetical protein	0.036	0.192	0.186	0.192	100.0	100.0	9.0
Syncc9902_2229	spermidine synthase	0.010	0.136	0.071	0.080	99.5	44.6	4.6
Syncc9902_2230	Putative agmatinase	0.016	0.138	0.114	0.113	100.0	66.8	6.0
Syncc9902_2231	hypothetical protein	0.018	0.148	0.125	0.120	100.0	36.1	4.5
Syncc9902_2232	Glycine cleavage system T protein (EC:2.1.2.10)	0.014	0.143	0.096	0.097	100.0	51.8	4.6
Syncc9902_2233	Aspartyl-tRNA synthetase	0.012	0.148	0.078	0.077	100.0	75.8	6.5
Syncc9902_2234	possible type II alternative RNA polymerase sigma factor	0.010	0.139	0.070	0.077	100.0	71.8	5.8
Syncc9902_2235	CTP synthetase (EC:6.3.4.2)	0.006	0.128	0.048	0.061	100.0	82.6	6.6
Syncc9902_2236	possible organic radical activating enzyme	0.041	0.167	0.247	0.233	100.0	51.9	6.8
Syncc9902_2237	possible 3-hydroxyacid dehydrogenase	0.029	0.152	0.189	0.203	100.0	96.7	10.5
Syncc9902_2238	ExsB	0.015	0.113	0.135	0.141	100.0	99.0	10.4
Syncc9902_2239	Anthranilate synthase (EC:4.1.3.27)	0.026	0.189	0.139	0.137	100.0	97.7	10.1
Syncc9902_2240	aminotransferase class IV	0.045	0.103	0.438	0.425	46.0	7.4	1.5
Syncc9902_2241	amino acid permease family protein	0.004	0.000	Inf	Inf	70.3	14.6	2.0
Syncc9902_2242	HAD-superfamily hydrolase YedP	0.000	0.000	NaN	NaN	88.9	0.0	1.2
Syncc9902_2243	Alpha amylase, catalytic subdomain	0.025	0.203	0.125	0.121	99.5	72.9	12.3
Syncc9902_2244	hypothetical protein	0.016	0.231	0.068	0.067	100.0	100.0	21.1
Syncc9902_2245	hypothetical protein	0.075	0.327	0.231	0.252	100.0	100.0	28.5
Syncc9902_2246	ATPase	0.016	0.250	0.066	0.065	100.0	100.0	27.9
Syncc9902_2247	ATPase	0.015	0.217	0.067	0.057	100.0	100.0	21.2
Syncc9902_2248	putative membrane protein of ABC transport system	0.026	0.305	0.086	0.092	100.0	100.0	15.9
Syncc9902_2249	putative urea ABC transporter	0.017	0.296	0.057	0.059	100.0	100.0	21.9
Syncc9902_2250	putative urea ABC transporter, urea binding protein	0.016	0.279	0.057	0.061	100.0	100.0	27.2
Syncc9902_2251	Urease accessory protein UreG	0.032	0.325	0.099	0.094	100.0	100.0	26.2
Syncc9902_2252	urease accessory protein F	0.034	0.231	0.149	0.148	100.0	100.0	20.2
Syncc9902_2253	urease accessory protein E	0.045	0.239	0.188	0.199	100.0	100.0	13.2
Syncc9902_2254	urease accessory protein D	0.049	0.203	0.239	0.244	100.0	100.0	17.1
Syncc9902_2255	Urease, gamma subunit region (EC:3.5.1.5)	0.009	0.260	0.035	0.035	100.0	100.0	23.0
Syncc9902_2256	Urease, beta subunit (EC:3.5.1.5)	0.023	0.238	0.096	0.092	100.0	100.0	21.4
Syncc9902_2257	Urease, alpha subunit (EC:3.5.1.5)	0.022	0.265	0.083	0.083	100.0	100.0	22.1

Table A5.2. dN/dS ratios of CC9902 genes, continued.

gene	description	dN	dS	dN/dS	dN/dS CC9902	% of gene covered 1X	% of gene covered 5X	avg. read depth
Syncc9902_2258	hypothetical protein	0.026	0.103	0.248	0.248	100.0	91.3	8.8
Syncc9902_2259	N-carbamoyl-L-amino acid amidohydrolase (EC:3.5.1.87)	0.012	0.075	0.161	0.161	100.0	99.3	10.1
Syncc9902_2260	Asparagine synthase, glutamine-hydrolyzing	0.014	0.105	0.132	0.131	100.0	100.0	12.3
Syncc9902_2261	putative aspartate carbamoyltransferase	0.019	0.154	0.123	0.123	100.0	99.9	13.7
Syncc9902_2262	hypothetical protein	0.014	0.025	0.568	0.619	100.0	98.8	11.4
Syncc9902_2263	sodium:solute transporter family, possible urea transporter	0.007	0.097	0.075	0.075	100.0	100.0	19.6
Syncc9902_2264	hypothetical protein	0.040	0.284	0.139	0.137	100.0	100.0	32.8
Syncc9902_2265	hypothetical protein	0.024	0.328	0.072	0.072	100.0	100.0	42.5
Syncc9902_2266	Transglutaminase-like	0.037	0.257	0.144	0.148	100.0	100.0	32.0
Syncc9902_2267	conserved hypothetical protein (ORFG Synechococcus sp.)	0.037	0.312	0.118	0.117	100.0	100.0	27.9
Syncc9902_2268	hypothetical protein	0.089	0.246	0.363	0.367	100.0	100.0	24.4
Syncc9902_2269	molybdenum cofactor biosynthesis protein A	0.053	0.201	0.263	0.255	100.0	100.0	26.4
Syncc9902_2270	molybdopterin-guanine dinucleotide biosynthesis protein A	0.067	0.278	0.242	0.241	100.0	100.0	21.7
Syncc9902_2271	Nitrate transporter	0.017	0.343	0.051	0.052	100.0	100.0	19.6
Syncc9902_2272	nitrate reductase	0.020	0.209	0.097	0.098	100.0	99.5	17.6
Syncc9902_2273	hypothetical protein	0.057	0.205	0.276	0.276	100.0	100.0	12.3
Syncc9902_2274	hypothetical protein	0.051	0.174	0.293	0.292	100.0	100.0	12.7
Syncc9902_2275	carbonic anhydrase	0.038	0.230	0.166	0.171	100.0	100.0	17.8
Syncc9902_2276	Molybdopterin cofactor biosynthesis MoaC region	0.026	0.161	0.162	0.166	100.0	100.0	12.6
Syncc9902_2277	Molybdopterin binding domain	0.040	0.200	0.197	0.196	100.0	99.9	12.8
Syncc9902_2278	hypothetical protein	0.042	0.145	0.287	0.252	100.0	75.4	8.6
Syncc9902_2279	molybdenum cofactor biosynthesis protein E	0.032	0.123	0.263	0.257	100.0	100.0	16.3
Syncc9902_2280	molydbenum cofactor biosynthesis protein D	0.012	0.171	0.069	0.076	100.0	100.0	17.9
Syncc9902_2281	Molybdopterin binding domain	0.036	0.232	0.156	0.154	100.0	100.0	18.1
Syncc9902_2282	Uroporphyrin-III C-methyltransferase-like	0.026	0.218	0.117	0.113	100.0	100.0	16.4
Syncc9902_2283	possible DNA mismatch repair proteins, MutS fa	0.053	0.239	0.223	0.225	100.0	100.0	22.7
Syncc9902_2284	ferredoxin--nitrite reductase	0.034	0.385	0.088	0.088	100.0	100.0	25.1
Syncc9902_2285	formate and nitrite transporters	0.014	0.278	0.050	0.052	100.0	100.0	29.6
Syncc9902_2286	hypothetical protein	0.077	0.244	0.313	0.299	100.0	100.0	23.2
Syncc9902_2287	cyanate hydratase (EC:4.2.1.104)	0.047	0.341	0.139	0.145	100.0	100.0	22.8
Syncc9902_2288	hypothetical protein	0.080	0.393	0.202	0.196	100.0	100.0	17.8
Syncc9902_2289	hypothetical protein	0.037	0.243	0.153	0.167	100.0	100.0	19.4
Syncc9902_2290	hypothetical protein	0.047	0.273	0.172	0.172	100.0	100.0	27.1
Syncc9902_2291	permease of the major facilitator superfamily	0.025	0.235	0.106	0.107	100.0	100.0	20.7
Syncc9902_2292	polyphosphate kinase (EC:2.7.4.1)	0.019	0.292	0.066	0.067	100.0	100.0	16.6
Syncc9902_2293	type II alternative RNA polymerase sigma factor, sigma-70 family	0.015	0.269	0.055	0.054	100.0	100.0	23.5
Syncc9902_2294	hypothetical protein	0.055	0.224	0.247	0.242	100.0	98.9	14.0
Syncc9902_2295	3-deoxy-7-phosphoheptulonate synthase (EC:2.5.1.54)	0.018	0.254	0.071	0.072	100.0	99.7	13.9
Syncc9902_2296	aconitate hydratase (EC:4.2.1.3)	0.013	0.221	0.061	0.065	100.0	100.0	17.2
Syncc9902_2297	putative chloride channel	0.026	0.239	0.110	0.105	100.0	100.0	22.5
Syncc9902_2298	Elongator protein 3/MiaB/NifB	0.026	0.246	0.106	0.095	100.0	100.0	22.4
Syncc9902_2299	hypothetical protein	0.049	0.264	0.186	0.180	100.0	98.8	20.9
Syncc9902_2300	Formyltetrahydrofolate deformylase	0.023	0.244	0.092	0.093	100.0	100.0	26.8
Syncc9902_2301	hypothetical protein	0.023	0.219	0.106	0.104	100.0	100.0	23.7
Syncc9902_2302	D-amino acid oxidase	0.060	0.224	0.268	0.269	100.0	100.0	17.0
Syncc9902_2303	molecular chaperone DnaK	0.010	0.267	0.038	0.038	100.0	100.0	14.3
Syncc9902_2304	shikimate / quinate 5-dehydrogenase	0.035	0.166	0.213	0.208	100.0	90.5	9.2
Syncc9902_2305	hypothetical protein	0.017	0.232	0.075	0.077	100.0	94.4	9.8
Syncc9902_2306	Ribosomal protein S6	0.038	0.337	0.112	0.115	100.0	100.0	14.6
Syncc9902_2307	hypothetical protein	0.073	0.251	0.291	0.287	100.0	100.0	10.2
Syncc9902_2308	Argininosuccinate synthase (EC:6.3.4.5)	0.012	0.224	0.053	0.048	100.0	100.0	11.7
Syncc9902_2309	hypothetical protein	0.022	0.154	0.145	0.141	100.0	98.7	8.0
Syncc9902_2310	hypothetical protein	0.008	0.211	0.039	0.041	100.0	95.6	5.7
Syncc9902_2311	Phospho-N-acetylmuramoyl-pentapeptide transferase (EC:2.7.8.13)	0.027	0.225	0.119	0.115	100.0	97.2	13.1
Syncc9902_2312	hypothetical protein	0.015	0.265	0.056	0.055	100.0	100.0	16.0
Syncc9902_2313	hypothetical protein	0.023	0.258	0.089	0.091	100.0	100.0	17.0
Syncc9902_2314	5'-phosphoribosylglycinamide transformylase	0.024	0.203	0.117	0.115	100.0	97.8	14.5
Syncc9902_2315	Sucrose-phosphate synthase (EC:2.4.1.14)	0.030	0.245	0.122	0.116	100.0	100.0	18.3
Syncc9902_2316	hypothetical protein	0.046	0.221	0.209	0.213	100.0	98.3	11.7
Syncc9902_2317	excinuclease ABC subunit A	0.013	0.275	0.049	0.050	100.0	94.3	10.8
Syncc9902_2318	DNA repair protein RecN	0.021	0.176	0.117	0.113	100.0	95.4	9.2
Syncc9902_2319	possible protein kinase	0.010	0.218	0.047	0.046	100.0	87.9	8.8
Syncc9902_2320	hypothetical protein	0.027	0.195	0.136	0.129	100.0	100.0	9.4
Syncc9902_2321	threonine synthase	0.007	0.134	0.049	0.047	100.0	87.1	7.1