# UC Davis

## UC Davis Previously Published Works


### Title

Genome Sequence of the PCE-Dechlorinating Bacterium Dehalococcoides ethenogenes


### Permalink

https://escholarship.org/uc/item/3tw4j26x



### Journal

Science, 307(5706)

### ISSN

0036-8075



### Authors

Seshadri, Rekha
Adrian, Lorenz
Fouts, Derrick E
et al.



### Publication Date

2005-01-07

### DOI

10.1126/science.1102226

Peer reviewed



eScholarship.org Powered by the California Digital Library
University of California

# Genome Sequence of the PCE-Dechlorinating Bacterium *Dehalococcoides ethenogenes*

Rekha Seshadri,[1]* Lorenz Adrian,[2] Derrick E. Fouts,[1] Jonathan A. Eisen,[1,3] Adam M. Phillippy,[1] Barbara A. Methe,[1] Naomi L. Ward,[1,4] William C. Nelson,[1] Robert T. Deboy,[1] Hoda M. Khouri,[1] James F. Kolonay,[1] Robert J. Dodson,[1] Sean C. Daugherty,[1] Lauren M. Brinkac,[1] Steven A. Sullivan,[1] Ramana Madupu,[1] Karen E. Nelson,[1] Katherine H. Kang,[1] Marjorie Impraim,[1] Kevin Tran,[1] Jeffrey M. Robinson,[1] Heather A. Forberger,[1] Claire M. Fraser,[1,5] Stephen H. Zinder,[6] John F. Heidelberg[1,4]

*Dehalococcoides ethenogenes* is the only bacterium known to reductively dechlorinate the groundwater pollutants, tetrachloroethene (PCE) and trichloroethene, to ethene. Its 1,469,720–base pair chromosome contains large dynamic duplicated regions and integrated elements. Genes encoding 17 putative reductive dehalogenases, nearly all of which were adjacent to genes for transcription regulators, and five hydrogenase complexes were identified. These findings, plus a limited repertoire of other metabolic modes, indicate that *D. ethenogenes* is highly evolved to utilize halogenated organic compounds and $H_2$. Diversification of reductive dehalogenase functions appears to have been mediated by recent genetic exchange and amplification. Genome analysis provides insights into the organism's complex nutrient requirements and suggests that an ancestor was a nitrogen-fixing autotroph.

Tetrachloroethene (PCE) and trichloroethene (TCE) are among the most pervasive organic groundwater pollutants, primarily because of their disposal after use as industrial cleaners. Aerobic biodegradation of these solvents is ineffective; however, under anaerobic conditions, they can serve as terminal electron acceptors for dehalorespiration by bacteria that sequentially dechlorinate PCE to less chlorinated ethenes. Most cultured anaerobic (de)halorespirers (for example, *Sulfurospirillum multivorans* and *Dehalobacter restrictus*) only perform incomplete reductive dechlorination, yielding the toxic end product *cis*-dichloroethene (DCE) (*1*). In contrast, *Dehalococcoides ethenogenes* performs complete reductive dechlorination of PCE and TCE to the nontoxic metabolite ethene (*2*). *D. ethenogenes* belongs to subphylum 2 of *Chloroflexi* (green nonsulfur bacteria) (*3*–*5*), whereas other dehalorespiring bacteria are members of *Proteobacteria* and *Firmicutes* (*6*).

*D. ethenogenes* strain 195 was derived from an anaerobic sewage digestor. Subsequent studies demonstrated that *Dehalococcoides* spp. occur at many contaminated groundwater sites, and their presence is correlated with dechlorination past the product DCE (*5*, *7*–*9*). *Dehalococcoides* spp. have also been shown to dechlorinate chlorobenzenes, chloronaphthalenes, polychlorinated biphenyls, and dibenzodioxins (*4*, *10*, *11*). *D. ethenogenes* exhibits an unusual metabolic specialization, using only $H_2$ as an electron donor and chlorinated compounds as electron acceptors to support growth (*2*). In culture, it also requires acetate, vitamin $B_{12}$, and extracts of mixed microbial cultures as nutrients, suggesting a relatively limited repertoire of biosynthetic capabilities.

The *D. ethenogenes* genome sequence provides a better understanding of dehalorespiration and its evolution. Because the organism is difficult to culture and largely intractable to study using contemporary molecular genetic methods, the genome sequence contributes to our understanding of the physiology of *D. ethenogenes* and the evolution of catabolic pathways.

The genome of *D. ethenogenes* is composed of a 1,469,720–base pair (bp) circular chromosome containing 1591 predicted coding sequences (CDSs) (table S1). The genome possesses large duplicated regions and several integrated elements (Table 1 and fig. S1). A 31-kb tandem duplication encodes enzymes for $CO_2$ fixation, corrinoid cofactor salvage, and transport, among others (fig. S2). Within this duplication, DET0674 may be a fusion of DET0708 (aspartyl–transfer RNA synthetase) and DET0640 (universal stress protein family). With the exception of a single frameshift in a histidine kinase gene (DET0696), no nucleotide polymorphisms were detected between the tandem duplications. The mechanism that generates this duplication is unclear because flanking regions have no repeats; however, nonhomologous illegitimate recombination (*12*) may have occurred. Although this duplication could represent a strain-specific event that may not prevail in the population, it seems likely that a selective advantage is conferred (*13*–*15*); for example, by increasing gene dosage, particularly for cofactor salvage.

A ~22-kb integrated element (IE), which contains phage-like genes and is flanked by putative attachment (att) sites, occurs in at least three copies (fig. S3 and Table 1) as a tandem duplication (elements III and IV) and a distal third copy (element VI). A very large number of clones correspond to the current assembly. However, because of the high redundancy of sequence data, the presence of additional tandem copies or episomal elements cannot be ruled out. These and other IEs (Table 1) represent ~13.6% of the genome of *D. ethenogenes*.

It has been speculated that anthropogenic compounds select for microbial populations that have acquired the ability to use them (*16*). Only in the past 50 years have substantial quantities of chlorinated ethene solvents been released into the environment. It appears that the amplification of dechlorination capabilities in this organism has been mediated by recent genetic exchange. For instance, four of the putative IEs (Table 1), which have atypical trinucleotide composition and encode site-specific recombinases, contain reductive dehalogenase (RD) genes. This suggests their recent acquisition (or perhaps consignment for dissemination). The gene that encodes TceA-RD, which essentially defines this organism's ability to dechlorinate chloroethenes past DCE (*17*), is located in an IE.

Previous studies have shown that *D. ethenogenes* reduces PCE to ethene using two RDs (*18*) belonging to a family of iron-sulfur corrinoid proteins (*19*, *20*). One of the RDs reduces PCE to TCE, and a second RD(TceA) reduces TCE and other chloroethenes to vinyl chloride and ethene (*17*). The genome sequence indicates 17 intact CDSs (table S2) that encode putative RDs (*21*), all possessing potential twin-arginine transport (TAT) signals for export to the periplasm

[1]The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA. [2]Fachgebiet Technische Biochemie, Institut für Biotechnologie, Technische Universität Berlin, Seestrasse 13, D-13353 Berlin, Germany. [3]Johns Hopkins University, Charles and 34th Streets, Baltimore, MD 21218, USA. [4]Center of Marine Biotechnology, University of Maryland Biotechnology Institute, Baltimore, MD 21202, USA. [5]Departments of Pharmacology and Microbiology and Tropical Medicine, George Washington University School of Medicine, 2300 Eye Street N.W., Washington, DC 20037, USA. [6]Section of Microbiology, Cornell University, 272 Wing Hall, Ithaca, NY 14853-5701, USA.

*To whom correspondence should be addressed. E-mail: rekha@tigr.org

(*22*), with adjacent genes that encode potential membrane-anchoring proteins. Additionally, two degenerate RD genes are present. One (DET0088) is truncated and lacks an adjacent anchoring protein, and the other (DET0162) has a point mutation. The membrane anchor (DET0163) of DET0162 is shorter, and flanking transposases (DET0165 and DET0166) suggest ongoing decay. Seventeen RD genes are located within 300,000 bp of the predicted origin of replication, and many are clustered (fig. S1); only DET1171 and DET0876 are located distally. With the exception of DET0079, which encodes TceA-RD, all RD operons are oriented in the direction of replication. Such a strong orientation bias was not seen with other gene groups.

*D. ethenogenes* has been shown to dehalogenate numerous halogenated substrates (*4*, *10*, *11*). Given the presence of multiple potential RDs, it is likely that some may have specificities for these alternate compounds. The *D. ethenogenes* RD genes are part of a gene family that includes confirmed and putative RDs from other dehalogenating organisms, sequences from uncultured Sargasso Sea microorganisms (*23*), and the *Silicibacter pomeroyi* genome (fig. S4).

With one exception (DET0079, *tceA*), all intact RD genes are found in close proximity to genes for transcription regulators, particularly two-component signal transduction systems (TCSs) (table S2), which suggests stringent regulation of RD activity. The RD genes with no proximal TCS (DET0079, DET0876, and DET0162) are located within putative integrated regions (Table 1). Typically, sensor histidine kinase (HK) components of TCS possess at least two transmembrane regions and function as periplasmic membrane receptors that detect environmental signals. However, all of the RD-associated HKs lack transmembrane helices, suggesting that these are soluble cytoplasmic proteins (*24*). In contrast, many other HK components are more typical, possessing two to four transmembrane regions (table S3). Phylogenetic analysis of these HK components suggests a major expansion of at least one group of HKs that are associated with RD genes (fig. S5).

The predicted cytoplasmic location suggests that these sensor HKs respond to intracellular rather than extracellular stimuli. One speculation is that they may detect redox levels of cytoplasmic electron carriers, based on the presence of ligand-binding PAS/PAC motifs (table S3). These electron carriers have been implicated in sensing cellular energy reserves through changes in proton motive force (PMF) or redox potential of the electron transport chain (*25*). This sensing of energy reserves would allow the cell to respond rapidly to changes in energy status, which is a critical function for an organism that has limited means for energy production. Another speculation is that soluble HKs could detect nonpolar chlorinated substrates that permeate the cell membranes, resulting in up-regulation of RD expression when the substrate becomes available.

The genome encodes five different putative membrane-bound multisubunit hydrogenases (Fig. 1 and table S4) (*26*), which are typically involved in energy transduction. Of these, only Hup hydrogenase possesses a TAT signal in the small catalytic subunit and is thus predicted to occur in the periplasm. This location makes Hup hydrogenase a good candidate as an uptake hydrogenase involved in energy conservation, because protons produced from $H_2$ oxidation will be outside the cell and will not dissipate a PMF (Fig. 1). The Ech hydrogenase is known to be a proton pump, and it can use a PMF to drive electrons thermodynamically uphill (reverse electron transport) (*27*). This uphill drive generates low-potential electrons that are needed to either reduce cobalt in the RDs to the $Co^{+1}$ state (*28*) or for biosynthetic reactions. Including those involved in hydrogenase maturation, over 40 genes in *D. ethenogenes* are potentially devoted to the simple reaction that interconverts $H_2$ with protons and electrons, signifying the importance of this reaction.

An 11-CDS operon encoding reduced nicotinamide adenine dinucleotide (NADH)–ubiquinone oxidoreductase (complex I, Nuo) lacks genes for a cytoplasmic input module (NuoEFG) that withdraws electrons from NAD(P)H. Instead, these genes are found associated with other oxidoreductase operons (table S4). It is not clear whether one of the distally located NuoEFG modules or a putative cofactor $F_{420}$ input module (FpoF) serves as the input domain of the *D. ethenogenes* Nuo-assembly, as described for Complex I assemblies of *Methanosarcina mazei* (*29*). Thus, it is difficult to predict the electron donor for this complex, and it is not clear whether the complex uses quinone as its electron acceptor.

The only operon potentially involved in the reduction of nonhalogenated electron acceptors is a putative membrane-bound molybdopterin-containing oxidoreductase. The molybdopterin-containing α subunit has a predicted periplasmic location and shares 28% identity with tetrathionate reductase from *Salmonella typhimurium*, although this low level of identity makes it difficult to predict its substrate.

*D. ethenogenes* has not been found to use formate as an electron donor (*2*); however, a putative membrane-bound formate dehydrogenase was detected. The methanogenic Archaeae *Methanocaldococcus jannaschii* and *M. thermoautrophicus* use only $H_2$ as an electron donor, but both possess genes homologous to those encoding formate dehydrogenases (*30*). If not involved in electron transport, the formate dehydrogenase may be involved in producing formate for biosynthesis.

**Table 1.** *D. ethenogenes* strain 195 putative IEs. Putative att sites were determined using MUMmer or Grasta software. The best matches were determined by Basic Local Alignment Search Tool (BLASTP) searches of a limited custom phage database. The putative attL/R sequence included here includes only the att site closest to recombinase.

| Integrated element | Region | Locus names | Size (bp) | Target for Integration | putative attL/R | Atypical trinucleotide? | %GC | Total # ORFS | # phage BLASTP hits | # conserved hypotheticals | # hypotheticals | Phage integrase | Resolvase | restriction endonuclease | methylase | DNA polymerase | DNA primase | Helicase | Transposase | *gin* -like invertase | *lexA* -like repressor | Reductive Dehalogenase |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I | 57850-88645 | DET0063-0091 | 30,796 | tRNA-Ala-1 | ctccggctc | yes | 46.7 | 28 | 15 | 2 | 11 | | 2 | | 1 | 1 | 1 | 1 | 1 | 1 | | DET0079 |
| II | 149104-165172 | DET0155-0169 | 16,069 | tRNA-Val-1 | cgcccacca | yes | 47.6 | 14 | 5 | 2 | 6 | 1 | 1 | | | 1 | | | 2 | 1 | | DET0162 |
| III | 240759-263123 | DET0251-0272 | 22,366 | TCTGATT | TCTGATT | yes | 45.7 | 22 | 6 | 1 | 15 | 1 | 1 | | | | 1 | | | 1 | 1 | |
| IV | 263124-285490 | DET0273-0295 | 22,366 | pstRNA-Lys | TAGGCCGTACAGGTCTCGAACCTGTGACCCTCTGATT | yes | 45.7 | 23 | 6 | 1 | 16 | 1 | 1 | | | | 1 | | | 1 | 1 | |
| V | 803127-810526 | DET0875-0883 | 7,400 | tRNA-Lys-2 | TAGGCCGTACAGGTCTCGAACCTGTGACCCTCTGATTAAAAG | yes | 38.9 | 9 | 2 | 3 | 2 | 1 | | | | | | | | | | DET0876 |
| VI | 810529-832902 | DET0884-0905 | 22,366 | tRNA-Lys-2 | TAGGCCGTACAGGTCTCGAACCTGTGACCCTCTGATTAAAAGTCA | yes | 45.7 | 22 | 6 | 1 | 15 | 1 | 1 | | | | 1 | | | 1 | 1 | |
| VII | 968488-1023745 | DET1066-1118 | 55,258 | unknown | unresolved | yes | 50.6 | 52* | 31 | 11 | 23 | 3 | | 2 | 2 | 1 | | 1 | | | | |
| VIII | 1325415-1338416 | DET1472-1478 | 13,002 | tRNA-Val-3 | TTTACACGCAGGGGGTCATAGGTTCGAATCCTATACCGCCCACCA | yes | 39.3 | 7 | 1 | 0 | 4 | 1 | | | | | | | | | | |
| IX | 1398183-1407943 | DET1552-1561 | 9,761 | tRNA-Ala-3 | CTCAGCTCCA | yes | 40.8 | 10 | 2 | 1 | 3 | | 1 | | | | | | | | | DET1559 |
| Total amount of DNA: | | | 199,384 | | | | | | | | | | | | | | | | | | | |
| % of chromosome: | | | 13.6% | | | | | | | | | | | | | | | | | | | |

*Region VII shows extensive synteny with the *Bacillus cereus* 10987 prophage.

Observations of enhanced growth in mixed culture (stimulated by amino acids and cell extracts), as opposed to poor growth of pure cultures of *D. ethenogenes*, suggest a dependent ecological interaction (symbiosis) with other species in an anaerobic consortium, as suggested previously (*31*). Aspects of the *D. ethenogenes* metabolic profile support this notion.

*D. ethenogenes* requires acetate as a carbon source (*2*). Acetyl–coenzyme A (CoA) synthetase, pyruvate-ferredoxin oxidoreductase, and gluconeogenesis enzymes indicate a pathway for acetate assimilation (fig. S6A). The tricarboxylic acid (TCA) cycle in anaerobes is usually incomplete and provides precursors for amino acid synthesis. Some genes encoding TCA-cycle enzymes were found in a single operon; however, the enzyme for a key step, citrate synthetase, which is needed to form glutamate family amino acids, is absent.

In the previously mentioned 31-kb tandem duplication, several CDSs involved in the Wood/Ljungdahl acetyl-CoA pathway of $CO_2$ fixation (*32*) were predicted. These include formyltetrahydrofolate ligase, a corrinoid iron-sulfur protein, and the α subunit of carbon monoxide dehydrogenase/acetyl-CoA decarbonylase-synthase (CODH/ACDS). Other essential components, such as methylene-tetrahydrofolate reductase and the β subunit of CODH/ACDS responsible for $CO_2$ reduction to CO (*32*), are missing (fig. S6B). Perhaps an ancestor of *D. ethenogenes* was an autotroph using the acetyl-CoA pathway, which may have been disrupted after genetic rearrangements. Phylogenomic analysis strongly supports the classification of *D. ethenogenes* as a relative of the green nonsulfur bacteria, all of which are photoautotrophic.

*D. ethenogenes* possesses a nitrogenase-encoding operon (*nifHI$_1$I$_2$DKENB*, DET1151-1158) and a distal gene, *nifV* (DET1614), which encodes homocitrate synthetase used in nitrogenase FeMo-cofactor biosynthesis. Phylogenetically, the nitrogenase structural genes belong to cluster III (*33*), which includes nitrogenases from diverse anaerobic Bacteria and Archaea. The *nif* cluster is adjacent to an adenosine triphosphate (ATP)–binding cassette (ABC) transporter for molybdenum, an element essential to nitrogenase. These findings suggest that *D. ethenogenes* can fix nitrogen, which is difficult to demonstrate in the undefined complex medium in which this organism grows and is surprising, in light of the bacterium's complex nutrient requirements.

Complete pathways for the biosynthesis of all of the amino acids were identified, with the exception of methionine and possibly glutamate. Pathways for the synthesis of some cofactors appear incomplete. *D. ethenogenes* requires the corrinoid vitamin $B_{12}$ in large amounts for growth (*2*). Corrinoids serve as a prosthetic group in bacterial RDs, homologous to those from *D. ethenogenes* (*34*), and inhibitor studies implicate corrinoids in *D. ethenogenes* RDs (*18*). Although no de novo corrinoid-ring biosynthetic genes are encoded, the 31-kb duplication (fig. S2) contains several genes (*cobCDSTU*) for corrinoid salvage (*35*) and an adjacent cobalamin/$Fe3^+$ ABC transporter. Amplification of these genes may have allowed for more efficient uptake and salvage of exogenously obtained corrin precursors. Because there is little nucleotide difference between the duplicates, the repeat appears quite recent and may have occurred during the period when the organism was cultured in the absence of added vitamin $B_{12}$ (*2*).

Similarly, genes for quinone modification, but not for quinone-ring synthesis, were found, although the role of quinones in electron transport in *D. ethenogenes* is unclear. Instead of genes for biotin synthesis, a puta-

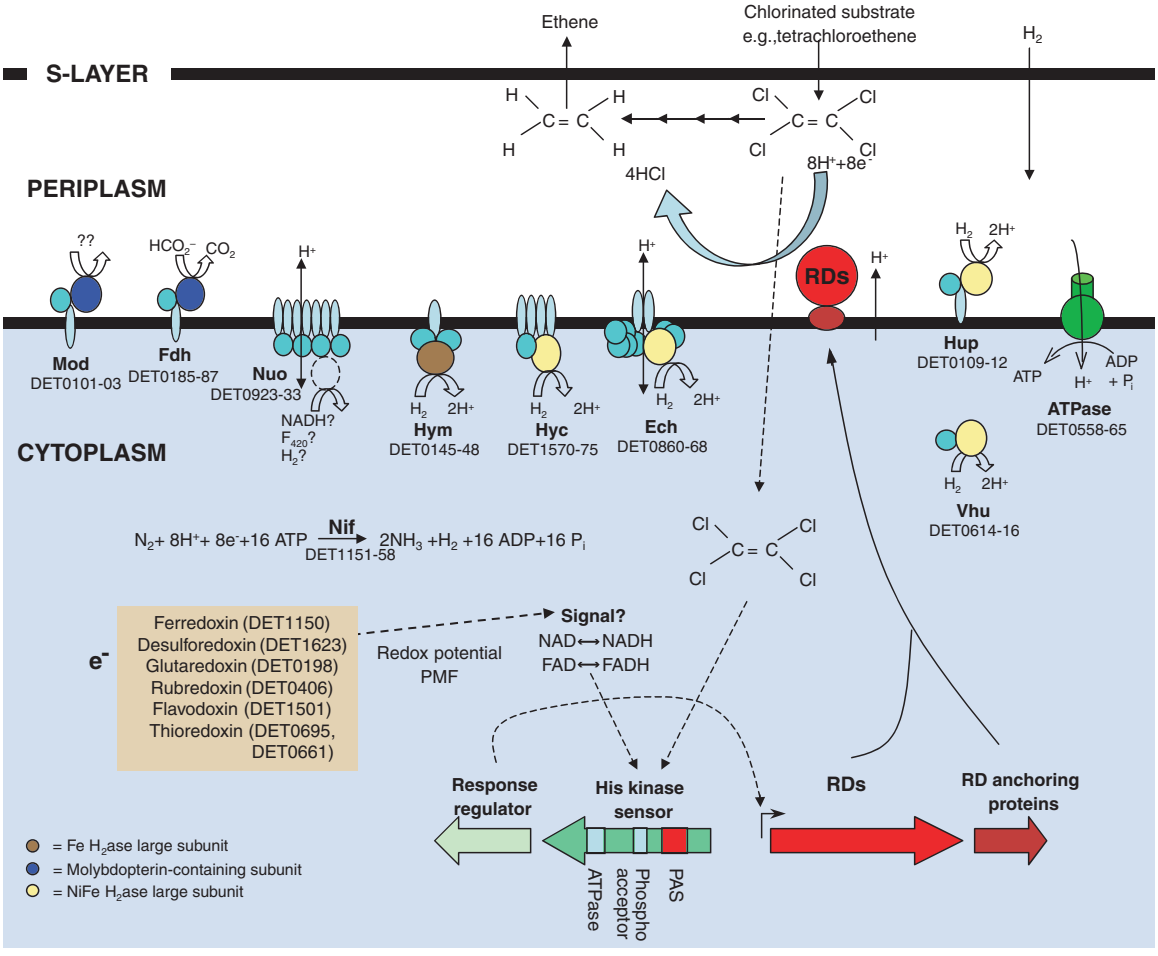


**Fig. 1.** Schematic representing the enzymes and regulatory network hypothesized from genome analysis to control dechlorination activities by *D. ethenogenes*. This network involves cytoplasmic two-component system regulators that sense various stimuli and up-regulate reductive dehalogenase activity in response to the redox status of the cell. $H_2$ase, hydrogenase. For a listing of loci encoding putative reductive dehalogenases and two-component systems, see tables S2 and S3, respectively.

tive *bioY* (DET1184) encoding a biotin transporter was found.

Thus, the genome sequence provides insight into the organism's complex nutrient requirements and its commitment to the dehalorespiratory process. The genome also suggests that an ancestor was a nitrogen-fixing autotroph. In the long term, genome data will serve as a foundation for the development of phylogenetic and functional marker probes, for detection and monitoring of *D. ethenogenes* in the environment, and for population genetic studies.


**References and Notes**

1. H. Smidt, W. M. De Vos, *Annu. Rev. Microbiol.* **58**, 43 (2004).
2. X. Maymó-Gatell, Y. Chien, J. M. Gossett, S. H. Zinder, *Science* **276**, 1568 (1997).
3. P. Hugenholtz, B. M. Goebel, N. R. Pace, *J. Bacteriol.* **180**, 4765 (1998).
4. L. Adrian, U. Szewzyk, J. Wecke, H. Görisch, *Nature* **408**, 580 (2000).
5. D. E. Fennell, A. B. Carroll, J. M. Gossett, S. H. Zinder, *Environ. Sci. Technol.* **35**, 1830 (2001).
6. S. H. Zinder, in *The Encyclopedia of Environmental Microbiology*, G. Bitton, Ed. (Wiley, New York, 2002), pp. 507–516.
7. F. E. Löffler, Q. Sun, J. Li, J. M. Tiedje, *Appl. Environ. Microbiol.* **66**, 1369 (2000).
8. E. R. Hendrickson *et al.*, *Appl. Environ. Microbiol.* **68**, 485 (2002).
9. D. W. Major *et al.*, *Environ. Sci. Technol.* **36**, 5106 (2002).
10. J. He, K. M. Ritalahti, K. L. Yang, S. S. Koenigsberg, F. E. Loffler, *Nature* **424**, 62 (2003).
11. D. E. Fennell, I. Nijenhuis, S. F. Wilson, S. H. Zinder, M. M. Häggblom, *Environ. Sci. Technol.* **38**, 2075 (2004).
12. S. D. Ehrlich *et al.*, *Gene* **135**, 161 (1993).
13. A. B. Reams, E. L. Neidle, *Annu. Rev. Microbiol.* **58**, 119 (2004).
14. A. B. Reams, E. L. Neidle, *Mol. Microbiol.* **47**, 1291 (2003).
15. A. B. Reams, E. L. Neidle, *J. Mol. Biol.* **338**, 643 (2004).
16. J. R. van der Meer, V. Sentchilo, *Curr. Opin. Biotechnol.* **14**, 248 (2003).
17. J. K. Magnuson, M. F. Romine, D. R. Burris, M. T. Kingsley, *Appl. Environ. Microbiol.* **66**, 5141 (2000).
18. J. K. Magnuson, R. V. Stern, J. M. Gossett, S. H. Zinder, D. R. Burris, *Appl. Environ. Microbiol.* **64**, 1270 (1998).
19. A. Neumann, H. Scholz-Muramatsu, G. Dieckert, *Arch. Microbiol.* **162**, 295 (1994).
20. B. A. van de Pas *et al.*, *J. Biol. Chem.* **274**, 20287 (1999).
21. R. Villemur, M. Saucier, A. Gauthier, R. Beaudet, *Can. J. Microbiol.* **48**, 697 (2002).
22. F. Sargent, *Trends Microbiol.* **9**, 196 (2001).
23. J. C. Venter *et al.*, *Science* **304**, 66 (2004).
24. M. Schobert, H. Gorisch, *Microbiol.* **147**, 363 (2001).
25. C. Wagner *et al.*, *Infect. Immun.* **70**, 6121 (2002).
26. P. M. Vignais, B. Billoud, J. Meyer, *FEMS Microbiol. Rev.* **25**, 455 (2001).
27. J. Meuer, H. C. Kuettner, J. K. Zhang, R. Hedderich, W. W. Metcalf, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5632 (2002).
28. W. Schumacher, C. Holliger, A. J. Zehnder, W. R. Hagen, *FEBS Lett.* **409**, 421 (1997).
29. S. Bäumer *et al.*, *J. Biol. Chem.* **275**, 17968 (2000).
30. D. R. Smith *et al.*, *J. Bacteriol.* **179**, 7135 (1997).
31. T. D. DiStefano, J. M. Gossett, S. H. Zinder, *Appl. Environ. Microbiol.* **58**, 3622 (1992).
32. T. I. Doukov, T. M. Iverson, J. Seravalli, S. W. Ragsdale, C. L. Drennan, *Science* **298**, 567 (2002).
33. J. P. Zehr, B. D. Jenkins, S. M. Short, G. F. Steward, *Environ. Microbiol.* **5**, 539 (2003).
34. A. Neumann *et al.*, *Arch. Microbiol.* **177**, 420 (2002).
35. M. J. Warren, E. Raux, H. L. Schubert, J. C. Escalante-Semerena, *Nat. Prod. Rep.* **19**, 390 (2002).


36. The complete genome sequence reported in this paper has been submitted to GenBank (accession number d_ethenogenes_195_68 CP000027). Supported by the U.S. Department of Energy, Office of Biological Energy Research, cooperative agreement DE-FC02-95ER61962. L.A. was supported by NSF grant MCB 0236044. We thank G. S. A. Myers, S. Salzberg, M. Pop, O. White, M. Heaney, S. Lo, M. Holmes, M. Covarrubias, J. Sitz, A. Resnick, J. Zhao, M. Zhurkin, S. Sengamalay, P. Sellers, R. Deal, R. Karamchedu, and V. Sapiro for informatics, database, and software support. Sequence data for *Silicibacter pomeroyi* were obtained from The Institute for Genomic Research Web site (www.tigr.org).


**Supporting Online Material**

www.sciencemag.org/cgi/content/full/307/5706/105/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S4
References

1 July 2004; accepted 4 November 2004
10.1126/science.1102226

# Y Chromosome of *D. pseudoobscura* Is Not Homologous to the Ancestral *Drosophila* Y

**Antonio Bernardo Carvalho[1]* and Andrew G. Clark[2]**

We report a genome-wide search of Y-linked genes in *Drosophila pseudoobscura*. All six identifiable orthologs of the *D. melanogaster* Y-linked genes have autosomal inheritance in *D. pseudoobscura*. Four orthologs were investigated in detail and proved to be Y-linked in *D. guanche* and *D. bifasciata*, which shows that less than 18 million years ago the ancestral *Drosophila* Y chromosome was translocated to an autosome in the *D. pseudoobscura* lineage. We found 15 genes and pseudogenes in the current Y of *D. pseudoobscura*, and none are shared with the *D. melanogaster* Y. Hence, the Y chromosome in the *D. pseudoobscura* lineage appears to have arisen de novo and is not homologous to the *D. melanogaster* Y.

The origin and evolution of the *Drosophila* Y seem to be different from that of the canonical (e.g., mammalian) Y chromosomes. Mammalian sex chromosomes originated from an ordinary pair of autosomes, so that when one of the homologs acquired a strong male-determining gene and became a Y, the other homolog became the X. Progressive gene loss from the Y resulted in a mostly degenerated chromosome (*1*, *2*). The main evidence for this common origin of the sex chromosomes is that most of the mammalian Y-linked genes and pseudogenes are shared with the X (*3*). On the other hand, in *D. melanogaster* none of the nine known single-copy Y-linked genes have homologs on the X. Instead, their closest homologs are autosomal (*4*, *5*), which strongly suggests that they were acquired from the autosomes by gene duplications, as has been shown for two mammalian Y genes (*3*). *D. melanogaster* X and Y chromosomes do share two multicopy genes, but it is doubtful that they represent common ancestry: *Ste-Su*(*Ste*) genes were recently acquired from an autosome (*6*), whereas *rDNA* genes are present in nonhomologous chromosomes in different *Drosophila* species (*5*, *7*, *8*). The lack of any clear sign of X-Y homology is consistent with the hypothesis that the *Drosophila* Y did not originate from the degeneration of an X-homolog, but rather from a supernumerary ("B") chromosome that acquired X-pairing and male-related genes, though it is also possible that the degeneration went so far that all signs of homology were erased (*5*, *9*). Whatever its true origin, the *melanogaster*-like *Drosophila* Y is at least 63 million years (My) old, dating back to the separation of the Sophophora and *Drosophila* subgenera (*10*), because at least three genes (*kl-2*, *kl-3*, and *kl-5*) are shared between the Y chromosomes of *D. melanogaster* (*4*, *5*) and *D. hydei* (*11*, *12*).

The assembled *D. pseudoobscura* genome sequence (*13*) now makes possible a genome-wide study of Y-linked genes in a second *Drosophila* species. Given that this species also belongs to the subgenus Sophophora, we expected it to share the ancestral Y chromosome with *D. melanogaster*, as does the more distant *D. hydei*. In fact, we found in the *D. pseudoobscura* genome orthologs for most *D. melanogaster* Y-linked genes (*kl-2*, *kl-3*, *kl-5*, *ORY*, *PPr-Y*, and *ARY*), as well as the orthologs of their autosomal parental genes (*14*). The orthology is strongly supported by phylogenetic analysis (Fig. 1 and fig. S1) and by the reciprocal best match criteria (*14*). However, when we tested for Y linkage by polymerase chain reaction (PCR) with genomic DNA from males and females, we found that all six genes are present in both sexes, ruling out Y linkage. A

[1]Departamento de Genética, Universidade Federal do Rio de Janeiro, Caixa Postal 68011, CEP 21944-970, Rio de Janeiro, Brazil. [2]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA.

*To whom correspondence should be addressed. E-mail: bernardo@biologia.ufrj.br