Title
Global Population Genetics and Evolution of Invasive Biotypes in the Whitefly Complex *Bemisia tabaci*

Author
Hadjistylli, Margarita

Publication Date
2010

Peer reviewed|Thesis/dissertation

Global Population Genetics and Evolution of Invasive Biotypes in the Whitefly Complex *Bemisia tabaci*

By

Margarita Hadjistylli

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Environmental Science, Policy, and Management

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor George K. Roderick, Chair
Professor Rosemary G. Gillespie
Professor Craig Moritz

Fall 2010

Global Population Genetics and Evolution of Invasive Biotypes in the Whitefly Complex *Bemisia tabaci*

by Margarita Hadjistylli

Abstract

Global Population Genetics and Evolution of Invasive Biotypes in the Whitefly Complex *Bemisia tabaci*

by

Margarita Hadjistylli

Doctor of Philosophy in Environmental Science, Policy, and Management

University of California, Berkeley

Professor George K. Roderick, Chair

The presence of morphologically cryptic lineages with divergent molecular, ecological, and physiological traits within a species is an intriguing evolutionary phenomenon that offers unique opportunities for evolutionary genetics studies. One such system is the whitefly species complex *Bemisia tabaci* (Hemiptera: Aleyrodidae), which comprises several cryptic lineages, known as "biotypes" with worldwide distribution, including two of the world's worst invasive pests. In this dissertation I take a population genetics approach to examine the global genetic structure of *B. tabaci* biotypes, with a focus on the origins, historical demography, and invasion pathways of the two invasive biotypes, known as "B" and "Q".

I begin with a historical overview of multilocus molecular markers used to examine aspects of the biology, ecology, and genetics of the *B. tabaci* species complex. The first markers employed were allozymes, particularly esterases, which became the basis for the biotype nomenclature, and were substantiated by ecological and biological data. The exploration of various DNA based markers has established that biotypes within *B. tabaci* are exceptionally diverse genetically, in spite of their identical morphologies.

Global population genetics analyses using microsatellite markers showed that well-characterized *B. tabaci* biotypes correspond to real genetic entities with strong geographic structure, and limited or no gene flow among them. The resulting genetic clusters from this analysis are in general agreement with the only well-resolved global phylogeny of the species, which is based on a single mitochondrial gene (*cytochrome oxidase I*). However, some cases of conflict in the two sets of markers do exist, perhaps associated with the different modes of inheritance, thus cautioning against the use of mitochondrial DNA as a single marker for species or subspecies delineation.

Analysis of genetic data with more sophisticated Bayesian coalescent-based approaches offers the opportunity to study both contemporary and ancestral invasion pathways. Using such an approach, I showed that divergence histories of the invasive biotypes B and Q coincided with periods of extensive human movement and trade of agricultural

goods in the Mediterranean, the Middle East, and Africa during the Iron and Bronze Ages, and the Roman period. Results also indicate that ancestral populations to the current B and Q biotypes had much larger effective sizes than those of emerging biotypes, a pattern consistent with expectations of diversification in invasive species.

In a contemporary context, I investigated the recent invasion history of biotype Q in the USA. I found that populations introduced into the USA originated from both the Western and Eastern Mediterranean, in at least three independent cryptic invasions, and spread directly from a single initial introduction site, likely through plant trade.

Findings from this dissertation underscore the practical importance of better monitoring invasions of this insect and other invasive pests at points of entry and dispersal through trade of plant material. From a theoretical perspective, this work adds insights into the origins of biotypes, both in the *B. tabaci* complex and more generally, emphasizing the demographic processes involved in diversification of invasive biotypes. The research highlights the potential to use *B. tabaci* in studies of broader applied as well as evolutionary significance.

Dedication

To my parents for all the sacrifices they made for my education and to my husband, Menelaos, for standing by me throughout the whole journey

Table of contents

Acknowledgements iv

Introduction 1

Chapter 1: Review of molecular markers used in population genetics studies of *Bemisia tabaci*

Abstract 8
Introduction 9
Protein markers 10
DNA-level markers 15
DNA-level PCR-based markers 17
Additional considerations and future directions 30
Conclusions 32
References 34
Tables 43

Chapter 2: Global population genetics of a worldwide pest: genetic diversity and population structure in the *Bemisia tabaci* cryptic species complex

Abstract 47
Introduction 48
Materials and Methods 50
Results 54
Discussion 57
References 64
Tables 70
Figures 75

Chapter 3: Historical demography and divergence of the invasive *Bemisia tabaci* biotypes B and Q associated with human movement and trade in the Mediterranean, Middle East, and Africa

Abstract 83
Introduction 84
Materials and Methods 85
Results 88
Discussion 89
References 93
Tables 97
Figures 100

Chapter 4: Evidence for multiple independent and cryptic invasions of the whitefly *Bemisia tabaci* biotype Q in the USA

Abstract 106
Introduction 107
Materials and Methods 109
Results 113
Discussion 117
References 122
Tables 127
Figures 133

Conclusions 144

Acknowledgements

I would like to begin by thanking George Roderick, my advisor and Chair of my Dissertation Committee. George was supportive of any initiative I pursued for my research and motivated me to keep up with my goals, while making sure that financial issues were the last thing to worry about during my studies. His encouragement for my participation in conferences and meetings worldwide helped me make connections and opened my mind to the academic world. I feel fortunate to have worked with George on an exciting research project and because through his mentoring I learned how to approach scientific questions and research in general.

I am greatly indebted to my Dissertation Committee members, Rosemary Gillespie and Craig Moritz, who gave me constructive comments and feedback from the beginning of my project to the final drafts of my dissertation. The opportunity to teach with, and be taught from Craig has been one of my most rewarding experiences at Berkeley, and inspired my interest in biology and evolution. I am also very grateful to Steve Welter and Vince Resh who have been great teachers and because our interactions helped shape my thinking as a researcher and teacher.

My research would not have been possible without Judith Brown of the University of Arizona, Tucson. As a professor with renowned whitefly expertise, she gave me constant feedback throughout my studies. In addition, Judith provided me with whitefly samples from her collection and gave me access to her lab, where I received my first molecular training from Donald Rogan. Furthermore, she helped me make connections with whitefly researchers from around the world. I am also indebted to many colleagues worldwide for contributing whitefly specimens, especially P. Bleeker, D. Bosco, C. Erdogan, P. De Barro, H. Delatte, N. Gauthier, K. Goodman, E. Hernandez, R. Horowitz, J. Imamura, and L. Papayiannis.

I would also like to acknowledge the funding I received from various sources: UC Berkeley, Department of Environmental Science, Policy and Management, Division of Organisms and Environment, Berkeley International Office, USDA-NRI, France-Berkeley Fund, NSF RCN: Invasive Insects, Moorea Biocode Project, Margaret C. Walker Fund Award for Systematic Entomology, William Carroll Smith Fellowship, Julius Freitag Memorial Award, and Robert L. Usinger Memorial Fund. I especially acknowledge the three-year educational grant from the A.G. Leventis Foundation in Cyprus. I would also like to thank the Ministry of Agriculture and the Government of Cyprus for the educational leave provided to me for doctoral studies.

At Berkeley, I would like to thank my labmates in the "Evolab" who have been great friends, and gave me wonderful moments to remember, from the lab bench, to the field, and to our nights out. I especially thank Pete Croucher whose guidance and help with genetic data analysis has been enormous; from him I learned how to be conscientious in research. I also thank Sean Schoville and Jeff Lozier for their help with molecular lab work. I feel very lucky that I had the opportunity to work with talented undergraduate students, especially Theresa Canavan, Jeff Sorokin, and Danielle Jones. Most of all however, I am grateful to have worked with Athena Lam, who helped me become a better teacher by being an ideal student. Her contribution to my research improved it significantly and her presence made my long hours in the lab more enjoyable.

Kari Roesch Goodman has been a key figure from the beginning to the end of my PhD studies and I feel fortunate to have gone through orals, teaching, conferences, lab work and data analysis with a friend close to me. Our dissertation group meetings, along with Pauline Kamath and Emily Rubidge helped me enormously – but above all I am thankful for our friendship. Natalia Choussou-Polydouri, Diamantis Sellis, Pete Oboyski, Jennifer Imamura, and Sal, have been great friends inside and outside the lab. Sotiria and Alan were family to me in Berkeley and supported me in every possible way; their presence in my life made home feel closer than it was.

I am thankful to my parents who always believed in me and supported all my goals. My sisters, Marina and Theodora, my best friend Vicky, and my sister in law Stella, kept my spirits up with phone calls and encouraging messages from Cyprus.

I cannot find words that are strong enough to express the gratitude, appreciation, and luck I feel for having Menelaos, my husband, by my side in every step of this process; from his help during my late-night hours in the lab, to editing drafts of my dissertation. Our friendship made difficult and challenging times less stressful and happy moments more joyful. Going through grad school and exploring California's parks together made the past five years the most rewarding and fulfilling of my life, academically and personally. I am grateful to have done this journey and most grateful to have done it with you.

Introduction

Studies focusing on genetically and ecologically divergent lineages within a species offer excellent opportunities for exploring the evolutionary and demographic processes in the early stages of speciation (Losos and Glor 2003). In genetic divergence leading to speciation, the role of geography and the environment may vary, with theory predicting that geographically isolated populations diverge and diversify as a result of random accumulations of genetic mutations (Dobzhansky 1936; Muller 1940) and selection driving adaptation to different environments (Schluter 2001). Evolutionary units in these early phases of differentiation, may represent "intermediate" stages between polymorphic populations and distinct species, are often described as "biotypes" (Dres and Mallet 2002). Biotypes have been documented widely across different animal taxa; particularly herbivorous insects may exhibit adaptations to different host-species, as well as differences in resistance to biotic and abiotic stressors. Biotypes have been reported from several insect orders including Hemiptera (e.g. *Nilaparvata* brown planthopper), Diptera (*Rhagoletis* apple maggot and *Anopheles* mosquitoes), Coleoptera (*Epilachna* ladybird beeteles), Hymenoptera (Platycampus sawfly) and Lepidoptera (e.g. *Mitoura* hairstreak butterflies) (see Dres and Mallet 2002 for a review).

A particularly informative case of genetically and ecologically divergent lineages is that of cryptic species. Cryptic species exhibit very little or no morphological variation – a phenomenon termed "morphological stasis" - in spite of extreme genetic, physiological and ecological diversification (Bickford et al. 2007). It has been suggested that morphological stasis can be maintained by stabilizing selection throughout the evolutionary history of cryptic species (Charlesworth et al. 1982). Not surprisingly, this phenomenon causes enormous confusion for taxonomic classification, with profound implications for conservation planning, biological control, and pest management (Bickford et al. 2007).

The sweetpotato whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae), is a prominent example of a cryptic species, comprising more than 25 cryptic biotypes worldwide. The species was first described as *Aleyrodes tabaci* by Gennadius in 1889 in samples collected from tobacco plants in Greece. Biotypes of *B. tabaci* exhibit phenotypic variation in pupal (nymph) stages, with diverse shapes, sizes, and colors of setae and pores, much of which is a response to leaf surface morphology (Neal and Bentz 1999). Because of the host-dependent plasticity of these characters, they cannot be used for taxonomic purposes. The lack of any other morphological variation complicates the taxonomy of this group (Gill 1990; Mound 1963).

Biological and ecological data from several biotypes have demonstrated profound differentiation in various traits, including host-specificity, plant virus transmission efficiency, and insecticide resistance. The extent to which these traits are expressed in different biotypes has defined the level of pest-status and ability for invasiveness. For example the most widespread member of this complex, biotype "B", once described as *Bemisia argentifolii* (Bellows et al. 1994), has exhibited extreme adaptation to environmental conditions, association with emergent plant viruses (Brown et al. 2000), high fecundity and resistance to insecticides, and a very wide host range (Brown et al. 1995b). More recently, the Q biotype has emerged as an invasive pest of comparable importance in Europe, Asia, and the North and Central America (Chu et al. 2006; Guirao

et al. 1997; Lee et al. 2008; Martinez-Carrillo and Brown 2007; Ueda and Brown 2006) through the trade of poinsettia plants. In contrast, there are other examples of *B. tabaci* biotypes that are host-specific, as is the case of the Jatropha biotype of Puerto Rico, which feeds exclusively on *Jatropha gossypifolia* and transmits the Jatropha mosaic virus (Bird 1957), and the T biotype in Italy, which is restricted to *Euphorbia characias* (Demichelis et al. 2005).

Because of their morphological conservatism, the identification of cryptic species relies primarily on molecular and genetic tools that have the potential to recover extreme underlying levels of diversity (e.g. Jacques et al. 2009; Novo et al. 2010). In the early 1990's, molecular data, such as allozymes (particularly esterases) and later DNA markers, brought to light significant genetic variation across worldwide populations of *B. tabaci* (Brown et al. 1995a; Brown et al. 1995b; Costa and Brown 1991; Costa et al. 1993; Frohlich et al. 1999). The genetic polymorphism was suggestive of species-level differences in some cases (Bellows et al. 1994; Brown et al. 1995a) and led to the conclusion that *B. tabaci* is a sibling or cryptic species complex. Biotypes were given alphabetical designations (A, B, C, D, etc) based on their distinctive esterase bands (Brown et al. 1995a; Perring 2001). In addition to molecular data, direct analyses of gene flow and mating experiments confirmed that some biotypes were largely reproductively isolated, further reinforcing the hypothesis of a cryptic species complex (Bedford et al. 1994; Brown et al. 2000; Byrne et al. 1995; Costa et al. 1993; De Barro et al. 2000).

This dissertation aims to provide a comprehensive picture of molecular and population genetics of the *B. tabaci* complex at a global scale, and at the same time address questions pertaining to the origins and diversification of biotypes. Chapter 1 provides a review of molecular markers used to study the *B. tabaci* system. Chapter 2 presents a global population genetics analysis including invasive and native biotypes. Chapters 3 and 4 deal specifically with the invasive biotypes B and Q, with Chapter 3 focusing on their evolutionary relationships and historical demography and Chapter 4 investigating the recent invasion of biotype Q in the USA. Details of each chapter follow.

In Chapter 1 I provide a detailed review of molecular markers used to study populations in the *B. tabaci* complex around the world from the early 90's to the present. I start with allozymes (esterases), the first molecular marker used to distinguish the invasive biotype B from the US native biotype A (Costa and Brown 1991), and continue with DNA based markers. I distinguish between those based on fragment separation in gels (restriction fragment length polymorphisms or RFLPs) and those based on the polymerase chain reaction (PCR); the randomly amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), and finally microsatellites. For each of these markers I provide a description of their function and application, and outline their relative advantages and limitations. I conclude by noting considerations for the selection of molecular markers and prospects for future studies of the *B. tabaci* species complex.

In Chapter 2, I undertake a global population genetics analysis of whiteflies from five continents using microsatellite markers. Although phylogenetic studies have been conducted previously to characterize relationships at the global level, our best knowledge relied on a single gene, the mitochondrial cytochrome oxidase I (mtCOI), which can be unreliable in depicting the true population histories. This study is the first attempt to

analyze, at the population level, multiple biotypes with various ecological and biological characters, from native host-restricted to invasive polyphagous members. The overall aim of this approach was to improve our understanding of biotype relationships in *B. tabaci*, while at the same time contrast results based on microsatellite nuclear DNA with those from mitochondrial DNA for the same biotypes.

Biotypes B and Q are not just interesting from an economic perspective; phylogenetic studies have shown that they form sister clades in the global phylogeny of the species, suggesting a more recent common ancestry compared to other biotypes. Since the suspected native range is the Middle East/ North-East Africa for biotype B (Perring 2001) and the Mediterranean for biotype Q (Brown 2007), this could point to a center of diversification around Africa, and the evolution of pest-related attributes in this area. In Chapter 3, I focus exclusively on biotypes B and Q by examining genetic diversity in several populations across their native and invaded range. Their relationships are examined by assessing genetic distances between their populations and those of their closest relatives in the Middle East (for B biotype) and Africa (for Q biotype). To examine their divergence histories, I estimate population demographic parameters, such as splitting times, effective population sizes and migration rates using the "Isolation with Migration" model (Nielsen and Wakeley 2001), and associate findings with historical periods of human movement in the area.

Contemporary invasions of species are becoming more and important because their rate has increased at alarming rates in the past 50 years as a result of extensive human traveling and global trade (Pimentel et al. 2005). Studies of species invasions, especially those of cryptic species, have been expanded with the use of genetic data and the application of sophisticated Bayesian approaches based on theoretical models of gene evolution (*i.e.* the coalescent) (Cornuet et al. 2008). For example, using such approaches, one can infer the invasion history and pathways of dispersal of invading populations (Guillemaud et al. 2010; Miller et al. 2005). In the fourth and final Chapter, I examine the history of the recent invasion of biotype Q in the USA. I analyze multiple populations sampled from several states around the time of its introduction and compare their genetic variation to populations from the Mediterranean native range. Using Bayesian assignment methods and approximate Bayesian computation approaches, I trace the sources of this invasion as well as possible routes of subsequent dispersal across the US. Such knowledge has important practical implications for managers aiming to achieve a more efficient quarantine and management of this biotype.

The four chapters of this dissertation present a synthesis of a historical overview of studies in the *B. tabaci* complex with molecular markers, and current analyses of global population structure, historical demography, and invasion history of biotypes. With this dissertation, I aim to contribute to our knowledge and understanding of the evolutionary history, current genetic structure, and future trajectories of this species complex. At the same time, from an applied perspective, I aim to provide insights into the pathways of introduction and dispersal of invasive biotypes, which will facilitate their more efficient management.

References

Bedford ID, Briddon RW, Brown JK, Rosell RC, Markham PG. 1994. Geminivirus transmission and biological characterization of *Bemisia tabaci* (Gennadius) biotypes from different geographic regions. *Ann. Appl. Biol.* 125: 311-325.

Bellows TS, Perring TM, Gill RJ, Headrick DH. 1994. Description of a species of *Bemisia* (Homoptera: Aleyrodidae). *Ann. Entomol. Soc. Am.* 87: 195-206.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I. 2007. Cryptic species as a window on diversity and conservation. *Trends Ecol. Evol.* 22: 148-155.

Bird J. 1957. A whitefly-transmitted mosaic of *Jatropha gossypifolia*. *Tech. Pap. P.R. Agric. Exp. Sta.* 22: 1-35.

Brown JK. 2007. The *Bemisia tabaci* complex: genetic and phenotypic variability drives begomovirus spread and virus diversification. Plant Disease Feature. http://www.apsnet.org/online/feature/btabaci Accessed: July 15, 2010.

Brown JK, Coats SA, Bedford ID, Markham PG, Bird J, Frohlich DR. 1995a. Characterization and distribution of esterase electromorphs in the whitefly, *Bemisia tabaci* (Genn) (Homoptera: Aleyrodidae). *Biochem. Genet.* 33: 205-214.

Brown JK, Frohlich DR, Rosell RC. 1995b. The sweetpotato or silverleaf whiteflies: Biotypes of *Bemisia tabaci* or a species complex? *Annu. Rev. Entomol.* 40: 511-534.

Brown JK, Perring TM, Cooper AD, Bedford ID, Markham PG. 2000. Genetic analysis of *Bemisia* (Hemiptera : Aleyrodidae) populations by isoelectric focusing electrophoresis. *Biochem. Genet.* 38: 13-25.

Byrne FJ, Cahill M, Denholm I, Devonshire AL. 1995. Biochemical identification of interbreeding between B-type and non B-type strains of the tobacco whitefly *Bemisia tabaci*. *Biochem. Genet.* 33: 13-23.

Charlesworth B, Lande R, Slatkin M. 1982. A neon-Darwinian commentary on macroevolution. *Evolution* 36: 474-498.

Chu D, Zhang YJ, Brown JK, Cong B, Xu BY, Wu QJ, Zhu GR. 2006. The introduction of the exotic Q biotype of *Bemisia tabaci* from the Mediterranean region into China on ornamental crops. *Fla. Entomol.* 89: 168-174.

Cornuet J-M, Santos F, Beaumont MA, Robert CP, Marin J-M, Balding DJ, Guillemaud T, Estoup A. 2008. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* 24: 2713–2719.

Costa HS, Brown JK. 1991. Variation in biological characteristics and esterase patterns among populations of *Bemisia tabaci*, and the association of one population with silverleaf symptom induction. *Entomol. Exp. Appl.* 61: 211-219.

Costa HS, Brown JK, Sivasupramaniam S, Bird J. 1993. Regional distribution, insecticide resistance, and reciprocal crosses between the A-biotype and B-biotype of *Bemisia tabaci*. *Insect Sci. Appl.* 14: 255-266.

De Barro PJ, Driver F, Trueman JWH, Curran J. 2000. Phylogenetic relationships of world populations of *Bemisia tabaci* (Gennadius) using ribosomal ITS1. *Mol. Phylogen. Evol.* 16: 29-36.

Demichelis S, Arno C, Bosco D, Marian D, Caciagli P. 2005. Characterization of biotype T of *Bemisia tabaci* associated with *Euphorbia characias* in Sicily. *Phytoparasitica* 33: 196-208.

Dobzhansky T. 1936. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* 21: 113-135.

Dres M, Mallet J. 2002. Host races in plant-feeding insects and their importance in sympatric speciation. *Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci.* 357: 471-492.

Frohlich D, Torres-Jerez I, Bedford I, Markham P, Brown J. 1999. A phylogeographic analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers. *Mol. Ecol.* 8: 1593-1602.

Gill RJ. 1990. The morphology of whiteflies. In *Whiteflies: Their Bionomics, Pest Status and Management*, ed. D Gerling, pp. 13-46. Andover: Intercept Ltd.

Guillemaud T, Beaumont MA, Ciosi M, Cornuet JM, Estoup A. 2010. Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity* 104: 88-99.

Guirao P, Beitia F, Cenis JL. 1997. Biotype determination of Spanish populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Bull. Entomol. Res.* 87: 587-593.

Jacques N, Mallet S, Casaregola S. 2009. Delimitation of the species of the *Debaryomyces hansenii* complex by intron sequence analysis. *Int. J. Syst. Evol. Microbiol.* 59: 1242-1251.

Lee M, Uhm KB, Lee S, Park H, Choi HS, Park JW. 2008. Occurrence of *Bemisia tabaci* B- and Q-biotypes in Korea. *J. Insect Sci.* 8: 27-27.

Losos JB, Glor RE. 2003. Phylogenetic comparative methods and the geography of speciation. *Trends Ecol. Evol.* 18: 220-227.

Martinez-Carrillo JL, Brown JK. 2007. First report of the Q biotype of *Bemisia tabaci* in southern Sonora, Mexico. *Phytoparasitica* 35: 282-284.

Miller N, Estoup A, Toepfer S, Bourguet D, Lapchin L, Derridj S, Kim KS, Reynaud P, Furlan L, Guillemaud T. 2005. Multiple transatlantic introductions of the western corn rootworm. *Science* 310: 992-992.

Mound LA. 1963. Host-correlated variation in *Bemisia tabaci* (Gennadius) (Homoptera: Aleyrodidae). *Proc. R. Entomol. Soc. Lond. Ser. A Gen. Entom.* 38: 171-180.

Muller HJ. 1940. Bearings of the Drosophila work on systematics. In *The New Systematics*, ed. J Huxley, pp. 185-268. Oxford: Clarendon Press.

Neal JW, Bentz JA. 1999. Evidence for the stage inducing phenotypic plasticity in pupae of the polyphagous whiteflies *Trialeurodes vaporariorum* and *Bemisia argentifolii* (Homoptera: Aleyrodidae) and the raison d'etre. *Ann. Entomol. Soc. Am.* 92: 774-787.

Nielsen R, Wakeley J. 2001. Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics* 158: 885-896.

Novo M, Almodovar A, Fernandez R, Trigo D, Cosin DJD. 2010. Cryptic speciation of hormogastrid earthworms revealed by mitochondrial and nuclear data. *Mol. Phylogen. Evol.* 56: 507-512.

Perring TM. 2001. The *Bemisia tabaci* species complex. *Crop Protect.* 20: 725-737.

Pimentel D, Zuniga R, Morrison D. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol. Econ.* 52: 273-288.

Schluter D. 2001. Ecology and the origin of species. *Trends Ecol. Evol.* 16: 372-380.

Ueda S, Brown JK. 2006. First report of the Q biotype of *Bemisia tabaci* in Japan by mitochondrial cytochrome oxidase I sequence analysis. *Phytoparasitica* 34: 405-411.

Chapter 1

Review of molecular markers used in population genetics studies of *Bemisia tabaci*

This review has been published previously and is reproduced here with kind permission of Springer Science and Business Media

Hadjistylli M, Brown JK, Roderick GK. 2010. Gene flow and population genetics of the *Bemisia tabaci* complex. In *Bemisia: Bionomics and Management of a Global Pest*, ed. PA Stansly, SE Naranjo, pp. 69-103. doi:10.1007/978-90-481-2460-2_3

Abstract

Molecular markers have been widely exploited to study the whitefly species complex *Bemisia tabaci*, ever since the first discovery of the extreme genetic diversity among its members in the early 1990's. The first markers employed were allozymes, particularly esterases, which became the basis for the nomenclature of the *B. tabaci* members, known as biotypes, with distinct esterase gel patterns corresponding to newly described biotypes. At the same time, these descriptions provided the foundation for ecological studies and assays, which correlated biological/ecological characteristics with distinct esterase characterizations. These discoveries were followed by the exploration of DNA based markers, such as restriction fragment length polymorphisms (RFLPs), and subsequently PCR-based markers such as randomly amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), and microsatellites. All these markers have established unequivocally that the *B. tabaci* complex consists of exceptionally diverse genetically, yet highly cryptic morphologically biotypes. In addition, more specific studies have helped address questions about biotype association with host or geography, gene flow, invasiveness, and possible links with virus epidemics, insecticide resistance, and mating incompatibilities. Here, we review these cases and provide an overview of each marker system, with a description of their function, relative advantages and limitations, and specific applications in *B. tabaci*. We conclude by noting considerations and prospects for the selection of molecular markers, and anticipate that this review will serve as a reference for further studies of the *B. tabaci* species complex.

Introduction

The use of molecular markers in the *Bemisia tabaci* complex has been a definitive step in identifying the enormous genetic diversity hidden behind the morphological similarity among its members (Gill and Brown 2010), and in determining interrelationships. The presence of biologically-based biotypes in *B. tabaci* was first realized in the 1950s by Bird (Bird 1957; Bird and Maramorosch 1978), who found that morphologically indistinguishable populations of the whitefly differed substantially in biological and ecological traits, including host range, adaptability to different hosts, and plant virus-transmission efficiencies. Later studies used ecological and biological experiments to examine mating compatibilities as well as differences among distinct populations in phytotoxic induction, insecticide resistance, and behavior (Brown et al. 1995b). The use of molecular markers, starting in the late 1980s with allozymes, was an attempt to assess variability at the molecular level and provide a tool to distinguish among biological and ecological variants. Allozymes were used for more than 10 years in *B. tabaci* studies for identification of certain variants in a region, and also served as the basis for biotype characterization and nomenclature, which is still largely in use today. Interest in the *B. tabaci* system grew following the first major outbreak of the invasive B biotype in the southwestern United States in 1991, and the subsequent displacement of the indigenous A biotype (Brown et al. 1995b). The worldwide expansion in range of biotype B in the years that followed, accompanied by significant economic consequences in the agricultural sector, prompted an international effort involving intensive studies aimed at understanding genetic variability and possible correlations with host use and/or geographic origins.

While the use of allozymes revealed significant polymorphism within *B. tabaci*, rapid technological advancements in the field of molecular genetics allowed the exploitation of new, neutral multilocus molecular markers. A wealth of literature continues to be produced by research laboratories worldwide using molecular tools to examine differences among biotypes in biological and ecological traits including host use, virus transmission efficiency, insecticide resistance, breeding incompatibility, endosymbiont composition, and finally, geographical distribution. Most of these studies have only been able to make associations between these characteristics and genetic variants leading to general conclusions about the genetic and geographic structure of biotypes.

The use of molecular markers in *B. tabaci* has not only revealed important biological information, but also has practical applications. In pest management for example, it is useful to identify specific biotypes that are associated with economically important traits, including invasiveness, ability to disperse, insecticide resistance, higher potential for plant damage and transmission of plant viruses. For example, specific molecular markers were developed to identify the invasive B biotype and track its spread into new regions. This ability to determine early the presence of a biotype possibly carrying resistance alleles in a new region has allowed timely adjustments in management practices in the US and elsewhere. Molecular markers have also been used to identify and track natural enemies of *B. tabaci* in its native range, making use of correlations between the biogeographic lineages, and pinpointing the most likely geographic locale to scour for natural enemies that co-evolved with their whitefly host (Kirk et al. 2000).

More complex issues can also be addressed with the use of genetic markers coupled with appropriate analyses. For example, one can ask questions concerning the colonization history of a biotype in an area, e.g. when was a biotype introduced in an area, where did it come from, and what route did it follow? One can also address the demographic history, including, the founding effective population size, the extent of a genetic bottleneck associated with the founding event, and the pattern of subsequent population growth. A final, and often neglected point is that different genetic markers may originate from different parts of the genome, and as such, may have distinct properties that will affect suitability for certain studies. The choice of markers should therefore depend on the question one is seeking to answer, the objectives of the study, the extent of genetic polymorphism required to address the specific questions, and the statistical approaches available for the application of each technique (Parker et al. 1998).

This review aims to provide a historical perspective on use of the types of multilocus molecular markers that are available to study population genetics of organisms in general. At the same time, we highlight studies of the *B. tabaci* sibling species group, which therein illustrates the paucity of available population genetics data from which profound or definitive conclusions can be drawn. The use of allozymes to categorize biotypes based on esterase bands is summarized, as well as the use of markers and microsatellite allele frequencies to detect fine scale population differentiation. We distinguish between markers based on protein variation, DNA sequence variation and frequency variation through the polymerase chain reaction (PCR). For each, we provide an overview of the marker system, outline advantages and limitations, and provide a review of the applications (Table 1). We conclude by noting considerations for the selection of molecular markers and prospects for future studies of the *B. tabaci* sibling species group, or complex.

Protein markers

Enzyme Electrophoresis

Allozymes are biochemical variants of an enzyme encoded by different alleles at the same locus, whereas the term isozyme refers to variants of an enzyme encoded by different loci (Lowe et al. 2004). Thus, the use of allozymes as markers relies on obtaining different allelic variants for the same enzyme when screening individuals from one or more populations. Detection of allozyme variation was introduced in the 1960's and represents the first use of true molecular markers (Schlötterer 2004).

Enzyme electrophoresis is based on the fact that non-denatured proteins with different net charges have differential mobility when moving through a gel to which an electrical current is applied (Avise 2004). If an enzyme marker is polymorphic, meaning it has different variants in the same or different individuals, then the data produced are visible in the gel as different bands, representing different alleles. Thus differences in the overall banding pattern obtained for different populations indicate that there is some degree of genetic differentiation between the populations, at least for that particular locus. In most cases, heterozygous individuals will be distinguished from homozygotes as

having two bands (alleles) for a specific locus, but the interpretation may be hindered by the quaternary structure of the protein (monomeric vs dimeric).

Uses, assumptions, and limitations

The first studies that utilized allozyme markers were undertaken by Harris (1966) to quantify variation in populations of humans and by Johnson and colleagues (1966) and Hubby and Lewontin (1966) to assess variation in natural *Drosophila* populations. Since then, allozymes have been used in many systems to examine and analyze the degree of genetic differentiation, inbreeding, genetic drift, and gene flow between populations (migration / hybridization) (Hoy 2003), as well as to provide evidence for polyploidy and introgression between species (Lowe et al. 2004).

Development of an array of methods of data analysis and interpretation for allozymes together with the large availability of protocols (Lowe et al. 2004) facilitated their use in studies of insects and other organisms. Also, the ease and low cost of application for large numbers of samples allowed their extensive use across laboratories and for different study systems. Although a large number of allozymes proved not to be polymorphic for some species, their initial use provided much more information about intraspecific variation than earlier genetic markers which relied on phenotypic Mendelian traits such as flower or fruit color (Parker et al. 1998).

The amount of polymorphism that can be revealed using allozymes is limited as they only allow detection of variation at the protein level, i.e. only what DNA encodes for, not variation in DNA itself (Schlötterer 2004). For example, mutations occurring in introns or synonymous substitutions in DNA sequences that code for the same amino acid and thus not reflected at the protein level cannot be scored with allozymes (Behura 2006). As a result, much of the polymorphism observed at the DNA level goes undetected with allozymes, leading to underestimation of true genetic variation. Furthermore, because allozymes are proteins encoded by genomic DNA, some of the observed polymorphism may not be neutral and thus could have been exposed to selective processes, potentially obscuring analysis and data interpretation. For example Karl and Avise (1992) showed that balancing selection acting upon enzyme loci rendered them poor markers for estimation of genetic variation in oysters and provided misleading evidence for gene flow whereas mitochondrial DNA and nuclear RFLPs showed high degrees of genetic structure and population subdivision.

Limitations during practical application may also be an issue with allozyme analysis, as it requires high quality samples that must be kept alive or deeply frozen until use. In addition some bands generated with allozyme analysis may not follow Mendelian inheritance (Roderick 1996), confounding data scoring and interpretation. Finally, enzyme markers provide little genealogical information, and as such they have limited applications for systematics and phylogenetic studies (Buth 1984).

Applications of allozyme analysis in *Bemisia tabaci*

The use of allozymes was the first attempt towards resolving the genetic diversity among whitefly species. The first study was conducted by Prabhaker et al. (1987), who exploited variation in nonspecific esterases (enzymes that hydrolyze esters) to differentiate three whitefly species. Subsequent studies were carried out in the late 80's - early 90's to look for polymorphism within *B. tabaci*. Wool et al. (1989) used general

esterases and α-Glycerophosphate dehydrogenase isozymes to assess polymorphism in populations from Colombia and Israel and showed that esterases were variable between the two *B. tabaci* populations.

The earliest detailed study of *B. tabaci* variants using general esterases was undertaken by Costa and Brown (1991), who identified the distinct "B" band and proposed the A and B biotype nomenclature. In this study, esterase variation was assessed in native polyacrylamide gels (PAGE) using α- and β- naphthyl acetate as substrates. The two B esterase bands, unique to a population derived from imported poinsettia plants (named the B type for the unique pattern in lane B), migrated faster than two major bands associated with a colony of *B. tabaci* from Arizona in the USA reared on cotton or pumpkin (later named the A type for the unique pattern in lane A). Importantly, the B type electromorph was associated with the squash silverleaf (SSL) phenotype in *Curcubita* spp. (Costa and Brown 1991), thus providing a second distinguishing trait used to separate B type from non-B type whiteflies. The SSL induction was later shown also to be associated with a variant of *B. tabaci* from Yemen (YC), which was characterized as having the two distinctive B type bands, plus one additional faster migrating band (Bedford et al. 1994). In further pursuit of the origin of the B biotype, additional possible correlations were sought. The SSL phenotype was found associated with a population of *B. tabaci* in Uganda (Sseruwagi et al. 2005) and another from the Indian Ocean islands (Byrne et al. 1995); both haplotypes are members of sister clades housing the B biotype but all are members of the major N. Africa-Mediterranean-Middle East major clade (Berry et al. 2004; Delatte et al. 2005; Sseruwagi et al. 2005; 2006). Although esterase profiles were not examined for either population, these results have revealed that the SSL phenotype is not strictly linked with the B biotype, but most likely, is a trait associated with it and some of its closest relatives (Brown 2007; Delatte et al. 2005; Sseruwagi et al. 2005).

Allozyme markers were also employed to test for host specificity of populations with different allozyme patterns. Burban et al. (1992) used isozyme electrophoresis in an attempt to identify host-associated differences among sympatric *B. tabaci* populations in the Ivory Coast of West Africa. The study revealed two different esterase profiles that corresponded to the host range of two *B. tabaci* biotypes, the polyphagous okra and monophagous cassava biotypes. In another study in Colombia, however, Wool et al. (1991) showed that esterase profiles were not correlated with specific host-plant species, but differed between geographic locations. Furthermore, in a study from Israel, esterase profiles failed to show correlation with host plant or geographical location (Wool et al. 1993), suggesting that, in general, esterases were useful markers for detecting variants based on banding patterns but not for determining host plant associations.

The presence of the poinsettia strain or B biotype in California, USA, which was suspected since 1990 (Costa and Brown 1990), was confirmed by Perring et al. (1992) with the use of isoelectric focusing (IEF), a technique that is based on the separation of molecules by their differing electric charges moving through a changing pH gradient. In this study, allelic variation in 18 loci encompassing 14 enzymes was assessed in whitefly populations from poinsettia, cotton, and broccoli. Perring et al. (1992) identified 6 loci that were polymorphic, with the cotton strain (A biotype, Western Hemisphere) expressing unique alleles, in contrast to the poinsettia and broccoli isolates of the B biotype (Eastern Hemisphere) that were identical in presence/ absence of alleles and in

induction of SSL symptoms in squash. This degree of divergence, together with evidence for reproductive isolation, was considered sufficient to warrant the description of the B biotype as a separate *Bemisia* species (see section below). However, subsequent examination of additional haplotypes representing different extant geographical origins revealed a far more complex scenario, with evidence for a range of differentiation across the complex (Brown et al 2000). Using the above criteria, together with evidence for reproductive isolation, many species would require recognition even though biological data are unavailable for most taxa investigated using molecular and population biology approaches.

Protein polymorphisms

General esterases were the first *B. tabaci* proteins explored in the search for genetic (biochemical) polymorphisms, providing corroborative support for observed phenotypic variation. The use of general esterases (allozymes), combined with biological and morphological studies facilitated the discovery that the A and B biotypes were genetically and phenotypically different. These results established the A and B biotype nomenclature as noted above. Additional polymorphic patterns were assigned alphabetical designations. Protein polymorphisms were supported by differences in host range and fecundity, and evidence that the poinsettia type induced phytotoxic silvering symptoms in the leaves of pumpkin, whereas, the Arizona cotton colony did not. The esterase polymorphisms were useful for revealing genetic differences between the two biologically distinct populations, thereafter, referred to as the A (endemic isolate) and B (poinsettia isolate) biotypes.

Allozyme frequencies were used by Perring et al. (1993) along with data from behavioral and crossing experiments to characterize the recently introduced (to North America) B biotype (common name, silverleaf whitefly) as a new species. Bellows et al. (1994) described the species as *Bemisia argentifolii* Bellows & Perring, distinct from *B. tabaci* based on mating incompatibilities, morphological, and behavioral differences, along with evidence for differential migration distances of allozymes for three enzymatic systems. Although now it is generally accepted that *B. tabaci* represents a sibling species complex (Brown et al. 1995a; 1995b), with the silverleaf whitefly or biotype B (Costa and Brown 1990; 1991) falling somewhere within a continuum, rather than as a separate species, these latter data together with distinctive esterase profiles (Brown et al. 1995a; Costa et al. 1993) provided the first insights into the unexpected extent of genetic diversity present within *B. tabaci*.

Studies that followed confirmed the extensive variation at the protein level among distinct *B. tabaci* populations, and esterase electromorphs were used to categorize variants into biotypes given alphabetical designations following Costa and Brown (1991). In addition, the use of the variable esterase patterns was expanded to include studies that investigated the biogeographic distribution of *B. tabaci* variants. For example, in a survey over the Americas and the Caribbean basin, Costa et al (1993) found the A biotype in Northern Mexico and Southwestern US, and the B biotype predominating in distribution and present throughout most of the Caribbean basin, the US, and Brazil. They also found, however, that Costa Rican and Nicaraguan populations of *B. tabaci* exhibited distinct esterase patterns, namely C and D type respectively. In another study, Brown et al. (1995a) examined esterase profiles for over 40 populations of *B. tabaci* and

found 12 non-A or non-B electromorphs, representing unique populations from Central America, Africa, Middle East, and India. The same study confirmed the widespread distribution and abundance of the B-type, and concluded that given the extreme amount of genetic diversity, *B. tabaci* was likely a species complex.

Extensive monitoring studies made it possible to identify and characterize more polymorphic variants based on distinctive esterase banding patters, and also to document the extensive genetic differentiation within *B. tabaci*. Brown et al. (2000) analyzed 21 populations in the *Bemisia* genus using isoelectric focusing electrophoresis, and found that 9 of these populations exhibited polymorphism in allozyme loci. They further analyzed these populations using 10 enzymes and were able to cluster these into 3 main groups: Western Hemisphere A- type variants, B-type variants and a single population from Benin, West Africa. The genetic distances (measures of the degree of dissimilarity between groups of individuals) among these variants ranged between 0.03-0.52. Based on Nei's (1976) genetic distance, the values calculated in this study were suggestive of species level boundaries, again suggesting that *B. tabaci* possibly represents a complex with more than one species (Brown et al. 2000). The limited availability of multiple informative markers has hindered progress greatly in advancing our thinking about the status of *B. tabaci* as a sibling species complex or a suite of separate species, or some of both. For example, a single informative nuclear marker (ITS-23S) and two mitochondrial markers (16SRNA, COI) have been explored in an attempt to advance one hypothesis or the other (see Brown 2010 for coverage of these molecular markers). Hopefully the necessary studies will be carried out in the next five or more years.

Esterases as markers

With the discovery of the invasive and highly damaging B biotype, and its associated esterase pattern (Costa and Brown 1991), the use of esterase electrophoresis proved to be a very useful and powerful tool in detecting and tracking the worldwide spread of the B type *via* trade of plant material, particularly poinsettias. Hence, the first surveys of this kind documented the presence of the B biotype in the United States (Costa and Brown 1991; Perring et al. 1992), in the Caribbean Basin and Brazil (Costa et al. 1993) in the Middle East, Central America, Northern Europe (Byrne and Devonshire 1993), in Australia (Gunning et al. 1995), in Africa, India, and many other locations throughout the world (Brown et al. 1995a).

Beyond their use as markers for characterizing and identifying biotype variants, esterases were also used to detect the presence of hybrids in the field or in the lab from crossing experiments, and provide evidence for interbreeding between some biotypes (Byrne et al. 1995; Caballero 2007; Gunning et al. 1997). This use of esterases relies on the fact that hybrids between biotypes will likely be heterozygotes and express allozyme bands from both parental biotypes, assuming of course that the enzymes under consideration follow Mendelian inheritance. Polymorphic esterases were also used to confirm that *B. tabaci* has a haplodiploid sex-determination system, that is, males are haploid, produced by unmated females and females are diploid, produced by mated females (Byrne and Devonshire 1996).

Insecticide resistant *B. tabaci* populations/biotypes have been shown to exhibit characteristically high esterase activity. However, the use of such markers may not be appropriate in monitoring or surveying populations with a recent history of exposure to

certain insecticides and even more in studying genetic structure of populations because local insecticide regimes may affect fitness at these allozyme loci.

Despite the limitations of esterases as markers, the ease and low cost of application of the esterase electrophoresis technique as well as its reliability in biotype determination has prompted continued use of the approach by some laboratories in subsequent years (Horowitz et al. 2003), especially for the purposes of simply detecting the presence of specific biotypes in an area (e.g. B and Q). However, as the interest in researching further the genetic variation within *B. tabaci* increased, several groups have focused their efforts on the use of DNA based markers that allow for finer resolution of genetic polymorphisms compared to allozymes.

DNA-level markers

RFLP Analysis

Restriction fragment length polymorphisms (RFLPs) reveal polymorphisms in DNA fragments of different size due to the absence or presence of restriction enzyme sites (Mitton 1994). Changes in the patterns of restriction fragments can occur from base pair substitutions, insertions or deletions (indels) in the restriction site, allowing assessment of DNA variation (Schlötterer 2004). The technique works by using restriction endonucleases, enzymes that cut the DNA strand in vitro at a specific nucleotide sequence 4-5 nucleotides in length. Restriction fragments can be separated by electrophoresis in agarose or acrylamide gels and detected by Southern blot hybridization (Botstein et al. 1980). Detecting the target sequences with Southern blotting requires binding to probes, which may be available from studies of related species or developed after cloning and sequencing species specific-DNA (Hoy 2003). Alternatively, if the target is small PCR-amplified DNA fragments, RFLP variation can be visualized directly in the agarose gel by staining with ethidium bromide after electrophoresis (Parker et al. 1998).

Uses, assumptions, and limitations

RFLPs were initially used for constructing a genetic linkage map of the human genome (Botstein et al. 1980). They were the first molecular markers that investigated within species variability at the DNA level and their advantages over allozymes, such as the ability to detect silent changes in protein sequences, or selectively neutral, non-coding polymorphisms (Karl and Avise 1992; Schlötterer 2004) were quickly appreciated. Thus, their use was subsequently extended in studies of population differentiation, hybridization, introgression, gene flow, autopolypoloidy and allopolyploidy, and for phylogeographic and phylogenetic inferences using animal mitochondrial and ribosomal DNA (Avise 2004; Lowe et al. 2004).

One of the major advantages of RFLP markers is that they are codominant (Lowe et al. 2004) meaning that homologous alleles can be detected in an electrophoretic gel, allowing the identification of both heterozygous and homozygous individuals. Codominant markers are preferred over dominant markers for population genetics, not only because they allow estimation of allele frequencies but also because they increase

the analytical power of the study with more alleles available for analysis in a given sample size (Lynch and Milligan 1994). Another advantage of RFLPs is that they can generate ordered data, *i.e.* the ancestral states mutate into derived states and the evolutionary direction can be identified (Lowe et al. 2004). Thus, they have a phylogenetic signal, allowing inferences on genealogies to be made. In addition, RFLP analysis allows a large amount of variation to be assessed in polymorphic loci and also provides highly repeatable results.

Since the RFLP technique was first developed before the invention of polymerase chain reaction (PCR), the first applications of RFLP markers relied on Southern blot hybridization, which is a time consuming, laborious, and expensive technique (Parker et al. 1998). Using a PCR step in RFLP analysis however eliminates some of its disadvantages (Karl and Avise 1992). Arbitrary PCR primers can be used to amplify random DNA sequences; from these, allele-specific PCR primers can be designed and used for amplification of the particular locus in other individuals. Subsequently, the product can be digested with restriction enzymes, and the fragments can then be separated on an electrophoretic gel and stained to visualize and identify the RFLPs (Hoy 2003). A modification of the RFLP technique, for example, is cleaved amplified polymorphic sequences (CAPS), which relies on restriction digestion of PCR-amplified DNA fragments (Lowe et al. 2004; Schlötterer 2004). In RFLP-PCR, no labeled probes are required, and it is faster, easier and less expensive than non PCR-amplified RFLP. However, it can only analyze DNA of very small specimens and is still relatively laborious as it requires constructing a genomic DNA library and DNA sequencing.

Applications of RFLP analysis in *Bemisia tabaci*

Despite their wide application for population genetics and phylogenetics in other organisms, RFLPs were not adequately explored as markers in studies of *B. tabaci*. This may be due to the extended use of allozymes and subsequently RAPDs by researchers, especially among collaborating laboratories that used similar protocols, which made their application more approachable. In addition, the need for designing allele-specific PCR primers from already studied sequences limits the feasibility of this approach.

Abdullahi et al. (2004) analyzed RFLPs of the ribosomal DNA internal transcribed spacer regions of *B. tabaci* in order to assess the genetic differentiation between cassava and non-cassava populations in Africa. They found that monophagous cassava-associated populations were genetically distinct from the polyphagous non-cassava populations as they clearly clustered in separate groups. Even though there was a significant subdivision of groups within the cassava cluster, these did not show any geographically associated structure. The results of this study are in agreement with previous studies that used other molecular techniques, e.g. allozymes (Burban et al. 1992), RAPDs and ribosomal DNA sequences (Abdullahi et al. 2003) to discriminate cassava from non-cassava populations. They authors suggest however that PCR-RFLP is a more reliable and cost-effective technique, especially when compared to RAPDs which failed to provide consistent patterns and tended to overestimate genetic differences between populations. Thus, for the purposes of routine identification and monitoring the spread of biotypes, particularly the B biotype, the use of the internal transcribed spacer RFLP markers is probably a much more practical approach than RAPDs or DNA sequencing (Abdullahi et al. 2004).

The utility of RFLP markers for quick identification of *B. tabaci* biotypes was also demonstrated recently by Bosco et al. (2006) in a survey of populations in the Mediterranean basin. The authors used a restriction enzyme to digest the PCR amplified cytochrome oxidase I mitochondrial gene and were able to successfully identify five biotypes from samples in the Mediterranean region. From these, biotypes B and Q were already known to be widespread in the area, whereas M, S, and T biotypes were considered to have a more restricted geographical and host-plant range. This approach provided clearly distinct and reproducible RFLP patterns for each biotype, demonstrating its potential for precise identification of the studied biotypes (Bosco et al. 2006).

Although novel and more informative molecular markers are being introduced that may be more appropriate for population genetics and phylogenetics studies in *B. tabaci*, RFLPs still have promise as a rapid and inexpensive means of monitoring for known biotypes at a regional scale.

DNA-level PCR-based markers

The invention of the polymerase chain reaction (PCR) in 1983 (Mullis et al. 1986; Saiki et al. 1988) revolutionized the field of molecular biology and amongst other advances it allowed the exploration of DNA-based markers for studies of molecular population genetics and systematics. PCR allows the selective amplification of a very small DNA fragment using specific primers that anneal to the regions flanking the locus of interest. The production of multiple copies of the genomic region under study *in vitro* allows analysis of numerous individuals without the need for cloning or isolating large amounts of pure genomic DNA (Schlötterer 2004). Thus, PCR enables studies with genetic markers to be undertaken quickly and at relatively low costs by almost any molecular laboratory. As a result, after the arrival of PCR molecular markers used for population studies largely shifted from protein based (allozymes) to DNA based (*e.g.,* randomly amplified polymorphic DNAs, microsatellites, amplified fragment length polymorphisms, single nucleotide polymorphisms, and DNA sequencing) (Schlötterer 2004).

RAPD analysis

RAPD markers were initially used by Williams et al. (1990) and Welsh and McClelland (1990) for the construction of genetic maps for different species, plant and animal breeding studies, intra-specific level identification (e.g. strains, varieties), epidemiology studies, as well as DNA fingerprinting and population genetics. The acronym RAPD stands for randomly amplified polymorphic DNA and the approach involves the use of short PCR arbitrary primers to amplify random DNA sequences in the genome (Avise 2004). The resulting PCR products represent polymorphic DNA segments that can be separated and visualized in an appropriate electrophoretic gel. Polymorphisms in RAPDs arise from single base substitutions, insertions, or deletions in primer recognition sites that result in a change in the pattern of amplified DNA segments (Williams et al. 1990).

Uses, assumptions, and limitations

Beyond their initial uses, RAPDs found applications in many organisms for studies of genetic diversity, systematics, and to investigate hybridization and introgression (Harris 1999). In a review paper, Hadrys et al. (1992) discuss the applications of RAPDs in molecular ecology and emphasize their usefulness for the determination of taxonomic identity, analysis of interspecific gene flow and hybrid speciation, paternity and kinship studies and analysis of mixed genome samples (e.g. sperm competition). In insect systems in particular, RAPDs have been used to study gene flow and genetic differentiation among insect biotypes or races. For example, Edwards and Hoy (1993) used RAPDs to analyze genetic variation in two Hymenopteran parasitoids, and were able to detect greater polymorphism than when using allozyme analysis. In another study, Black et al. (1992) were able to detect high genetic variation in RAPDs among biotypes, populations and color morphs of four aphid species that exhibited very little allozyme variability. The authors suggested that the method would be particularly successful for rapid species diagnostics, especially in groups where adult or larval characters are poor for species identification. Clearly, the ability to amplify small, potentially polymorphic DNA segments through PCR, made RAPD analysis a much more powerful tool for studying intraspecific variation compared to allozyme analysis.

The RAPD method was initially widely adopted because of technical advantages, such as low cost of application and technical simplicity. In addition, it is a fast method that allows multiple variable loci to be analyzed at once. The main advantage however is that it does not require any prior knowledge of the sequence of the target organism and a universal set of primers can be used for genomic analysis in a wide variety of species (Welsh and McClelland 1990; Williams et al. 1990).

RAPDs however have also been severely criticized and their drawbacks limited their extended application as new markers were explored. Their main limitation is that, like AFLPs, they segregate as dominant markers so they are scored as present or absent in a gel, thus making it impossible to distinguish between homozygote and heterozygote individuals (Black et al. 1992; Harris 1999). Thus the estimation of genetic diversity and the way it is partitioned can only be determined indirectly (Harris 1999). Another severe drawback, not shared by AFLPs, is that RAPDs can be unreliable and very difficult to reproduce or even provide comparable results among different laboratories that use different reaction conditions, different types of *Taq* polymerase or even different thermal cycling machines (Black 1993; Jones et al. 1997; Schierwater and Ender 1993). Hadrys et al. (1992) also discuss the limitation of co-migration of bands with similar size or the formation of fragments from non-specific priming. In addition, like RFLPs and AFLPs, RAPDs provide much lower resolution than microsatellites which target short tandem repeats in the DNA sequence, so they cannot give as much insight into within-population genetic diversity. Finally, their interpretation relies on the assumption that DNA fragments (alleles) with the same position on a gel are homologous, *i.e.* identical by descent, and not the result of size homoplasy, *i.e*., the similarity in size because of convergent mutations (Black 1993; Harris 1999). Beyond the issues regarding the interpretation of data produced by RAPDs, there may be difficulties in publishing results obtained from these markers. For example, due to the problems of reproducibility, dominance, and homology, the journal *Molecular Ecology* rarely accepts for peer review

papers based on RAPDs for population genetic studies, and encourages the use of other markers, that do not suffer from these limitations.

As a result of the issues discussed above, despite the early promise of RAPDs as a quick, easy and reliable technique, the problems of reproducibility limited their extended use in systematics, phylogeographic and phylogenetics studies (Harris 1999), especially in molecular systematics above the intraspecific level (Black 1993). Their popularity in later years was surpassed by the extensive application of microsatellites which are also highly polymorphic but are codominant, allowing the identification of homozygotes and heterozygotes, and therefore more genetically informative (Avise 2004).

Applications of RAPD analysis in *Bemisia tabaci*

RAPD markers were used extensively for the identification and analysis of genetic diversity in biotypes of *B. tabaci.* Their application was particularly focused in studies of biotype B whose spread over the continents in the past two decades required extensive monitoring for the application of more efficient control measures. The first study in this direction was undertaken by Gawel and Bartlett (1993), who were able to distinguish between the A and B biotype based on genetic differences revealed by 20 RAPD markers. The authors concluded that, based on their RAPD data, the degree of similarity between the two biotypes is similar to that between other whitefly species, even genera. However, they suggested that RAPDs are probably of limited value for taxonomic determination of biotypes without the use of additional genetic, morphological, and physiological data. Since RAPDs were the first DNA-based markers used in *B. tabaci*, this study was also the first to demonstrate the advantages of DNA-based markers over allozymes – the ability to analyze dead specimens preserved for years in alcohol and the flexibility of using a small amount of material of any whitefly stage. Perring et al. (1993) used RAPDs to detect variation in A and B biotypes, and showed that populations of either biotype shared 80-100% genetic similarity in bands whereas between them similarity was only 10%. The authors used these data along with differences in allozyme frequencies, and behavioral and mating experiments to propose that the B biotype should be re-classified as a new species, which was later named *Bemisia argentifolii* (Bellows et al. 1994). This extreme genetic differentiation among worldwide populations as shown in RAPD and other molecular markers was the basis for suggesting that *B. tabaci* is a group or complex of species, (instead of a sibling species group) (Brown et al. 1995a; Perring 2001) of which biotype B is a member rather than a different species.

Most studies that followed used RAPDs similarly to allozymes, that is, as a means of detecting the presence of biotypes in an area, or the distribution of biotypes within and between populations in different geographic regions, host plants, and in indoor versus outdoor cultivations. De Barro and Driver (1997), for example, used RAPDs to compare B biotype populations from Australia, the Cook Islands, Israel, the Netherlands, New Caledonia, and the US, and showed that they had similar banding patterns but differed from other native biotypes in Australia. They also confirmed the flexibility of the technique to utilize alcohol preserved material and analyze all stages of the whitefly. In a subsequent survey of populations from Australia and 18 Pacific island countries, De Barro et al. (1998) showed that biotypes of *B. tabaci* were present and widespread in each country surveyed. Analysis of RAPD profiles of these samples revealed the presence of

three distinct biotypes in these regions. The first was the widespread B biotype and was found in islands with strong links with France or the USA, suggesting introduction through the plant trade. The authors also detected a type with unique RAPD bands which was widely distributed in 13 countries and was termed the Nauru biotype, and a third biotype, termed Australasian, which is most probably native to the region as it was reported previously from Australia (De Barro and Driver 1997).

In a study in the Iberian Peninsula, Guirao et al. (1997) used both esterases and RAPDs, along with physiological examinations of silverleaf induction to characterize Spanish populations of *B. tabaci*. This study demonstrated that two genetically distinct populations existed in Spain and the Canary Islands, the already known B biotype and a new variant with unique RAPD and esterase patterns, referred to by others as the Q biotype (Rosell et al. 1997). This was the first use of the Q biotype designation in the Mediterranean region, corresponding to endemic haplotypes from Spain and Sudan populations that were previously detected (referred to in Kirk et al. 2000). A cluster analysis based on RAPD band sharing sorted the sampled populations into four groups. The first included only the A biotype from Arizona, the second included the Spanish Q type, the third included populations of the B biotype, similar by 90%, from eight different countries, and the fourth included populations from Pakistan, India, and Turkey. Genetic similarity estimates (% of RAPD bands shared) used in this study showed that the new genetic type found in Spain was closer to B biotype than to any other non-B type, but only shared 55% of bands observed. An interesting conclusion in this study was that the new biotype had not been displaced as a result of competitive interaction with the B biotype, as was the case in the US with the A biotype (Guirao et al. 1997). Some have speculated that the Q biotype and its closest relatives are endemic in the North African-Mediterranean-Middle East, and that the Q biotype in Spain competitively displaced the exotic B biotype there with the intervention of neonicotinoid insecticides. This was the inverse to the permanent displacement of the endemic A biotype in the Southwestern U.S.A. by the introduced B biotype, which was susceptible to locally applied insecticides including pyrethroids, to which the B biotype harbored inherent resistance (Coats et al. 1994; Costa et al. 1993), although the degree to which insecticides played a role in this displacement is unknown. The widespread displacement of indigenous biotypes by the invasive B has been documented worldwide, for example a recent study showed competitive displacement of indigenous biotypes in Australia and China by biotype B, both in the field and in laboratory settings (Liu et al. 2007).

In a more detailed analysis of *B. tabaci* populations from the Iberian Peninsula, Moya et al. (2001) were able to identify six genetically distinct clusters consisting of either biotype Q, or B, or a mixture of the two. In contrast to previous studies that used RAPDs to estimate genetic relationships among haplotypes, Moya et al. (2001) used analysis of molecular variance (AMOVA) of RAPD data to investigate the genetic differentiation among populations. They found that genetic variance is partitioned more between biotypes than among populations within the same biotype suggesting that gene flow is restricted between the two biotypes even in mixed field populations. This means that the two biotypes are genetically isolated under the existing ecological conditions in the area, which confirms previous studies of the B and Q biotypes. In addition, populations of the Q biotype were shown to be more genetically polymorphic than populations of B which suggested that the Q is more ancestral, existing long before the

introduction of B in the region (Moya et al. 2001). The hypothesis that biotype Q and its variants are possibly the indigenous Mediterranean genetic type is also supported by studies showing its predominance in fields in Italy, in contrast to B biotype which is widespread in greenhouses, possibly introduced through the plant trade (Demichelis et al. 2000).

Similar to the Mediterranean studies, Lima et al. (2000) used RAPD markers to survey *B. tabaci* populations and identify the distribution of biotypes in Brazil. They showed that the B biotype was well established and distributed across 20 Brazilian states in the northeast, east and midwest of the country. In contrast to the Spanish surveys, Lima et al. found that B was the dominant genetic type in Brazil, but also coexisted in some areas with BR, a native Brazilian, non-B biotype with distinctive RAPD patterns. In a subsequent study of Brazilian populations, Lima et al. (2002) used analysis of molecular variance of RAPD data and demonstrated that there is a significant genetic differentiation between B and the Brazilian biotype (BR). The B biotype was found to predominate over the A or the BR biotype throughout Brazil, even in regions or hosts where BR was dominant before. It also exhibited considerable genetic variability despite its relatively recent introduction into Brazil in the early 1990s, perhaps due to multiple founder events or because of the differential selection regimes caused by variable insecticide applications across different crops (Lima et al. 2002).

In a study of *B. tabaci* in association with the cassava mosaic disease (CMVD) in East Africa, Maruthi et al. (2002) used 10 RAPD markers along with mating and life history studies to test whether a new biotype was responsible for the spreading of the severe pandemic across East Africa. In a comparison of populations between pandemic and non-pandemic zones, the authors found no difference in RAPD patterns or life history traits, and no mating incompatibilities, suggesting that the CMVD pandemic in East Africa may not be associated with a new genetically distinct *B. tabaci* biotype. In this study, the RAPD variability data that grouped pandemic and non-pandemic colonies in a single cluster were confirmed by mating and life history tests. However, the results of this study contradicted those of Legg et al. (2002) who collected samples across several years and in transects across the spreading epidemic zone. This revealed the presence of both a western and eastern type African haplotype, suggesting that a mixture of two distinct *B. tabaci* haplotypes was associated with the epidemic of CMVD in Uganda. The subsequent reversion to the east African haplotype resulted in the hypothesis that the invasive population may represent a unidirectional hybrid between the eastern and western haplotypes, based on mitochondrial DNA phylogenetic data (Sseruwagi et al. 2004). The contradictory findings may be due to the limited power of RAPDs to discriminate phylogenetic lineages compared to mtDNA. Even so, the lack of association of a distinct genotype with the CMVD epidemic was subsequently supported by assays that showed no difference in transmission efficiencies among sampled populations (Maruthi et al. 2002). Conversely, assays performed with the field populations collected early in the outbreak and maintained in colonies used in the transmission experiments may not necessarily have represented collections from the field assessed by mtCOI analysis as the dynamic epidemic continued to unfold. Indeed, the mtCOI was not assessed for the lab colonies (Maruthi et al. 2001), nor were transmission studies performed with the field haplotypes identified later in the epidemic zone, and so it has not been possible to explain the differing results. Further studies employing typed

colonies and possibly new markers, together with transmission experiments will be necessary to elucidate the genetic status of cassava-associated populations in relation to the spreading pandemic.

Similarly to studies initiated by Burban et al. (1992), who looked for associations between distinct biotypes (based on allozyme patterns) and host range, Abdullahi et al. (2003) analyzed RAPD data using AMOVA and showed that samples collected from cassava and non-cassava populations represent two genetically isolated groups, supporting the hypothesis of a distinct cassava specific lineage in Africa. This finding was also in agreement with ribosomal DNA data although RAPDs exhibited a greater ability to differentiate geographically distinct populations. More importantly, these results were supported by physiological experiments showing that cassava populations were monophagous and restricted to cassava whereas populations from other hosts were polyphagous but could not colonize cassava. These results suggest that host specialization in this case may drive isolation, leading to genetically differentiated populations. More extensive studies of cassava populations based on mitochondrial CO1 sequence analyses supported the hypothesis of host-specificity but also revealed a high variation within cassava populations in Sub-Saharan Africa, with a strong phylogeographic basis (Berry et al. 2004). However, it is clear that more studies are needed, especially focusing on comparisons with non-cassava populations in the continent in order to provide an overall picture of host association within *B. tabaci*. Furthermore, to support the hypothesis of host-specificity, it will be critical to ensure that individuals collected from the host actually colonize and breed on that host. Indeed, one recent study in Uganda addressed this issue by sampling pupae from leaves of cassava and non-cassava hosts (Sseruwagi et al. 2006). In this study, phylogenetic analysis of mtCO1 indicated that, while cassava is only colonized by the cassava *B. tabaci* (Ug1 genotype), this type is also capable of colonizing other non-cassava plant species suggesting that it is not monophagous but can possibly survive on other hosts when cassava is not available. This work highlights the importance of the sampling procedures used for such studies, but also illustrates that more molecular markers and new analyses could be explored to identify host-associated genotypes in the continent.

In another study from south India, Rekha et al. (2005) analyzed RAPD diversity in *B. tabaci* populations and identified two indigenous genotypic clusters, a southern and a northern cluster, as well as a third group represented by the B biotype. These results point to a geographic, rather than host-associated, structuring of populations, perhaps related to distinct cropping regimes and climatic conditions prevailing in the two areas. The unique RAPD banding patterns of the B biotype made it possible to identify its distribution in the region, which has been rapidly expanding since it was first reported in India in 1999 (Banks et al. 2001; Rekha et al. 2005). Additional phylogenetic analysis using mitochondrial CO1 sequences revealed the presence of three distinct genotypes that are possibly indigenous to India but were also found in other Asian countries from which reference samples were used. These indigenous types were found in higher frequencies in areas where the B type is considered to have invaded recently, but were probably excluded in regions where B has been present for at least two years and where it represents the dominant type now (Rekha et al. 2005).

Along the same lines, Delatte et al. (2005), identified a biotype indigenous to the South-West Indian Ocean islands based on distinct RAPD bands and sequencing of the

mitochondrial CO1 gene. This new biotype, Ms, occurs along with B on the island and, although considered to have lower fecundity and restricted host range, is also capable of inducing silverleaf symptoms in *Cucurbita sp.* and acquiring and transmitting TYLCV, (Delatte et al. 2005). Subsequent studies with microsatellites (Delatte et al. 2006) showed that although Ms is the common biotype in weeds across the island, the invasive B biotype has spread in vegetable plantations, possibly interbreeding at some parts with the Ms biotype.

Besides their use in population genetics, RAPDs found wide application as markers for quick detection of a biotype in a region. For example, in Israel, the Q biotype was first identified by RAPD and esterase profiles and was found in sympatry with the B biotype in three crops, although the Q biotype predominated in numbers (Horowitz et al. 2003). The authors suggested that the dynamic distribution of biotypes could have been a result of the differential inherent levels of resistance exhibited by the two biotypes. In a survey of *B. tabaci* biotypes across Italy, Simon et al. (2003) detected the biotypes B and Q (which was almost identical to the Spanish Q) using RAPDs, squash silverleaf assays, and esterases. More importantly they were able to identify a genetically distinct population, namely biotype T, which unlike B and Q is geographically isolated, monophagous, restricted to *Euphorbia characias* at high altitudes and possibly endemic to the island of Sicily. In a survey of *B. tabaci* from crop fields and weeds in Florida, RAPD patterns showed only the presence of biotype B, suggesting that the other native biotypes have been displaced from the agricultural ecosystems in the area (McKenzie et al. 2004). Khasdan et al. (2005) further demonstrated the usefulness of RAPDs as diagnostic markers by developing a new protocol for distinguishing between B and Q biotypes, based on two techniques: SCAR (sequence characterized amplified regions), which requires sequencing of RAPD fragments and designing sequence specific PCR primers, and CAPS (cleaved amplified polymorphic sequences), which uses restriction endonucleases to further cleave the RAPD amplified DNA fragments and reveal polymorphisms.

Despite the limitations of RAPD markers, they were used extensively for almost 15 years in population genetics studies and molecular identification in *B. tabaci*. The use of RAPDs provided insights into the genetic structure of biotypes across the world. More importantly, RAPDs revealed aspects of the dynamic distribution and rapid spreading of the invasive B biotype in several regions in relation to indigenous biotypes, which were generally less competitive and less damaging than the B biotype. The use of RAPDs in the future will no doubt decline, as they are replaced by new, more informative markers, such as microsatellites, DNA sequencing, single nucleotide polymorphisms (SNPs) and expressed sequence tags (ESTs) that allow for high throughput genotyping in large scale population genetics studies (Behura 2006). Due to the ease and low cost of application however, it is possible that RAPDs will continue to be used in studies of detection and monitoring for the presence of a biotype in a region.

AFLP analysis

The analysis of amplified fragment length polymorphisms (AFLP) was developed as a technique for DNA fingerprinting in 1995 (Vos et al. 1995). The AFLP protocol

involves three steps. First, genomic DNA is digested with restriction enzymes and oligonucleotide double stranded adapters are ligated to the ends of the restriction fragments. Subsequently, the fragments are selectively amplified with PCR primers that are complementary to the adaptor and the restriction site fragments. Finally, the amplified fragments are separated through gel electrophoresis and visualized or scored in an automated sequencing machine. Polymorphisms in AFLPs are scored as differences in the lengths of the amplified sequences which may be caused by base substitutions within or near the restriction sites or by insertions or deletions of sequences (Avise 2004). The AFLP technique has elements of both RFLP and RAPD analysis, and as such it is regarded as a clever combination of steps from the two older marker systems (Bensch and Akesson 2005).

Uses, assumptions, and limitations

AFLP markers have been used for molecular characterization of closely related species or strains, for population and conservation genetics, fingerprinting, and parentage analysis (Mueller and Wolfenbarger 1999). They also found application in the construction of genetic linkage maps in insects and other organisms with unexplored genomes (e.g. Parsons and Shaw 2002). The AFLP technique combines the ease of RAPDs and the reliability of RFLPs (Behura 2006). One major advantage is that they do not require previous knowledge of primer sequences in the target species (Schlötterer 2004), thus there is no need to obtain or design species-specific primers for the analysis. In addition, with AFLPs a large number of marker loci drawn from the entire genome (typically 50-100 fragments per reaction), are selectively amplified and can be analyzed at once (Vos et al. 1995). Although some of these characteristics are shared by other markers, AFLPs are more time and cost-efficient, and provide higher resolution of genetic variability when compared to allozymes, RFLPs and RAPDs (Mueller and Wolfenbarger 1999). The higher reliability and reproducibility of AFLPs compared to RAPDs can be attributed to the fact that they are amplified by primers of longer sequence with precise annealing to their target sequence, rather than by short, arbitrary primers (Lowe et al. 2004).

The application of the AFLP protocol may be restricted by the need for high technical laboratory skills and thus may not be approachable to less experienced researchers. Another issue that emerges from studies with AFLPs is the need for large amounts of high quality DNA (Lowe et al. 2004), which may restrict the number and variety of samples that can be analyzed in a study. AFLPs also suffer from data interpretation problems, such as size homoplasy, an issue also discussed in RAPDs, or uncertain locus/allele designations (Harris 1999).

The major drawback of AFLPs for population studies is that, like RAPDs, AFLPs are dominant markers and so PCR amplification reveals only the presence or absence of bands, making the technique less suitable for studies that require analysis of allelic diversity or heterozygosity (Mueller and Wolfenbarger 1999). There have been cases, however, where codominant AFLPs were detected, possibly due to insertions/deletions of repeats in microsatellites often present in AFLP fragments, producing multiple variable alleles for the same locus (Wong et al. 2001). If this is the case, with the vast number of fragments generated it would be impossible to determine whether alleles in different individuals belong to the same AFLP locus. Because dominance is assumed in AFLP

analysis, the production of codominant AFLPs may compromise their use in population genetics studies as the estimation of parameters such as allele frequencies and heterozygosity will be biased (Wong et al. 2001). For example, Yan et al. (1999) showed that the average heterozygosity and population differentiation (F_{ST}) in populations of the yellow fever mosquito in Trinidad were underestimated when AFLP markers were used, compared to data obtained from RFLP markers. Nevertheless, for large data sets, with the vast number of markers typically scored in ALFP analysis, the frequency of codominant bands may be minimal, thus reducing the resulting bias (Parsons and Shaw 2002). In addition, it is possible to detect codominance in AFLP data following several procedures that have been developed for this purpose (Jansen et al. 2001).

Despite their high promise as molecular markers for population genetics, phylogenetics and quantitative trait loci (QTL) mapping (Mueller and Wolfenbarger 1999), AFLPs are largely underrepresented in animal studies in these fields. On the contrary, the method has been widely adopted and applied in studies of plants, bacteria, and fungi (Bensch and Akesson 2005). In the limited number of animal studies, AFLPs were shown to be powerful in delineating genetic relationships among cryptic sympatric species (Parsons and Shaw 2002). Their use in phylogenetic and phylogeographic studies on the other hand, may be limited by the fact that they are not ordered markers, i.e., they do not carry genealogical information, and thus their use for such purposes is controversial (Robinson and Harris 1999). Bensch & Akesson (2005) strongly advocated for the advantages of ALFPs and suggested that the drawback of dominance can be overcome by employing more AFLP loci; they also strongly urged their further exploitation in studies of genetic diversity and population structure in animals.

Applications of AFLP Analysis in *Bemisia tabaci*

Similar to other animal systems, AFLPs are also poorly represented in the *B. tabaci* literature. Although from a theoretical perspective it seems that these markers could provide further insights into the genetic diversity within the *B. tabaci* sibling species group, there are only a few published studies in the literature on AFLP analysis of *B. tabaci* populations. In one such study, Cervera et al. (2000) analyzed the genetic diversity of nine biotypes, two field populations of *B. tabaci* (from cowpea and cassava in Africa), and two field populations of two other *Bemisia* species, *B. medinae* and *B. afer*. Based on the AFLP profiles, the nine esterase types (Brown et al. 1995a; Costa et al. 1993) (A, B, H, K, M, Q, S, Pakistan I, and Pakistan II) were grouped together and separately from the other two *Bemisia* species. This suggests that all genetic types (erroneously referred to as biotypes) studied are more closely related to each other than to any of the two other species, which is consistent with morphological differentiation and in agreement with the existing literature. A further grouping, within *B. tabaci*, separated the nine esterase types into four clusters: a Near East and Indian subcontinent cluster (H, K, M biotypes), a cluster with biotypes B, Q, and a Nigerian cowpea population, a third group consisting only of New World A biotype, and finally a cluster with the S biotype and a Nigerian cassava population. The results from the AFLP analysis were in agreement with previous studies based on RAPDs and provided similar resolution of genetic differentiation; AFLPs, however, offered better quality and reproducibility (Cervera et al. 2000). These findings have also been further supported by mitochondrial sequence analysis (Berry et al. 2004; Qiu et al. 2007).

In a recent study of *B. tabaci* in China, Zhang et al. (2005) used both AFLP data and mitochondrial cytochrome oxidase I (mtCO1) sequences to assess genetic diversity in geographically discrete populations. The AFLP analysis indicated that at least four genetically distinct groups of *B. tabaci* are present in China, and their genetic isolation is most likely being maintained by geographic barriers. The use of mtDNA sequences allowed the assignment of the two groups to known mtCO1 clades - the first represented by the widespread B biotype and the second by Q biotype populations from Spain and Israel. The third and fourth groups were assigned to a non-B/Q clade as they were significantly differentiated from either biotype. The populations belonging to B biotype were the most widespread throughout the country, while the Q type was reported for the first time from China and was found to be much more restricted in range (Zhang et al. 2005).

With the growing use of new molecular markers in the literature for population level genetic studies, we expect that research in *B. tabaci* will also shift towards new approaches, especially as more information on the whitefly genome becomes available. Traditional markers that are reliable and cost-effective, however, like AFLPs, still will have a place in small scale studies and may reveal new and useful information on population differentiation from previously unexplored parts of the genome of *B. tabaci.*

Microsatellite analysis

Microsatellites are tandem repeats of usually 1-6 nucleotide bases that are randomly distributed throughout the eukaryotic nuclear genome, and are also known as simple sequence repeats (SSR), variable number tandem repeats (VNTR), and short tandem repeats (STR) (Selkoe and Toonen 2006). The high degree of polymorphism of microsatellites in natural populations was first appreciated in the late 1980's (Tautz 1989) and they have since become particularly attractive as markers for population genetics studies (Ellegren 2004). Their high popularity relies partly on their Mendelian inheritance and codominance, meaning that, in contrast to RAPD's and AFLP's, heterozygotes can be distinguished from homozygotes in the electrophoretic gel (Lowe et al. 2004). In addition, their high degree of polymorphism allows for fine resolution in within population level studies in contrast to other markers discussed here.

The high polymorphism of microsatellites reflects the great variability in repeat length, commonly between 5 and 40 repeats, a result of high rates of mutation. The most common types of microsatellites used in molecular genetics studies are dinucleotide, trinucleotide, and tetranucleotide repeats (Selkoe and Toonen 2006). Microsatellites are thought to gain and lose repeat units typically through polymerase slippage, an error during DNA replication which results in an increase or decrease in the number of repeats (Levinson and Gutman 1987; Schlötterer and Tautz 1992), leading to the formation of different alleles. The mutation rate of microsatellites has been estimated at between 10^{-6} and 10^{-2} per generation (Schlötterer 2000), although it may vary among different loci, especially between di- and tetra-nucleotides (Weber and Wong 1993). The flanking regions, the DNA sequences adjacent to either side of the microsatellite locus, are generally conserved across individuals of the same species or between closely related species and thus can serve as primers for polymerase chain reaction (PCR) amplification

of the microsatellite locus in con-generic species (Selkoe and Toonen 2006). The assumption is that any mutations occurring will likely be at the microsatellite locus as insertions or deletions of a repeat and not at the more conserved flanking region (Schlötterer 2001; Schlötterer and Pemberton 1994; Weber and Wong 1993).

Microsatellite data are obtained following amplification in a PCR with species-specific primers, and subsequent genotyping and allele scoring in an automated DNA sequencer. Microsatellite analysis is based on a locus of known size and subsequent alleles are characterized by the number of repeats that differ from the known locus. Thus, depending on the questions being asked, the type of data generated can be raw alleles for parentage analysis or allele frequencies per locus and estimates of heterozygosity for population studies. Statistical analysis of microsatellite data can be performed using maximum likelihood, Bayesian, or coalescent methods in a variety of software freely available on the internet (Excoffier and Heckel 2006; Luikart and England 1999).

Uses, assumptions, and limitations

Microsatellites have been used extensively in the past decade to test for inbreeding, parentage and relatedness, to examine genetic diversity and demographic patterns, as well as to investigate population differentiation in many insect systems (e.g. Endersby et al. 2005; Kim and Sappington 2006). Contrary to allozymes, several thousand potentially polymorphic markers are available with microsatellites (Schlötterer 2000) and their fast rate of evolution compared to mitochondrial genes or other nuclear loci makes them more suitable for resolving the genetic diversity among closely related, recently founded populations (Davies et al. 1999). The relatively large number of microsatellites usually available allows multi-locus analyses, providing much more information and statistical power at a fine scale.

However, the use of microsatellite DNA in population genetics requires knowledge and understanding of the mutation processes they undergo and their rate of evolution (Schlötterer 2000). Additionally, there are other issues and limitations that need to be considered before using microsatellite markers, some of which are discussed here:

a) *Size homoplasy*. Microsatellite analysis methods assume that alleles identical in size have identical genealogical history, or common descent, although this may not hold true because of size convergence as discussed earlier for RAPDs and AFLPs. When mutation rates are high, homoplasy can inflate estimates of gene flow and give misleading inferences of population structure (Viard et al. 1998). However, Estoup et al. (2002) suggest that high polymorphism often compensates for the homoplasious evolution of microsatellites, although it may still be an issue in studies of large populations and at loci with high mutation rate.

b) *Complex pattern of mutation*: Mutation rates may be complex and may vary considerably among loci, complicating the use of models (e.g. infinite allele model, stepwise mutational model) to infer statistics of population differentiation based on estimates of allele frequencies such as F_{ST} and R_{ST} (Selkoe and Toonen 2006).

c) *Neutrality and selection.* Most methods used to analyze microsatellite data assume a neutral rate of evolution, *i.e.,* that the locus used as a marker is not under selection and evolves at a more or less constant rate. Although microsatellites are generally considered

selectively neutral markers, this assumption is sometimes questionable (Zhang and Hewitt 2003) as microsatellites have been detected in loci under selection or linked to such loci (Schlötterer and Pemberton 1994).
d) *Scoring errors*. Genotyping errors or mistyping during allele scoring in microsatellite analysis may obstruct data interpretation and bias the conclusions of the study. DeWoody et al. (2006) review three common scoring errors that may occur during microsatellite analysis: stuttering, large allele dropout, and presence of null alleles, and discuss their implications and potential ways to circumvent these problems. Null alleles represent the most problematic scoring error and also the best studied (Callen et al. 1993; Dakin and Avise 2004; Pemberton et al. 1995). These are alleles that fail to amplify in a PCR reaction due to a mutation in the primer–binding region of a microsatellite locus, leading to either no amplification at all in homozygous individuals or amplification of only one allele in heterozygotes (Dakin and Avise 2004). Presence of null alleles could result in the mistyping of heterozygotes as homozygotes in the analysis, and consequently in heterozygote deficiency, leading to underestimation of genetic diversity and population structure. Several approaches are available to test for null alleles (Dakin and Avise 2004). For example deviations from Hardy-Weinberg equilibrium in some loci but not in others might suggest the presence of null alleles. Additionally, testing for suspected null alleles could be done by directly sequencing PCR products. Also, new analytical methods are better at dealing with null alleles (e.g., Geneland, Guillot et al. 2008).
e) *Limited use for genealogical studies*. Microsatellites, like most of the other markers discussed here, produce mostly unordered data which means that they do not contain much ancestral information because of the multiple allele states often present. Thus, unlike nuclear or mitochondrial DNA sequence data, microsatellite data will not be very useful for phylogenetic inferences (Zhang and Hewitt 2003), although it has been demonstrated that phylogenetic inference is possible in some cases (Schlötterer 2001 for a review; Wilson and Balding 1998).
f) *Mendelian inheritance*. Although microsatellites have generally been considered to be markers with Mendelian inheritance (Jarne and Lagoda 1996), some researchers have questioned this assumption (e.g. Dobrowolski et al. 2002; Smith et al. 2000) and suggest that evaluation of microsatellite loci for this property in parent-offspring pairs should be done before using microsatellites in analyses (Smith et al. 2000).
g) *Standardization of microsatellites among laboratories*. Because of the variability in allelic reads, arising from using different thermocyclers, DNA sequencers, and genotyping programs, as well the subjective scoring of alleles, microsatellite data cannot be compared accurately or shared among different laboratories.

Applications of microsatellite analysis in *Bemisia tabaci*

Microsatellites have been used as molecular markers for almost 20 years (Schlötterer 2004). Still, their application to studies of *B. tabaci* populations was only recently explored. Microsatellite loci were first isolated in *B. tabaci* in 2003 by two independent groups (De Barro et al. 2003; Tsagkarakou and Roditakis 2003), and more recently by others (Dalmon et al. 2008; Delatte et al. 2006; Gauthier et al. 2008; Schwartz unpublished; Tsagkarakou et al. 2007). Studies undertaken so far using microsatellites have examined the structure of *B. tabaci* populations at a regional scale.

With the use of 15 microsatellite loci, De Barro (2005) showed that *B. tabaci* in the Asia-Pacific region consists of six genetically distinct populations—more than previously revealed by mitochondrial (De Barro et al. 2005) and ribosomal (De Barro et al. 2000) DNA markers—with little or no gene flow between them. This study also demonstrated the presence of strong geographic structure possibly maintained by physical barriers and/or competitive interactions, leading to allopatric divergence in most populations. However, De Barro (2005) did not find evidence for strong host-plant associated structure, suggesting that the polyphagous nature of *B. tabaci* is common among most populations, preventing host-based structuring of genotypes.

Delatte et al. (2006) used a set of eight microsatellite loci to examine the distribution of the invasive biotype B in the island of La Réunion and its interactions with the indigenous MS biotype. They demonstrated that the invasive biotype has spread throughout the island, overlapping in range with the indigenous biotype. Their segregation however into different hosts, with the invasive B occupying mostly vegetable crops and the indigenous MS colonizing weeds, determines their relative distribution in distinct parts of the island. The examination of shared alleles among populations indicated the presence of a cluster that could represent a hybrid population, a result of asymmetrical and locus specific introgression of MS alleles into B genotypes possibly from a unidirectional cross between B females and MS males. In addition, despite the estimated recent invasion of the B biotype, the putative hybrid population exhibited considerable genetic variation and levels of geographic structure similar to the local biotype, suggesting that multiple introductions may have contributed to its invasion into the island.

More recently, Tsagkarakou et al. (2007) used six microsatellite loci and mitochondrial cytochrome oxidase I DNA sequences (mtCOI) to study the genetic polymorphism of *B. tabaci* populations from Greece. Analysis of the mtCOI sequences failed to differentiate the populations into discrete clusters; the use of microsatellite loci however provided better resolution of the genetic structure, even at this small geographic scale, and differentiated the samples into at least two distinct genetic populations. It was also possible to use two polymorphic microsatellite loci to discriminate Q from B biotype on the basis of shared and unique alleles detected in reference populations and accordingly to characterize the Greek samples as belonging only to the Q biotype.

A more recent study examined the microsatellite variation of biotypes B and Q in the Mediterranean region by examining a number of populations from different areas (Simon et al. 2007). This study looked specifically for signatures of recent bottleneck events - sudden reductions in population size - and they were able to detect such genetic signatures in Iberian Q, Canarian Q and Egyptian B populations. In addition they tested for genetic differentiation between pairs of populations and inferred close relationships among Q biotype populations from Israel, the Iberian Peninsula and Italy, and likewise between B biotype populations from Egypt and Israel. However, the origins of the Mediterranean populations were not resolved.

Finally, another study looked at the genetic structure of 22 *B. tabaci* populations sampled from vegetable and ornamental glasshouses in southern France (Dalmon et al. 2008). This study tested the hypothesis that the enclosed environment and patchy distribution of glasshouses in combination with the limited outdoor survival of whiteflies in the winter would promote genetic differentiation among populations. Using seven

microsatellite loci and mtCO1 sequences, two genetic groups were detected, corresponding to B and Q-like biotypes. While the Q group was predominant across the sampled areas, the B type group showed limited distribution. The authors suggested that extreme temperature tolerance and insecticide resistance might have favored Q over B in glasshouses in southern France. Within the Q group, the microsatellite analysis indicated limited genetic differentiation among glasshouse populations, suggesting a recent colonization event and dispersal over large distances likely facilitated by movement of plant material between glasshouses. As the authors discuss, high genetic similarity among localities resulting from human trade is also evident at a worldwide scale, with individuals on different continents exhibiting 100% nucleotide identity of CO1 sequences.

Although not fully explored in *B. tabaci*, the high degree of polymorphism displayed by microsatellites makes them promising for studies measuring gene flow, identifying genetic structure and inferring past demographic histories of populations. Future studies employing microsatellites likely will focus on associating populations with disease epidemics, investigating host-associated structure, and untangling complex population histories.

Additional considerations and future directions

To date, research using multilocus molecular markers in the *B. tabaci* complex system has targeted identifying the current distribution of biotypes, patterns of gene flow and population structure, host-associations, and surveying for the presence or evidence of introductions of specific biotypes in a region. Information about genetic differentiation among populations/biotypes can be obtained by calculating "summary statistics" such as the fixation index $F_{ST.}$ Although such analyses are informative for presenting a current perspective, they do not distinguish between the histories that could have produced that level of genetic differentiation, *i.e.,* between contemporary and historical gene flow (Bossart and Prowell 1998). Genetic data obtained from molecular markers, particularly DNA sequences (see Brown Chapter 2) and microsatellites can reveal genealogical information and answer questions pertaining to the evolutionary history of biotypes. One can ask for example: how long ago did biotypes in the *B. tabaci* complex diverge from one another? How large was the population that diverged to give rise to a new biotype? Was the population constant in size throughout its history or did it fluctuate as a result of bottlenecks, founder events, or expansions? Have some biotypes continued to exchange migrant whiteflies at some parts of their history or did they become completely geographically isolated following the classical model of allopatric diversification? All these questions are important in understanding how biotypes and species diverge and evolve, especially while maintaining cryptic and often overlapping morphologies. In the case of *B. tabaci*, these questions can also help us understand the long-term histories of invasive "pest" biotypes and how past demographic processes (*e.g.,* selection, gene flow, fluctuations in population size) have shaped their current high genetic diversity, ability to adapt to extreme conditions and expand their range.

Employing methods and programs that are based on the theory of the coalescent can help approach questions in population biology (Marjoram and Tavaré 2006). The

coalescent refers to the process by which, looking backward in time, the genealogies of two alleles merge at a common ancestor. Thus, according to this approach, the descent of any given sample of alleles can be traced back in history, with a coalescent event occurring at every node of common ancestry. By developing a model for the time intervals between each coalescent event for a given sample of genes, the ancestral history can be inferred. Any historical changes in population size affects the probability distribution of the coalescent times, with stable, exponentially expanding or declining population sizes leaving distinct signatures on the pattern of the coalescence (Hartle and Clark 2007). Demographic histories of populations can be inferred using this approach from samples of gene sequences or alleles obtained from extant populations (Emerson et al. 2001) and interesting scenarios about the evolution of the species/organism under consideration can be hypothesized. The *B. tabaci* complex of biotic and genetic variants –the most well studied of which have been named as biotypes– which are quite likely at the point of divergence into distinct species, represents an excellent model for answering such theoretical questions that can nevertheless serve as a basis for addressing more applied issues, such as understanding the biology and spread of invasive organisms.

The use of molecular markers in ecological genetics has other considerations often neglected, the most important being the natural history of a chosen molecular marker. Different markers have different rates of mutation, with nuclear DNA typically evolving slower than mitochondrial or ribosomal DNA. In addition, the mtDNA does not undergo recombination and is only inherited through the mother, therefore only the maternal lineage is traced in mtDNA analysis. As a consequence, the effective population size of mtDNA is smaller than the nuclear because fewer copies of mtDNA are passed to each generation. These issues may have implications for phylogenetic studies as the inferred genealogical histories will only reflect the history of the mtDNA locus without that necessarily corresponding to the species histories (Ballard and Whitlock 2004). To provide an overall and non-biased perspective of population and species histories, a set of multiple non-linked markers, both from the nuclear and mitochondrial genome must be investigated. Complementing markers from these two different genomes having different natural histories can also help detect past hybridization events between biotypes by testing for discrepancies between the generated nuclear and mitochondrial phylogenies (Lowe et al. 2004). Moreover, use of microsatellite loci can add significant information because they are examined as multilocus markers and a good number of polymorphic loci are already available in the literature for *B. tabaci* (Dalmon et al. 2008; De Barro et al. 2003; Delatte et al. 2006; Gauthier et al. 2008; Schwartz unpublished; Tsagkarakou and Roditakis 2003; Tsagkarakou et al. 2007). Even though microsatellites are nuclear DNA, they evolve faster than mtDNA owing to the frequent error of polymerase slippage during replication. Although not as appropriate for phylogenetic inference as other DNA markers, microsatellites have been used successfully to infer population demographic histories and their high polymorphism allows for analysis at a finer scale. These observations illustrate that selection of a marker must be evaluated upon the natural history of the marker and the type of question being asked, whether the focus is short-term or long-term estimations of population parameters.

A final consideration is the ability to produce connectable data with such markers. The analysis of allozymes, AFPLs, RFLPs, RAPDs, and microsatellites cannot be

standardized across different laboratories due to the variability in the produced fragment sizes arising from using different equipment and the user subjectivity in allele scoring. Therefore, it is difficult to share data and compare results accurately among different laboratories. In contrast, the use of sequences allows such connectivity in the produced data, and researchers can easily share this information, even in publicly available databases such as GenBank. Technological advances in sequencing methods will allow the exploration of more genomic/genetic data that will hopefully provide the opportunity to share data in more collaborative efforts among scientists.

With the impending sequencing of the *B. tabaci* genome, functional genomics approaches will make it possible to identify genes associated with extreme phenotypes, such as invasiveness, insecticide resistance, and fecundity, as well as the basis for reproductive isolation among biotypes. This advancement would make an enormous contribution to our knowledge of the *B. tabaci* complex and its interactions with host, viruses, endosymbionts, and its environment. In addition, the exploration of "next generation" sequencing, *e.g*. pyrosequencing and 454-sequencing technology (Ellegren 2008, Metzker 2010), and recent advances such as the RAD approach (Baird et al 2008) will allow the rapid development of large numbers of single nucleotide polymorphisms (SNPs) at relatively low cost. These markers are ideal for studies of population histories with marked advantages over other markers such as microsatellites (Brumfield et al 2003). However, up to now, their development was dependent on genome data, which were only available for model organisms, such as humans and *Drosophila* (Hartle and Clark 2007, Hughes et al 2008, Pool et al 2010). Such applications could shed light on complex demographic histories of biotypes in *B. tabaci*, especially parameters involved in invasions such as changes in population size and pathways of introduction and dispersal. Directing these analyses towards some interesting reference biotypes, such as the invasive B and Q, as well as the less damaging A biotype and the monophagous T biotype, would help pinpoint genes associated with resistance, polyphagy, and ability for invasiveness.

Conclusions

Since the first appreciation of the high genetic polymorphism within the *B. tabaci* sibling species group in the mid-80's, a wealth of literature has been produced, with studies focusing on a range of molecular markers, from allozymes, to RFLP, RAPD, AFLP, and more recently to microsatellites. These studies addressed questions ranging from host associations and geographic structure of biotypes, to mating compatibilities between biotypes and provided a worldwide perspective of these topics. Our knowledge of biotype biology and ecology was complemented with genetic information through the use of molecular markers and allowed us to appreciate the unique nature of the *B. tabaci* sibling species group—a complex of cryptic variants that exhibit genetic variation comparative to species level in other organisms, while having maintained the same morphologies throughout their evolutionary history.

In this overview, our aim was to provide a historical overview of the studies that employed multilocus markers to answer these questions, by focusing on the main findings of each study, while also providing some background information about the

markers used. We also consider possible future studies that can help enhance our knowledge not only of the current status and distribution of biotypes, but also of the long-term history that contributed to the evolution of the present biotype variants. With these thoughts, we hope that future research will focus on the *B. tabaci* sibling species group as a model system with which to address broad evolutionary questions.

References

Abdullahi I, Atiri GI, Thottappilly G, Winter S. 2004. Discrimination of cassava-associated *Bemisia tabaci* in Africa from polyphagous populations, by PCR-RFLP of the internal transcribed spacer regions of ribosomal DNA. *J. Appl. Entomol.* 128:81-87.

Abdullahi I, Winter S, Atiri GI, Thottappilly G. 2003. Molecular characterization of whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae) populations infesting cassava. *Bull. Entomol. Res.* 93:97-106.

Anthony N, Brown JK, Markham PG, ffrench-Constant RH. 1995. Molecular analysis of cyclodiene resistance-associated mutations among populations of the sweetpotato whitefly *Bemisia tabaci*. *Pestic. Biochem. Physiol.* 51: 220-228.

Avise JC. 2004. *Molecular Markers, Natural History, and Evolution*. Sunderland, MA: Sinauer Associates, Inc.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3, e3376, doi:3310.1371/journal.pone.0003376.

Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Mol. Ecol.* 13:729-744.

Banks GK, Colvin J, Chowda Reddy RV, Maruthi MN, Muniyappa V, Venkatesh HM, Kiran Kumar M, Padmaja AS, Beitia FJ, Seal SE 2001. First report of the *Bemisia tabaci* B biotype in India and an associated tomato leaf curl virus disease epidemic. *Plant Dis.* 85:231.

Bedford ID, Briddon RW, Brown JK, Rosell RC, Markham PG. 1994. Geminivirus transmission and biological characterization of *Bemisia tabaci* (Gennadius) biotypes from different geographic regions. *Ann. Appl. Biol.* 125:311-325.

Behura SK. 2006. Molecular marker systems in insects: current trends and future avenues. *Mol. Ecol.* 15:3087-3113.

Bellows TS, Perring TM, Gill RJ, Headrick DH. 1994. Description of a species of *Bemisia* (Homoptera: Aleyrodidae). *Ann. Entomol. Soc. Am.* 87:195-206.

Bensch S, Akesson M. 2005. Ten years of AFLP in ecology and evolution: why so few animals? *Mol. Ecol.* 14:2899-2914.

Berry SD, Fondong VN, Rey C, Rogan D, Fauquet CM, Brown JK. 2004. Molecular evidence for five distinct *Bemisia tabaci* (Homoptera: Aleyrodidae) geographic haplotypes associated with cassava plants in sub-Saharan Africa. *Ann. Entomol. Soc. Am.* 97:852-859.

Bird J. 1957. A whitefly-transmitted mosaic of *Jatropha gossypifolia*. *Tech. Pap. P.R. Agric. Exp. Sta.* 22:1-35.

Bird J, Maramorosch K. 1978. Viruses and virus diseases associated with whiteflies. *Adv. Virus Res.* 22:55-110.

Black WC. 1993. PCR with arbitrary primers: approach with care. *Insect Mol. Biol.* 2:1-6.

Black WC, Duteau NM, Puterka GJ, Nechols JR, Pettorini JM. 1992. Use of the random amplified polymorphic DNA - polymerase chain reaction (RAPD-PCR) to detect

DNA polymorphisms in aphids (Homoptera: Aphididae). *Bull. Entomol. Res.* 82:151-159.

Bossart JL, Prowell DP. 1998. Genetic estimates of population structure and gene flow*:* limitations*,* lessons and new directions*. Trends Ecol. Evol.* 13:202–206.

Bosco D, Loria A, Sartor C, Cenis JL. 2006. PCR-RFLP identification of *Bemisia tabaci* biotypes in the Mediterranean basin. *Phytoparasitica* 34:243-251.

Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32:314-331.

Brown JK. 2007. The *Bemisia tabaci* complex: genetic and phenotypic variability drives begomovirus spread and virus diversification. Plant Disease Feature. http://www.apsnet.org/online/feature/btabaci Accessed: July 15, 2010.

Brown JK, Bird J. 1992. Whitefly-transmitted geminiviruses in the Americas and the Caribbean Basin: past and present. *Plant Dis*. 76: 220-225.

Brown JK, Bird J. 1996. Introduction of an exotic whitefly (*Bemisia*) vector facilitates secondary spread of Jatropha mosaic virus, a geminivirus previously vectored exclusively by the Jatropha biotype. In *Bemisia 1995: Taxonomy, Biology, Damage, Control and Management*, ed. D Gerling, RT Mayer, pp. 351-353. Wimborne: Intercept Publications.

Brown JK, Coats SA, Bedford ID, Markham PG, Bird J, Frohlich DR. 1995a. Characterization and distribution of esterase electromorphs in the whitefly, *Bemisia tabaci* (Genn) (Homoptera: Aleyrodidae). *Biochem. Genet.* 33:205-214.

Brown JK, Fletcher D, Bird J. 1993. First report of *Passiflora* leaf mottle caused by a whitefly-transmitted geminivirus in Puerto Rico. *Plant Dis*. 77:1264.

Brown JK, Frohlich DR, Rosell RC. 1995b. The sweetpotato or silverleaf whiteflies: Biotypes of *Bemisia tabaci* or a species complex? *Annu. Rev. Entomol.* 40:511-534.

Brown JK, Perring TM, Cooper AD, Bedford ID, Markham PG. 2000. Genetic analysis of *Bemisia* (Hemiptera: Aleyrodidae) populations by isoelectric focusing electrophoresis. *Biochem. Genet.* 38:13-25.

Brown JK, 2010. Phylogenetic biology of the *Bemisia tabaci* sibling species group. In *Bemisia: Bionomics and Management of a Global Pest*, ed. PA Stansly and SE Naranjo. Dordrecht: Springer Science. In press.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV. 2003. The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol. Evol.* 18:249-256

Burban C, Fishpool LDC, Fauquet C, Fargette D, Thouvenel JC. 1992. Host-associated biotypes within West African populations of the whitefly *Bemisia tabaci* (Genn), (Homoptera: Aleyrodidae). *J. Appl. Entomol.* 113:416-423.

Buth DG. 1984. The application of electrophoretic data in systematic studies. *Annu. Rev. Ecol. Syst.* 15:501-522.

Byrne FJ, Cahill M, Denholm I, Devonshire AL. 1995. Biochemical identification of interbreeding between B type and non B-type strains of the tobacco whitefly *Bemisia tabaci*. *Biochem. Genet.* 33:13-23.

Byrne FJ, Devonshire AL. 1993. Insensitive acetylcholinesterase and esterase polymorphism in susceptible and resistant populations of the tobacco whitefly *Bemisia tabaci* (Genn). *Pestic. Biochem. Physiol.* 45:34-42.

Byrne FJ, Devonshire AL. 1996. Biochemical evidence of haplodiploidy in the whitefly *Bemisia tabaci*. *Biochem. Genet.* 34:93-107.

Byrne FJ, Gorman KJ, Cahill M, Denholm I, Devonshire AL. 2000. The role of B-type esterases in conferring insecticide resistance in the tobacco whitefly, *Bemisia tabaci* (Genn). *Pest Manage. Sci.* 56: 867-874.

Caballero Rafael De J. 2007. Systematics of the *Bemisia tabaci* complex and the role of endosymbionts in reproductive compatibility. *PhD Thesis*. The University of Arizona, 107 pp.

Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, Sutherland GR. 1993. Incidence and origin of null alleles in the (AC)n microsatellite markers. *Am. J. Hum. Genet.* 52:922-927.

Cervera MT, Cabezas JA, Simon B, Martinez-Zapater JM, Beitia F, Cenis JL. 2000. Genetic relationships among biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae) based on AFLP analysis. *Bull. Entomol. Res.* 90:391-396.

Coats SA, Brown JK, Hendrix DL. 1994. Biochemical characterization of biotype specific esterases in the whitefly, *Bemisia tabaci* Genn (Homoptera: Aleyrodidae). *Insect Biochem. Mol. Biol.* 24:723-728.

Costa HS, Brown JK. 1990. Variability in biological characteristics, isozyme patterns and virus transmission among populations of *Bemisia tabaci* (Genn.) in Arizona. *Phytopathology* 80:888.

Costa HS, Brown JK. 1991. Variation in biological characteristics and esterase patterns among populations of *Bemisia tabaci*, and the association of one population with silverleaf symptom induction. *Entomol. Exp. Appl.* 61:211-219.

Costa HS, Brown JK, Sivasupramaniam S, Bird J. 1993. Regional distribution insecticide resistance, and reciprocal crosses between the A-biotype and B-biotype of *Bemisia tabaci*. *Insect Sci. Appl.* 14:255-266.

Dakin EE, Avise JC. 2004. Microsatellite null alleles in parentage analysis. *Heredity* 93:504-509.

Dalmon A, Halkett F, Granier M, Delatte H, Peterschmitt M. 2008. Genetic structure of the invasive pest *Bemisia tabaci*: evidence of limited but persistent genetic differentiation in glasshouse populations. *Heredity* 100:316-325.

Davies N, Villablanca FX, Roderick GK. 1999. Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics. *Trends Ecol. Evol.* 14:17-21.

De Barro PJ. 2005. Genetic structure of the whitefly *Bemisia tabaci* in the Asia-Pacific region revealed using microsatellite markers. *Mol. Ecol.* 14:3695-3718.

De Barro PJ, Driver F. 1997. Use of RAPD PCR to distinguish the B biotype from other biotypes of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae). *Aust. J. Entomol.* 36:149-152.

De Barro PJ, Driver F, Trueman JWH, Curran J. 2000. Phylogenetic relationships of world populations of *Bemisia tabaci* (Gennadius) using ribosomal ITS1. *Mol. Phylogenet. Evol.* 16:29-36.

De Barro PJ, Liebregts W, Carver M. 1998. Distribution and identity of biotypes of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) in member countries of the Secretariat of the Pacific Community. *Aust. J. Entomol.* 37:214-218.

De Barro PJ, Scott KD, Graham GC, Lange CL, Schutze MK. 2003. Isolation and characterization of microsatellite loci in *Bemisia tabaci*. *Mol. Ecol. Notes* 3:40-43.

De Barro PJ, Trueman JWH, Frohlich DR. 2005. *Bemisia argentifolii* is a race of *B. tabaci* (Hemiptera: Aleyrodidae): the molecular genetic differentiation of *B. tabaci* populations around the world. *Bull. Entomol. Res.* 95:193-203.

Delatte H, David P, Granier M, Lett JM, Goldbach R, Peterschmitt M, Reynaud B. 2006. Microsatellites reveal extensive geographical, ecological and genetic contacts between invasive and indigenous whitefly biotypes in an insular environment. *Genet. Res.* 87:109-124.

Delatte H, Reynaud B, Granier M, Thornary L, Lett JM, Goldbach R, Peterschmitt M. 2005. A new silverleaf-inducing biotype Ms of *Bemisia tabaci* (Hemiptera: Aleyrodidae) indigenous to the islands of the south-west Indian Ocean. *Bull. Entomol. Res.* 95:29-35.

Demichelis S, Bosco D, Manino A, Marian D, Caciagli P. 2000. Distribution of *Bemisia tabaci* (Hemiptera: Aleyrodidae) biotypes in Italy. *Can. Entomol.* 132:519-527.

DeWoody J, Nason JD, Hipkins VD. 2006. Mitigating scoring errors in microsatellite data from wild populations. *Mol. Ecol. Notes* 6:951-957.

Dobrowolski MP, Tommerup IC, Blakeman HD, O'Brien PA. 2002. Non-mendelian inheritance revealed in a genetic analysis of sexual progeny of *Phytophthora cinnamomi* with microsatellite markers. *Fungal Genet. Biol.* 35:197-212.

Edwards OR, Hoy MA. 1993. Polymorphism in two parasitoids detected using random amplified polymorphic DNA polymerase chain reaction. *Biol. Control* 3:243-257.

Ellegren H. 2004. Microsatellites: Simple sequences with complex evolution. *Nat. Rev.: Genetics* 5:435-445.

Ellegren H. 2008. News and Views: Sequencing goes 454 and takes large-scale genomics into the wild. *Mol. Ecol.* 17:1629-1631.

Emerson BC, Paradis E, Thebaud C. 2001. Revealing the demographic histories of species using DNA sequences. *Trends Ecol. Evol.* 16:707-716.

Endersby NM, McKechnie SW, Vogel H, Gahan LJ, Baxter SW, Ridland PM, Weeks AR. 2005. Microsatellites isolated from diamondback moth, *Plutella xylostella* (L.), for studies of dispersal in Australian populations. *Mol. Ecol. Notes* 5:51-53.

Estoup A, Jarne P, Cornuet JM. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol. Ecol.* 11:1591-1604.

Excoffier L, Heckel G. 2006. Computer programs for population genetics data analysis: a survival guide. *Nature Reviews: Genetics* 7:745-758.

Gauthier N, Dalleau-Clouet C, Bouvret M-E. 2008. Twelve new polymorphic microsatellite loci and PCR multiplexing in the whitefly, *Bemisia tabaci*. *Mol. Ecol. Resources* 8:1004-1007.

Gawel NJ, Bartlett AC. 1993. Characterization of differences between whiteflies using RAPD-PCR. *Insect Mol. Biol.* 2:33-38.

Gill RJ, Brown JK. 2010. Systematics of *Bemisia* and *Bemisia* relatives: can molecular techniques solve the *Bemisia tabaci* complex conundrum – a taxonomist's

viewpoint. In *Bemisia: Bionomics and Management of a Global Pest*, ed. PA Stansly and SE Naranjo. Dordrecht: Springer Science. In press.
Guillot G, Santos F, Estoup A. 2008. Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface. *Bioinformatics* 24:1406-1407.
Guirao P, Beitia F, Cenis JL. 1997. Biotype determination of Spanish populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Bull. Entomol. Res.* 87:587-593.
Gunning RV, Byrne FJ, Conde BD, Connelly MI, Hergstrom K, Devonshire AL. 1995. First report of B-biotype *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) in Australia. *J. Aust. Entomol. Soc.* 34:116-120.
Gunning RV, Byrne FJ, Devonshire AL. 1997. Electrophoretic analysis of non-B and B-biotype *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) in Australia. *Aust. J. Entomol.* 36:245-249.
Hadrys H, Balick M, Schierwater B. 1992. Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Mol. Ecol.* 1:55-63.
Harris H. 1966. Enzyme polymorphisms in man. *Proc. R. Soc. Ser. B: Biol. Sci.* 164:298-310.
Harris SA. 1999. RAPDs in systematics - a useful methodology? In *Molecular Systematics and Plant Evolution*, ed. PM Hollingsworth, RM Bateman, RJ Gornall, pp. 211-228. London: Taylor & Francis.
Hartle D, Clark A. 2007. *Principles of Population Genetics*. Sunderland MA: Sinauer Associates.
Horowitz AR, Denholm I, Gorman K, Cenis JL, Kontsedalov S, Ishaaya I. 2003. Biotype Q of *Bemisia tabaci* identified in Israel. *Phytoparasitica* 31:94-98.
Hoy M. 2003. *Insect Molecular Genetics: An Introduction to Principles and Applications*. San Diego: Academic Press.
Hubby JL, Lewontin RC. 1966. A molecular approach to study of genic heterozygosity in natural populations .I. Number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* 54:595-609.
Hughes AL, Welch R, Puri V, Matthews C, Haque K, Chanock SJ, Yeager M. 2008. Genome-wide SNP typing reveals signatures of population history. *Genomics* 92:1-8.
Jansen RC, Geerlings H, Van Oeveren AJ, Van Schaik RC. 2001. A comment on codominant scoring of AFLP markers. *Genetics* 158:925-926.
Jarne P, Lagoda PJL. 1996. Microsatellites, from molecules to populations and back. *Trends Ecol. Evol.* 11:424-429.
Johnson FM, Kanapi CG, Richardson RH, Wheeler MR, Stone WS. 1966. An analysis of polymorphisms among isozyme loci in dark and light *Drosophila ananassae* strains from American and Western Samoa. *Proc. Natl. Acad. Sci. USA* 56:119-125.
Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, van de Wiel C, Bredemeijer G, Vosman B, Matthes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevschi A, Marmiroli N, Aert R, Volckaert G, Rueda J, Linacero R, Vazquez A, Karp A. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Mol. Breed.* 3:381-390.

Karl SA, Avise JC. 1992. Balancing selection at allozyme loci in oysters - implications from nuclear RFLPs. *Science* 256:100-102.

Khasdan V, Levin I, Rosner A, Morin S, Kontsedalov S, Maslenin L, Horowitz AR. 2005. DNA markers for identifying biotypes B and Q of *Bemisia tabaci* (Hemiptera: Aleyrodidae) and studying population dynamics. *Bull. Entomol. Res.* 95:605-613.

Kim KS, Sappington TW. 2006. Molecular genetic variation of boll weevil populations in North America estimated with microsatellites: Implications for patterns of dispersal. *Genetica* 127:143-161.

Kirk AA, Lacey LA, Brown JK, Ciomperlik MA, Goolsby JA, Vacek DC, Wendel LE, Napompeth B. 2000. Variation in the *Bemisia tabaci* s. l. species complex (Hemiptera: Aleyrodidae) and its natural enemies leading to successful biological control of *Bemisia biotype* B in the USA. *Bull. Entomol. Res.* 90:317-327.

Legg JP, French R, Rogan D, Okao-Okuja G, Brown JK. 2002. A distinct *Bemisia tabaci* (Gennadius) (Hemiptera : Sternorrhyncha : Aleyrodidae) genotype cluster is associated with the epidemic of severe cassava mosaic virus disease in Uganda. *Mol. Ecol.* 11:1219-1229.

Levinson G, Gutman GA. 1987. Slipped-strand mispairing - a major mechanism for DNA-sequence evolution. *Mol. Biol. Evol.* 4:203-221.

Lima LHC, Campos L, Moretzsohn MC, Navia D, de Oliveira MRV. 2002. Genetic diversity of *Bemisia tabaci* (Genn.) populations in Brazil revealed by RAPD markers. *Genet. Mol. Biol.* 25:217-223.

Lima LHC, Navia D, Inglis PW, de Oliveira MRV. 2000. Survey of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) biotypes in Brazil using RAPD markers. *Genet. Mol. Biol.* 23:781-785.

Liu SS, De Barro PJ, Xu J, Luan JB, Zang LS, Ruan YM, Wan FH. 2007. Asymmetric mating interactions drive widespread invasion and displacement in a whitefly. *Science* 318:1769-1772.

Lowe A, Harris S, Ashton P. 2004. *Ecological Genetics: Design, Analysis, and Application.* Oxford, UK, Blackwell Publishing Ltd.

Luikart G, England PR. 1999. Statistical analysis of microsatellite DNA data. *Trends Ecol. Evol.* 14:253-256.

Lynch M, Milligan BG. 1994. Analysis of population genetic structure with RAPD markers. *Mol. Ecol.* 3:91-99.

Marjoram P, Tavaré S. 2006. Modern computational approaches for analyzing molecular genetic variation data. *Nature Reviews: Genetics* 7:759-770.

Maruthi MN, Colvin J, Seal S. 2001. Mating compatibility, life-history traits, and RAPD-PCR variation in *Bemisia tabaci* associated with the cassava mosaic disease pandemic in East Africa. *Entomol. Exp. Appl.* 99:13-23.

Maruthi MN, Colvin J, Seal S, Gibson G, Cooper J. 2002. Co-adaptation between cassava mosaic geminiviruses and their local vector populations. *Virus Res.* 86:71-85.

McKenzie CL, Anderson PK, Villarreal N. 2004. An extensive survey of *Bemisia tabaci* (Homoptera: Aleyrodidae) in agricultural ecosystems in Florida. *Fla. Entomol.* 87:403-407.

Metzker ML. 2010. Sequencing technologies — the next generation. *Nat. Rev. Genet.* 11:31-46.

Mitton JB. 1994. Molecular approaches to population biology. *Annu. Rev. Ecol. Syst.* 25:45-69.
Moya A, Guirao P, Cifuentes D, Beitia F, Cenis JL. 2001. Genetic diversity of Iberian populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae) based on random amplified polymorphic DNA-polymerase chain reaction. *Mol. Ecol.* 10:891-897.
Mueller UG, Wolfenbarger LL. 1999. AFLP genotyping and fingerprinting. *Trends Ecol. Evol.* 14:389-394.
Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. 1986. Specific enzymatic amplification of DNA *in vitro* - the Polymerase Chain Reaction. *Cold Spring Harbor Symp. Quant. Biol.* 51:263-273.
Nei M. 1976. Mathematical models of speciation and genetic distance. In *Population Genetics and Ecology*, ed. S Karlin, E Neve, pp. 723-765. New York: Academic Press.
Parker PG, Snow AA, Schug MD, Booton GC, Fuerst PA. 1998. What molecules can tell us about populations: Choosing and using a molecular marker. *Ecology* 79:361-382.
Parsons YM, Shaw KL. 2002. Mapping unexplored genomes: A genetic linkage map of the Hawaiian cricket *Laupala*. *Genetics* 162:1275-1282.
Pemberton JM, Slate J, Bancroft DR, Barrett JA. 1995. Nonamplifying alleles at microsatellite loci - a caution for parentage and population studies. *Mol. Ecol.* 4:249-252.
Perring TM. 2001. The *Bemisia tabaci* species complex. *Crop Prot.* 20:725-737.
Perring TM, Cooper A, Kazmer DJ. 1992. Identification of the poinsettia strain of *Bemisia tabaci* (Homoptera: Aleyrodidae) on broccoli by electrophoresis. *J. Econ. Entomol.* 85:1278-1284.
Perring TM, Cooper AD, Rodriguez RJ, Farrar CA, Bellows TS Jr. 1993. Identification of a whitefly species by genomic and behavioral studies. *Science* 259:74-77.
Pool JE, Hellmann I, Jensen JD, Nielsen R. 2010. Population genetic inference from genomic sequence variation. *Genome Res.* 20:291-300
Prabhaker N, Coudriet DL, Meyerdirk DE. 1987. Discrimination of 3 whitefly species (Homoptera: Aleyrodidae) by electrophoresis of nonspecific esterases. *J. Appl. Entomol.* 103:447-451.
Qiu BL, Coats SA, Ren SX, Dris AM, Xu CX, Brown JK. 2007. Phylogenetic relationships of native and introduced Bemisia tabaci (Homoptera: Aleyrodidae) from China and India based on mtCOI DNA sequencing and host plant comparisons. *Prog. Nat. Sci.* 17:645-654.
Rekha AR, Maruthi MN, Muniyappa V, Colvin J. 2005. Occurrence of three genotypic clusters of *Bemisia tabaci* and the rapid spread of the B biotype in south India. *Entomol. Exp. Appl.* 117:221-233.
Robinson JP, Harris SA. 1999. Amplified fragment length polymorphisms and microsatellites: a phylogenetic perspective. In *Which DNA Marker for Which Purpose?*, ed. EM Gillet, pp. 1-27. Göttingen: Institut für Forstgenetik und Forstpflanzenzüchtung, Universität Göttingen.
Roderick GK. 1996. Geographic structure of insect populations: gene flow, phylogeography, and their uses. *Annu Rev Entomol* 41:325-352.

Rosell RC, Bedford ID, Frohlich DR, Gill RJ, Markham PG, Brown JK. 1997. Analyses of morphological variation in distinct populations of *Bemisia tabaci*. *Ann. Entomol. Soc. Am*. **90**:575-589.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA-polymerase. *Science* 239:487-491.

Schierwater B, Ender A. 1993. Different thermostable DNA polymerases may amplify different RAPD products. *Nucleic Acids Res.* 21:4647-4648.

Schlötterer C. 2000. Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109:365-371.

Schlötterer C. 2001. Genealogical inference of closely related species based on microsatellites. *Genet. Res.* 78:209-212.

Schlötterer C. 2004. The evolution of molecular markers - just a matter of fashion? *Nature Reviews: Genetics* 5:63-69.

Schlötterer C, Pemberton J. 1994. The use of microsatellites for genetic analysis of natural populations. In *Molecular Ecology and Evolution: Approaches and Applications*, ed. B Schierwater, B Streit, GP Wagner, R DeSalle, pp. 203-214. Basel: Birkhauser Verlag.

Schlötterer, C, Tautz D. 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* 20: 211-215.

Selkoe KA, Toonen RJ. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol. Lett.* 9:615-629.

Simon B, Cenis JL, De La Rua P. 2007. Distribution patterns of the Q and B biotypes of *Bemisia tabaci* in the Mediterranean Basin based on microsatellite variation. *Entomol. Exp. Appl.* 124:327-336.

Simon B, Cenis JL, Demichelis S, Rapisarda C, Caciagli P, Bosco D. 2003. Survey of *Bemisia tabaci* (Hemiptera: Aleyrodidae) biotypes in Italy with the description of a new biotype (T) from Euphorbia characias. *Bull. Entomol. Res.* 93:259-264.

Smith KL, Alberts SC, Bayes MK, Bruford MW, Altmann J, Ober C. 2000. Cross-species amplification, non-invasive genotyping, and non-Mendelian inheritance of human STRPs in Savannah baboons. *Am. J. Primatol.* 51:219-227.

Sseruwagi P, Legg JP, Maruthi MN, Colvin J, Rey MEC, Brown JK. 2005. Genetic diversity of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) populations and presence of the B biotype and a non-B biotype that can induce silverleaf symptoms in squash, in Uganda. *Ann. Appl. Biol.* 147:253-265.

Sseruwagi P, Maruthi MN, Colvin J, Rey MEC, Brown JK, Legg JP. 2006. Colonization of non-cassava plant species by cassava whiteflies (Bemisia tabaci) in Uganda. *Entomol. Exp. Appl.* 119:145-153.

Sseruwagi P, Rey MEC, Brown JK, Legg J. 2004. The cassava mosaic geminiviruses occuring in Uganda following the 1990s epidemic of severe mosaic disease. *Ann. Appl. Biol.* 145:113-121.

Tautz D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* 17:6463-6471.

Tsagkarakou A, Roditakis N. 2003. Isolation and characterization of microsatellite loci in *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Mol. Ecol. Notes* 3:196-198.

Tsagkarakou A, Tsigenopoulos CS, Gorman K, Lagnel J, Bedford ID. 2007. Biotype status and genetic polymorphism of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Greece: mitochondrial DNA and microsatellites. *Bull. Entomol. Res.* 97:29-40.

Viard F, Franck P, Dubois MP, Estoup A, Jarne P. 1998. Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. *J. Mol. Evol.* 47:42-51.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M. 1995. AFLP - a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23:4407-4414

Weber JL, Wong C. 1993. Mutation of human short tandem repeats. *Hum. Mol. Genet.* 2:1123-1128.

Welsh J, McClelland M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* 18:7213-7218.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18:6531-6535.

Wilson IJ, Balding DJ. 1998. Genealogical inference from microsatellite data. *Genetics* 150:499-510.

Wong A, Forbes MR, Smith ML. 2001. Characterization of AFLP markers in damselflies: prevalence of codominant markers and implications for population genetic applications. *Genome* 44:677-684.

Wool D, Gerling D, Bellotti A, Morales F, Nolt B. 1991. Spatial and temporal genetic variation in populations of the whitefly *Bemisia tabaci* (Genn) in Israel and Colombia - an interim report. *Insect Sci. Appl.* 12:225-230.

Wool D, Gerling D, Bellotti AC, Morales FJ. 1993. Esterase electrophoretic variation in *Bemisia tabaci* (Genn) (Homoptera: Aleyrodidae) among host plants and localities in Israel. *J. Appl. Entomol.* 115:185-196.

Wool D, Gerling D, Nolt BL, Constantino LM, Bellotti AC, Morales FJ. 1989. The use of electrophoresis for identification of adult whiteflies (Homoptera: Aleyrodidae) in Israel and Colombia. *J. Appl. Entomol.* 107:344-350.

Yan G, Romero-Severson J, Walton M, Chadee DD, Severson DW. 1999. Population genetics of the yellow fever mosquito in Trinidad: comparisons of amplified fragment length polymorphism (AFLP) and restriction fragment length polymorphism (RFLP) markers. *Mol. Ecol.* 8:951-963.

Zhang DX, Hewitt GM. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Mol. Ecol.* 12:563-584.

Zhang LP, Zhang YJ, Zhang WJ, Wu QJ, Xu BY, Chu D. 2005. Analysis of genetic diversity among different geographical populations and determination of biotypes of *Bemisia tabaci* in China. *J. Appl. Entomol.* 129:121-128.

Table 1. Multilocus molecular markers used in studies of *Bemisia tabaci*

A. Markers and properties, reproduced from Lowe et al. (2004) by permission of the authors and publisher

	Allozymes	**RFLPs**	**RAPDs**	**AFLPs**	**Microsatellites**
Polymorphism	Charged amino acid substitutions	Nucleotide substitutions, indels, inversions	Nucleotide substitutions, indels, inversions	Nucleotide substitutions, indels, inversions	Repeat number changes
Level of polymorphism	Low	Medium	Medium	Medium	High
Abundance in genome	Low	High	Very high	High	Medium
Dominance	Usually codominant	Codominant	Dominant	Codominant/dominant	Codominant
Development costs	Low	Medium	Low	Medium	High
Start up costs	Low	Medium/high	Low	Medium	High
Automation	No	Limited	Yes	Yes	Yes
Reproducibility	Medium/high	High	Low	Medium	High

B. Applications in *Bemisia tabaci*

	Allozymes	**RFLPs**	**RAPDs**	**AFLPs**	**Microsatellites**
Host associations	Brown et al. (1995a); Brown and Bird (1996); Burban et al. (1992); Costa and Brown (1991); Wool et al. (1991, 1993)	Abdullahi et al. (2004)	Abdullahi et al. (2003)	-	De Barro (2005)
Insecticide resistance	Anthony et al. (1995); Byrne et al. (2000); Coats et al. (1994)	-	-	-	-
Sex determination, haplodiploidy	Byrne and Devonshire (1996)	-	-	-	-
Interbreeding	Bedford et al. (1994); Byrne et al. (1995); Caballero (2007); Costa et al. (1993); Gunning et al. (1997)	-	-	-	Delatte et al. (2006)
Virus epidemic associations	Bedford et al. (1994); Brown and Bird (1992); Brown (2007); Brown and Bird (1996); Brown et al. (1993); Legg et al. (2002)	-	Maruthi et al. (2002)	-	-

B. Applications in *Bemisia tabaci*

	Allozymes	**RFLPs**	**RAPDs**	**AFLPs**	**Microsatellites**
Biotype/Haplotype/Species identification	Costa and Brown (1990, 1991); Costa et al. (1993); Perring et al. (1992); Brown et al. (1995a); Brown et al. (2000)	Bosco et al. (2006)	Gawel and Bartlett (1993); Guirao et al. (1997); Delatte et al. (2005); Horowitz et al. (2003); Simon et al. (2003); McKenzie et al. (2004); Khasdan et al. (2005)	Cervera et al. (2000)	Tsagkarakou et al. (2007)
Geographic structure	-	-	De Barro and Driver (1997); De Barro et al. (1998); Rekha et al. (2005)	Cervera et al. (2000); Zhang et al. (2005)	De Barro (2005)
Population differentiation	Costa et al (1993); Brown et al. (1995a); Brown et al. (2000); Wool et al. (1989)	Abdullahi et al. (2004)	Perring et al. (1993); De Barro and Driver (1997); De Barro et al. (1998); Guirao et al. (1997); Moya et al. (2001); Lima et al. (2000)	Cervera et al. (2000); Zhang et al. (2005)	De Barro (2005); Delatte et al. (2006); Tsagkarakou et al. (2007); Simon et al. (2007); Dalmon et al. (2008)

Chapter 2

Global population genetics of a worldwide pest: genetic diversity and population structure in the *Bemisia tabaci* cryptic species complex

Abstract

The evolution of cryptic species remains one of the most intriguing and challenging questions in evolutionary biology. The whitefly species complex *Bemisia tabaci* (Hemiptera: Aleyrodidae) comprises several morphologically cryptic lineages, known as "biotypes" with worldwide distribution, including two of the world's worst invasive pests. *Bemisia tabaci* lineages exhibit enormous diversity in their ecology and biology and extensive genetic studies in the last 30 years have revealed substantial diversity in the mitochondrial DNA (mtCOI) with other markers being less variable. However, our view of global relationships in this complex is largely limited to a single marker without the ability to assess gene flow and genetic structure at the population level. In this study we attempted to untangle the global population structure of *B. tabaci* using microsatellite markers. We examined a large set of worldwide collections representing monophagous and polyphagous as well as invasive and local lineages. We found that the well-characterized biotypes and other geographic lineages represent highly differentiated genetic clusters with little or no gene flow among them. Surprisingly, invasive biotypes exhibited relatively high levels of genetic diversity compared to other biotypes, suggesting both large founding populations with ancestral variation maintained despite the homogenizing effects of human-mediated gene flow among their populations. Analyses of other, non-invasive biotypes also revealed interesting results, demonstrating the wealth of knowledge still to be gained from extensive studies of this species complex. The microsatellite results are in general agreement with the mtCOI phylogeny, although some cases of conflict in the two markers, perhaps associated with the different modes of inheritance, caution against the use of mtCOI as a single marker for species delineation in this taxon. Our study supports the hypothesis that the worldwide biotypes are very old entities and adds yet another independent piece of information to our understanding of cryptic diversity in the *B. tabaci* species complex.

Introduction

The evolutionary processes leading to speciation in ecologically and genetically divergent populations have attracted the interest of many researchers (Coyne and Orr 2004; Schluter 2001; Turelli et al. 2001), with much focus in the study of speciation in phytophagous insects (Bush and Butlin 2004; Dres and Mallet 2002; Feder et al. 1994; Funk et al. 2002). Speciation can be perceived as a continuous process from polymorphic populations to distinct species, with "biotypes" or "ecological races" acting as intermediate stages (Diehl and Bush 1984; Dres and Mallet 2002). Different mechanisms and complex processes play into the formation of the "intermediate" lineages, with variable roles of ecology (e.g. host fidelity in apple maggots, Feder et al. 1994), geography (allopatry *vs* sympatry, Bush 1969; Mayr 1963), and sexual selection (Lande 1981; Lande and Kirkpatrick 1988). For example sympatric populations of a cichlid fish in a Nicaraguan crater lake have evolved into two morphologically distinct species in less than 10,000 yr, the process driven by both resource partitioning (ecological diversification) and mate choice (sexual selection) (Barluenga et al. 2006).

At the other extreme, we have several empirical examples of morphologically conserved lineages, apparently reproductively isolated for millions of years that are currently either allopatric (e.g. Dolman and Moritz 2006; Elmer et al. 2007) or have shifted into sympatric (e.g. Stuart et al. 2006) or parapatric ranges (e.g. Brown et al. 2007). Such cases of morphologically conserved lineages that are genetically divergent and reproductively isolated are referred to as cryptic species because of their previous misclassification into a single taxon based on identical morphologies. Cryptic species are more common than previously thought and have been found across major metazoan taxa and biogeographical regions (Pfenninger and Schwenk 2007). With advances in molecular and genetic techniques the discovery and description of cryptic species has increased exponentially in the past two decades (Bickford et al. 2007).

The view that cryptic lineages are the result of recent speciation events has been contested in light of recent studies suggesting ancestral divergence of morphologically cryptic lineages, in some cases dating back to the Oligocene, 24 million years ago (Beheregaray and Caccone 2007; Bickford et al. 2007; Elmer et al. 2007 and several references therein). It has been suggested that behavioral, physiological and developmental plasticity may allow organisms to compensate for environmental perturbations without the need to change morphologically (Wake et al. 1983), leading to "morphological stasis". Persisting morphologies may be maintained by stabilizing selection (Charlesworth et al. 1982), while divergence at other traits (behavioral, ecological, genetic) and speciation proceeds at a "normal" pace; such selection pressures may be imposed by conditions experienced at extreme environments (Bickford et al. 2007).

Studies of multiple co-distributed cryptic lineages, combining phylogeographic and population genetics approaches provide an excellent framework to appreciate cryptic biodiversity and understand the processes leading to cryptic speciation (Beheregaray and Caccone 2007). One such system, comprising more than 25 cryptic lineages worldwide, some of which currently overlap in geographic range, is the whitefly species complex *Bemisia tabaci* (Hemiptera: Aleyrodidae), first described by Gennadius as *Aleyrodes tabaci* in 1889. The morphologically conserved lineages, known as *B. tabaci* "biotypes"

show extreme variation in biological and ecological traits such as plant virus transmission efficiency, resistance to insecticides, fecundity, dispersal ability, and mating behavior, with some being invasive polyphagous pests while others are monophagous and restricted in geographic range (Brown 1995; Perring 2001). Ever since molecular techniques were used to study this system in the late 1980's, it became evident that this complex also exhibits extreme variation at the genetic level (for a review see Chapter 1, Brown 2010), with the latest studies recognizing 20-25 cryptic variants around the world (Boykin et al. 2007; Brown 2010; Dinsdale et al. 2010). Which ecological factors have contributed to the extreme diversification in *B. tabaci* is not yet understood, but it has been suggested that lineages of *B. tabaci* diverged millions of years ago following separation of continental landmasses (Gill and Brown 2010). Evidence supporting this hypothesis stems from the extreme divergence in mtCOI sequences among clades (up to 26%, Brown 2010) and the robust phylogeographic pattern of the mtCOI gene.

Understanding the boundaries of biological species in this complex requires a thorough assessment of different biological and ecological characters as well as crossing experiments to determine which lineages are reproductively isolated. However, establishing which lineages are reproductively isolated, or constitute "distinct gene pools" can be assessed through indirect estimates of gene flow, by studying differentiation in nuclear markers (Slatkin 1985). Genetic analyses not only can provide information about contemporary gene flow among lineages, but also shed light on historical demographic processes such as ancestral population expansions and time of divergence from a common ancestor. Analyses with several markers and DNA sequences of a number of gene regions in *B. tabaci* have provided us with a general picture of the biogeographic distribution of lineages, with the mtCOI gene being the most informative and other nuclear loci showing much less divergence (De Barro et al. 2000; Frohlich et al. 1999). Thus, our current view of the global phylogenetic and phylogeographic relationships in *B. tabaci* relies solely on the mtCOI (Boykin et al. 2007; Brown 2010; Dinsdale et al. 2010; Frohlich et al. 1999); a marker that despite its widespread use, has been severely criticized for application in phylogeographic studies and species boundaries delineation (Galtier et al. 2009). Furthermore, this marker may not be variable enough at the population level to be used efficiently for assessment of population demographic histories and migration.

Microsatellite markers are traditionally used for the inference of genetic relationships between populations of the same species or closely related species because of their high polymorphism owing to their high mutation rates (Schlotterer 2001). Multiple microsatellite loci have been isolated from *B. tabaci* since 2003 (Dalmon et al. 2008; De Barro et al. 2003; Delatte et al. 2006; Gauthier et al. 2008; Schwartz unpublished; Tsagkarakou and Roditakis 2003; Tsagkarakou et al. 2007) and were used in a number of studies to examine local population structure and differentiation in biotypes (see Chapter 1 for a review). These studies have provided interesting results, from dissecting the roles of host and geography in structuring of populations (De Barro et al. 2008), to detecting hybridization among sympatric invasive and indigenous biotypes (Delatte et al. 2006), and to identifying the sources and routes of dispersal of invasive biotypes in new areas (Chapter 4). However, they have not yet been employed to determine genetic structure of worldwide lineages and biotypes despite their potential to provide useful information both within and between biotypes.

In this study, we used 13 microsatellite loci to untangle the global population structure of this cryptic species complex and add further insight into the current classification of *B. tabaci* lineages. The main goal was to determine whether "biotypes" that were initially described on the basis of esterase markers and biological/ecological data are real genetic entities, that is, whether they form monophyletic groups with restricted or no gene flow among them. By analyzing a wide range of populations representing monophagous, polyphagous, widespread invasive and local biotypes, we aimed to examine any differential patterns of genetic diversity across the species complex. Finally, we attempted to compare our results with the mtCOI phylogeny and determine if the two markers from different parts of the genome provide similar or contrasting views of the *B. tabaci* species history. We anticipate that this study will add yet another independent piece of information to our understanding of biotype and cryptic lineage divergence in the *B. tabaci* complex.

Materials and Methods

Bemisia tabaci samples and populations

A total of 839 female whiteflies from 50 collections around the world were genotyped in this study (Table 1). Adult females were used for genetic analysis because whiteflies are haplodiploid with diploid females and haploid males (Byrne and Devonshire 1996). The samples span distinct geographic sampling locations representative of the worldwide distribution of *B. tabaci* (Fig 1) and were obtained through direct field sampling and laboratory collections (JK Brown) (Table 1). The samples encompass characterized and well-studied biotypes and major clades of the mtCOI phylogeny of the species complex (Brown unpublished).

DNA extraction

Whiteflies were preserved in 95% ethanol. For DNA extraction, whole adult whiteflies were homogenized in individual 1.5 ml microcentrifuge tubes and genomic DNA was extracted using the Qiagen DNeasy DNA Blood and Tissue kit following the manufacturer's protocol. A modification to this protocol was to perform a final elution of the DNA from the column in 80µl buffer AE followed by a second elution in 20µl AE, with the two eluates combined in a single microcentrifuge tube and stored at -20° C.

Microsatellite loci

A total of 13 microsatellite loci (Table 2) were amplified using polymerase chain reaction (PCR) (see protocols below). We used loci developed both in our laboratory (Schwartz unpublished) and from several literature sources (Dalmon et al. 2008; De Barro et al. 2003; Gauthier et al. 2008; Tsagkarakou and Roditakis 2003; Tsagkarakou et al. 2007). The loci we used were isolated from different *B. tabaci* biotypes and had different repeat motifs, likely evolving at different rates (Table 2). Out of a total of 65 microsatellite primer pairs we tested and screened, we selected 13 loci that cross-amplified in most of our populations. Samples that failed to amplify at any given locus were genotyped at least 3 times before scoring those genotypes as missing data.

PCR protocols and microsatellite genotyping

Microsatellite forward primers were labeled with a fluorescent dye with the reverse primer unlabeled (Table 2). PCRs were initially performed as single primer-pair reactions in 10 µl volumes with 1 µl of genomic DNA, 0.75 mM dNTPs mix, 0.5 µM of each primer, 1 µl of 0.1mg/ml bovine serum albumin (BSA, New England BioLabs), 1.25 µl 10x PCR Buffer (Applied Biosystems), 1.125 mM $MgCl_2$ (ABI), and 0.5 U of AmpliTaq® Gold DNA Polymerase (ABI). Amplification of all loci was done using a touchdown PCR protocol with the following thermocycling conditions: one cycle of 95°C for 10 min followed by 2 cycles of 94°C for 1 min, 60°C for 1 min, 70°C for 35 s followed by 18 cycles of 45 s at 93°C, 45 s at 59°C (lowering 0.5°C each cycle), and 45 s at 70°C. The final amplification step consisted of 20 cycles of 30 s at 92°C, 30 s at 50°C, and 1 min at 70°C followed by a final extension step at 70°C for 5 min. Once we had an assessment of allele ranges for each population, subsequent amplifications of loci across all populations were done by multiplexing multiple primer pairs in single reactions (3-5 pairs per reaction, labeled with different fluorescent dyes or having discrete allele size ranges). The multiplex PCRs were done in 96-well plates using the Multiplex PCR Kit (Qiagen) in 15 µl reaction volumes containing 6 µl of the Qiagen PCR Master Mix (1X), 2 µM of each primer, 3 µl RNase-free water and 1 µl of genomic DNA. Thermocycling conditions were as follows: 15 min at 95°C followed by 30 cycles of 30 s at 94°C, 90 s at 60°C, 60 s at 72°C, with a final extension step of 60°C for 30 min. Final PCR products were mixed with a cocktail of 48:1 Hi-Di formamide (ABI): LIZ500 size standard (ABI) (0.5 µl PCR product, 0.2 µl LIZ, 9.3 µl formamide) and were denatured at 95°C for 5 min. Fragments were ran on an ABI 3730 DNA sequencer and genotypic data were visualized and scored manually using the software GeneMapper version 4.0 (ABI). All runs included negative and multiple positive controls to ensure consistency in allele scoring. We used the program Flexibin (Amos et al. 2006) to facilitate binning and help detect miscalled microsatellite alleles.

Basic population genetics statistics

Prior to any analyses we tested for significant differentiation among populations sampled from neighboring locations using the exact test as implemented in Genepop (version 4.0.10) (Raymond and Rousset 1995; Rousset 2008). Significance levels were corrected for multiple comparisons using the standard Bonferroni correction at the 0.05 level. Populations that were not significantly differentiated were pooled into a single sample for further analyses (see Table 1). We used the program GenAlEx 6.1 (Peakall and Smouse 2006) to calculate locus-specific statistics (number of alleles, total expected heterozygosity, mean expected and observed heterozygosities and *F* statistics) across all populations.

Because not all of our samples amplified at all loci, to obtain unbiased estimates of genetic diversity we used a reduced dataset based on 7 loci (see Table 2), excluding any individuals or whole populations (Uganda-sweetpotato, China- Hainan, *Jatropha*-P.Rico, Guatemala, Mozambique) that had any missing data, resulting in a reduced dataset with 690 individuals. We chose loci isolated from different biotype sources for these calculations (two loci from biotype B, two from biotype Q, two from Asian populations, one from an Australian population) to avoid bias since we observed that microsatellites were more variable in the biotype they were isolated from. Observed (*H*o)

and expected (H_E) heterozygosities per population were calculated in GenAlEx 6.1. For presentation purposes we averaged *H*o and H_E across populations for each biotype or geographic group based on the results of neighbor joining tree (NJ) and clustering analyses as explained below. Allelic richness (number of alleles) was calculated using the rarefaction method implemented in the program HPrare (Kalinowski 2005) to correct for differences in the size and number of populations per biotype / region. Rarefaction standardized samples to the minimum eight genes per population and one population per region. We used the hierarchical sampling scheme offered in HPrare to group populations into known biotypes or geographic regions as determined from the NJ tree.

Tests for significant genotypic linkage disequilibrium (LD) among all pairs of loci and for significant deviations from Hardy-Weinberg equilibrium (HWE) were done in the program Genepop (version 4.0.10) using the Markov chain method and with the default settings. The significance levels were adjusted with a Bonferroni correction at the 0.05 level.

Null alleles

Null alleles can potentially bias estimates of genetic differentiation by reducing the genetic diversity within populations thereby increasing and overestimating the inter-population genetic differentiation (estimates of F_{ST} and genetic distance) (Chapuis and Estoup 2007; Paetkau et al. 1997). In order to address this issue we used the program FreeNa (Chapuis and Estoup 2007) in our full dataset to calculate unbiased global and pairwise F_{ST} of Weir (1996) and Cavalli-Sforza and Edwards chord distance (D_C) (1967) corrected for null alleles using 1000 bootstrap replications over all 13 loci. We then used pairwise D_C corrected and uncorrected for null alleles to construct neighbor-joining (NJ) trees using the *neighbor* program in the software package Phylip-3.69 (Felsenstein 2005). The resulting trees were compared in order to assess the degree to which null alleles could bias our other analyses.

Individual-based analyses

Neighbor Joining Tree

We used the program Populations version 1.2.30 (Langella 1999) to construct an unrooted NJ tree based on D_C among individuals using information from the full dataset. D_C is less biased by the presence of null alleles compared to Nei's (1978) standard genetic distance (D_S) (Chapuis and Estoup 2007). In addition, D_C does not make the assumption of constant population size or constant mutation rates among loci and performs better than other genetic distances in recovering correct tree topologies (Chapuis et al. 2008). The NJ analysis was done at the individual rather than the population level, because this allowed detection of genetic structure in sample populations and migration of individuals between populations. The output obtained from the program Populations was visualized using the Interactive Tree Of Life (iTOL) online tool (Letunic and Bork 2006).

Principal Coordinates Analysis (PCA)

We used PCA as an alternative analysis of the individual genotypes to compare consistency with results from other methods. This technique allows detection of the major patterns in a multivariate dataset, such as a microsatellite dataset with multiple

samples and loci (Peakall and Smouse 2006) by transforming and condensing the multilocus genotype information into a smaller number of derived variables. We used the Excel based program GenAlEx 6.1 (Peakall and Smouse 2006) to calculate a genetic distance matrix [option codominant-genotypic, method described in Smouse and Peakall (1999)] and to convert this into a covariance matrix with data standardization for the PCA. The genetic distance matrix estimation is based on a pairwise individual-by-individual calculation without taking into account the population individuals belong to. For a more simple display of individuals from all populations into a single PCA we color-coded individuals belonging in the same biotype or geographic group (as determined from the NJ tree). For this analysis we used the seven loci that amplified across most populations and an additional locus (Table 2) and individuals with very few missing data that added valuable information to the PCA, resulting in a dataset with 712 individuals.

Bayesian clustering analysis to assess worldwide population structure

In order to identify major genetic clusters in the worldwide populations of *B. tabaci* we used the Bayesian clustering approach implemented in the program STRUCTURE 2.3.3 (Falush et al. 2003; 2007; Hubisz et al. 2009; Pritchard et al. 2000). Since the model used in this method assumes Hardy-Weinberg and linkage equilibria within populations we excluded the three loci that significantly deviated from HWE in more than one population out of 41 (Table 3 and Figs. 6A,B). We did not exclude any population from the analysis since there was no consistent pattern of deviation from HWE at multiple loci. We ran five replicates, each using a burn-in length of 100,000 and a run length of 1,000,000 steps, with the admixture and the correlated allele frequencies models since some of the populations are likely admixed and have shared allele frequencies, without using prior population information (geographic sampling location).

Because after multiple runs we observed that STRUCTURE could identify the strongest signals of genetic partition (among very divergent genetic groups) but could not detect lower levels of differentiation, we subdivided the dataset to examine substructure within each of the inferred clusters. For each of these substructure runs, we excluded loci that had missing data or deviated from Hardy-Weinberg equilibrium for those populations (Fig. 6B). Each run was done using 500,000 iterations following a 100,000 burn-in period, with 3 replicates, and with the admixture and correlated allele frequencies models.

The initial number of clusters (K) to be tested for each run is given below the plots (Fig. 6A,B). To determine the best K explained by the data for all runs we examined both the posterior probabilities of the data for each K and the ΔK estimator described by Evanno et al. (2005). Results from replicates for the inferred K from each run were analyzed in the program CLUMPP (Jakobsson and Rosenberg 2007) to produce averaged matrices of individual and population cluster membership coefficients. Finally, we used the program distruct v1.1 (Rosenberg 2004) to produce graphical displays of the resulting barplots. The program BAPS (Corander and Marttinen 2006) was used as an alternative Bayesian clustering approach to compare and check for consistency with the results from STRUCTURE.

Results

Basic population genetics statistics

Out of the 10,907 genotypes we attempted to obtain (13 loci x 839 samples), 1,829 failed to amplify consistently in certain populations (16.8 % of dataset) with only 136 missing genotypes (1.2%) arising from random amplification problems (*i.e*. not due to the presence of null alleles). The full dataset of 13 loci was used only for estimation of null alleles and the NJ tree, with other analyses done with reduced datasets, excluding loci deviating from HWE and missing data, as indicated in each case.

Tests for population differentiation showed that six sets of samples, each collected from neighboring locations, had non-significant population differentiation ($P > 0.05$ after Bonferroni correction) and were pooled together resulting in a total of 41 populations to be further analyzed (Table 1).

We found high allelic richness in all microsatellites (Table 3), with 10 to 42 alleles per locus resulting in a total of 271 alleles across the 41 populations. When patterns of deviation were examined at the locus level, three loci deviated significantly from HWE in more than one population (Table 3). No significant linkage disequilibrium was detected between loci.

We found different patterns of genetic diversity among biotypes and geographic groups. Fig. 2 shows the number of alleles after rarefaction averaged across seven loci for each of the different biotypes or geographic groups. Populations from the New World (including biotype A) and biotype S had the lowest number of alleles, while the invasive biotypes B and Q (Eastern and Western Mediterranean) had an intermediate number of alleles. The highest number of alleles was observed in the Yemen population (closest relative of biotype B) and in populations from Sub-Saharan West and N. and W. Africa (closest relatives of biotype Q). Estimates of heterozygosity showed similar patterns as allelic richness: the introduced to Spain S biotype and New World populations had the lowest heterozygosity; the invasive biotypes B and Q had moderate to high heterozygosity, whilst the highest levels were observed in their closest relatives (Yemen and N & W Africa), and biotype Ms (Fig. 3).

Null alleles

Although analysis in the program FreeNa showed that null alleles were present in our dataset, the estimates obtained from the corrected and non-corrected dataset were very similar. In the case of $F_{ST,}$ the global estimate across all loci and populations was 0.54 (95% CI: 0.46-0.62) before correction for null alleles and 0.53 (0.45-0.61) after correction. Per locus estimates of F_{ST} with and without the correction were also very similar, with only six out of 13 loci having slightly overestimated values before the correction and the largest difference between corrected - uncorrected dataset at a locus being 0.03 (locus WF1B06). The NJ trees built based on D_C calculated with both the corrected and uncorrected for null alleles dataset gave identical topologies (not shown), that were consistent with the groupings recovered from the individual-based NJ tree (Fig. 4).

Individual-based analyses

Neighbor Joining tree

The NJ tree based on individuals gave a clear picture of genetic structure associated with biotype or geographic origin with the exception of introduced and invasive populations (Fig. 4). A total of 18 groups (including seven known biotypes) showed clear genetic structure, with individuals within each group being more related to each other than to individuals from any other group. Within biotype Q there was a clear split into populations originating from the eastern part of the Mediterranean (including the described Israel-Q) and those sampled from the western part of the Mediterranean (including the described Spain-Q). The Eastern Mediterranean Q seems more structured than West Mediterranean Q and consists of well differentiated populations (Cyprus, Greece, Israel, and Turkey), whereas there seems to be more gene flow within the Western Mediterranean group (Spain, Morocco, France, Canary Islands, and an invasive population from China). Biotype B formed a monophyletic group with evidently high migration among its five populations with the exception of the Arizona B population that seemed well differentiated from the rest. Other known biotypes that formed monophyletic groups across the tree were biotype S (from Spain), biotype Ms, biotype T, and the *Jatropha* biotype, whereas populations of biotype A (Arizona A and Riverside A) did not seem to be differentiated from other New World populations from C. America. There were also cases of individuals not belonging to known biotypes but that nevertheless formed clear and well differentiated groups: samples from Burkina Faso (in "Biotype Q relatives N & W Africa" clade), Sudan and Sudan Q-like (in "Biotype Q-like (Sudan)" clade), Moorea-FP (sister clade to "Sub-Saharan West Africa" clade), Yemen (B relative clade), as well as samples from China-Hainan, India, Pakistan, and the Uganda-sweetpotato population. Individuals from Mozambique, South Africa, and Uganda-cassava (the last two sampled from cassava) were not well differentiated from each other but formed one large clade with biotype S nested within it, which we call Sub-Saharan East Africa. There was a similar pattern with individuals from Cameroon and Ivory Coast, which we have called Sub-Saharan West Africa. When we examined patterns within populations, we observed high diversity among individuals within populations of biotype B despite the lack of structure among populations. The exact opposite pattern was observed in the *Jatropha* biotype, with the 28 individuals we examined forming a distinct clade but being almost genetically identical to each other. Repeating the analysis using an 8-loci dataset gave a tree with much less resolution within biotypes B and Q and between those and their closest relatives from Yemen and N. and W. Africa respectively (since those excluded five loci amplified mostly in these populations). Otherwise, the analysis with eight loci did not affect the overall tree topology and resulted in the same groupings in other biotypes and geographic regions (results not shown).

Principal Coordinates Analysis

The 8-loci dataset we used for the PCA had very few missing data (out of 5,696 possible genotypes only 82 or 1.4% failed to amplify). The analysis showed that the first three components explained a cumulative 72.3% of the total variance (Fig. 5) of the data. Overall the PCA analysis gave similar major patterns as the individual-based NJ tree (Fig. 4). There were 7 clear clusters that corresponded to individuals from the New

World, Asia, Sub-Saharan East Africa and biotype S, biotype T, biotype Ms, biotype B, and biotype Q. In biotype Q there was a split with some overlap between the Eastern and Western Mediterranean at the level of the second PCA axis, which was visible when the axes were rotated. The Yemen individuals were loosely clustered between biotypes B and Ms. A similar pattern was observed for individuals from the Q biotype-related populations from W. and N. Africa (Burkina Faso, Sudan) and Sudan-Q like which were scattered between African populations and biotype Q. Likewise, individuals from Cameroon and Ivory Coast (Sub-Saharan West Africa) did not form a cluster, but were scattered and mostly overlapped with the Q-relatives from W. and N. Africa. Overall, this analysis showed that the most differentiated cluster with the largest genetic distance from all others at all levels of the PCA was the New World cluster. Biotypes B, Q and the Asian cluster were also very distinct, occupying the furthermost positions of the PCA axes.

Bayesian clustering analysis to assess worldwide population structure

The initial results from the clustering analysis with all individuals in STRUCTURE revealed the presence of 9 genetic groups that in general corresponded to biotype designations (Fig. 6A). However some populations such as India, T biotype and Uganda - sweet potato had mixed estimated membership coefficients and unclear assignments into clusters between replicate runs. In addition, in the multiple runs we attempted, the correct value of K varied, with the Evanno et al. (2005) procedure indicating the presence of multiple peaks in ΔK at K=4, 9, 11 and 13. Here we chose to present K=9 because when we plotted the posterior probabilities of the data [Pr(K)] against K, Pr(K) seemed to plateau at K=9.

After subdividing the dataset to biotype or geographic group level (based on results from the NJ tree and *a priori* knowledge of population affiliations with biotypes), the genetic structure was clearer and it was possible to identify the correct number of clusters K by examining both the posterior probabilities of the data against K and the ΔK estimator. Results from these sub-structure runs revealed some interesting patterns (Fig. 6B). In the Western Mediterranean Q biotype split in four clusters, consisting of China, France, Morocco and Spain, and Canary Islands, with the last two clusters seemingly sharing migrants. France and the population introduced in China were well differentiated, while individuals from Canary Islands, Morocco and Spain did seem to share a fraction of their genotypes with France and China. A different picture was observed in the Eastern Mediterranean Q biotype plot, with clear genetic structure among all four populations. Between the two Q biotype groups (Eastern and Western) there seems to be very little gene flow (Fig. 6A), with only the Cypriot and Greek populations from Eastern Mediterranean sharing ancestry with the Western Mediterranean cluster. There was a similar pattern in the Asian populations and their relative T biotype from Italy, with all populations well differentiated and no evidence of gene flow among them. In the analysis of African populations, we excluded Mozambique (which in other analyses clustered with Sub-Saharan East African populations) in order to use as many loci as possible since this population deviated from HWE in an additional two loci. Biotype S, S. Africa and Uganda (cassava) formed a single cluster, populations from B. Faso and Sudan another, while Cameroon and Ivory Coast grouped together as a separate cluster. The Ugandan sweet potato and Sudan Q-like populations were genetically

distinct from all other clusters. Despite the well-defined structure in this plot, some admixture and migration was detected between the Western and Eastern African clusters. Genetic structure was also observed within the New World, with the *Jatropha* population forming a distinct cluster and a clear differentiation between the Arizona A and the Riverside A biotype, with the first being genetically similar to the Guatemalan population and the second with the Mexican populations. We also observed moderate structure in the B biotype analysis. The Arizona B population was genetically different from the Egypt, Greece-1 and Cyprus-2006 populations, which all formed a single cluster, with only fractions of individuals' genotypes sharing ancestry with Arizona B. Populations from Israel and Panama were admixed between the Arizona B cluster and the cluster consisting of Egypt, Greece, and Cyprus. Finally, the population from Yemen was genetically distinct from all B biotype populations.

Analysis using BAPS indentified 28 genetic clusters worldwide, which were generally in agreement with the STRUCTURE results when examined at the biotype/ geographic region level (results now shown). Because in BAPS we included all populations, this analysis also identified as genetically distinct the samples from Reunion Ms, Reunion Ms-2009, Cyprus-Ork, Moorea-FP, and Australia, which were excluded from the sub-structure runs in STRUCTURE because they did not belong to a larger geographic group.

Discussion

Genetic diversity and population differentiation at the worldwide level

Estimates of genetic diversity, measured as allelic richness and heterozygosity, showed variable patterns in different biotypes. The highest levels of allelic richness and heterozygosity were found in the populations from Sudan & Burkina Faso and Yemen (Figs. 2,3) that are the closest relatives of biotypes B and Q respectively, as shown in previous studies based on mtCOI (Brown, unpublished; Frochlich et al 1999). High levels of heterozygosity were also observed in biotype Ms from the Reunion island, also found to be a relative of biotype B in this study (Figs. 4, 5) and sister clade of biotypes B and Q in the mtCOI phylogeny (Delatte et al. 2005). Overall these results show that populations from Africa and the Middle East have the highest microsatellite diversity within and among populations, suggesting that they could represent older lineages that gave rise to the more recent biotypes, at least to biotypes B and Q. This finding agrees with the high divergence of mtCOI haplotypes (~16-26%) found in the African clade, pinpointing the origins and diversification of *B. tabaci* in this continent (Boykin et al. 2007; De la Rua et al. 2006; Legg et al. 2002; Sseruwagi et al. 2005; Sseruwagi et al. 2006).

Levels of allelic richness and heterozygosity in populations we examined may be influenced by two additional factors: the effects of inbreeding in samples obtained from lab colonies and the fact that microsatellite loci can be more variable in biotypes from which they were isolated (Ellegren et al. 1995; Hutter et al. 1998). In the first case we did find some trends for low diversity in some samples that originated from lab colonies (biotype A from Riverside and Arizona, biotype S from Spain); however field collected populations from other New World locations (Mexico) had similar estimates as biotype A

that originated from a lab colony. Furthermore, samples from other lab colonies (Arizona B, Israel Q, Italy T) had moderate to high allelic richness and heterozygosity, similar to field populations from the same region or biotype. Thus, although laboratory rearing of some populations may play a role in the estimates we obtained, it is likely that demographic histories of these populations (such as ancestral population bottlenecks) account mostly for the pattern we observed. For example, the low genetic diversity in New World populations (lab or field collected) is consistent with data from the mtCOI phylogeny which shows strikingly less within-clade divergence (~8%) in New World compared to Old World clades (14-26%) (Brown 2007; 2010). The species-specific variability of microsatellites is compensated by the fact that we used loci from 4 different genetic groups/biotypes (Asia, Australia, biotype B, biotype Q), thereby minimizing this effect in our results and conclusions.

The microsatellite analysis of *B. tabaci* populations revealed large genetic distances at the worldwide level and suggests that this taxon consists of very divergent cryptic lineages that represent old entities (see Results and Figs 4, 5, 6). These findings are in line with previous studies of cryptic species that suggested divergence dating back to millions of years ago, despite morphological conservatism (Beheregaray and Caccone 2007; Bickford et al. 2007; Elmer et al. 2007). The results based on nuclear microsatellites show that these lineages correspond to described and well-studied "biotypes" that have been characterized on the basis of allozyme differences, phylogenetic analysis of the mtCOI gene, and biological/ ecological assays such as host-feeding, virus transmission and crossing experiments (Boykin et al. 2007; Brown 2010; Brown et al. 1995; Dinsdale et al. 2010; Frohlich et al. 1999; Perring 2001). Furthermore, other distinct genetic groups identified by our analysis have a geographic basis, with the exception of invasive and introduced populations, and are in general agreement with the mtCOI phylogeny of the species (see Table 1 for population affiliations with mtCO1 clades). The PCA showed that the most divergent lineage of those we examined is that consisting of New World populations, followed by biotypes B, Q, and the Asian populations, with no evidence for gene flow among them (Fig. 5). Although these results suggest older divergence among these groups compared to any others, divergence date estimates would require molecular dating analysis of phylogenetic clades since microsatellites are unsuitable for inferring deep phylogenetic relationships.

Another indication of the extreme divergence in *B. tabaci* emerging from our results is the fact that out of 13 loci we used, seven failed to amplify across some worldwide populations. Non-amplification of microsatellite loci during PCR is caused by poor primer specificity due to mutations in regions flanking the microsatellite repeat sequence, resulting to what is known as "null alleles" (Callen et al. 1993; Pemberton et al. 1995). Although null alleles can occur at a low frequency at the species level, this frequency increases with increasing phylogenetic distance at the genus level (Chapuis and Estoup 2007). This in fact, is in line with our results as microsatellite loci that failed to amplify did so systematically within a biotype or geographic group, consistent with the presumed genetic relationships among populations, with non-amplifying loci occurring mostly in individuals of the most divergent lineages (*e.g.* New World biotypes) (Fig 5). Despite the occurrence of null alleles in our data, genetic differentiation estimates (F_{ST}) using corrected and uncorrected datasets were very similar. In addition, NJ trees

constructed based on D_c corrected for null alleles gave the same tree topologies as uncorrected trees, suggesting that null alleles had minimal impact in our analyses and did not affect our overall conclusions.

The invasive biotypes (B and Q)

The most commonly known biotypes in the *B. tabaci* complex are the invasive biotypes B and Q. Biotype B, once described as *B. argentifolii* (Perring et al. 1993), has expanded its range to a worldwide scale in the past 30 years, while the Q biotype has only become invasive in the last 6 years in Asia (Ueda and Brown 2006; Zhang et al. 2005), America (Dalton 2006; Martinez-Carrillo and Brown 2007), and New Zealand (Drayton et al. 2009; Scott et al. 2007). Many other biotypes and genetically distinct populations around the world, such as the cassava and sweet potato populations found in Africa (Legg 1996; Sseruwagi et al. 2005) are highly devastating local pests and plant virus vectors but they have not yet earned the status of a worldwide invasive pest.

What has favored the worldwide expansion and extreme adaptation of biotypes B and Q has yet to be determined; but probably the direction of global trade of infested plant products has facilitated these invasions. From this and previous studies (Boykin et al. 2007), it does not seem that other non-B or Q populations have expanded their range worldwide and become invasive. It is likely therefore that inherent characteristics of these two biotypes B and Q have contributed to their successful invasions and displacement of local biotypes such as the A biotype displacement by B in the US (Brown et al. 1995). For example both B and Q biotypes are known to be highly resistant to insecticides (Horowitz et al. 2005). Invasion of these biotypes to new areas where local biotypes are still susceptible and manageable with certain chemicals gives the invaders a competitive advantage and with time their populations built up, driving local biotypes to extinction if they cannot interbreed with them.

In our study we found that the invasive biotypes B and Q did not have lower genetic diversity compared to all other biotypes (Figs. 2 and 3); the expected pattern in invasive populations under bottlenecks from multiple founder effects and directional selection from intensive agriculture (e.g. insecticide applications). In fact, despite the relatively low structure that we observed in biotype B and the Western Mediterranean biotype Q clades (Fig. 4), we found substantial variation among individuals, especially in Q, with all multilocus genotypes being unique. The moderate diversity in biotype B, indicates a large effective population size and an ancestral lineage, and suggests that the presence of ancestral variation likely resists the homogenizing effects of human-mediated gene flow among populations. This is the opposite of what we observed for example in the monophagous *Jatropha* biotype where almost all individuals were identical (Fig 4), possibly a result of its restricted range on this host. It seems that despite the genetic bottlenecks induced by founder effects in these invasive biotypes, genetic diversity remains high, possibly either due to persistent ancestral variation or due to migrations from multiple diverse source populations from different regions. This latter phenomenon has been observed in other organisms such as invasive lizards (Kolbe et al. 2004) and in the case of *B. tabaci* it would be facilitated by human-mediated transportation of infested plants around the globe from multiple source populations.

We also observed some interesting patterns of population differentiation within the two invasive biotypes B and Q. Within biotype Q there was a large split into Western

and Eastern Mediterranean populations. The two lineages correspond to what are known as Spain-Q (Western) and Israel-Q (Eastern) based on mtCOI haplotypes (Horowitz et al. 2003; Rosell et al. 1997), and while they are both described as biotype Q, they seem to represent distinct and divergent lineages with minimal gene flow between them. It will be interesting to determine which of the two lineages has become more widespread outside their native Mediterranean range. In this study the invasive population from China we examined was assigned to the Spanish Q lineage but preliminary results from another study suggest that at least the US invasion of biotype Q occurred from both the Spain and the Israel Q lineages (Chapter 4). Whether the two lineages have biological and ecological differences remains to be studied; however it seems that they both possess high inherent resistance to insecticides (Horowitz et al. 2003; Rauch and Nauen 2003).

Within biotype B there was high gene flow among populations with only Arizona B being well differentiated from others. Arizona B originated from a lab colony in Arizona (JK Brown laboratory), which was founded from the first occurrence of biotype B in the US in the 1990's and was the reference population for the esterase-based characterization of biotype B (Costa and Brown 1991). Thus, the differentiation we see from the other B populations may simply be due to temporal changes in allele frequencies since 1990; however we did not see much differentiation in populations from Egypt and Cyprus, which were sampled with a gap of five years. The founding effect in the Arizona B colony may be associated with this differentiation, as only a subset of the introduced population was selected and bred for multiple generations. The shared ancestry of Arizona B with the population from Israel suggests that the B biotype introductions in the USA may have originated from this area of the Mediterranean.

The exact origins of the two invasive biotypes have yet to be determined, but it has been suggested that the Q biotype originated around the North/West Africa or the Mediterranean while the B biotype diversified in the eastern African Sahel region or the Middle East (Brown 2010). Consistent with these hypotheses are our findings of genetic similarities of biotype Q with N. and W. African populations and of biotype B with the Yemeni population. Understanding where these biotypes originated from and when they diverged from their closest relatives may be critical in identifying attributes that facilitated their invasiveness and their high pest status. In a separate study, we examine populations from both biotypes and their relatives to determine whether the timing of divergence and their evolution is associated with human activities, movement, and trade of agricultural products (Chapter 3).

Other notable biotypes

The *B. tabaci* complex represents an excellent system for studies of cryptic speciation where evolution has favored a wide array of genetically, ecologically, and biologically diverse lineages around the world. While most whitefly scientists are interested in applied research with most studies focusing on *B. tabaci* as an agricultural pest, we argue that there is a lot to be learned and to contribute to evolutionary theory by studying some of the less known but ecologically and biologically interesting biotypes. This knowledge may in turn prove useful to understanding any differences between invasive and non-invasive biotypes, and how these have shaped their evolution and adaptation. We discuss some interesting non-invasive biotypes here with our findings and anticipate that some of these will be the subjects of future extensive studies.

Biotype T was first identified in Italy in 2003 developing only on *Euphorbia characias* in a high altitude area (Bosco et al. 2006; Demichelis et al. 2005; Simon et al. 2003). The population we examined from Puglia, Italy from this biotype appears to be a distinct *B. tabaci* lineage, but genetically related to the Asian populations, which agrees with existing literature based on mtCOI (Boykin et al. 2007; De la Rua et al. 2006). Biotype T likely represents an ancestral introduction of populations into the Mediterranean from Asia, or remnant populations of a wider historical distribution of Asian lineages whose range later contracted. Another finding supporting these two hypotheses is that the *B. tabaci* first sampled and described by Gennadius in Greece in 1889 (DNA obtained from pupal specimens from the paratype collection at the British Museum), also belongs to the Asian clade of the mtCOI phylogeny (Gill and Brown 2010). An extensive sampling of high altitude areas in the Mediterranean, away from agricultural areas where biotypes B and Q are likely to be found, could provide more information to test these hypotheses. Indeed in one of our sampling efforts in the island of Cyprus in 2008, we found 2 individuals in a mountainous area far from agricultural fields, which formed a clade sister to the Asian clade which includes the T biotype, while in a mtCOI tree they form a clade sister to the T biotype clade, with ~8% divergence from the T biotype haplotype (unpublished results). Further collecting in such areas to obtain individuals for biological and ecological assays would determine whether the population we sampled these whiteflies from represents a distinct biotype, relative to the Asian populations, like biotype T. Likewise, larger sampling efforts from non-agricultural areas from the whole Mediterranean basin, especially Greece, would help elucidate the origins of these Asia-related biotypes, including the *B. tabaci* population first described by Gennadius in Greece. With appropriate genetic analyses of such samples we would also be able to date their divergence from Asian relatives and determine whether they represent an introduction or remnants of an ancestral widespread distribution in the Mediterranean.

Biotype S which was first described from the weed *Ipomoea indica* in Spain in 1995 (Banks et al. 1999) is considered to be a relative of African populations that was introduced in Spain but did not become widespread in the region (Cervera et al. 2000). Our microsatellite analysis showed that this population forms a clade nested within the Sub-Saharan East African clade which includes Mozambique and cassava-associated populations from Uganda, and South Africa, consistent with previous studies based on mtCOI, (De Barro et al. 2005, De la Rua et al. 2006), ITS I sequences and RAPD analysis (Abdullahi et al. 2003), as well as AFLP analysis (Cervera et al. 2000). Genetic diversity estimates showed that this population had lower allelic richness compared to other African populations, which suggests that a subset of the source population was introduced and survived in the sampled area, likely undergoing a genetic bottleneck. Although this biotype had only been found in Spain until now, we found five individuals that seem to belong to this biotype (after we sequenced and BLASTed part of the mtCOI gene, unpublished results) in a sample of 30 individuals from the French Q biotype population we examined. What is more interesting is that the microsatellite genotypes from these individuals were not clearly assigned to the biotype S population, but seemed to have admixed ancestry between the Q and S biotypes (unpublished results). This suggests sex-biased admixture, with females from the S biotype retaining their maternal mtDNA but exchanging genes with males from the Q biotype, perhaps induced by the

maternally inherited *Wolbachia* endosymbionts, known to cause uni-directional gene flow between infected and non-infected populations (Hoffmann and Turelli 1997; Stouthamer et al. 1999; Werren 1997). The S biotype therefore seems to occur in small populations in the Western Mediterranean, and in some cases in co-existence with the Q biotype. A question that arises is why has this biotype not become invasive after its introduction in the area, like biotype B? Likely, the competition with biotype Q and/or lower levels of insecticide resistance does not allow its populations to built-up and expand. It is also possible that like its cassava-restricted relatives in Africa (Abdullahi et al. 2003), biotype S cannot easily adapt to other hosts and become an invasive pest in the region. The distribution of the S biotype in the Mediterranean and its ecological and genetic interactions with the other prevalent biotypes in the area makes a very interesting case to be studied extensively in the future.

Whitefly population structure inferred using nuclear microsatellites *vs* mitochondrial DNA

Although quantitative comparisons between the results from the microsatellite and the mitochondrial DNA cannot be made, we can evaluate some general patterns that have emerged from these different markers. Overall, there was agreement in the major lineages/clades/biotypes identified from our microsatellite analysis with those obtained from mitochondrial DNA sequences in several studies in the literature (Boykin et al. 2007; Dinsdale et al. 2010) and from our unpublished data (Table 1, Fig. 4). The relationships between populations in different continents (e.g. Americas, Asia), as well as host-associated structure (e.g. African cassava – associated populations) corroborates the general picture we had of worldwide structure of *B. tabaci* populations.

We did however observe some discrepancies that point out that caution should be taken when using mitochondrial DNA as a single marker because it may not provide an accurate picture of the history of this species complex. For example, in this microsatellite analysis we found that the population from France we examined clearly clusters with the Western Mediterranean Q biotype (Spain Q), with evident gene flow with the Spanish population. When we examined the mtDNA of these populations however (Chapter 4) the French population had the mitochondrial haplotype of Eastern Mediterranean (same as Cyprus, Turkey, Israel). This implies female-biased admixture, between Eastern Mediterranean females and Western Mediterranean males with offspring retaining their maternal mtDNA, but with gene flow evident in their nuclear DNA. While it is possible that the mtDNA of Eastern Mediterranean has a competitive advantage over the Western Mediterranean, the first hypothesis sounds more plausible as it can be easily explained with the involvement of *Wolbachia* infections. The important point is that a mtDNA phylogeny would erroneously assign France to the Eastern Mediterranean clade, ignoring the background nuclear gene flow and would provide a misleading picture of the species and population histories.

Similar patterns were observed in other populations we examined (from the Greek Q biotype and S biotype) indicating that this was not an isolated case. In fact the high prevalence of *Wolbachia* infections in *B. tabaci* (Nirgianaki et al. 2003; Zchori-Fein and Brown 2002) suggests that this may be a fairly common phenomenon in this insect. Furthermore, the influence of endosymbionts might be even more profound if populations are infected with different bacterial strains. Several lines of evidence suggest that

maternally inherited endosymbionts can have huge effects on mtDNA evolution, by either decreasing inter-population genetic diversity (if there is a selective sweep induced by a symbiont through a population, with the mtDNA hitchhiking along) or by increasing diversity following fixation of different symbiont strains in different populations (which would show high haplotype differentiation despite ongoing nuclear gene flow) (Hurst and Jiggins 2005). It is therefore possible that the enormous diversity that we see in the *B. tabaci* mtCOI, with some populations with as much as 8% divergence interbreeding and producing offspring (Legg et al. 2002; Maruthi et al. 2001), is largely driven by variable endosymbiont infections throughout the species history.

This study has demonstrated that in spite of their cryptic nature, biotypes in *B. tabaci* likely represent very old lineages, perhaps isolated for millions of years as has been suggested for other cryptic taxa. Our analysis using nuclear microsatellite markers from multiple worldwide populations belonging to several biotypes has demonstrated extreme levels of genetic differentiation. Although the pattern of genetic/geographic structure is generally consistent to that obtained with the mtCOI phylogeny, some cases of strong conflict between the two markers were observed. In light of these results, we argue that care should be taken when using mtDNA as a single marker for phylogenetic reconstruction in *B. tabaci*, since it only represents a single evolutionary aspect and may not accurately reflect the true species phylogeny. Results from multiple genetic markers from different parts of the genome can help us corroborate biologically based differences in populations and provide information about the relative time of divergence among biotypes. However, a multifaceted approach complementing biological/ ecological/ behavioral assays, crossing experiments, along with genetic information should be used in order to understand and determine species boundaries in a complex system like *Bemisia tabaci*.

References

Abdullahi I, Winter S, Atiri GI, Thottappilly G. 2003. Molecular characterization of whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae) populations infesting cassava. *Bull. Entomol. Res.* 93: 97-106.

Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS. 2006. Automated binning of microsatellite alleles: problems and solutions. *Mol. Ecol. Notes* 7: 10-14.

Banks GK, Bedford ID, Beitia FJ, Rodriguez-Cerezo E, Markham PG. 1999. A novel geminivirus of *Ipomea indica* (Convolvulaceae) from southern Spain. *Plant Dis.* 83: 486.

Barluenga M, Stolting KN, Salzburger W, Muschick M, Meyer A. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* 439: 719-723.

Beheregaray LB, Caccone A. 2007. Cryptic biodiversity in a changing world. *J.Biol.* 6: 9.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I. 2007. Cryptic species as a window on diversity and conservation. *Trends Ecol. Evol.* 22: 148-155.

Bosco D, Loria A, Sartor C, Cenis JL. 2006. PCR-RFLP identification of *Bemisia tabaci* biotypes in the Mediterranean basin. *Phytoparasitica* 34: 243-251.

Boykin LM, Shatters RG, Rosell RC, McKenzie CL, Bagnall RA, De Barro P, Frohlich DR. 2007. Global relationships of *Bemisia tabaci* (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Mol. Phylogenet. Evol.* 44: 1306-1319.

Brown DM, Brenneman RA, Koepfli KP, Pollinger JP, Mila B, Georgiadis NJ, Louis Jr EE, Grether GF, Jacobs DK, Wayne RK. 2007. Extensive population genetic structure in the giraffe. *BMC Biology* 5: 57.

Brown JK. 2007. The *Bemisia tabaci* complex: genetic and phenotypic variability drives begomovirus spread and virus diversification. Plant Disease Feature. http://www.apsnet.org/online/feature/btabaci Accessed: July 15, 2010.

Brown JK. 2010. Phylogenetic biology of the *Bemisia tabaci* sibling species group. In *Bemisia: Bionomics and Management of a Global Pest*, ed. PA Stansly, SE Naranjo. Dordrecht: Springer Science. In press.

Brown JK, Frohlich DR, Rosell RC. 1995. The sweetpotato or silverleaf whiteflies: Biotypes of *Bemisia tabaci* or a species complex? *Annu. Rev. Entomol.* 40: 511-534.

Brown JK, Perring TM, Cooper AD, Bedford ID, Markham PG. 2000. Genetic analysis of *Bemisia* (Hemiptera: Aleyrodidae) populations by isoelectric focusing electrophoresis. *Biochem. Genet.* 38:13-25.

Bush GL. 1969. Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera: Tephritidae). *Evolution* 23:237-251.

Bush GL, Butlin RK. 2004. Sympatric speciation in insects. In *Adaptive Speciation, ed.* U Dieckmann, M Doebeli, JAJ Metz, D. Tautz, pp. 229-248. Cambridge: Cambridge University Press.

Byrne FJ, Devonshire AL. 1996. Biochemical evidence of haplodiploidy in the whitefly *Bemisia tabaci*. *Biochem. Genet.* 34: 93-107.

Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, Sutherland GR. 1993. Incidence and origin of null alleles in the (AC)n microsatellite markers. *Am. J. Hum. Genet.* 52: 922-927.

Cavalli-Sforza LL, Edwards AWF. 1967. Phylogenetic analysis: models and estimation procedures. *Evolution* 32: 550–570.

Cervera MT, Cabezas JA, Simon B, Martinez-Zapater JM, Beitia F, Cenis JL. 2000. Genetic relationships among biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae) based on AFLP analysis. *Bull. Entomol. Res.* 90: 391-396.

Chapuis MP, Estoup A. 2007. Microsatellite null alleles and estimation of population differentiation. *Mol. Biol. Evol.* 24: 621-631.

Chapuis MP, Lecoq M, Michalakis Y, Loiseau A, Sword GA, Piry S, Estoup A. 2008. Do outbreaks affect genetic population structure? A worldwide survey in *Locusta migratoria*, a pest plagued by microsatellite null alleles. *Mol. Ecol.* 17: 3640-3653.

Charlesworth B, Lande R, Slatkin M. 1982. A neon-Darwinian commentary on macroevolution. *Evolution* 36: 474-498.

Corander J, Marttinen P. 2006. Bayesian identification of admixture events using multilocus molecular markers. *Mol. Ecol.* 15: 2833-2843.

Costa HS, Brown JK. 1991. Variation in biological characteristics and esterase patterns among populations of *Bemisia tabaci*, and the association of one population with silverleaf symptom induction. *Entomol. Exp. Appl.* 61: 211-219.

Coyne J, Orr H. 2004. *Speciation*. Sunderland, MA: Sinauer Associates, Inc.

Dalmon A, Halkett F, Granier M, Delatte H, Peterschmitt M. 2008. Genetic structure of the invasive pest *Bemisia tabaci*: evidence of limited but persistent genetic differentiation in glasshouse populations. *Heredity* 100: 316-325.

Dalton R. 2006. The Christmas invasion. *Nature* 443: 898-900.

De Barro PJ, Driver F, Trueman JWH, Curran J. 2000. Phylogenetic relationships of world populations of *Bemisia tabaci* (Gennadius) using ribosomal ITS1. *Mol. Phylogenet. Evol.* 16: 29-36.

De Barro PJ, Hidayat SH, Frohlich D, Subandiyah S, Ueda S. 2008. A virus and its vector, pepper yellow leaf curl virus and *Bemisia tabaci*, two new invaders of Indonesia. *Biol. Invasions* 10: 411-433.

De Barro PJ, Scott KD, Graham GC, Lange CL, Schutze MK. 2003. Isolation and characterization of microsatellite loci in *Bemisia tabaci*. *Mol. Ecol. Notes* 3: 40-43.

De Barro PJ, Trueman JWH, Frohlich DR. 2005. *Bemisia argentifolii* is a race of *B. tabaci* (Hemiptera: Aleyrodidae): the molecular genetic differentiation of *B. tabaci* populations around the world. *Bull. Entomol. Res.* 95: 193-203.

De la Rua P, Simon B, Cifuentes D, Martinez-Mora C, Cenis JL. 2006. New insights into the mitochondrial phylogeny of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in the Mediterranean Basin. *J. Zool. Syst. Evol. Res.* 44: 25-33.

Delatte H, David P, Granier M, Lett JM, Goldbach R, Peterschmitt M, Reynaud B. 2006. Microsatellites reveal extensive geographical, ecological and genetic contacts between invasive and indigenous whitefly biotypes in an insular environment. *Genet. Res.* 87: 109-124.

Delatte H, Reynaud B, Granier M, Thornary L, Lett JM, Goldbach R, Peterschmitt M. 2005. A new silverleaf-inducing biotype Ms of *Bemisia tabaci* (Hemiptera: Aleyrodidae) indigenous to the islands of the south-west Indian Ocean. *Bull. Entomol. Res.* 95: 29-35.

Demichelis S, Arno C, Bosco D, Marian D, Caciagli P. 2005. Characterization of biotype T of *Bemisia tabaci* associated with *Euphorbia characias* in Sicily. *Phytoparasitica* 33: 196-208.

Diehl SR, Bush, GL. 1984. An evolutionary and applied perspective of insect biotypes. *Annu Rev Entomol* 29:471-504.

Dinsdale A, Cook L, Riginos C, Buckley YM, De Barro P. 2010. Refined global analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial cytochrome cxidase I to identify species level genetic boundaries. *Ann. Entomol. Soc. Am.* 103: 196-208.

Dolman G, Moritz, C. 2006. A multilocus perspective on refugial isolation and divergence in rainforest skinks (*Carlia*). *Evolution* 60: 573-582.

Drayton GM, Teulon DAJ, Workman PJ, Scott IAW. 2009. The Christmas dispersal of *Bemisia tabaci* (Gennadius) in New Zealand. *N. Z. Plant Protect.* 62: 310-314.

Dres M, Mallet J. 2002. Host races in plant-feeding insects and their importance in sympatric speciation. *Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci.* 357: 471-492.

Ellegren H, Primmer CR, Sheldon BC. 1995. Microsatellite "evolution": directionality or bias? *Nat. Genet.* 11: 360-362.

Elmer KR, Davila JA, Lougheed SC. 2007. Cryptic diversity and deep divergence in an upper Amazonian leaflitter frog, *Eleutherodactylus ockendeni*. *BMC Evol. Biol.* 7: 247.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164: 1567-1587.

Falush D, Stephens M, Pritchard JK. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* 7: 574-578.

Feder JL, Opp SB, Wlazlo B, Reynolds K, Go W, Spisak S. 1994. Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly. *Proc. Natl. Acad. Sci. USA* 91: 7990-7994.

Felsenstein J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle. http://evolution.genetics.washington.edu/phylip.html

Frohlich D, Torres-Jerez I, Bedford I, Markham P, Brown J. 1999. A phylogeographic analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers. *Mol. Ecol.* 8: 1593-1602.

Funk DJ, Filchak KE, Feder JL. 2002. Herbivorous insects: model systems for the comparative study of speciation ecology. *Genetica* 116: 251-267.

Galtier N, Nabholz B, Glemin S, Hurst GDD. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol. Ecol.* 18: 4541-4550.

Gauthier N, Dalleau-Clouet C, Bouvret ME. 2008. Twelve new polymorphic microsatellite loci and PCR multiplexing in the whitefly, *Bemisia tabaci*. *Mol. Ecol. Res.* 8: 1004-1007.

Gill RJ, Brown JK. 2010. Systematics of *Bemisia* and *Bemisia* relatives: can molecular techniques solve the *Bemisia tabaci* complex conundrum – a taxonomist's viewpoint. In *Bemisia: Bionomics and Management of a Global Pest*, ed. PA Stansly, SE Naranjo. Dordrecht: Springer Science. In press.

Hoffmann AA, Turelli M. 1997. Cytoplasmic incompatibility in insects. In *Influential Passengers: Inherited Microorganisms and Arthropod Reproduction*, ed. SL O'Neill, AA Hoffman, JH Werren, pp. 42-80. Oxford: Oxford University Press.

Horowitz AR, Denholm I, Gorman K, Cenis JL, Kontsedalov S, Ishaaya I. 2003. Biotype Q of *Bemisia tabaci* identified in Israel. *Phytoparasitica* 31: 94-98.

Horowitz AR, Kontsedalov S, Khasdan V, Ishaaya I. 2005. Biotypes B and Q of *Bemisia tabaci* and their relevance to neonicotinoid and pyriproxyfen resistance. *Arch. Insect Biochem. Physiol.* 58: 216-225.

Hubisz MJ, Falush D, Stephens M, Pritchard JK. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* 9: 1322-1332.

Hurst GDD, Jiggins FM. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc. R. Soc. Lond., Ser. B: Biol. Sci.* 272: 1525-1534.

Hutter CM, Schug MD, Aquadro CF. 1998. Microsatellite variation in *Drosophila melanogaster* and *Drosophila simulans*: A reciprocal test of the ascertainment bias hypothesis. *Mol. Biol. Evol.* 15: 1620-1636.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801-1806.

Kalinowski ST. 2005. HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. *Mol. Ecol. Notes* 5: 187-189.

Kolbe JJ, Glor RE, Schettino LRG, Lara AC, Larson A, Losos JB. 2004. Genetic variation increases during biological invasion by a Cuban lizard. *Nature* 431: 177-181.

Lande R. 1981. Models of speciation by sexual selection on polygenic traits. *Proc. Natl. Acad. Sci. USA* 78: 3721-3725.

Lande R, Kirkpatrick M. 1988. Ecological speciation by sexual selection. *J. Theor. Biol.* 133: 85-98.

Langella O. 1999. Populations 1.2.30: Population genetic software (individuals or populations distances, phylogenetic trees). http://bioinformatics.org/~tryphon/populations/

Legg JP. 1996. Host-associated strains within Ugandan populations of the whitefly *Bemisia tabaci* (Genn) (Homoptera: Aleyrodidae). *J. Appl. Entomol.* 120: 523-527.

Legg JP, French R, Rogan D, Okao-Okuja G, Brown JK. 2002. A distinct *Bemisia tabaci* (Gennadius) (Hemiptera: Sternorrhyncha: Aleyrodidae) genotype cluster is associated with the epidemic of severe cassava mosaic virus disease in Uganda. *Mol. Ecol.* 11: 1219-1229.

Letunic I, Bork P. 2006. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23: 127-128.
Martinez-Carrillo JL, Brown JK. 2007. First report of the Q biotype of *Bemisia tabaci* in southern Sonora, Mexico. *Phytoparasitica* 35: 282-284.
Maruthi MN, Colvin J, Seal S. 2001. Mating compatibility, life-history traits, and RAPD-PCR variation in *Bemisia tabaci* associated with the cassava mosaic disease pandemic in East Africa. *Entomol. Exp. Appl.* 99: 13-23.
Mayr E. 1963. *Animal Species and Evolution*. Cambridge, MA: Harvard University Press.
Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*. 89: 583-590.
Nirgianaki A, Banks GK, Frohlich DR, Veneti Z, Braig HR, Miller TA, Bedford ID, Markham PG, Savakis C, Bourtzis K. 2003. *Wolbachia* infections of the whitefly *Bemisia tabaci*. *Curr. Microbiol.* 47: 93-101.
Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck C. 1997. An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics* 147: 1943-1957.
Peakall R, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6: 288-295.
Pemberton JM, Slate J, Bancroft DR, Barrett JA. 1995. Nonamplifying alleles at microsatellite loci - a caution for parentage and population studies. *Mol. Ecol.* 4: 249-252.
Perring TM. 2001. The *Bemisia tabaci* species complex. *Crop Prot.* 20: 725-737
Perring TM, Cooper AD, Rodriguez RJ, Farrar CA, Bellows TS, Jr. 1993. Identification of a whitefly species by genomic and behavioral studies. *Science* 259: 74-77.
Pfenninger M, Schwenk K. 2007. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evol. Biol.* 7: 121.
Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
Rauch N, Nauen R. 2003. Identification of biochemical markers linked to neonicotinoid cross resistance in *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Arch. Insect Biochem. Physiol.* 54: 165-176.
Raymond M, Rousset F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* 86: 248-249.
Rosell RC, Bedford ID, Frohlich DR, Gill RJ, Brown JK, Markham PG. 1997. Analysis of morphological variation in distinct populations of *Bemisia tabaci* (Homoptera: Aleyrodidae). *Ann. Entomol. Soc. Am.* 90: 575-589.
Rosenberg NA. 2004. DISTRUCT: a program for the graphical display of population structure. *Mol. Ecol. Notes* 4: 137-138.
Rousset F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.
Schlotterer C. 2001. Genealogical inference of closely related species based on microsatellites. *Genet. Res.* 78: 209-212.
Schluter D. 2001. Ecology and the origin of species. *Trends Ecol. Evol.* 16: 372-380
Scott, IAW, Workman PJ, Drayton GM, Burnip GM. 2007. First record of *Bemisia tabaci* biotype Q in New Zealand. *N. Z. Plant Protect.* 60: 264-270.

Simon B, Cenis JL, Demichelis S, Rapisarda C, Caciagli P, Bosco D. 2003. Survey of *Bemisia tabaci* (Hemiptera: Aleyrodidae) biotypes in Italy with the description of a new biotype (T) from *Euphorbia characias*. *Bull. Entomol. Res.* 93: 259-264.

Slatkin M. 1985. Gene flow in natural populations. *Annu. Rev. Ecol. Syst.* 16: 393-430.

Smouse PE, Peakall R. 1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* 82: 561-573.

Sseruwagi P, Legg JP, Maruthi MN, Colvin J, Rey MEC, Brown JK. 2005. Genetic diversity of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) populations and presence of the B biotype and a non-B biotype that can induce silverleaf symptoms in squash, in Uganda. *Ann. Appl. Biol.* 147: 253-265.

Sseruwagi P, Maruthi MN, Colvin J, Rey MEC, Brown JK, Legg JP. 2006. Colonization of non-cassava plant species by cassava whiteflies (*Bemisia tabaci*) in Uganda. *Entomol. Exp. Appl.* 119: 145-153.

Stouthamer R, Breeuwer JAJ, Hurst GDD. 1999. *Wolbachia pipientis*: Microbial manipulator of arthropod reproduction. *Annu. Rev. Microbiol.* 53: 71-102.

Stuart BL, Inger RF, Voris HK. 2006. High level of cryptic species diversity revealed by sympatric lineages of Southeast Asian forest frogs. *Biol. Lett.* 2: 470-474.

Tsagkarakou A, Roditakis N. 2003. Isolation and characterization of microsatellite loci in *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Mol. Ecol. Notes* 3: 196-198.

Tsagkarakou A, Tsigenopoulos CS, Gorman K, Lagnel J, Bedford ID. 2007. Biotype status and genetic polymorphism of the whitefly *Bemisia tabaci* (Hemiptera : Aleyrodidae) in Greece: mitochondrial DNA and microsatellites. *Bull. Entomol. Res.* 97: 29-40.

Turelli M, Barton NH, Coyne JA. 2001. Theory and speciation. *Trends Ecol. Evol.* 16: 330-343.

Ueda S, Brown JK. 2006. First report of the Q biotype of *Bemisia tabaci* in Japan by mitochondrial cytochrome oxidase I sequence analysis. *Phytoparasitica* 34: 405-411.

Wake DB, Roth G, Wake MH. 1983. On the problem of stasis in organismal evolution. *J. Theor. Biol.* 101: 211-224.

Weir BS. 1996. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data (2nd ed.)*. Sunderland, MA: Sinauer Associates, Inc.

Werren JH. 1997. Biology of *Wolbachia*. *Annu. Rev. Entomol.* 42: 587-609.

Zchori-Fein E, Brown JK. 2002. Diversity of prokaryotes associated with *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae). *Ann. Entomol. Soc. Am.* 95: 711-718.

Zhang LP, Zhang YJ, Zhang WJ, Wu QJ, Xu BY, Chu D. 2005. Analysis of genetic diversity among different geographical populations and determination of biotypes of *Bemisia tabaci* in China. *J. Appl. Entomol.* 129: 121-128.

Tables

Table 1. Whitefly collections with information on presumed status as native, extinct, introduced, or invasive, collection year, host together with rearing information if cultured in the laboratory, affiliation with clade of the mitochondrial COI phylogeny, and number of adult females genotyped. Population names refer to the geographic sampling location and were given biotype code (e.g. ARIZONA A) if they represent a reference collection previously characterized (Brown et al. 2000). Samples with the same number were collected from neighboring locations and were pooled together for analysis after exact tests of population differentiation showed non-significant differentiation.

Pooled pop. number	Population name	Status	Collection year	Original host and rearing information for lab cultured populations	Mitochondrial COI clade	n
1	Arizona A	Native, extinct	1997	Cotton. Lab colony on poinsettia since 1988	Americas/ Caribbean	29
2	Arizona B	Invasive		Cotton. Lab colony on poinsettia since 1987	N.Africa/ Mediterranean/M.East	36
3	Australia	Native	2000	*Euphorbia cyathophora*	Australia-Asian (outlier to Asia)	4
4	Burkina Faso 3	Native	1999	Cotton	N.Africa/ Mediterranean/M.East	12
4	Burkina Faso 4	Native	1999	Tomato	N.Africa/ Mediterranean/M.East	9
4	Burkina Faso 6	Native	2000	Tomato	N.Africa/ Mediterranean/M.East	4
5	Cameroon 2	Native	2004	Cassava	Sub-Saharan Africa	6
5	Cameroon 14	Native	2004	Cassava	Sub-Saharan Africa	6
6	CanIsl-1 (Canary Islands)	Native	2009	Tomato	N.Africa/ Mediterranean/M.East	5
6	CanIsl-3	Native	2009	Tomato	N.Africa/ Mediterranean/M.East	6
6	CanIsl-4	Native	2009	Tomato	N.Africa/ Mediterranean/M.East	11
6	CanIsl-6	Native	2009	*Solanum nigrum*	N.Africa/ Mediterranean/M.East	10
7	China	Invasive	2007	Unknown	Asia	14
8	China-Hainan	Native	1994	Cotton	Asia	28
9	Cyprus 2006	Invasive	2006	*Cucumis sativum*	N.Africa/ Mediterranean/M.East	30
10	Cyprus 2008	Undetermined	2008	Gerbera	N.Africa/ Mediterranean/M.East	30
11	Cyprus 5	Undetermined	2008	Unidentified plant	Asia	2 (♀,♂)
12	Egypt	Invasive	2001	Cucumber	N.Africa/ Mediterranean/M.East	30
13	France	Undetermined	2007	Soybean	N.Africa/ Mediterranean/M.East	25
14	Greece-1	Invasive	2006	Melon	N.Africa/ Mediterranean/M.East	9
14	Greece-3	Invasive	2006	Bean	N.Africa/ Mediterranean/M.East	4

Pooled pop. number	Population name	Status	Collection year	Original host and rearing information for lab cultured populations	Mitochondrial COI clade	n
15	Greece-2	Undetermined	2006	Tomato	N.Africa/ Mediterranean/M.East	17
16	Guatemala	Native	1994	Cotton	Americas/ Caribbean	6
17	India	Native	1995	Watermelon	Asia	30
18	Israel Q	Native	2003	Cotton. Lab colony since 1992	N.Africa/ Mediterranean/M.East	29
19	Israel-3 B	Undetermined	2007	Pepper	N.Africa/ Mediterranean/M.East	3
19	Israel-4 B	Undetermined	2007	Cotton	N.Africa/ Mediterranean/M.East	4
20	Italy T	Introduced	2007	*Euphorbia characias* (monophagous) Lab colony since 2004	Asia	29
21	Ivory Coast	Native	2001	Okra	N.Africa/ Mediterranean/M.East	9
22	Jatropha-Puerto Rico	Native	Unknown	*Jatropha* (monophagous) Lab colony since 1988	Americas/ Caribbean	28
23	Mexico	Native	1993	Wild *Solanum* spp.	Americas/ Caribbean	9
24	Mexico-Culiacan	Native	1997	Cotton	Americas/ Caribbean	30
25	Moorea-French Polynesia	Undetermined	2008	Unidentified plant	Groups with Uganda-sweetpotato haplotype but is distinct	7
26	Morocco	Introduced	2003	Tomato	N.Africa/ Mediterranean/M.East	9
27	Mozambique	Native	1998	Wild sp./ unknown	East/South/Central Africa	18
28	Pakistan K	Native	1995	Cotton	Asia	30
29	Panama	Invasive	1992	Tomato	N.Africa/ Mediterranean/M.East	6
30	Reunion-MS	Native	2008	Annual poinsettia (*Euphorbia heterophylla*), Lab colony since 2005	N.Africa/ Mediterranean/M.East	10
31	Reunion-MS-2009	Native	2009	Annual poinsettia	N.Africa/ Mediterranean/M.East	29
32	Riverside A	Native, extinct	2006	Cotton. Lab colony since 1980	Americas/ Caribbean	30
33	South Africa	Native	1999	Cassava	East/South/Central Africa	25
34	Spain S	Introduced	2001	*Ipomoea indica*. Lab colony since 1995	Sub-Saharan Africa	29
35	Spain	Native	2008	Zucchini	N.Africa/ Mediterranean/M.East	28
36	Sudan	Undetermined	1994	Beans	N.Africa/ Mediterranean/M.East	6
37	Sudan Q-like L	Native	1994	Cotton	N.Africa/ Mediterranean/M.East	30
38	Turkey M	Native	1994	Cotton	N.Africa/ Mediterranean/M.East	29

Pooled pop. number	Population name	Status	Collection year	Original host and rearing information for lab cultured populations	Mitochondrial COI clade	n
39	Uganda-cassava	Native	2000	Cassava	East Africa (UG-1 haplotype)	9
40	Uganda-sweetpotato	Native	1994	Sweet potato	Outlier to *B. tabaci* sibling species group / *B. tabaci* complex probably	30
41	Yemen 1 – B_2	Native	2006	Tomato	N.Africa/ Mediterranean	5
41	Yemen 2 – B_2	Native	2006	Tomato	N.Africa/ Mediterranean	5

Table 2. Characteristics and sources of microsatellite loci used in this study

Microsatellite locus	Allele size range (bp)	Fluorescent dye color[1]	Repeat motif	Isolation source	Reference
***WF1B11**	101-178	PET	$(CTTGA)_{12}$imp	Biotype B	Schwartz unpublished
WF2C01	110-225	PET	$(GTTT)_{11}$imp	Biotype B	Schwartz unpublished
WF2H06	141-214	NED	$(TTTG)_{11}$	Biotype B	Schwartz unpublished
WF1B06	128-170	PET	$(ACTC)_8$	Biotype B	Schwartz unpublished
WF2E11	159-264	PET	$(GATT)_{27}$imp	Biotype B	Schwartz unpublished
***BEM6**	161-236	6-FAM	$(CA)_8$imp	Australia	De Barro et al. (2003)
***BEM15**	166-238	6-FAM	$(CAA)_6(CAG)_4(CAA)_4$	Asia (Indonesia)	De Barro et al. (2003)
***BEM31**	105-142	HEX	$(GCT)_4(GTT)_2$	Asia (Vietnam)	De Barro et al. (2003)
***BT-b103**	118-151	HEX	$(AC)_8(TC)_3$	Biotype Q	Tsagkarakou and Roditakis (2003)
***BT-e49**	266-390	6-FAM	$(TTG)_{12}(TTC)_{11}$	Biotype Q	Tsagkarakou et al. (2007)
BtIs1.2	256-368	6-FAM	$(CA)_{13}N_{14}(CA)_8$	B/Q	Gauthier et al. (2008)
****MS145**	171-225	PET	$(AC)_9$	Biotype B	Dalmon et al. (2008)
***MS177**	233-278	6-FAM	$(CA)_7$	Biotype B	Dalmon et al. (2008)

[1] 6-FAM, PET, NED (Applied Biosystems), 6-FAM, HEX (SIGMA-ALDRICH)
* denotes loci used in the 7-locus analysis (**: additional locus used in the PCA)
imp=imperfect

Table 3. Locus specific statistics across populations: number of alleles (N), mean expected (H_E) and observed (H_O) heterozygosities (averaged across populations) (±SE), inbreeding coefficient within individuals relative to the population (F_{IS}), and number of populations that deviated from HWE in each locus out of a total of 41 populations.

Locus	N	Mean H_E	Mean H_O	F_{IS}	No pop. out of HWE
WF1B11	19	0.318 (0.048)	0.318 (0.050)	-0.001	-
WF2C01	19	0.367 (0.053)	0.306 (0.050)	0.167	1
WF2H06	19	0.234 (0.046)	0.255 (0.050)	-0.091	-
WF1B06	12	0.108 (0.026)	0.081 (0.022)	0.250	-
WF2E11	22	0.334 (0.051)	0.306 (0.050)	0.084	-
BEM6	19	0.200 (0.039)	0.031 (0.013)	0.847	8
BEM15	23	0.455 (0.041)	0.472 (0.049)	-0.039	-
BEM31	10	0.242 (0.036)	0.288 (0.047)	-0.190	-
BT-b103	14	0.327 (0.039)	0.234 (0.035)	0.284	6
BT-e49	41	0.530 (0.048)	0.469 (0.047)	0.115	3
BtIs1.2	42	0.372 (0.054)	0.332 (0.051)	0.105	1
MS145	13	0.275 (0.041)	0.176 (0.036)	0.360	1
MS177	18	0.361 (0.042)	0.314 (0042)	0.128	1

Figures

Fig. 1. Map showing the general geographic sampling locations of *B. tabaci* populations used in this study (in black dots) with the presumed range of the geographic origins of the invasive B and Q biotypes and their closest relatives (in dotted circles). Numbers in parentheses represent the different collections we had from the same region, which were pooled into a single population for analysis

Fig. 2. Allelic richness (corrected for unequal sample size after rarefaction) in different biotypes and geographic groups

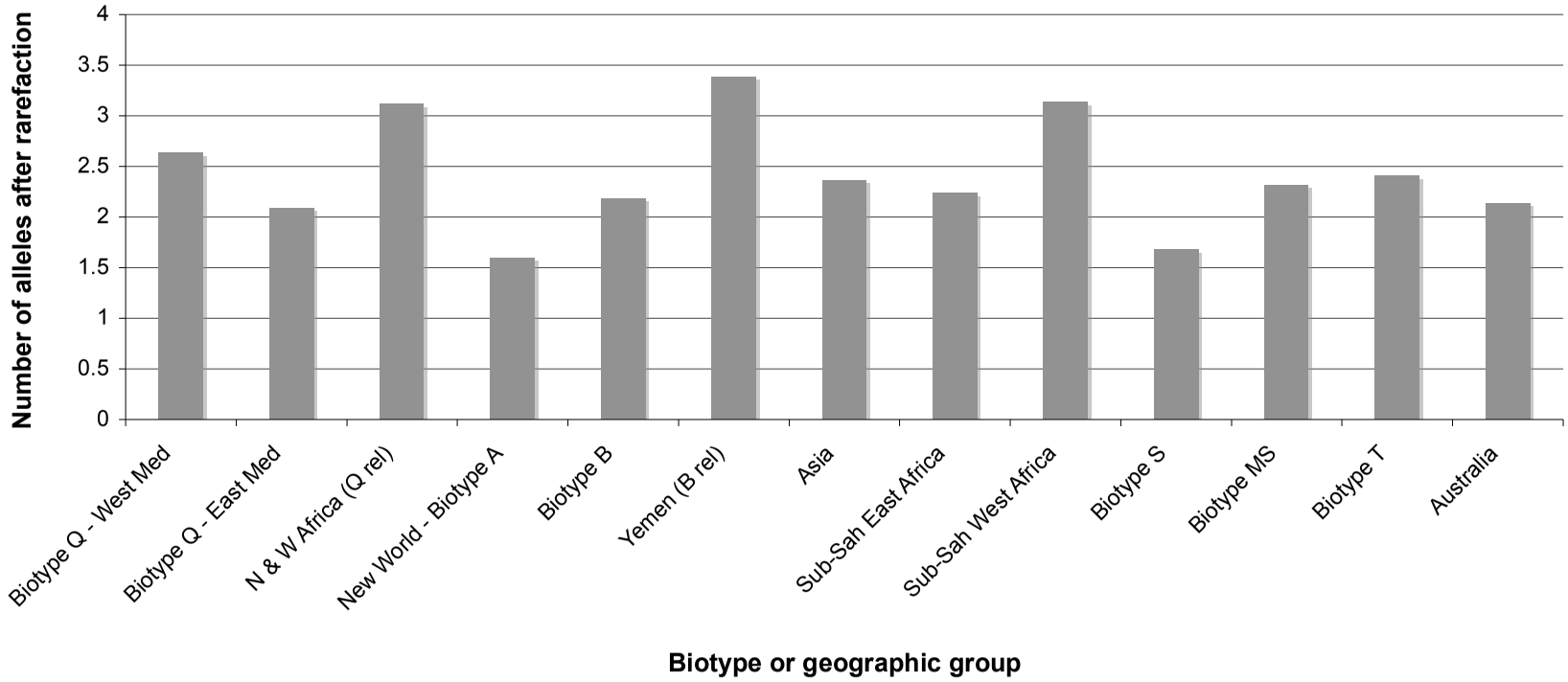


Fig. 3. Observed (*Ho*) and expected (*He*) heterozygosity in different biotypes and geographic groups

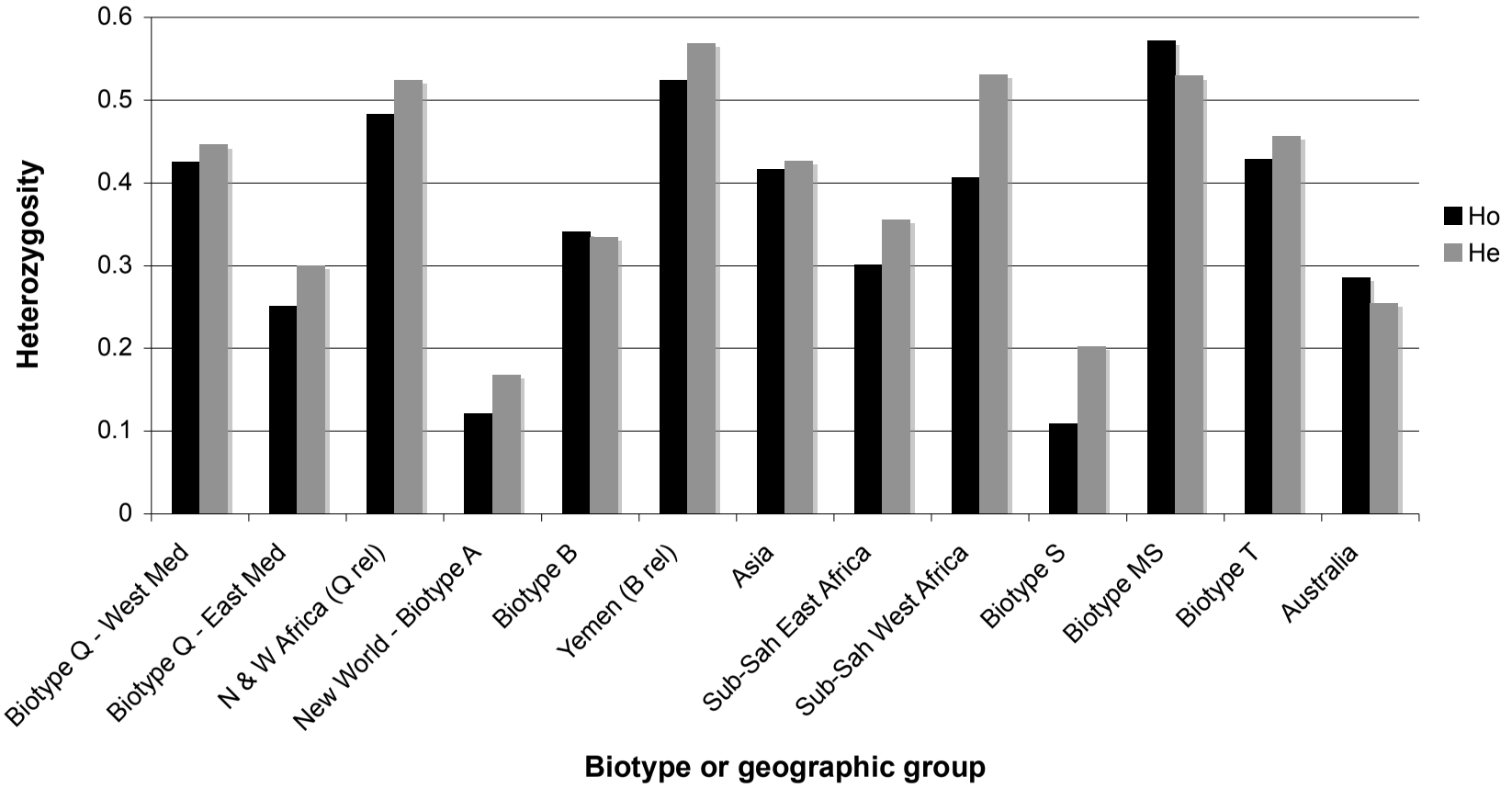


Fig. 4. Unrooted NJ tree based on Cavalli-Sforza & Edwards chord distance (D_C). Colored clades and branches represent biotypes previously characterized based on mtCOI data and biological/ ecological information. Other groupings were named according to geographic structuring of populations.

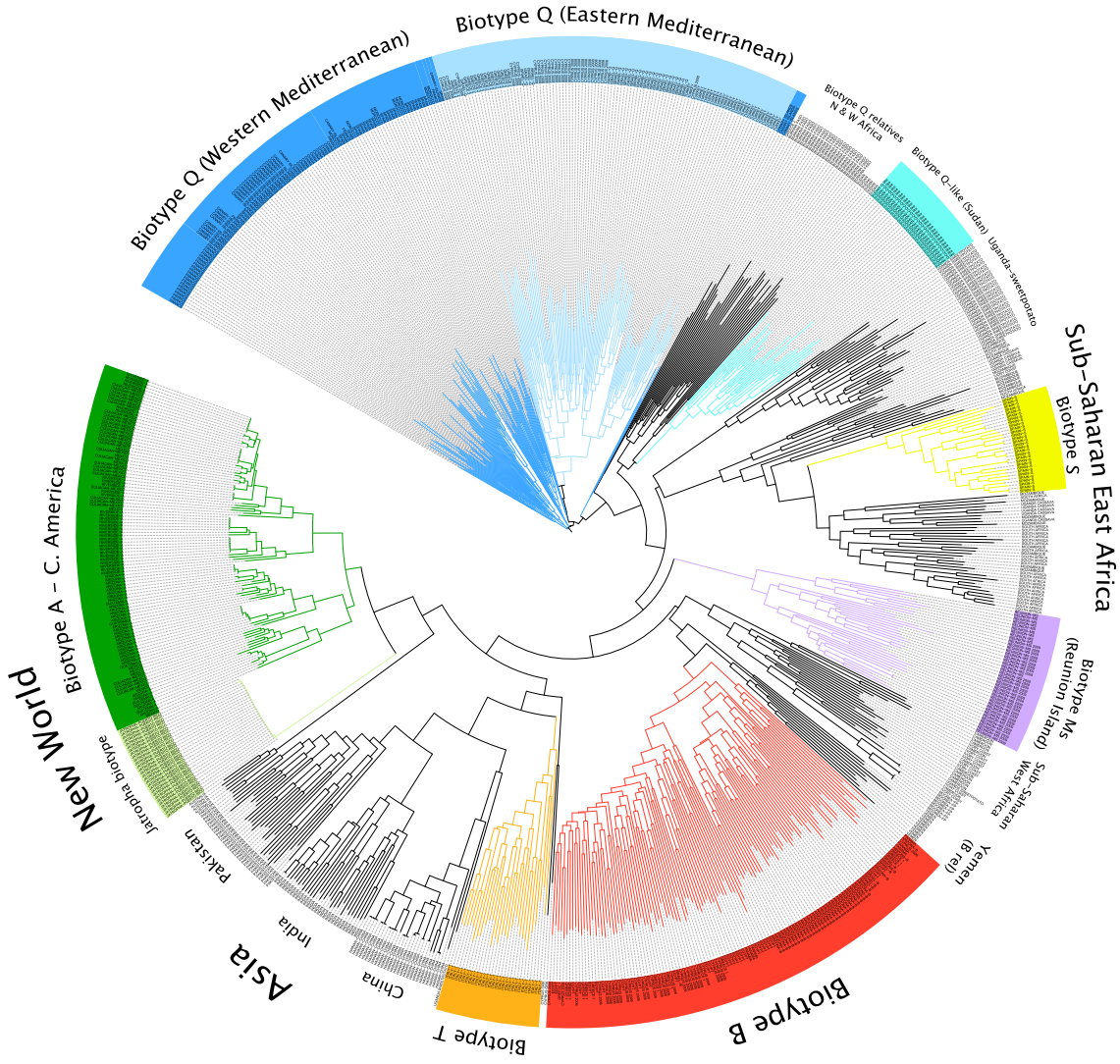


Fig. 5. Three-dimensional plot of a Principal Coordinates Analysis based on individual microsatellite genotypes. Individuals are color-coded according to biotype or geographic group they belong to based on results from NJ tree with corresponding colors

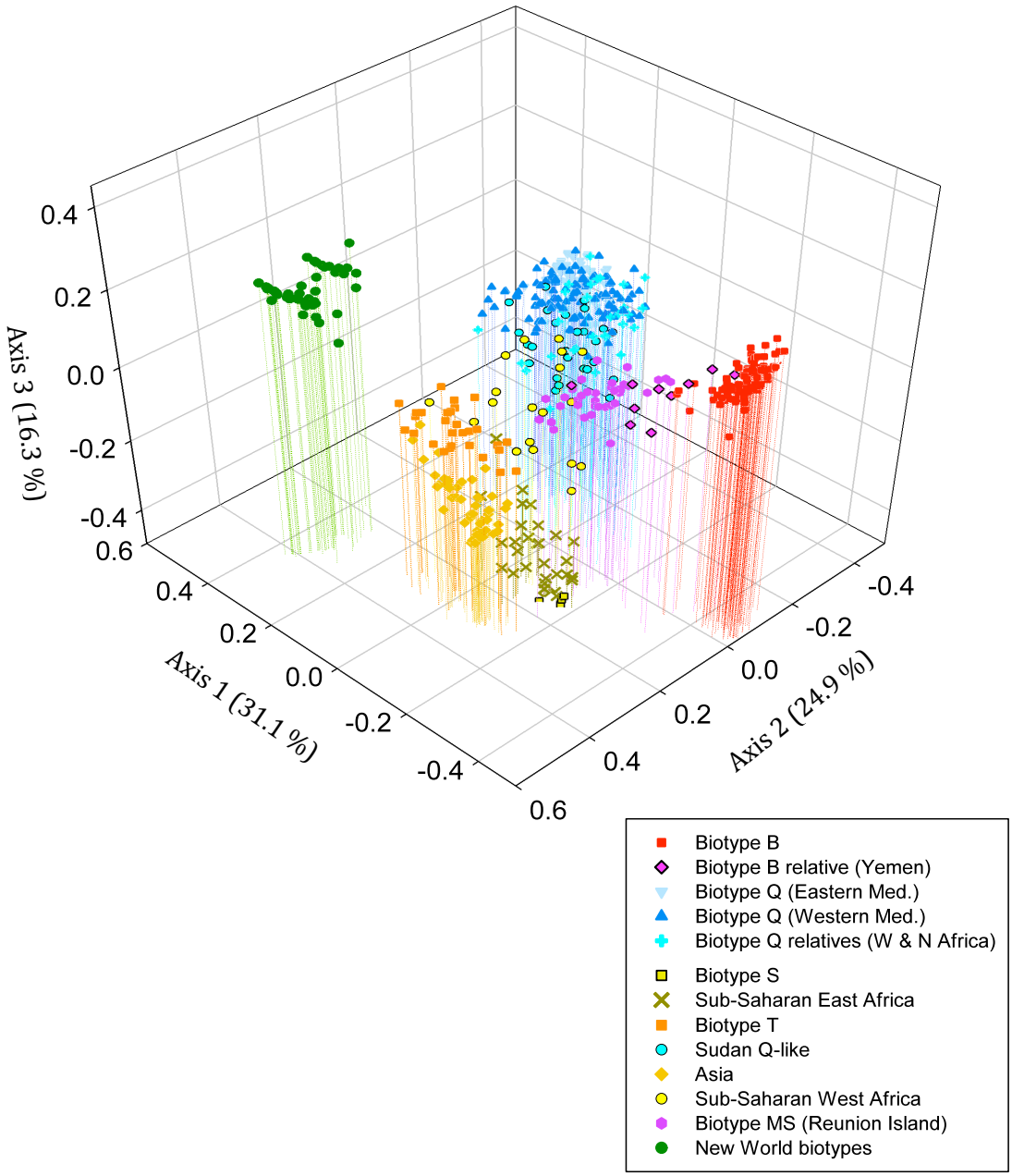
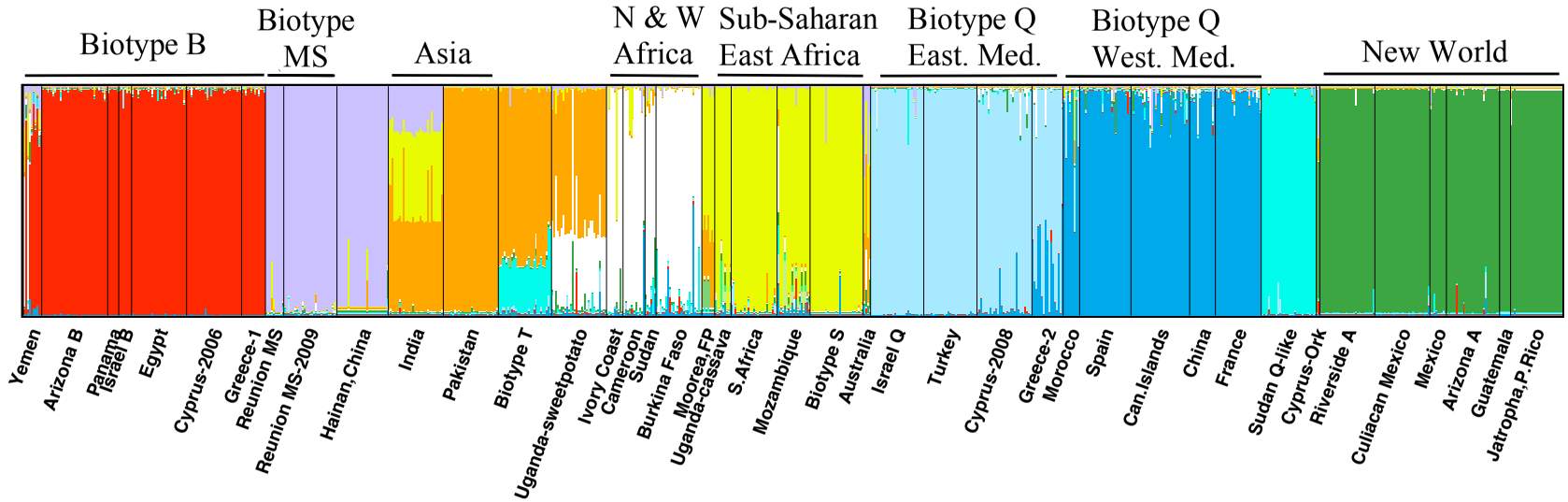


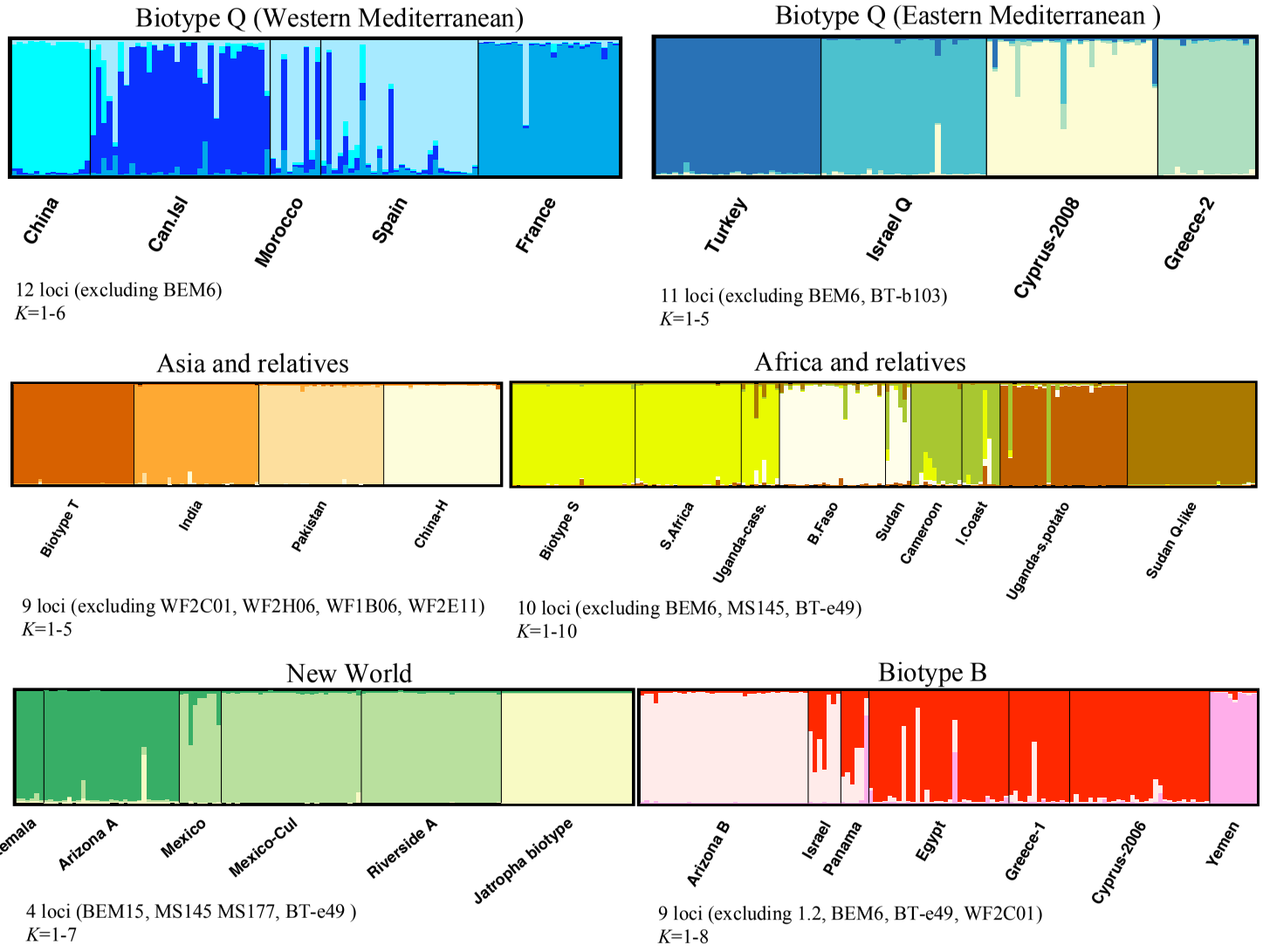
Fig. 6. Bayesian clustering analysis results of worldwide multilocus genotypes of *B. tabaci* performed in STRUCTURE. Individuals are arranged on the x-axis, each represented by a thin vertical line and partitioned into each of 9 inferred clusters (*K*) with their estimated membership fractions on the y-axis. Labels below the plot represent the sampled populations and above the plot the biotypes or geographic groups. Clusters are colored according to groupings identified in other analyses (PCA, NJ tree). Fig. 6A shows results of worldwide analysis and Fig. 6B shows results of sub-structure analysis for different biotypes / geographic regions. The number of *K* specified and the loci used in each run are indicated below each plot.

A)



10 loci (excluding BEM6, BT-b103, BT-e49)
K=1-15

B)


Biotype Q (Western Mediterranean)
China
Can.Isl
Morocco
Spain
France
12 loci (excluding BEM6)
K=1-6
Biotype Q (Eastern Mediterranean)
Turkey
Israel Q
Cyprus-2008
Greece-2
11 loci (excluding BEM6, BT-b103)
K=1-5
Asia and relatives
Biotype T
India
Pakistan
China-H
9 loci (excluding WF2C01, WF2H06, WF1B06, WF2E11)
K=1-5
Africa and relatives
Biotype S
S.Africa
Uganda-cass.
B.Faso
Sudan
Cameroon
I.Coast
Uganda-s.potato
Sudan Q-like
10 loci (excluding BEM6, MS145, BT-e49)
K=1-10
New World
Guatemala
Arizona A
Mexico
Mexico-Cul
Riverside A
Jatropha biotype
4 loci (BEM15, MS145 MS177, BT-e49)
K=1-7
Biotype B
Arizona B
Israel
Panama
Egypt
Greece-1
Cyprus-2006
Yemen
9 loci (excluding 1.2, BEM6, BT-e49, WF2C01)
K=1-8

Chapter 3

Historical demography and divergence of the invasive *Bemisia tabaci* biotypes B and Q associated with human movement and trade in the Mediterranean, Middle East, and Africa

Abstract

Understanding the processes involved in the early stages of isolation between populations is an intriguing and important topic in evolutionary biology, yet, still poorly understood. Studies of diverging populations, or biotypes, can help us examine the role of various demographic processes such as migration, isolation and parameters such as population sizes, while at the same time they allow us to make inferences about historical associations, such as the role of humans or geological events in the isolation process. In this study, we utilize the Isolation with Migration (IM) model to study the divergence history of two invasive biotypes, termed "B" and "Q" in a complex of cryptic species, the whitefly *Bemisia tabaci*. Spatial and temporal patterns of molecular variation suggest that biotype B diverged from its relative in Yemen about 3600 years ago, while biotype Q diverged into a Western and an Eastern Mediterranean lineage about 1400 years ago. These periods coincide with extensive human movement and trade of agricultural goods in the Mediterranean, the Middle East, and Africa during the Iron and Bronze Ages and the Roman period. Consistent with expectations of diversification of invasive species, our results indicate that ancestral populations had much larger effective sizes than those of emerging biotypes, but derived populations in both biotypes had similar effective sizes. Gene flow during divergence was found to be asymmetrical, with higher rates from biotype B to its Yemen relative and from Western to Eastern Mediterranean for biotype Q. Using a combination of microsatellite and mitochondrial DNA markers, we also document different levels of genetic diversity, both among biotypes as well as between the different markers, suggesting complex population histories in these biotypes. Based on our findings, we propose a scenario of diversification and evolution of biotypes B and Q associated with different movements of humans out of Africa. This study highlights the role of humans in the emergence of invasive insect populations, not only in contemporary, but also in ancestral timescales, mediated by trade of agricultural products.

Introduction

Studies of past demographic histories of diverging populations can enhance our understanding of the processes involved in the early stages of reproductive isolation (Emerson et al. 2001). The role of gene flow and other demographic processes in divergence and diversification can be studied by examining populations within a species that still exchange genes, yet are in the phase of isolation into distinct gene pools. Some excellent cases for such studies are "biotypes", which are viewed as intermediate stages along the continuous route to speciation (Diehl and Bush 1984; Dres and Mallet 2002). Biotypes exhibit heritable differences in biological and ecological traits such as virulence, host preference and survival, and can be reproductively isolated as has been reported for biotypes of the rice brown planthopper, *Nilaparvata lugens* (Claridge et al. 1985).

Geographic isolation has been implicated as a primary force in the initial separation of populations and their subsequent divergence into biotypes. Apart from ancestral vicariance events (*e.g.* separation of landmasses leading to breaks in the range of a taxon), more recent cases of geographic isolation of populations have been facilitated through hitchhiking during human migrations and the trade of agricultural products. There are numerous examples of human-mediated transportation of organisms from prehistoric times, as is the case of Polynesians moving across the Pacific and carrying along plants and animals (Gillespie et al. 2008; Matisoo-Smith 2007), to the recent, large-scale species invasions *via* international trade in the past 50 years (Pimentel et al. 2005).

Invasive biotypes in the whitefly complex *Bemisia tabaci* (Hemiptera: Aleyrodidae) represent a notable example of an organism that has been transported around the globe by humans. The species has a broad distribution with remarkable genetic variation at the global level. Among the 25+ biotypes described so far (Boykin et al. 2007; Brown 2010; Dinsdale et al. 2010), biotypes B and Q are notorious invasive pests that damage agricultural crops worldwide. The two biotypes are indistinguishable morphologically, but have pronounced differences at the genetic level (Chapter 2), as well as in ecological and biological traits, such as ability to vector viruses, insecticide resistance, and temperature tolerance (Brown 2010). Biotype B was first reported as an economically important pest in the US in 1986 (Hamon and Salguero 1987), and since then has expanded its range across the world. In the last 10 years, the Q biotype has emerged as an equally important invasive pest and has spread from its native range in the Mediterranean to Asia, New Zealand, North and Central America, and Europe.

The origins of the two biotypes are not known precisely, but analyses of mitochondrial COI have suggested that the B biotype diversified in East Sub-Saharan Africa (Eastern Sahel region) or the Middle East (Boykin et al. 2007; Brown 2010; Frohlich et al. 1999; Kirk et al. 2000), while the Q biotype originated around North, West and Central Africa and the Mediterranean (Brown 2010). This is based on findings that native populations from Yemen, Eritrea, Ethiopia, and Uganda are the closest known relatives of biotype B (Brown unpublished, Frohlich et al. 1999) and native populations from Burkina Faso, Mali, Senegal and Sudan are the closest relatives of biotype Q (Brown unpublished). Moreover, because the Q biotype has been found throughout the Mediterranean with substantial genetic structure compared to the B biotype, it has been suggested that this area represents its native range (Brown 2007).

From a theoretical perspective, the *B. tabaci* system provides excellent opportunities to address questions of evolutionary significance, such as historical demography and evolutionary processes leading to isolation and divergence. The exploitation of genetic data and methodological advances based on the coalescent theory (Kingman 1982) have provided the means to approach such questions with applications to several species (Hey and Machado 2003), from humans (Hey 2005; Laan and Paabo 1997) and other primates (Hey 2010b; Won and Hey 2005) to insects (Hey and Nielsen 2004; Lozier and Mills 2009; Roderick 1996; Schoville and Roderick 2009). From an applied standpoint, prevention and management of future invasions can be facilitated by understanding how humans have been contributing to the emergence of invasive pest populations, from historical to contemporary times.

In the present study, we aimed to elucidate when and how the two invasive biotypes, B and Q in this species complex diversified from their hypothesized closest relatives in the Mediterranean and Africa (Chapter 2), and whether their divergence was associated with human movement and agriculture. We used genetic data to estimate historical demographic parameters, including splitting times, effective population sizes, and rates of pairwise migration by fitting our data from two pairs of populations from biotypes B and Q to an "Isolation with Migration" model (Nielsen and Wakeley 2001).

Materials and Methods

Samples and DNA extraction

We analyzed a total of 547 female *Bemisia tabaci* whiteflies belonging to biotypes B, Q, and their closest relatives sampled from 32 locations in the USA, Africa, Asia, the Mediterranean, and the Middle East (Table 1). Results from previous analyses (Chapters 2, 4) showed that biotype Q consists of two genetically divergent lineages, one in the Western and the other in the Eastern part of the Mediterranean. Therefore, populations from this biotype were labeled as Q_{WEST} and Q_{EAST} according to our previous results (Table 1). Populations from Yemen (closest relative of biotype B – Chapter 2) and Burkina Faso and Sudan (closest relatives of biotype Q- Chapter 2) have been labeled B relative and Q relative respectively.

Whiteflies were preserved in 95% ethanol prior to DNA extraction. Genomic DNA was extracted from whole whiteflies homogenized in 1.5 ml microcentrifuge tubes, using the Qiagen DNeasy DNA Blood and Tissue kit following the manufacturer's protocol. This protocol was slightly modified to perform a final elution from the column in 80µl buffer AE followed by a second elution in 20µl AE, with the two eluates combined in a single microcentrifuge tube and stored at -20° C.

Genetic data

We genotyped whiteflies using the polymerase chain reaction (PCR) for a total of 13 microsatellite loci (Table 2) obtained from previous work (Dalmon et al. 2008; De Barro et al. 2003; Gauthier et al. 2008; Tsagkarakou and Roditakis 2003; Tsagkarakou et al. 2007) as well as our own (Schwartz unpublished). Forward primers were labeled with a fluorescent dye. Methods for multiplex and single primer pair PCRs, as well as fragment analysis and genotyping are described in Chapter 4.

A region of approximately 800 bp of the mitochondrial cytochrome oxidase

subunit I (COI) gene was amplified from a total of 62 individuals from biotypes B, Q_{EAST} and Q_{WEST} (Table 1) and sequenced in both directions using the primers C1-J-2195 (5-TTGATTTTTTGGTCATCCAGAAGT-3) and L2-N-3014 (5-TCCAATGCACTAATCTGCCATATTA-3) (Simon et al. 1994). PCRs were performed as described in Chapter 4, with different annealing temperature for each population as indicated in Table 1. PCR purification, sequencing, and sequence editing were performed as described in Chapter 4.

Genetic diversity and relationships among biotypes B, Q and their relatives

Prior to any analysis, we performed exact tests of population differentiation using Genepop (version 4.0.10) with the default settings (Raymond and Rousset 1995; Rousset 2008), to test whether populations from neighboring collection locations were differentiated. Those sampled populations that were not significantly differentiated after standard Bonferroni correction for multiple comparisons (at a = 0.05) were pooled into a single "sample population" (Table 1). In the same program, we calculated pairwise F_{ST} values in all populations using the default settings. To assess population differentiation within each lineage (biotype B, Q_{WEST}, Q_{EAST}, B relative, Q relative) we averaged all pairwise F_{ST} values among populations for each lineage. Pairwise F_{ST} values across all populations were used to construct a neighboring-joining (NJ) tree using the *neighbor* program in the software package Phylip-3.69 (Felsenstein 2005). The resulting tree was visualized using TreeView (Page 1996).

A *Principal Coordinates Analysis (PCA)* was performed on a genetic distance matrix of the individual multilocus genotypes (converted into a covariance matrix with data standardization) using the Excel based program GenAlEx 6.1 (Peakall and Smouse 2006). To visualize better the relationships among biotypes and their closest relatives, individuals from each population were pooled into the respective groupings prior to analysis and color-coded accordingly in the PCA. This does not affect the clustering of individuals since the genetic distance matrix is estimated based on pairwise individual-by-individual calculation without taking into account the population or grouping they belong to. In order to obtain estimates of genetic distances among lineages (biotypes and their closest relatives), we pooled populations into each lineage and estimated pairwise F_{ST}s among these groupings using Genepop.

Estimating divergence times and migration rates in the B and Q biotypes using the Isolation with Migration (IM) model

We examined the divergence history between biotypes B and Q and their closest relatives by fitting an Isolation with Migration (IM) model implemented in the program IMa2 (Hey 2010a; Hey 2010b; Hey and Nielsen 2007; Nielsen and Wakeley 2001). Divergence times and migration rates were estimated for two pairs of populations: 1) Cyprus (biotype B) and Yemen (B relative) and 2) Cyprus (biotype Q_{EAST}) and Spain (biotype Q_{WEST}).

The IMa2 model implements a Bayesian Markov Chain Monte Carlo approach to estimate posterior probability densities of parameters for effective population sizes, gene flow, and time of population splitting. In its simplest form, this model assumes that an ancestral population diverged at some time *t* in the past, giving rise to two daughter populations, which have may have exchanged genes since their divergence. When a two-population model is used, IMa2 estimates six demographic parameters: effective sizes of

the ancestral and derived populations N_A, N_0, N_1, time *t* since divergence, and migration rates between derived populations in both directions $m_{0\rightarrow1}$ and $m_{1\rightarrow0}$.

One of the model assumptions is that there are no other populations that are more closely related to the sampled populations than they are to each other (http://genfaculty.rutgers.edu/hey/software). For this reason, we applied this method only between a population pair from B biotype and Yemen and between a pair from Q_{WEST} and Q_{EAST} lineages, and not between biotypes B and Q. This decision was based on results from earlier analyses (Chapter 2) and previous work (Frohlich et al. 1999) which showed that the Yemen population is the closest relative of biotype B, and the Q_{WEST} and Q_{EAST} lineages are each others' closest relatives. We used a total of 12 loci for the biotype B-Yemen analysis and 10 loci for the Q_{WEST}-Q_{EAST} analysis (Table 2). We selected loci that did not seem to deviate from the stepwise mutation model (SMM), based on the distribution of the alleles. We converted allele sizes to repeat numbers and rounded to the closest integer alleles for which we obtained non-integer repeat numbers.

We followed the guidelines recommended in the IMa2 documentation (http://genfaculty.rutgers.edu/hey/software) to select upper bounds for the parameter prior distributions, and after multiple preliminary runs, we adjusted the priors for the final run according to posterior probability distributions and set upper bounds for population size at *q*=200, splitting time *t*=5, migration rate *m*=20 for the Q_{WEST}-Q_{EAST} run, or *m*=10 for the biotype B-Yemen run. After multiple trial runs we chose geometric heating schemes (ha=0.96, hb=0.9) that achieved good mixing, following examination of the resulting trendlines and Effective Sample Size (ESS) values (>120). For the final analyses, we ran the program using 50 chains, sampling more than 100,000 genealogies per locus (saving every 100 steps) following a burn-in of 500,000 steps. Two independent runs with different random number seeds were done for each analysis, resulting to a total of 269,821 genealogies for the biotype B-Yemen run and 255,667 for the Q_{WEST}-Q_{EAST} run. These resulting genealogies were loaded to the L (Load Genealogies) mode run in IMa2 for each analysis to estimate joint-posterior densities of parameters. To obtain demographic parameter estimates from model parameter estimates we set the microsatellite mutation rate estimated previously for *Drosophila melanogaster* (Schug et al. 1998) as $1x10^{-4}$ (with a range of $1x10^{-5}$ – $1x10^{-2}$) per year for each locus. We set the generation time at 0.1, which is 10 generations per year, an approximate estimate of generation time per year in this species (Butler et al. 1983). Finally, the supplementary program IMfig (http://genfaculty.rutgers.edu/hey/software) was used to generate figures of the model from the IMa2 output file.

Mitochondrial DNA sequence diversity

We calculated haplotype diversity (*h*), nucleotide diversity (π) (Nei 1987), Tajima's *D* statistic, and Fu's F_S statistic in sequences from the three populations using the program DNAsp v.5 (Rozas et al. 2003). Tajima's *D* and Fu's F_S are statistics used to test for population growth / purifying selection (significant negative values), and population bottlenecks / balancing selection (significant positive values). To examine the distribution of mtDNA sequence diversity in the three populations we calculated a median-joining (MJ) network (Bandelt et al. 1999) using a variable sites matrix in the program NETWORK 4.5.1.6. We used the default settings of equal connection weights (transition: transversion ratio 1:1) to all sites and epsilon value = 0.

Results

Genetic diversity and relationships among biotypes B, Q and their relatives

The PCA showed a clear clustering of individuals into biotypes B, and Q, confirming the split between the Q_{WEST} and Q_{EAST} lineages (Fig. 1). The B relative group, represented only by a single population (Yemen) and consisted of individuals scattered loosely between the Q relative cluster and B biotype, closer to the latter. The Q relative cluster was between the Q_{WEST} and Q_{EAST} clusters. Pairwise F_{ST}s between populations averaged within each group showed that the Q_{WEST}, Q_{EAST}, and the Q relative group had very similar values, from 0.26 (Q West) to 0.29 (Q relative). By contrast, the B biotype group had much lower average F_{ST} (0.12) between its populations. F_{ST} values calculated between groups (after populations were pooled into their respective groups) showed that the smallest genetic distance was between the Q_{WEST} and Q_{EAST} group, followed by biotype B with its relative (Yemen) (Table 3). The Q relative group was closer genetically to the Q_{WEST} than the Q_{EAST} group. The next closest groups were the B relative with the Q relative, Q_{WEST}, and Q_{EAST}, followed by the Q relative with B biotype. The groups with the highest genetic distance were the Q_{WEST} and Q_{EAST} with biotype B. Similar results were obtained from NJ tree analysis using the population pairwise F_{ST} estimates (Fig. 2). The Yemen population was closely related to a group consisting of B biotype populations. Populations from the Eastern Mediterranean (Greece, Cyprus, Israel, Turkey) as well as an introduced to the US population formed a group sister to another group that consisted of populations from the Western Mediterranean (Spain, Morocco, Canary Islands, and France) as well as introduced populations to the US and China. Populations from Burkina Faso and Sudan were genetically closer to the Q than the B biotype group.

Estimating divergence times and migration rates in the B and Q biotypes using the Isolation with Migration model

The results from the analysis of divergence using the isolation with migration model in IMa2 are shown in Fig. 3, with the Q_{WEST} - Q_{EAST} divergence graph on the left and the biotype B-Yemen divergence graph on the right. The Q biotype analysis showed that the populations representing the Eastern (Q_{EAST}) and Western (Q_{WEST}) lineages diverged about 0.0014 million (MYR), or 1400 years ago. The ancestral population was estimated to have a much larger effective size ($1.4x10^6$) than the daughter populations, which had similar sizes. Gene flow was asymmetrical, with a much higher and statistically significant estimated population migration rate from the Q_{WEST} to the Q_{EAST} lineage. The migration rate from Q_{EAST} to Q_{WEST} was not statistically significant.

Results from the B biotype analysis showed that the population representing biotype B and that of its closest relative from Yemen split approximately 3600 years ago, from an ancestral population with an effective size of $1.6 x10^6$. This was much larger than the derived populations, which, as in the biotype B-Yemen run, had similar sizes. Although some migration was detected from the Yemen to the B biotype population, it was non-significant and 13 times smaller than gene flow in the opposite direction. The rate of gene flow from the B biotype into the Yemen population was 1.3 effective genes per generation and was statistically significant ($P<0.001$). A comparison between the two analyses indicates that although the biotype B-Yemen split happened much earlier,

gene flow was almost twice as high between these populations compared to gene flow between Q_{WEST} - Q_{EAST}.

Mitochondrial DNA sequence diversity

Analysis of a segment of the mitochondrial COI showed the presence of only a single haplotype in each of the populations representing the Q_{EAST} lineage (Cyprus Q) and the B biotype (Cyprus B) in the 27 and 17 individuals sequenced respectively from each population (Table 4, Fig. 4). In contrast, in the Spain Q population (from the Q_{WEST} lineage) we found a total of six haplotypes in a sample of 18 individuals (Fig. 4), and a high haplotype and nucleotide diversity. Tajima's *D* and Fu's *Fs* were negative but non-significant, although the Tajima's *D* estimate was close to significance ($P<0.10$).

The haplotype network showed a clear split of haplotypes representing the three lineages. Biotype B was separated from both the Q_{EAST} and Q_{WEST} lineages by 34 substitutions, an approximate 5% sequence divergence. The two Q lineages were separated by 11 mutations or about 1.6% sequence divergence. As shown by diversity estimates, the Q_{WEST} lineage had high haplotype diversity with 2 haplotypes being more common, with frequencies of 0.22 and 0.56 and the rest represented by only a single individual (Table 4, Fig. 4).

Discussion

Genetic diversity and relationships among biotypes B, Q and their relatives

We found that the invasive biotypes B and Q are clearly genetically differentiated, with no evidence for gene flow between them. Moreover, we detected a clear distinction between Eastern and Western Mediterranean lineages of biotype Q. These findings corroborate previous analyses on the degree of differentiation among biotypes B, Western Mediterranean Q (originally described from Spain), and Eastern Mediterranean Q (originally described from Israel) (Chapters 2 and 4; Ahmed et al. 2009; Chu et al. 2008; Tsagkarakou et al. 2007). Moreover, our NJ tree analysis based on pairwise F_{ST}s confirmed the presumed genetic relationships between these biotypes and their closest relatives, the populations from Sudan, Burkina Faso, and Yemen.

We found that the two Q lineages and their closest relatives in Africa had high and similar levels of pairwise F_{ST}s among their populations, suggesting a high degree of genetic structure. This is consistent with expectations; biotype Q is considered to be native in the region our samples came from (Brown 2010). A comparison with its invaded range in the USA in a separate study showed that the invading populations had lower pairwise F_{ST}s, a typical effect of recently introduced populations (Chapter 4).

Biotype B had much lower pairwise F_{ST}s, an indication that its populations are not well differentiated across our sampling locations. This could be a result of the smaller number of populations we examined from biotype B; however our samples spanned collections from five countries, including the USA, and were collected over a large period of time (1987 to 2007). This suggests that biotype B, which has been introduced in most regions worldwide since the 80's, is much more homogeneous genetically compared to biotype Q, likely a result of its widespread invasion. These results may have important implications in the management of these populations: genetically diverse

populations may have inherently different characteristics related to invasiveness, adaptation, and insecticide resistance than genetically homogeneous populations. Such populations or strains may require a more strategic management program for their control. Furthermore, the enforcement of efficient quarantine measures to avoid introduction and spread of genetically diverse and potentially more invasive and adaptable strains becomes much more important in such cases.

Our analysis involving the closest relatives of B and Q biotypes gave some interesting results. The NJ tree based on pairwise F_{ST}s and the PCA of individual genotypes showed that the B and Q relatives (Yemen, Sudan and Burkina Faso) are more closely related than biotypes B and Q are to each other. This suggests that the B and Q biotypes had a more ancestral divergence than their relatives, or that the B-Q relatives have been exchanging genes more recently in the Central Africa region. If indeed Africa is the place of origin of the *B. tabaci* species, including biotypes B and Q (Chapter 1; Boykin et al. 2007; Brown 2010), these results may point to an interesting scenario: Central Africa may have been the common place of origin of the B-Q relatives populations, which later diversified towards the Eastern Africa (Q biotype relatives) and the Western Africa (B biotype relatives) (Fig. 5).

Estimating divergence times and migration rates in the B and Q biotypes using the Isolation with Migration model

The analysis of isolation with migration for the two biotypes suggest that ancestral populations had much larger effective sizes than those of emerging biotypes. The very large effective population size we found is not surprising for insect populations that can reach extreme sizes in a single culture (Peters and Barbosa 1977). However, the much smaller sizes of the daughter populations for both biotypes suggest that their divergence was likely accompanied by a strong bottleneck, resulting in a small fraction of the original population colonizing new areas, and giving rise to populations of similar, small sizes.

We found that the timing of the origins of both B and Q biotypes coincides with periods of human migration, as has been suggested for the domestication of other insect pests (Nardi et al. 2010). The divergence analysis places the split of the B biotype from its relative in Yemen at 1600 BC (3600 ya) coinciding with the "Bronze Age trade" that happened in the Fertile Crescent starting from 3600 to 1400 BC (Sherratt 2004). These dates suggest that the populations split at the time humans moved from West Africa (Eritrea, Ethiopia, Somalia) - the hypothesized origin of B biotype - to the Middle East (including Yemen), the Mediterranean, and Eastern Europe. Migration rates in the two populations were found to be asymmetrical, with much higher and significant gene flow from the Cyprus B population to Yemen. These migration rates represent average estimates across time since the populations diverged; thus the direction of the original colonization cannot be inferred. It is possible that the Yemen population has been receiving more migrants from the Cyprus (biotype B) population than *vice versa* because biotype B is invasive and has expanded its range across the world in the last 30 years; likely much earlier than that in the Mediterranean and Middle East where it originates from.

The Q_{WEST} and Q_{EAST} lineages were found to have diverged from their common ancestor around 600 AD (1400 ya). This period coincides with the end of the Iron Age in

the Mediterranean and the Roman period (Sherratt 2004), perhaps reflecting trade of plants and other goods that took place from North/West Africa (location of the Q closest relative) across the Mediterranean. The much higher and significant gene flow in the direction of Q_{WEST} to Q_{EAST} suggests that the Eastern Q lineage has been receiving migrants from the West since their divergence, perhaps due to ancestral and contemporary trade of whitefly infested plants from West to East. The much more recent split between Q_{WEST}-Q_{EAST} compared to biotype B-Yemen is interesting, especially given the much higher migration rates in the latter pair. This suggests a divergence with ongoing gene flow between biotype B and Yemen, and likely populations remained reproductively compatible, in contrast to the Q lineages that seem to have had more restricted gene flow despite a more recent splitting time. These results are in line with our knowledge of the status of these two biotypes: biotype B has been the invasive biotype for a longer time with populations pretty much distributed around the world (Brown 2000), likely exchanging genes with other reproductively compatible populations. By contrast, biotype Q has expanded its range recently and was found to be much more structured genetically in the Mediterranean (Chapter 2).

A particularly interesting research focus would be to study the interactions between the two Q biotype lineages in the field and the laboratory. Such studies could reveal whether they are reproductively isolated or have any ecological differences such as virus transmission efficiency, host adaptability, and insecticide resistance. For example, previous studies have demonstrated the competitive displacement of biotype B by Q in Spain (Moya et al. 2001), while the Cyprus and Israel Q populations (Eastern Q) were found to coexist with B biotype (Chapter 2; Hadjistylli 2003; Horowitz et al. 2003). In addition, the Q populations from Israel have not shown the extreme levels of insecticide resistance to novel insecticides (neonicotinoids) as is the case of the Spanish Q populations (Nauen et al. 2002). It is possible that the Q_{WEST} - Q_{EAST} lineages differ in many more attributes other than genetics, which may define their dynamic distribution in the Mediterranean and their sympatric occurrence with biotype B.

Based on our findings and the presumed regions of origin of biotypes B and Q, we hypothesize that the common ancestor of these two biotypes and their relatives diversified in Central Africa. The split between the ancestral B and Q lineages could have given rise to populations east of Africa (B relatives) and west of Africa (Q relatives), which later could have expanded towards the Middle East and the Western Mediterranean respectively (Fig. 5). Based on the high genetic differentiation between B and Q biotypes and their reproductive isolation in the field (Elbaz et al. 2010), it seems that some form of barrier prevented their sympatric occurrence in the Mediterranean for a long period of time. The closer genetic relationships between the B and Q relatives may indicate that their populations were in secondary contact in Africa after their isolation. Such hypotheses can be tested by examining divergence histories between the B and Q relatives from multiple regions across Africa, and associate splitting times with human movements or environmental/geological events. For the Q biotype for example, it has been suggested that the formation of the Sahara desert may have imposed a geographic barrier on biotype Q and a related biotype, J, found only south of the desert (De la Rua et al. 2006). Indeed, this hypothesis does not contradict our findings since it would place the divergence of the Q relatives about 5,500 years ago, when retreating monsoons resulted in desiccation of the Sahara (Kuper and Kropelin 2006). Biotype B relatives in

Africa extend as south as Uganda (Sseruwagi et al. 2005) and a closely related biotype (Ms) has also been found in the Reunion island, Madagascar, and Seychelles (Delatte et al. 2005). Divergence analysis using Bayesian coalescence-based approaches such as IM, or phylogenetic dating of a tree with the closest common relatives of B and Q, could shed light into the timing of their divergence and clarify associations with human migration and geological events.

Mitochondrial DNA sequence diversity

Analysis of mitochondrial DNA in populations from B, Q_{EAST} and Q_{WEST} biotypes showed that B biotype is much more divergent from the two Q lineages than they are with each other, which is consistent with microsatellite DNA (PCA analysis, F_{ST} distances). However, the levels of genetic diversity in these three lineages differed substantially at the two genetic markers. Although the Q_{EAST} lineage had slightly higher pairwise F_{ST} values among its populations compared to Q_{WEST}, its mitochondrial DNA diversity was zero in the samples we examined, with a single haplotype detected in 27 individuals. A strikingly high diversity (6 haplotypes in a sample of 17 individuals) was observed on the other hand in the Q_{WEST} lineage. This discordance in the genetic signature of nuclear and mitochondrial DNA is not uncommon and has been documented numerous times from several organisms (e.g. Brito 2007). A number of reasons could explain these patterns, such as a) selection on mtDNA, b) sex-biased dispersal or admixture, c) homoplasy of microsatellite markers may have reduced their signal of differentiation d) smaller effective population size of the mitochondrial DNA (Larmuseau et al. 2010). A plausible explanation in this case would be that the more ancestral mitochondrial diversity has been lost in B biotype and Q_{EAST} because of a selective advantage and adaptation of the persisting haplotypes. Another possibility is that these populations have gone through genetic bottlenecks during invasions in their older histories, but recurrent gene flow is evident in their nuclear genes. Finally, it is also likely that involvement of maternally inherited *Wolbachia* endosymbionts, may have contributed to female-biased admixture, with some females exchanging genes with the rest of the population but carrying distinct mitochondrial haplotypes in Q_{WEST}. The differential levels of mitochondrial diversity between the Eastern and Western Q lineage has been documented before; however by combining microsatellite markers, this analysis demonstrates that mitochondrial diversity does not always correspond to nuclear diversity and *vice-versa*.

Studies of demographic histories of populations can offer significant insights into evolutionary processes leading to the formation of distinct genetic pools; the first stages toward speciation from reproductively isolated populations. This study focused on three such lineages that represent invasive biotypes within the whitefly *B. tabaci,* a complex of cryptic species. Our findings associate divergence histories with periods of human movement and trade, consistent with other studies of diversification in insect pests. Results from this study highlight the role of humans in the diversification of invasive insect populations, not only in contemporary, but also in ancestral times of human-associated trade of agricultural products.

References

Ahmed MZ, Shatters RG, Ren SX, Jin GH, Mandour NS, Qiu BL. 2009. Genetic distinctions among the Mediterranean and Chinese populations of *Bemisia tabaci* Q biotype and their endosymbiont *Wolbachia* populations. *J. Appl. Entomol.* 133: 733-741.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16: 37-48.

Boykin LM, Shatters RG, Rosell RC, McKenzie CL, Bagnall RA, De Barro P, Frohlich DR. 2007. Global relationships of *Bemisia tabaci* (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Mol. Phylogen. Evol.* 44: 1306-1319.

Brito PH. 2007. Contrasting patterns of mitochondrial and microsatellite genetic structure among Western European populations of tawny owls (*Strix aluco*). *Mol. Ecol.* 16: 3423-3437.

Brown JK. 2000. Molecular markers for the identification and global tracking of whitefly vector-Begomovirus complexes. *Virus Res.* 71: 233-260.

Brown JK. 2007. The *Bemisia tabaci* complex: genetic and phenotypic variability drives begomovirus spread and virus diversification. Plant Disease Feature. http://www.apsnet.org/online/feature/btabaci Accessed: July 15, 2010.

Brown JK. 2010. Phylogenetic biology of the *Bemisia tabaci* sibling species group. In *Bemisia: Bionomics and Management of a Global Pest*, ed. PA Stansly, SE Naranjo. Dordrecht: Springer Science. In press.

Butler GD, Henneberry TJ, Clayton TE. 1983. *Bemisia tabaci* (Homoptera: Aleyrodidae) - development, oviposition, and longevity in relation to temperature. *Ann. Entomol. Soc. Am.* 76: 310-313.

Claridge MF, den Hollander J, Morgan JC. 1985. The status of weed-associated populations of the brown planthopper, *Nilaparvata lugens* (Stal): host race or biological species? *Zool. J. Linn. Soc.* 84: 77–90.

Chu D, Wan FH, Tao YL, Liu GX, Fan ZX, Bi YP. 2008. Genetic differentiation of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) biotype Q based on mitochondrial DNA markers. *Insect Sci.* 15: 115-123.

Dalmon A, Halkett F, Granier M, Delatte H, Peterschmitt M. 2008. Genetic structure of the invasive pest *Bemisia tabaci*: evidence of limited but persistent genetic differentiation in glasshouse populations. *Heredity* 100: 316-325.

De Barro PJ, Scott KD, Graham GC, Lange CL, Schutze MK. 2003. Isolation and characterization of microsatellite loci in *Bemisia tabaci*. *Mol. Ecol. Notes* 3: 40-43.

De la Rua P, Simon B, Cifuentes D, Martinez-Mora C, Cenis JL. 2006. New insights into the mitochondrial phylogeny of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in the Mediterranean Basin. *J. Zool. Syst. Evol. Res.* 44: 25-33.

Delatte H, Reynaud B, Granier M, Thornary L, Lett JM, Goldbach R, Peterschmitt M. 2005. A new silverleaf-inducing biotype Ms of *Bemisia tabaci* (Hemiptera: Aleyrodidae) indigenous to the islands of the south-west Indian Ocean. *Bull. Entomol. Res.* 95: 29-35.

Diehl SR, Bush GL. 1984. An evolutionary and applied perspective of insect biotypes. *Ann. Rev. Entomol.* 29: 471–504.
Dinsdale A, Cook L, Riginos C, Buckley YM, De Barro P. 2010. Refined global analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. *Ann. Entomol. Soc. Am.* 103: 196-208.
Dres M, Mallet J. 2002. Host races in plant-feeding insects and their importance in sympatric speciation. *Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci.* 357: 471-492.
Elbaz M, Lahav N, Morin S. 2010. Evidence for pre-zygotic reproductive barrier between the B and Q biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Bull. Entomol. Res.* In press. doi:10.1017/S0007485309990630
Emerson BC, Paradis E, Thebaud C. 2001. Revealing the demographic histories of species using DNA sequences. *Trends Ecol. Evol.* 16: 707-716.
Felsenstein J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle. http://evolution.genetics.washington.edu/phylip.html
Frohlich DR, Torres-Jerez I, Bedford I, Markham P, Brown J. 1999. A phylogeographic analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers. *Mol. Ecol.* 8: 1593-1602.
Gauthier N, Dalleau-Clouet C, Bouvret ME. 2008. Twelve new polymorphic microsatellite loci and PCR multiplexing in the whitefly, *Bemisia tabaci*. *Mol. Ecol. Res.* 8: 1004-1007.
Gillespie RG, Claridge EM, Goodacre SL. 2008. Biogeography of the fauna of French Polynesia: diversification within and between a series of hot spot archipelagos. *Phil. Trans. R. Soc. B*. 363: 3335-3346.
Hadjistylli M. 2003. Mating interactions between two *Bemisia tabaci* biotypes and their implications for pest and resistance management. MSc Thesis. Imperial College, University of London, London. 80 pp.
Hamon AB, Salguero V. 1987. *Bemisia tabaci*, sweet potato whitefly, in Florida (Homoptera: Aleyrodidae: Aleyrodinae). *Entomology Circular, Division of Plant Industry, Florida Department of Agriculture and Consumer Services*: 2 pp.
Hey J. 2005. On the number of New World founders: A population genetic portrait of the peopling of the Americas. *PLoS Biol.* 3: 965-975.
Hey J. 2010a. Isolation with Migration models for more than two populations. *Mol. Biol. Evol.* 27: 905-920.
Hey J. 2010b. The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Mol. Biol. Evol.* 27: 921-933.
Hey J, Machado CA. 2003. The study of structured populations - New hope for a difficult and divided science. *Nat. Rev. Genet.* 4: 535-543.
Hey J, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167: 747-760.
Hey J, Nielsen R. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc. Natl. Acad. Sci. USA* 104: 2785-2790.

Horowitz AR, Denholm I, Gorman K, Cenis JL, Kontsedalov S, Ishaaya I. 2003. Biotype Q of *Bemisia tabaci* identified in Israel. *Phytoparasitica* 31: 94-98.
Kingman JFC. 1982. The coalescent. *Stoch. Proc. Appl.* 13: 235-248.
Kirk AA, Lacey LA, Brown JK, Ciomperlik MA, Goolsby JA, Vacek DC, Wendel LE, Napompeth B. 2000. Variation in the *Bemisia tabaci* s.i. species complex (Hemiptera: Aleyrodidae) and its natural enemies leading to successful biological control of *Bemisia biotype* B in the USA. *Bull. Entomol. Res.* 90: 317-327.
Kuper R, Kropelin S. 2006. Climate-controlled Holocene occupation in the Sahara: Motor of Africa's evolution. *Science* 313: 803-807.
Laan M, Paabo S. 1997. Demographic history and linkage disequilibrium in human populations. *Nat. Genet.* 17: 435-438.
Larmuseau MHD, Raeymaekers JAM, Hellemans B, Van Houdt JKJ, Volckaert FAM. 2010. Mito-nuclear discordance in the degree of population differentiation in a marine goby. *Heredity*: 1-11.
Lozier JD, Mills NJ. 2009. Ecological niche models and coalescent analysis of gene flow support recent allopatric isolation of parasitoid wasp populations in the Mediterranean. *Plos One* 4(6): e5901. doi:10.1371/journal.pone.0005901
Matisoo-Smith E. 2007. Animal translocations, genetic variation and the human settlement of the Pacific. In *Genes, Language and Culture History in the Southwest Pacific,* ed. JS Friedlaender, pp. 147-157. Oxford: Oxford University Press.
Moya A, Guirao P, Cifuentes D, Beitia F, Cenis JL. 2001. Genetic diversity of Iberian populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae) based on random amplified polymorphic DNA-polymerase chain reaction. *Mol. Ecol.* 10: 891-897.
Nardi F, Carapelli A, Boore JL, Roderick GK, Dallai R, Frati F. 2010. Domestication of olive fly through multi-regional host shifts on cultivated olives: complete mitochondrial genomes. *Mol. Phylogen. Evol.* In press.
Nauen R, Stumpf N, Elbert A. 2002. Toxicological and mechanistic studies on neonicotinoid cross resistance in Q-type Bemisia tabaci (Hemiptera : Aleyrodidae). *Pest Manage. Sci.* 58: 868-875.
Nei M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia University Press.
Nei M, Maruyama T, Chakraborty R. 1975. Bottleneck effect and genetic variability in populations. *Evolution* 29: 1-10.
Nielsen R, Wakeley J. 2001. Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics* 158: 885-896.
Page RDM. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12: 357-358.
Peakall R, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6: 288-295.
Peters TM, Barbosa P. 1977. Influence of population density on size, fecundity, and developmental rate of insects in culture. *Annu. Rev. Entomol.* 22: 431-450.
Pimentel D, Zuniga R, Morrison D. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol. Econ.* 52: 273-288.
Raymond M, Rousset F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* 86: 248-249.

Roderick GK. 1996. Geographic structure of insect populations: gene flow, phylogeography, and their uses. *Annu Rev Entomol* 41: 325-352.
Rousset F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.
Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496-2497.
Schoville SD, Roderick GK. 2009. Alpine biogeography of Parnassian butterflies during Quaternary climate cycles in North America. *Mol. Ecol.* 18: 3471-3485.
Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TFC, Aquadro CF. 1998. The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. *Mol. Biol. Evol.* 15: 1751-1760.
Sherratt A. 2004. Trade routes: The growth of global trade. http://www.archatlas.org/Trade/Trade.php Accessed: July 31, 2010.
Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial-gene sequences and a compilation of conserved polymerase chain-reaction primers. *Ann. Entomol. Soc. Am.* 87: 651-701.
Sseruwagi P, Legg JP, Maruthi MN, Colvin J, Rey MEC, Brown JK. 2005. Genetic diversity of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) populations and presence of the B biotype and a non-B biotype that can induce silverleaf symptoms in squash, in Uganda. *Ann. Appl. Biol.* 147: 253-265.
Tsagkarakou A, Roditakis N. 2003. Isolation and characterization of microsatellite loci in *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Mol. Ecol. Notes* 3: 196-198.
Tsagkarakou A, Tsigenopoulos CS, Gorman K, Lagnel J, Bedford ID. 2007. Biotype status and genetic polymorphism of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Greece: mitochondrial DNA and microsatellites. *Bull. Entomol. Res.* 97: 29-40.
Won YJ, Hey J. 2005. Divergence population genetics of chimpanzees. *Mol. Biol. Evol.* 22: 297-307.

Tables

Table 1. Whitefly populations used in this study with information on collection site, biotype status, number of individuals genotyped (microsatellites) and sequenced (mtCOI), host plant, collection year, and annealing temperature (Ta) for sequencing of the mtCOI. Populations with the same number were collected from neighboring locations and were pooled together for analysis after exact tests of population differentiation showed non-significant differentiation (P>0.05).

Pop No	Pop. ID (collection site)	Biotype	Number of individuals genotyped	Host Plant	Collection year	Number of individuals sequenced	Ta
1	Arizona, USA	B	36	Cotton	Lab colony on poinsettia since 1987	-	
2	Cyprus*	B	30	*Cucumis sativum*	2006	17	52
3	Egypt	B	30	Cucumber	2001	-	
4	Greece_1	B	9	Melon	2006	-	
4	Greece_3	B	4	Bean	2006	-	
5	Israel_3	B	3	Pepper	2007	-	
5	Israel_4	B	4	Cotton	2007	-	
6	Yemen*	B relative	10	Tomato	2006	-	
7	Burkina Faso_3	Q relative	12	Cotton	1999	-	
7	Burkina Faso_4	Q relative	9	Tomato	1999	-	
7	Burkina Faso_6	Q relative	4	Tomato	2000	-	
8	Sudan-Q_like	Q relative	30	Cotton	1994	-	
9	Sudan	Q relative	6	Beans	1994	-	
10	Cyprus*	Q_{EAST}	30	Gerbera	2008	27	51
11	Greece_2	Q_{EAST}	17	Tomato	2006	-	
12	Israel_1992	Q_{EAST}	29	Cotton	1992 (colony until 2003)	-	
13	Israel_1	Q_{EAST}	6	Sunflower	2007	-	
13	Israel_2	Q_{EAST}	4	Cotton	2007	-	
14	Louisiana	Q_{EAST}	40	*Hibiscus*	2005	-	
15	Turkey	Q_{EAST}	29	Cotton	1994	-	
16	Alabama, USA	Q_{WEST}	31	Poinsettia	2005	-	
17	Arizona, USA	Q_{WEST}	8	Not available	2005	-	
18	Canary Islands_1	Q_{WEST}	5	Tomato	2009	-	
18	Canary Islands_2	Q_{WEST}	6	Tomato	2009	-	
18	Canary Islands_3	Q_{WEST}	11	Tomato	2009	-	
18	Canary Islands_4	Q_{WEST}	10	*Solanum nigrum*	2009	-	
19	China	Q_{WEST}	14	Not available	2007	-	
20	France	Q_{WEST}	25	Soybean	2007	-	
21	Morocco	Q_{WEST}	9	Tomato	Unknown	-	
22	Spain_2005	Q_{WEST}	30	Tomato	2005 (colony until 2008)	-	
23	Spain_2008*	Q_{WEST}	28	Zucchini	2008	18	48
23	Spain_2008	Q_{WEST}	28	Pepper	2008	-	

* denotes populations used in the IMa2 analysis

Table 2. Genetic properties of microsatellite loci used in each of the data analyses and literature source they were obtained from.

Locus	Allele size range (bp) (number of alleles)	Fluorescent dye color[1]	Repeat motif	Analysis used in	Reference
WF2C01	137-237 (20)	PET	$(GTTT)_{11}$imp	All	Schwartz unpublished
WF2H06	158-204 (15)	NED	$(TTTG)_{11}$	All except IMa2 (Q_{East}-Q_{West})	Schwartz unpublished
WF1B06	128-170 (12)	PET	$(ACTC)_8$	All except IMa2 (Q_{East}-Q_{West})	Schwartz unpublished
WF2E11	159-264 (22)	PET	$(GATT)_{27}$imp	All except IMa2 (Q_{East}-Q_{West})	Schwartz unpublished
WF1B11	103-178 (13)*	PET	$(CTTGA)_{12}$imp	Only IMa2 (biot. B-Yemen)	Schwartz unpublished
BEM6	161-230 (13)	6-FAM	$(CA)_8$imp	All except IMa2 (biot. B-Yemen)	De Barro et al (2003)
BEM15	186-219 (9)*	6-FAM	$(CAA)_6(CAG)_4(CAA)_4$	Only IMa2 (biot. B-Yemen, Q_{East}-Q_{West})	De Barro et al (2003)
BEM31	105-136 (7)	HEX	$(GCT)_4(GTT)_2$	All	De Barro et al (2003)
BT-b103	131-151 (10)	HEX	$(AC)_8(TC)_3$	All	Tsagkarakou & Roditakis (2003)
BT-e49	272-390 (34)	6-FAM	$(TTG)_{12}(TTC)_{11}$	All	Tsagkarakou et al (2007)
Bt*Is*1.2	273-368 (41)	6-FAM	$(CA)_{13}N_{14}(CA)_8$	All	Gauthier et al (2008)
MS145	173-225 (13)	PET	$(AC)_9$	All	Dalmon et al (2008)
MS177	233-278 (16)	6-FAM	$(CA)_7$	All	Dalmon et al (2008)

[1] 6-FAM, PET, NED (Applied Biosystems), 6-FAM, HEX (SIGMA-ALDRICH)

* allele size range corresponds only in populations used in the IMa2 analysis

Table 3. F_{ST} values between pairs of biotypes (from multiple pooled populations) in ascending order. The asterisk indicates the pair of populations used in the IMa2 analysis.

Biotype/lineage pair	F_{ST}
*Q_{EAST} - Q_{WEST}	0.2142
*B relative - B	0.2264
Q relative - Q_{WEST}	0.2588
Q relative - Q_{EAST}	0.2721
B relative - Q relative	0.2790
B relative - Q_{WEST}	0.3531
B relative - Q_{EAST}	0.3556
Q relative - B	0.3847
Q_{EAST} - B	0.4046
Q_{WEST} - B	0.4351

Table 4. Summary statistics of the mtCOI for the three populations from biotypes Q_{WEST}, Q_{EAST}, and B.

	n	Haplotype diversity *h* (±SD)	Nucleotide diversity π	Tajima's *D*	Fu's *Fs*
Spain "Q" (Q_{WEST})	18	0.667	0.002	-1.720*	-1.564
Cyprus "Q" (Q_{EAST})	27	0	0	0	0
Cyprus "B"	17	0	0	0	0

*P<0.10

Figures

Fig. 1. Principal Components Analysis of individual genotypes from biotypes B, Q_{WEST}, Q_{EAST}, B relative, and Q relative. F_{ST}s ±SE represent values averaged from pairwise comparisons among all populations within each group.

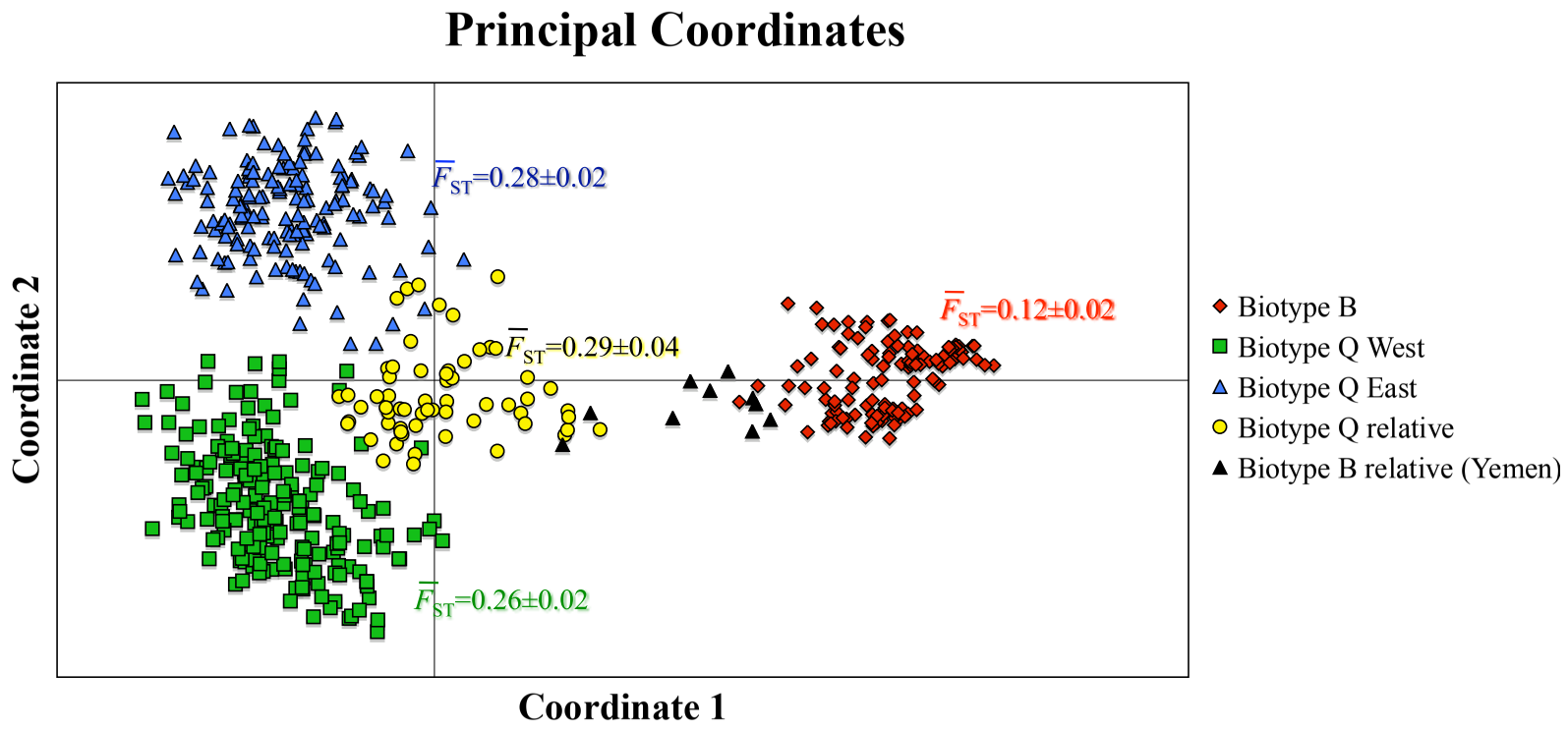


Fig. 2. Unrooted neighbor-joining tree based on F_{ST} values calculated among all pairs of populations used in this study.

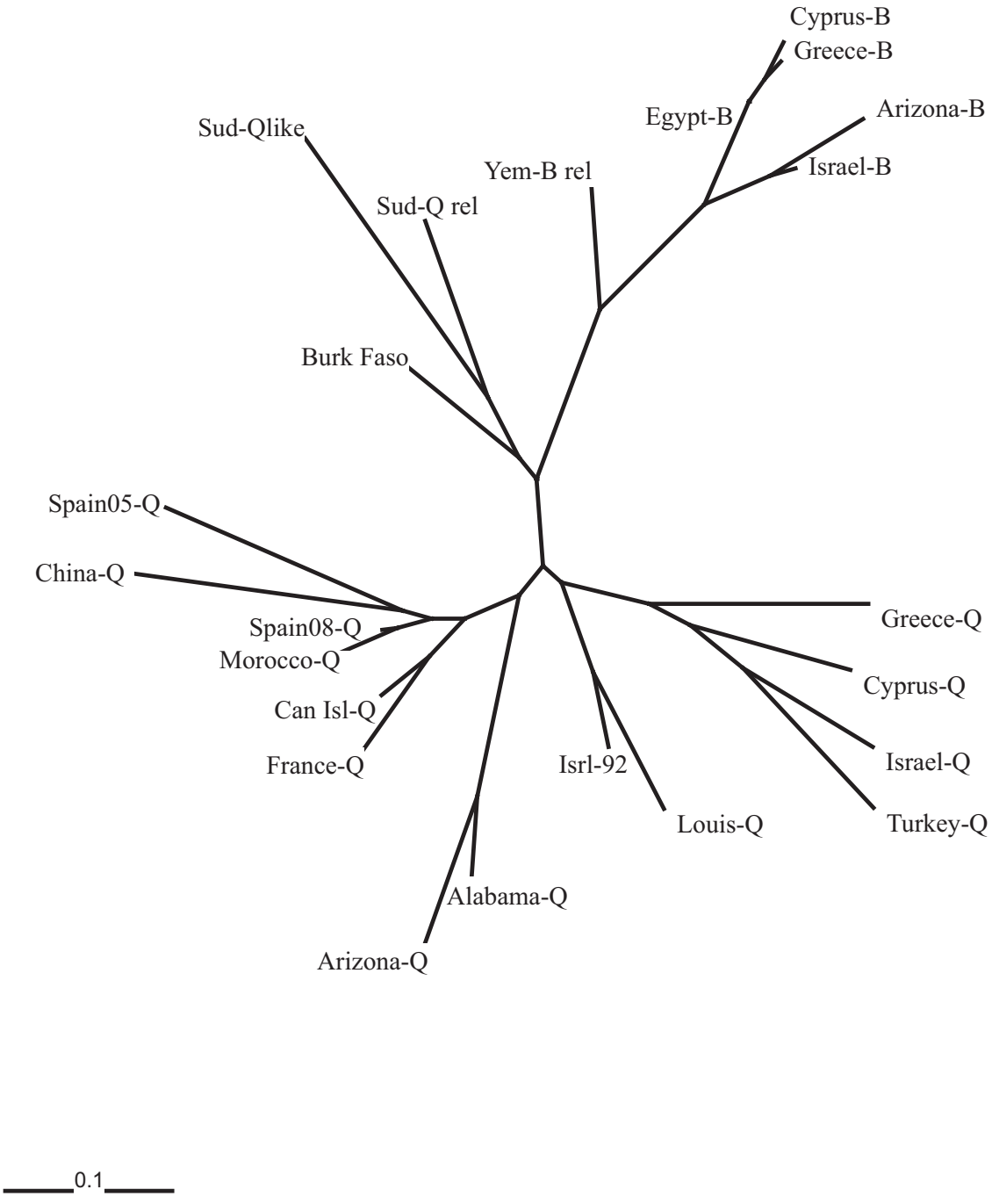


Fig. 3. Divergence history of biotypes B and Q using the IM model. The left figure shows results for the Q_{WEST} - Q_{EAST} analysis, and the right figure for biotype B-Yemen. Sampled and ancestral (*Ne*) populations are shown as dark blue boxes. Horizontal grey lines show splitting times (*t*) (million years) and the red curved arrows show migration rates (2*NM*) from the source population to the receiving population forward in time. The 95% highest posterior density intervals are represented with arrows and light blue boxes for population sizes and as grey dotted lines for splitting times. The red asterisks show 2*NM* rates found to be statistically significant by the likelihood-ratio test of Nielsen & Wakeley (2001): ***P*<0.01, ****P*<0.001.

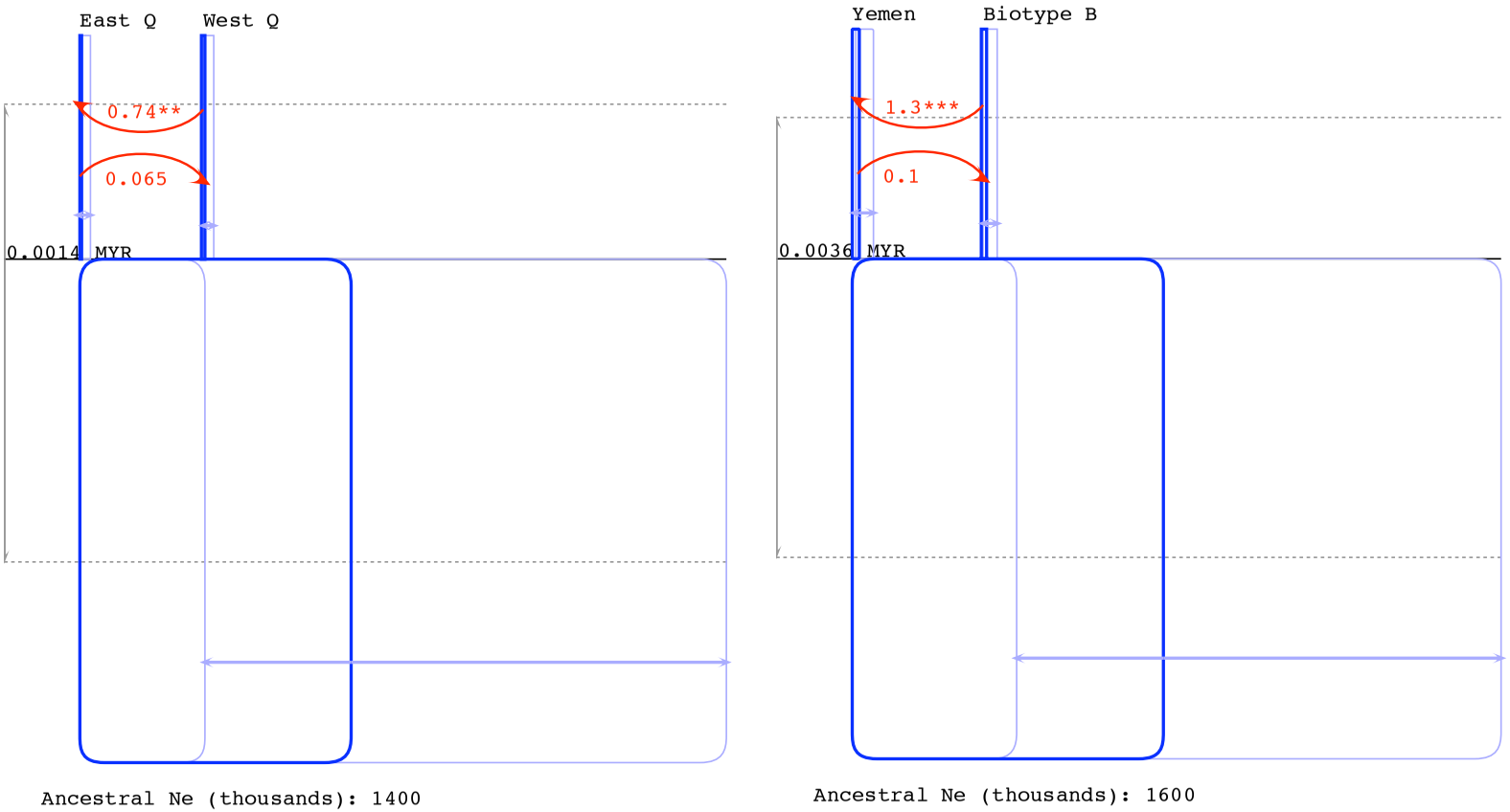


Fig. 4. Median-Joining (MJ) haplotype network of the mitochondrial COI sequences of individuals from Cyprus B, Cyprus Q, and Spain Q populations. The red numbers connecting branches represent substitutions in the mtCOI sequence. The small red diamonds are median vectors; hypothesized (often ancestral) sequences required to connect existing sequences within the network with maximum parsimony (NETWORK 4.5.1.6 manual). The number of individuals per haplotype are given either inside or next to the haplotype.

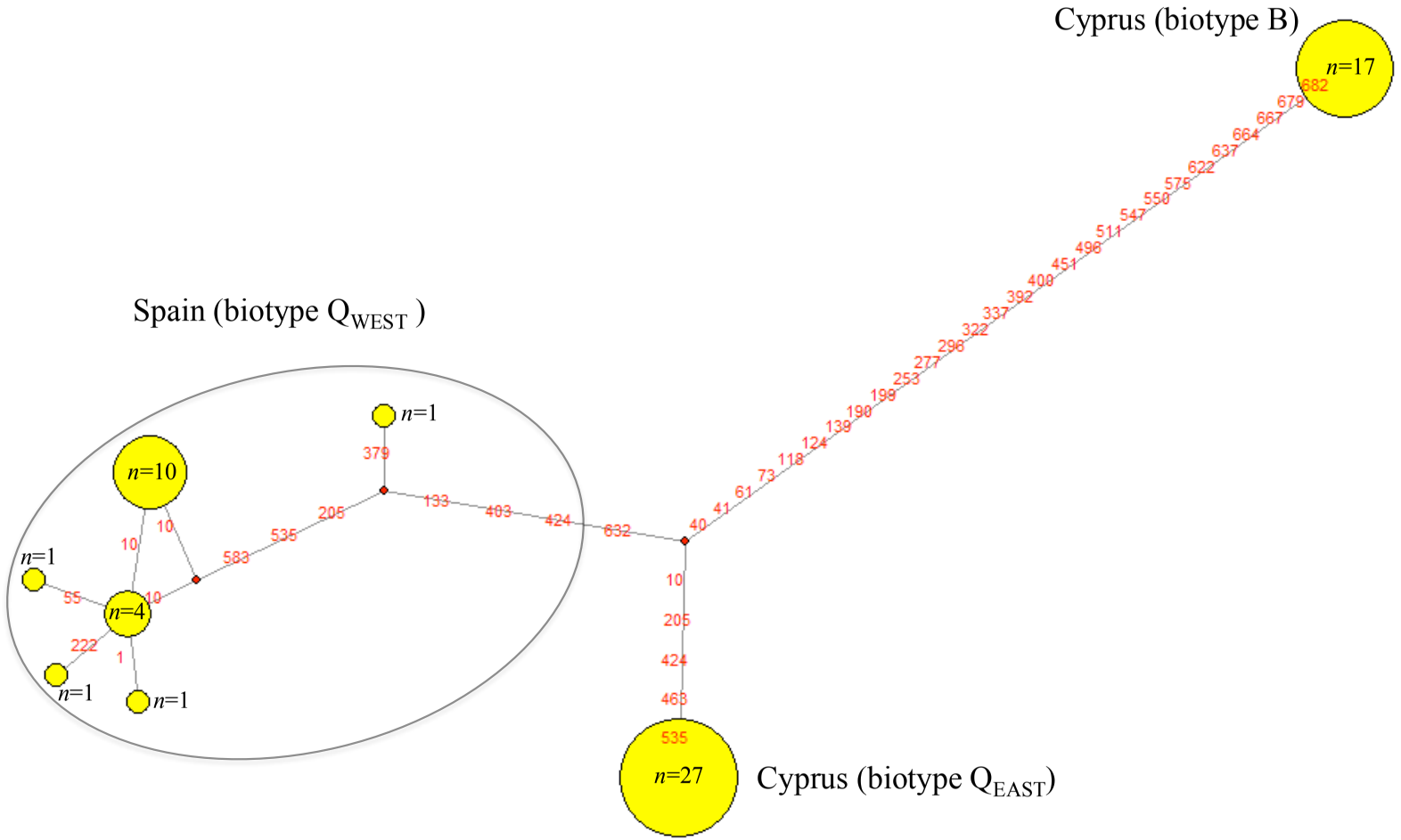
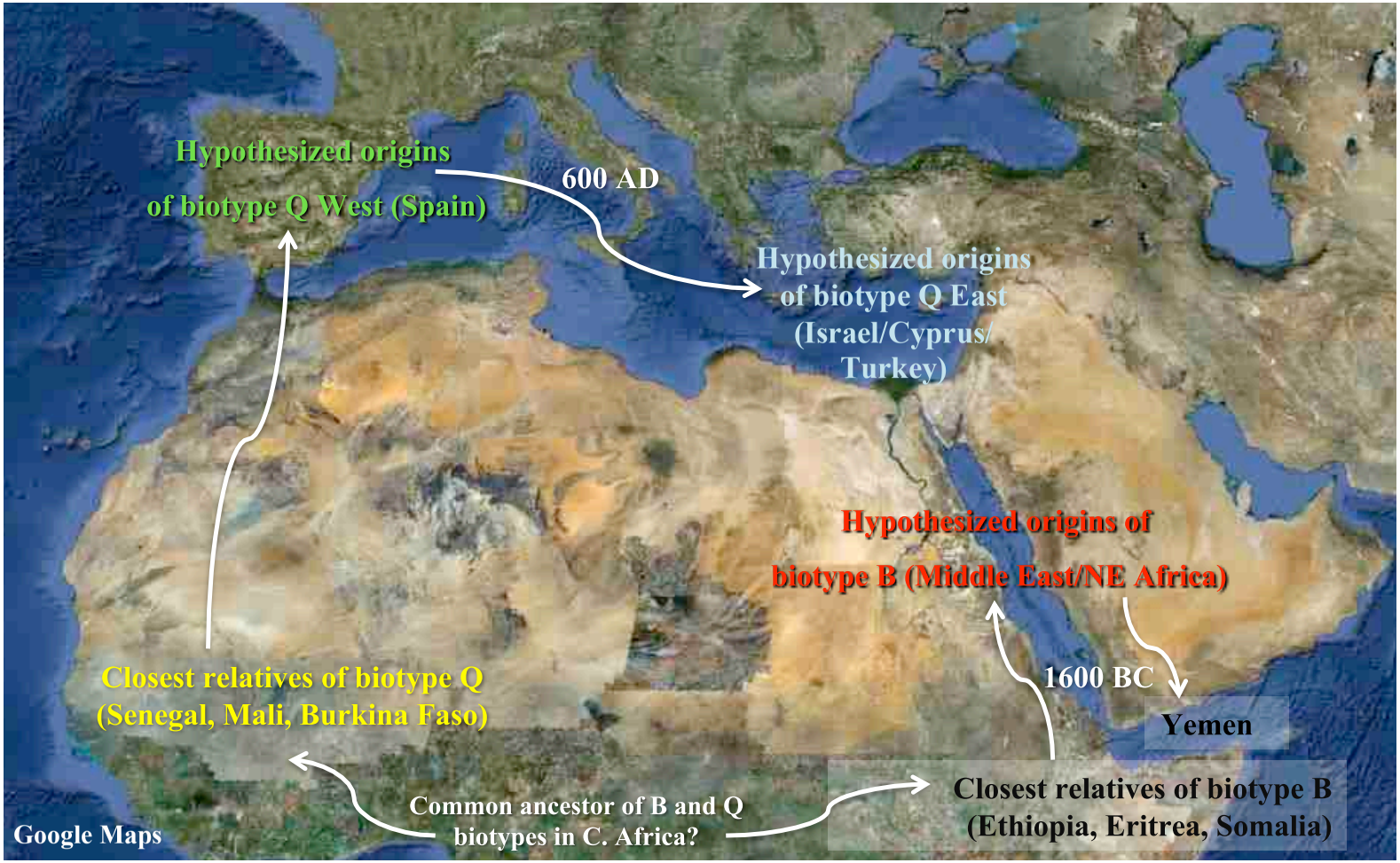


Fig. 5. Map showing the presumed origins of biotypes B and Q and their closest relatives in the Mediterranean. White arrows show the hypothesized patterns and routes of diversification of these biotypes and their closest relatives from their ancestors.



Chapter 4

Evidence for multiple independent and cryptic invasions of the whitefly *Bemisia tabaci* biotype Q in the USA

Abstract

The field of invasion genetics has emerged as a central research area in the invasive species literature, providing insights into the processes and genetic changes that facilitate the successful establishment of introduced populations. Bayesian assignment methods, as well as the recently explored approximate Bayesian computation (ABC) approaches in this field have made it possible to address challenging questions about the complex demographic histories of invaded populations. We applied these approaches to examine the invasion history of the whitefly *Bemisia tabaci* biotype "Q" in the USA. This biotype was first detected in Arizona in 2004; a finding that prompted a collaborative effort of multiple agencies to a nationwide survey that found this biotype in 25 states in greenhouses and nurseries. The high alertness against this pest stemmed from the earlier detrimental invasion of another biotype ("B"), in the 1980's that led to losses of more than $1 billion in agriculture. This study aimed to trace the invasion history of this biotype in the USA by examining samples collected from eight states in 2005, as well as potential source populations sampled from the Mediterranean native range. We assessed genetic variation in 13 microsatellite loci using Bayesian assignment approaches and in the mitochondrial cytochrome oxidase I (COI) gene using maximum parsimony network analyses. We found that these populations originated from both the Western and Eastern Mediterranean regions, through at least three independent cryptic invasion events, likely from Spain, and the area around Israel and Cyprus. The relatively high genetic diversity in these populations compared to those from the native range suggest that these invasions are likely quite recent, and have not yet resulted in severe genetic bottlenecks. The ABC analysis suggested that, if the Western Mediterranean invasions were first introduced to California as prior information suggested, they most likely spread across other states directly in independent events, not from state to state. Because a single grower in California handles 70% of the US poinsettia production and our US samples came primarily from these plants, these findings implicate the trade of poinsettias as the means of dispersal of this pest. Our results highlight the need for monitoring *B. tabaci* and other insect pest and vector invasions at points of entry and dispersal of associated host plants. In addition, they broaden our understanding of how insects travel on plant material through international corridors, spread from the initial sites of introduction, and colonize new habitats.

Introduction

The prevalence and importance of invasive species has increased dramatically in the past 50 years as a result of human population growth, movement and traveling of people, global trade, and environmental change (Pimentel et al. 2005). In contrast to natural colonizations that are usually untraceable, human-induced invasions provide the opportunity to study ecological, evolutionary, and genetic changes in the introduced range from the initial stage of an invasion. Understanding how genetic diversity contributes to potential for colonization, geographic expansion, and success of the invading populations may facilitate and improve their successful management (Sakai et al. 2001). Such studies in the newly emerged field of "invasion genetics" have expanded with the exploitation of highly variable markers (Estoup et al. 2004) and population genetics approaches. For example, research has shown that a typical introduced population suffers from decreased genetic diversity; a result of the bottleneck induced by founder events (Nei 2005; Nei et al. 1975). This reduction in diversity however, may be compensated by repeated invasions from multiple locations, which result in admixture of genetic variation (Kolbe et al. 2007), increasing the evolutionary potential for adaptation in the introduced range (Lee 2002; Sakai et al. 2001). Such cases have been documented in several taxa, including lizards (Kolbe et al. 2004), fish (Collins et al. 2002), grasses (Novak and Mack 1993), trees (Pairon et al. 2010), nematodes (Herrmann et al. 2010), and insects (Thomas et al. 2010).

More recently, sophisticated population genetics approaches, such as Approximate Bayesian Computation (ABC) (Beaumont et al. 2002) have been applied to problems of invasion genetics, paving the way to address more complex and interesting questions about the demographic history of biological invasions (Cornuet et al. 2008; Guillemaud et al. 2010). For example, with the application of such model-based methods one can ask: where did an invasive population come from (e.g. Lombaert et al. 2010; Miller et al. 2005)? Did it originate from a single or multiple geographically discrete source populations? ABC approaches can also address whether the invasion was the result of a single isolated invasion wave or serial independent events and can help trace invasion pathways (e.g. Lombaert et al. 2010; Miller et al. 2005; Pascual et al. 2007). Such knowledge is interesting for evolutionary biologists and geneticists, as well as managers working on the implementation of efficient quarantine measures. Examples of successful application of such approaches come primarily from studies of invasive insects – beetles (Lombaert et al. 2010) and *Drosophila* flies (Miller et al. 2005) – but also amphibians (Estoup et al. 2004) and birds (Estoup and Clegg 2003).

One of the world's worst invasive insects that has caused unprecedented damage in the last 30 years to vegetable and ornamental crops globally is the whitefly *Bemisia tabaci* Gennadius (Hemiptera: Aleyrodidae). With a host range of more than 500 plant species from 74 families (Brown et al. 1995), this pest injures plants by both direct feeding and by transmission of plant viruses. The species comprises a complex of more than 25 cryptic biotypes at a global scale (Boykin et al. 2007; Brown et al. 1995; Frohlich et al. 1999; Perring 2001). Of all biotypes characterized so far based on esterases, mtDNA, and biological/ecological assays, two are commonly known as worldwide invasive pests: biotypes B and Q. Biotype B was first reported as an economically important pest in the US in 1986 (Hamon and Salguero 1987). Possibly an introduction

from Israel through Florida, this invasion resulted in losses of more than half a billion dollars in US agriculture in 1991 (Perring et al. 1993). Since then, and following the development of molecular techniques that discriminated it from other biotypes (Costa and Brown 1991), its spread was confirmed across most continents, rendering it one of the most economically important pests worldwide.

The first description of biotype Q came in 1996, following genetic surveys in the Iberian Peninsula, which showed distinct RAPD and esterase patterns in populations from Majorca, Spain, and Portugal, compared to sympatric populations of biotype B (Banks et al. 1998; Guirao et al. 1997). Subsequent studies showed that the Q biotype was associated with elevated insecticide resistance and within 10 years, it was detected across the Mediterranean basin, from the Canary Islands (Beitia et al. 1997) to Israel (Horowitz et al. 2003) and as far as The Netherlands in Central Europe (Chu et al. 2008). Because the Q biotype was initially found in the Mediterranean only, it has been suggested that this area represents its native range and subsequent detections around the world represent recent invasions (Brown 2007). In the last five years, this biotype was reported from China (Chu et al. 2006), Japan (Ueda and Brown 2006), Korea (Lee et al. 2007), New Zealand (Scott et al. 2007), Mexico (Martinez-Carrillo and Brown 2007), Guatemala (Bethke et al. 2009) and the USA (Dennehy et al. 2006).

In the United States, biotype Q was first discovered in at a retail nursery in Tucson, Arizona in December 2004, and was identified independently by two different laboratories (Dennehy et al. 2005; Dennehy et al. 2006). Bioassays of these whiteflies demonstrated strikingly reduced susceptibility to novel insecticides that are still efficient against the already established invasive biotype B, leading to great concerns about its potential impact in agriculture (Dennehy et al. 2006). Thus, this discovery prompted a nationwide survey in a collaborative effort of several agencies (Universities of California, Arizona, and Florida, and USDA ARS) that led to its detection in 25 states since 2005 (http://mrec.ifas.ufl.edu/LSO/bemisia/bemisia.htm) in retail shops, greenhouses, and nurseries (Fig 1). Initial assessments from researchers at The University of Arizona suggested that this biotype was introduced through California from poinsettia cuttings prepared in Guatemala, Honduras, or Mexico and shipped to the US (Bethke et al. 2009; Dalton 2006). The first infested poinsettias found in the Tucson market were in fact traced back to California growers, supporting the scenario that California was the entry point of this biotype in the US (Dalton 2006). While the immediate source of introduction was suspected to be Central America, this area was likely just an intermediate source since this biotype was only recently reported from there (Bethke et al. 2009) and its native range is the Mediterranean. A possible scenario is that whiteflies from the Mediterranean were introduced into Central America on poinsettia cuttings, and subsequently transported to California. A recent study that examined samples across Florida showed the presence of three unique mitochondrial haplotypes, two of which were found in mixed populations, suggesting that biotype Q entered Florida through at least two independent introductions (McKenzie et al. 2009). However, the exact sources of introduction from the Mediterranean and pathways of dispersal across the US have not been thoroughly examined. The use of polymorphic multi-locus markers such as microsatellites, analyzed with population genetics approaches allows for more robust inferences to be made in answering such questions (Davies et al. 1999).

The aim of this study was to determine the origins of the US invasion of biotype Q and the pathways of subsequent dispersal across several states. We approached this problem using a combination of mitochondrial DNA and nuclear microsatellite markers in a population genetics framework. We assessed genetic variation in several samples collected from nurseries and greenhouses across eleven locations in the US, fourteen locations in the Mediterranean native range and a location in China. The US samples were collected around the time of introduction in 2005. We compared levels of genetic diversity in the introduced and native range to evaluate the effects of founder events in the invading populations. With the application of Bayesian assignment methods as well as ABC approaches we traced the invasion history of biotype Q by determining the sources of origin of the invading populations as well as possible routes of dispersal across the US. Because our sampling does not include whiteflies from Central America, we do not address the hypothesis that the California introduction was *via* Central America; therefore our aim was to associate introduced populations in the US with potential original sources from the Mediterranean native range, even if Central America acted as an intermediate source of this invasion.

Materials and Methods

Samples and DNA extraction

A total of 410 female whiteflies from 27 locations were used in this study (Table 1). Whiteflies were preserved in 95% ethanol until extraction. Genomic DNA was extracted from whole whiteflies homogenized in 1.5 ml microcentrifuge tubes, using the Qiagen DNeasy DNA Blood and Tissue kit following the manufacturer's protocol. This protocol was slightly modified to perform a final elution from the column in 80µl buffer AE followed by a second elution in 20µl AE, with the two eluates combined in a single microcentrifuge tube and stored at -20° C.

Microsatellite loci

We genotyped all individuals across 13 microsatellite loci obtained from several literature sources (Dalmon et al. 2008; De Barro et al. 2003; Gauthier et al. 2008; Schwartz unpublished; Tsagkarakou and Roditakis 2003; Tsagkarakou et al. 2007) using polymerase chain reaction (PCR). Forward primers were labeled with a fluorescent dye (Table 2). PCRs were done either as single primer-pair reactions or by multiplexing 3-5 primer pairs with different fluorescent dyes or amplifying alleles with discrete size ranges, using a multiplex PCR scheme. Single primer pair reactions were done in a total of 10 µl volume with 1 µl of genomic DNA, 0.75 mM dNTPs mix, 0.5 µM of each primer, 1 µl of 0.1mg/ml bovine serum albumin (BSA, New England BioLabs), 1.25 µl 10x PCR Buffer (Applied Biosystems), 1.125 mM $MgCl_2$ (ABI), and 0.5 U of AmpliTaq® Gold DNA Polymerase (ABI).

Amplification was done using a touchdown protocol with the following thermocycling conditions: 95°C for 10 min (1 cycle) followed by 94°C for 1 min, 60°C for 1 min, 70°C for 35 s (2 cycles), followed by 93°C for 45 s, 59°C (lowering 0.5°C each cycle) for 45 s, and 45 s at 70°C (18 cycles). The final amplification step consisted of 20 cycles at 92°C for 30 s, 50°C for 30 s, and 70°C for 1 min, followed by a final

extension step at 70°C for 5 min. The multiplex PCRs were done in 96-well plates using the Multiplex PCR Kit (Qiagen) in 15 µl reaction volumes containing 1 µl of genomic DNA, 6 µl of the Qiagen PCR Master Mix (1X), 2 µM of each primer and 3 µl RNase-free water. Thermocycling conditions for this protocol were: 15 min at 95°C (1 cycle) followed by 94°C for 30 s, 60°C for 90 s, and 72°C for 60 s (30 cycles), with a final extension step of 60°C for 30 min.

For fragment analysis the PCR products were added to a mix of 48:1 Hi-Di formamide (ABI): LIZ 500 size standard (ABI) (0.5 µl PCR product, 0.2 µl LIZ, 9.3 µl formamide) and were denatured at 95°C for 5 min, followed by 5 min on ice. Fragment analysis was performed in a 3730 DNA sequencer and data were visualized and scored using the software GeneMapper version 4.0 (ABI). All PCR runs included negative and multiple positive controls to ensure consistency in allele scoring. The program Flexibin (Amos et al. 2006) was used to facilitate binning and help detect miscalled microsatellite alleles.

Mitochondrial DNA sequencing

A region of approximately 800 bp of the mitochondrial cytochrome oxidase subunit I (COI) gene was amplified from 131 individuals randomly selected from 21 populations (Table 1) and sequenced in both directions using the primers C1-J-2195 (5-TTGATTTTTTGGTCATCCAGAAGT-3) and L2-N-3014 (5-TCCAATGCACTAATCTGCCATATTA-3) (Simon et al. 1994). Reactions were performed in 25 µl volumes using 2 µl of genomic DNA, 0.8 mM dNTPs mix, 0.5 µM of each primer, 2.5 µl 10x PCR Buffer (Applied Biosystems), 1.2-1.5 mM $MgCl_2$ (ABI), and 0.4 U of AmpliTaq® Red DNA Polymerase (ABI). Thermocycling conditions were as follows: 2 min at 95°C (1 cycle) followed by 30 cycles of 95°C for 1 min, 40-52°C (different annealing temperature for each population is specified in Table 1) for 1 min, and 72°C for 1 min, with a final extension step of 72°C for 20 min. PCR products were purified using *ExoSAP-IT* (USB Corporation) and sequenced in both directions in 10 µl volumes using a cycle-sequencing protocol which included 2 µl purified PCR product, 1 µl BigDye v3.1 (ABI), 1.5 µl 5X sequencing buffer, 5.1 µl deionized purified water and 0.4 µl primer (forward or reverse depending on direction to be sequenced). Sequencing was performed in a 3730 ABI DNA sequencer, after purification of cycle-sequencing products using Sephadex (Fisher Scientific). All sequences were manually edited using the software Sequencher v4.6 (Gene Codes Corporation) and Geneious and aligned using MacClade 4.06 OS X (Maddison and Maddison 2003). After alignment, all sequences of the COI gene were trimmed to 688bp to eliminate missing data. Sequences will be submitted to GENBANK.

Genetic comparison of populations in the introduced and native range

Prior to any microsatellite analysis, exact tests of population differentiation were performed on populations collected from neighboring locations around the same time, using the method implemented in Genepop (version 4.0.10) with the default settings (Raymond and Rousset 1995; Rousset 2008). Populations that were not significantly differentiated after standard Bonferroni correction for multiple comparisons (at the 0.05 level) were pooled into a single "sample population" (Table 1). Locus specific statistics across all populations (number of alleles, total expected heterozygosity, mean observed

H_O and expected H_E heterozygosities, and F statistics) were calculated using the program GenAlEx 6.1 (Peakall and Smouse 2006). Pairwise F_{ST}s were calculated in the program Arlequin version 3.5 (Excoffier et al. 2005) using 10000 permutations for the significance test. We tested for significant genotypic linkage disequilibrium (LD) among all pairs of loci and for significant deviations from Hardy-Weinberg equilibrium (HWE) in the program Genepop (version 4.0.10) using the Markov chain method and with the default settings with 10000 iterations. Significance levels were adjusted with a Bonferroni correction at the 0.05 level. The program FreeNa (Chapuis and Estoup 2007) was used to calculate unbiased global and pairwise F_{ST} of Weir (1996) using the ENA correction for null alleles with 1000 bootstrap replications over loci. The calculated global F_{ST} values before and after the ENA correction were compared to assess the degree to which null alleles could bias our other analyses.

To assess genetic diversity of populations in the introduced (US and China) and the native range (Mediterranean) we compared levels of heterozygosity (H_E, H_O) and number of private alleles (alleles unique to a single population as calculated in GenAlEx). We examined private alleles found in the Mediterranean populations and assessed which of those were found in the introduced populations to determine potential sources of invasion. Because estimates of allelic richness can be biased by the number of samples per population, we calculated allelic richness (number of alleles) using the rarefaction method implemented in HPrare (Kalinowski 2005) to correct for different sample sizes by standardizing samples to the minimum eight genes per population. For this analysis we excluded one locus (WF1B06) with missing data due to poor amplification in populations from the Eastern Mediterranean. We compared the number of alleles, *He*, and *Ho* between populations using one-way analysis of variance (ANOVA) followed by Tukey's multiple comparison test, using the statistical software R v. 2.10.1 (R core development team 2009).

We also performed analyses of molecular variance (AMOVA) in the introduced and native range separately, to determine hierarchical levels of genetic structure within populations, among populations, and among groups of populations sampled from different geographic locations. The among groups level of structure was based on analysis in STRUCTURE as explained below that showed an evident geographic structure in two clusters in the native range (Eastern and Western Mediterranean) and clear assignment of introduced samples in either of the two clusters. Analyses were done using the locus-by-locus AMOVA option based on the number of different alleles (F_{ST}-like) in the program Arlequin (Excoffier et al. 2005) with 10,000 permutations. Calculations were done using 12 loci (excluding locus WF1B06 that had more than 5% missing data in populations from the Eastern Mediterranean).

Population genetic structure and mitochondrial DNA diversity

We used the Bayesian clustering approach implemented in the program STRUCTURE 2.3.3 (Falush et al. 2003; 2007; Hubisz et al. 2009; Pritchard et al. 2000) to identify levels of genetic differentiation in all populations together (introduced and native range). For this analysis we used 10 microsatellites, excluding three loci that significantly deviated from HWE in more than one population (BEM6, BTb103, MS11). We ran the analysis using a burn-in length of 100,000 and a run length of 1,000,000 steps, with the admixture and the correlated allele frequencies models without prior

population information, specifying the number of clusters *K* to be tested from 1 to 15, with five replicates per *K*. Because we observed that this method can detect the higher levels of genetic structure and may not detect substructure if there is a strong signal of genetic differentiation among major clusters, we also ran the program independently for each of the major clusters detected in the first analysis. The loci that deviated from HWE in those populations used for each run were excluded as described under each plot in the results. The *K* that best described the observed structure in the data for all runs was inferred by examining both the posterior probabilities of the data for each *K* (where the highest difference in LnP(D) is observed between successive K values and the plot starts to plateau) and the ΔK estimator (highest ΔK) described by Evanno et al. (2005). The program Harvester (Earl 2009) was used to assist the ΔK estimation and produce the relevant plots. Results from the five replicates for the inferred *K* were analyzed in the program CLUMPP (Jakobsson and Rosenberg 2007) to produce averaged matrices of individual and population cluster membership coefficients. Finally, graphical displays of the resulting barplots were produced using the program distruct v1.1 (Rosenberg 2004).

In the mitochondrial DNA sequence data, we calculated haplotype diversity (h) and nucleotide diversity (π) (Nei 1987) in all samples (sequences pooled into each collection region) using the program DNAsp v. 5 (Rozas et al. 2003). In order to examine the distribution of mtDNA sequence diversity in the introduced and native range we calculated a maximum parsimony (MP) median-joining (MJ) network (Bandelt et al. 1999) using a variable sites matrix in the program NETWORK 4.5.1.6. We used the default settings of equal connection weights (transition : transversion ratio 1:1) to all sites and epsilon value = 0.

Potential sources of invasion in the USA

In order to identify potential source populations for each of the individuals in the introduced range we performed assignment tests using the Bayesian assignment approach of Rannala & Mountain (1997) as implemented in the program GENECLASS 2 (Piry et al. 2004) with a threshold of 0.05. The likelihoods of assignment scores for each individual's multi-locus genotype *i* into sample *s* ($L_{i\rightarrow s}$ expressed on a –log scale) were averaged over the sampled population to determine the most likely source of each of the introduced populations (the one with the highest $L_{i\rightarrow s}$). We also examined the actual individual assignments (before averaging them) to each of the putative sources and present results as the number of individuals assigned to the different native range source populations. Because the assignment methods assume that the actual source population has been sampled and will attempt to assign all individuals to one of the reference populations included in the analysis, exclusion tests can also be used as a stricter test to exclude a population as a potential source of an individual. For this test we used the simulation algorithm of Paetkau et al. (2004) in GENECLASS 2 because it takes into account the sample size of the reference population during resampling (Paetkau et al. 2004; Piry et al. 2004), and we ran the program using 1,000,000 simulations and an alpha level of 0.05. All assignment and exclusion tests were done using 12 loci (excluding WF1B06 with >5% missing data).

Identification of potential dispersal pathways of introduced populations in the US using Approximate Bayesian Computation

We used the program diyABC v1.0 (Cornuet et al. 2010; Cornuet et al. 2008) to compare seven competing scenarios of dispersal among five introduced populations across the US. These five populations were chosen based on pairwise F_{ST}s and results from STRUCTURE, which suggested that they originated from a single invasion event. Of these, populations N.York and N.Carolina-1 were pooled as one sample because of negative pairwise F_{ST}. The scenarios to be compared included a scenario of independent dispersal events from California to Michigan, Alabama, and N.York/N.Carolina-1, *vs* 6 scenarios of all possible combinations of sequential dispersal from California to each of the other states (Fig. 1). We chose the default prior distributions and the Generalized Stepwise Mutation model, and uniform distributions for the parameters "time" (t) and "effective population size" (N). For these parameters we set intervals at (time) with intervals at t=1-100 and N=1-10000. We selected the following summary statistics to measure similarity between simulated and observed data: a) mean number of alleles, b) mean gene diversity across loci, c) mean allele size variance across loci, d) mean M index across loci, e) F_{ST} between two samples, f) mean index of classification (two samples) (for details on each statistic see diyABC manual, Cornuet et al. 2010). We simulated 1,000,000 data per scenario. The resulting scenarios were compared by computing their posterior probabilities, selecting the 500 of the simulated data closest to the observed for the direct estimate approach and 70,000 for the logistic regression approach. The first estimate is the relative proportion of each scenario in the simulated datasets whereas the second is the logistic regression of each scenario probability on the deviations between the simulated and observed summary statistics (Cornuet et al. 2010 and references therein). Both estimates are useful in assessing support for the competing scenarios since the logistic regression is more discriminant but can be unstable with low numbers of closest datasets, and the direct estimate is more sensitive to the number of closest simulated datasets. We assessed support for each scenario by examining both the direct estimate output graph as well as the logistic regression graph.

Results

Genetic comparison of populations in the introduced and native range

We found high allelic richness across the 13 loci, with number of alleles per locus ranging from 5 to 28, resulting in a total of 155 alleles across 410 individuals (Table 2). Across the 5330 genotypes we tried to amplify only 148 were coded as missing after at least two failed amplification attempts. Most of these failed amplifications (85%) occurred at a single locus (WF1B06). Specifically, this locus failed to amplify completely in certain populations from the Eastern part of the Mediterranean (Israel-2007, Turkey, Cyprus) and two US populations (Louisiana, N. Carolina-2) but amplified successfully in all individuals from Israel-Q and in a few individuals from Greece. This locus was excluded from analyses where missing genotypes at more than >5% of the data are not allowed, as described in the Materials and Methods section.

Tests for population differentiation revealed four sets of samples, each collected from neighboring locations, with non-significant differentiation ($P > 0.05$ after

Bonferroni correction) and were pooled together resulting in a total of 19 populations to be further analyzed (Table 1). Tests for HWE showed that five loci deviated from equilibrium expectations, but only three of those deviated in more than one population (Table 2) and were excluded from analyses that assume HWE. When patterns of deviation were examined at the population level, eight populations deviated significantly from HWE in at least one locus (Table 1). No significant linkage disequilibrium was detected between pairs of loci.

Analysis for null alleles showed that despite the presence of null alleles in our dataset, estimates of global F_{ST} using the corrected and uncorrected datasets were very similar. The F_{ST} across all loci based on the corrected dataset was 0.31 (95% CI: 0.23-0.40) whereas the estimate based on the non-corrected dataset was 0.30 (95% CI: 0.23-0.39). Per locus F_{ST} estimates were also very similar between corrected and uncorrected datasets, with the highest difference at a locus being equal to 0.04 at locus WF1B06. This locus was also problematic due to poor amplification in some populations and as explained earlier it was excluded from several analyses.

Although populations from the Mediterranean native range of this biotype had generally higher levels of H_E across loci than those from the introduced range (US and China), differences were not statistically significant ($F_{18,228}$=1.32, P=0.17) (Fig. 2). The lowest levels of H_E were observed in the Arizona population (H_E=0.2) from the introduced range and the highest in Greece (H_E=0.48) from the native range. Populations from Louisiana and N. Carolina-2 in the introduced range had relatively high levels of H_E compared to all other populations. Observed heterozygosities (H_o) were also generally lower in the introduced range, with the exception of Louisiana and N. Carolina-2 that had the highest levels of H_O among all populations analyzed in both the introduced and native range. Differences in H_O were statistically significant ($F_{18,218}$=2.14, P=0.005) with the Arizona population having significantly lower H_O compared to populations from the Canary Islands, Cyprus, France, Israel-Q, Louisiana, Morocco, N.Carolina-2, and Spain-2008 (P<0.05) (Fig. 2).

Examination of the average number of alleles across loci showed similar patterns. With the exception of Louisiana, N.Carolina-2, and China, all other populations from the introduced range had lower number of alleles compared to the native range but the differences were not statistically significant ($F_{28,199}$=1.16, P=0.27) (Fig. 3). Private alleles absent from the native range populations were detected in three populations; one in Louisiana (frequency=0.037), one in N.Carolina-2 (frequency=0.071) and two (at two different loci) in Arizona (at frequencies of 0.125 and 0.625). Examination of private alleles from the native range (private alleles detected when only the native range populations were analyzed) showed their presence in introduced populations from Michigan, Louisiana, and N. Carolina-2, with some of them in high frequencies (Table 3). The respective Mediterranean populations bearing these private alleles were Spain-2005, Spain-2008, France, Cyprus, Israel-Q, and Canary Islands.

Analysis of molecular variance showed high and similar levels of within population differentiation in the introduced (F_{ST}=0.35) and native range (F_{ST}=0.36). The remainder of the genetic variance however was mostly partitioned between the 2 genetic clusters (Eastern and Western Mediterranean lineages) in the introduced range (28.23%) with only 6.9% of the variance explained by differences among populations within these clusters (Table 4a). In contrast, in the native range populations that variance was almost

equally partitioned between the "among groups" and "among populations within groups" components (18.69% and 17.08% respectively), with the remainder variance explaining the "within populations" component (Table 4b).

Pairwise F_{ST}s between populations introduced from the Western Mediterranean showed very low and non-significant differentiation between California and all other populations except for Arizona, and between N.York and all other populations except for Alabama and Arizona (Table 5). All populations had high and significant F_{ST}s with Arizona, ranging from 0.11 to 0.32. Pairwise F_{ST} between N.Carolina-2 and Louisiana (introduced from the Eastern) Mediterranean was 0.182.

Population genetic structure and mitochondrial DNA diversity

Results from the STRUCTURE analysis including all populations showed a clear split into two genetic clusters (K=2) (Fig. 4, Fig. 5a,b). In the native range these two clusters correspond geographically to the Western (Canary Islands, Morocco, Spain-2005, Spain-2008, France) and Eastern (Greece, Cyprus, Israel-Q, Israel-2007, Turkey) part of the Mediterranean, with some levels of admixture between them. While most individuals had admixed ancestry (Q) greater than 0.8, some individuals from China, Greece, Morocco, and the Canary Islands had Q values between 0.5-0.7. Individuals sampled from the introduced range (US and China) were clearly assigned to each of the two genetic clusters with individual assignments Q>0.96. Specifically, individuals from Alabama, Arizona, California, Michigan, N.York, N.Carolina-1 and China were assigned to the Western Mediterranean cluster. Individuals from Louisiana and N.Carolina-2 were assigned to the Eastern Mediterranean cluster, with the exception of one individual from Louisiana assigned to the Western Mediterranean cluster (Q=0.99) (Fig. 4). There was no evidence for admixture between the two clusters in populations from the US, as was observed in the native range populations, with the exception of China that shared some ancestry with the Eastern Mediterranean cluster.

When genetic structure was examined within each of the two clusters to examine possible substructure there was a clearer picture of population structuring. Within the Western Mediterranean cluster (that included the six US populations and China) we found evidence for two levels of substructure by examining the ΔK plot, one at K=2 and one at K=5 (Fig. 6a). Because the L(K) plot suggests that K=5 is more likely (Fig. 6b), we present those results here but also report the results from K=2. For K=5, the five genetic clusters corresponded to: a) the US populations, b) France, c) Spain-2005, d) Spain-2008 and Morocco (with Spain-2008 sharing ancestry with Spain-2005 and Canary Islands), and e) Canary Islands and China (Fig. 7a). With the exception of Alabama that seemed to share some ancestry with Spain-2008, Spain-2005, and Morocco, the US populations did not have evident admixture. In the Mediterranean, Spain-2008 was highly admixed, sharing ancestry with all other Mediterranean populations, and the Canary Island population that shared ancestry with Spain-2008 and Spain-2005. For K=2, the split was observed between a) the US populations and b) the Mediterranean populations and China (not shown).

Within the Eastern Mediterranean, both the L(K) and ΔK plots suggested that K=5 best explains the genetic structure in the data (Fig. 8a,b). The five clusters in this analysis corresponded to a) the US populations (N.Carolina-2 and Louisiana – excluding the individual that was previously assigned to Western Mediterranean), b) Cyprus and

Israel-2007, c) Greece, d) Israel-Q, and e) Turkey (Fig. 7b). There was a stronger pattern of genetic structure within the Eastern compared to the Western Mediterranean cluster and lower levels of admixture among populations, although the US populations did seem to have some genotypes from Cyprus/Israel-2007 and Israel-Q (but not *vice-versa*).

A total of 11 haplotypes with 16 polymorphic sites across 688 bp were recovered in the 131 individuals we sequenced from the introduced (38) and native (93) range (Table 6). Haplotypes 1 and 2 were the most frequent (n=69 and n=39 respectively) whereas the other 9 haplotypes had frequencies from n=1 to n=13. The MJ network showed that haplotypes 1 and 2 were separated by six substitutions and corresponded geographically to Eastern Mediterranean (individuals from Cyprus, Israel-Q, Turkey, Israel-2007, Israel-Q) and Western Mediterranean (individuals from Spain-2005, Spain-2008) (Fig. 9). An exception to this pattern was the Greek individuals that possessed Hap1 (Eastern) and Hap2 (Western), as well as two other haplotypes closer to Western Mediterranean, and French individuals that had only Hap1. In the 38 individuals we sequenced from the introduced range we only recovered the two most common haplotypes. Individuals from Alabama, Arizona, California, Michigan, N.Carolina-1, N.York and one individual from Louisiana possessed Hap2, whereas individuals from N.Carolina-2 and the rest of the individuals from Louisiana possessed Hap1 (Table 6).

Potential sources of invasion in the USA

Bayesian assignment tests using the Rannala & Mountain (1997) approach showed that most individuals from the majority of the introduced populations except for Louisiana and N.Carolina-2 were assigned to Spain-2008 (Fig. 10). A smaller number of individuals were assigned to the Canary Islands population, whereas the other potential source populations from Western Mediterranean had very small to zero number of individual assignments. Most individuals from the Louisiana population were assigned to Israel-2007, but a smaller number of individuals were also assigned to Israel-Q, Cyprus, Greece and Turkey from the Eastern Mediterranean, and to the Canary Islands, Morocco, and Spain-2008 from the Western Mediterranean. Finally, individuals from N. Carolina-2 were mostly assigned to Cyprus, with a few of those assigned to Israel-Q. The highest mean assignment likelihoods (averaged score over all individual assignments for a given introduced population) are given in Table 7. These scores show that the most likely source of Louisiana and N. Carolina-2 is Israel-2007 and Cyprus respectively, and the most likely source of the rest of the US introductions and China is Spain-2008.

Results from the stricter assignment test using the Monte Carlo resampling algorithm by Paetkau et al (2004) showed relatively low probabilities of individual assignments, with zero values in some cases. This method does not assume that the actual source population has been sampled and allows computation of the probability that an individual belongs to each source population. The highest probabilities of an individual from each introduced population and the associated potential source populations are given in Table 7. The highest individual probability of assignment was observed in a single individual from Louisiana (87%), which was assigned to Spain-2008. Individuals from Alabama, Michigan, and California also had relatively high probabilities of assignment and were also associated with Spain-2008, whereas the lowest individual assignment probabilities were found in Louisiana (13%) associated with Israel-2007

(except for individual mentioned above) and in N.Carolina-2, associated with Cyprus (1%).

Identification of potential dispersal pathways of introduced populations in the US using Approximate Bayesian Computation

For this analysis we excluded the Arizona population, which was well differentiated from the other populations (Table 5), as well as populations from N.Carolina 2 and Louisiana that were shown to have invaded the US from the Eastern Mediterranean.

We chose to use California as starting point for all scenarios because: a) it had the highest assignment score to a potential source population, b) in pairwise F_{ST}s, it had the lowest value with all other populations than either of them with each other (Table 5) (with the exception of N.York and N.Carolina-1 which had negative pairwise F_{ST}s and were pooled in to a single sample), c) prior information indicated that this invasion may have occurred through California (Dalton 2006)**.** The analysis showed that the best-supported scenario is the one describing a route of independent dispersal of all three populations from California to each of the states they were sampled from (Fig. 11). The posterior probability of the other six scenarios of successive dispersal across states was much lower with the direct estimate and close to zero with the logistic regression.

Discussion

Genetic comparison of populations in the introduced and native range

Our results showed no statistically significant differences in allelic richness and heterozygosity between populations recently introduced in the US and China and populations from the native Mediterranean range (Figs 2 and 3). Theory and empirical data suggest that populations that have been introduced into a new area are typically genetically impoverished, a result of founder events and associated genetic bottlenecks (Dlugosch and Parker 2008; Nei 2005; Nei et al. 1975). The loss of genetic diversity is expressed as a reduction of the total number of alleles (allelic richness) and heterozygosity in the introduced populations (Dlugosch and Parker 2008 and references therein; Leberg 1992). However, a recent meta-analysis of multiple invasion studies, including insects, suggested that the reduction in genetic diversity follows a *U*-shaped pattern, such that a decline in diversity occurs over several decades after the initial introduction (Dlugosch and Parker 2008). Our findings that recently introduced populations had no significant differences in genetic diversity are in line with these expectations. The only exception to this pattern was the Arizona population, which had the lowest H_O among all populations (Fig. 2). It is possible that the population that invaded Arizona consisted of a small number of individuals such that it underwent a severe bottleneck resulting in significant loss of genetic diversity.

Analysis of molecular variance showed that although levels of within population differentiation (F_{ST}) were similar in populations from the introduced and native range, partitioning of the remaining genetic variance differed substantially in the two regions. The grouping of populations into geographic origin (from Eastern or Western Mediterranean) accounted for most of the genetic variance in the introduced range, while

within these groupings there was little differentiation. This supports the results of the STRUCTURE analysis that showed strong differentiation between the two genetic clusters. In contrast, variability among populations accounted for a small percentage of the genetic variance in the introduced range, suggesting that populations dispersed quite rapidly across the US following introduction, without time for much diversification among them. This is supported by the low pairwise F_{ST} values among these populations, except for Arizona. In the native range on the other hand, variance among populations within each group (Eastern or Western Mediterranean) was quite high and almost as high as the "among groups" variance, suggesting strong differentiation. This pattern of genetic variation is expected since these populations likely diversified in this region and despite the ongoing trade between Mediterranean countries, there is still strong genetic structure (Ahmed et al. 2009; Moya et al. 2001; Tsagkarakou et al. 2007).

Population genetic structure and mitochondrial DNA diversity

Our analysis in STRUCTURE revealed a clear geographic structure with populations from Spain, France, Morocco, and the Canary Islands forming one cluster (Western Mediterranean) and populations from Greece, Cyprus, Israel, and Turkey comprising another (Eastern Mediterranean). Further analysis of each of these two lineages separately in STRUCTURE showed that they both consist of well-differentiated populations. Our results are consistent with studies focused on the distribution of biotype Q in the Mediterranean in the past 10 years, that demonstrated substantial genetic differentiation among its populations in nuclear and mitochondrial DNA markers (Ahmed et al. 2009; Dalmon et al. 2008; De la Rua et al. 2006; Guirao et al. 1997; Moya et al. 2001; Simon et al. 2007; Tsagkarakou et al. 2007). These findings are also in agreement with recent studies, which suggested that biotype Q comprises two subclades, a Western (Q1 or MedBasin 1) and an Eastern (Q2 or MedBasin 2) Mediterranean subclade (Ahmed et al. 2009; Chu et al. 2008). This high genetic structure in native populations is particularly useful in identifying sources of introduction of this biotype and tracking its worldwide invasion.

Although analysis of the mitochondrial DNA strongly supported the large split into a Western and Eastern Mediterranean lineage (Fig. 9), these data showed a less clear geographic association of populations. The inconsistency compared to microsatellites was observed in samples from France, which had the Eastern Mediterranean haplotype and in samples from Greece, which had both Eastern and Western Mediterranean haplotypes. The presence of both haplotypes as well as admixture in nuclear microsatellites in the Greek individuals suggests that there may be range overlap and introgression between the two lineages in Greece. The conflict between the uni-parentally inherited nuclear and maternal mitochondrial markers points to a complex demographic history of these populations, perhaps associated with *Wolbachia* infections that can cause female-biased admixture (Chapter 2; Hoffmann and Turelli 1997; Stouthamer et al. 1999; Werren 1997).

Potential sources of invasion in the USA

Examination of both microsatellite (Fig. 4) and mitochondrial DNA (Fig. 9) suggested that the origins of biotype Q in the US are both the Western and Eastern Mediterranean regions. Specifically, our STRUCTURE analysis and the mitochondrial

haplotype network showed that populations sampled from Louisiana and N. Carolina 2 originated from the Eastern Mediterranean, whereas populations from Alabama, Michigan, California, N. Carolina 1, N. York, and Arizona originated from the Western Mediterranean. Assignment tests, as well as examination of private alleles, showed that the most likely source of the Eastern Mediterranean invasions is Israel or Cyprus and the most likely source of the Western Mediterranean invasions is Spain. However, the low probabilities of assignment of some individuals and the clustering results from STRUCTURE (Fig. 7) suggest either that we do not have the actual source population in our samples or that allele frequencies have shifted in the invaded range. This is a common characteristic of invading populations where rare alleles that persist may become dominant and shifts in allele frequencies may occur (Dlugosch and Parker 2008).

The hypothesis of an Eastern and a Western Mediterranean invasion front in the US was also proposed by Chu et al. (2008) and Ahmed et al. (2009) who showed the presence of mtDNA haplotypes from both lineages in individuals sampled from Georgia and California. Their finding of a Western Mediterranean origin of the China invasion is also consistent with our results. A possible scenario that emerges is that a single introduction wave originated from Spain and that whiteflies from this invasion colonized Alabama, California, Michigan, N. Carolina 1, and N. York. Previous information suggested that this biotype was introduced into the US via Central America from a large poinsettia distributor in southern California (Dalton 2006). Since the California population had the highest $L_{i \rightarrow s}$ to Spain, it is possible that Guatemala acted as an intermediate stage for the US invasion that began in California. The California population had the lowest differentiation (F_{ST}) with three out of the four populations of Western Mediterranean invasions, adding further evidence that it was the source population for the invasion of the other states.

Although the Arizona individuals were also assigned to Spain in both the assignment tests and STRUCTURE, some lines of evidence suggest that this population either originated from an independent invasion event, or experienced a much strong founder event compared to other populations. First, this population had the lowest heterozygosity and allelic richness compared to all introduced populations. Second, it had much higher and significant pairwise F_{ST} values compared to all other population pairs of Western Mediterranean origin. This high F_{ST} could be an artifact of reduced heterozygosity, which would be consistent with a founder event. Finally, we found two private alleles at high frequencies in this population, not detected in the Western Mediterranean invasions or in the other introduced populations. Although this latter argument is less compelling, it provides some indication of an interesting case to be further examined. These results suggest that an independent cryptic invasion may have occurred at an earlier time, likely from Spain as suggested by results from the STRUCTURE analysis; however additional sampling from other states and their inclusion in the analysis would clarify this hypothesis.

The Eastern Mediterranean invasions must have originated from a separate invasion event. This would be plausible since the Spanish Q populations have not been detected in Israel/Cyprus or *vice versa* in our study or previous studies. Assignment tests showed that the most likely source of the Louisiana population was Israel_2007, whereas the most likely source of the N. Carolina 2 population was Cyprus, but the individual assignment scores were low. Although these two populations clustered together in

STRUCTURE, they both had private alleles and much higher pairwise F_{ST} compared to the Western Mediterranean invasions. One possibility is that they originated from different sources as two independent cryptic invasions. Another explanation for their differentiation is that these populations exchanged genes with the Western Mediterranean populations in their invaded localities; this would be likely since one individual from the Louisiana population and our second population from N. Carolina had Western Mediterranean origin. Although no hybrids or admixed individuals were found in the invaded range, our results suggest that introgression does occur from the Western to the Eastern Mediterranean lineage in the invaded range: first, we found private alleles from Spain and France in Louisiana and N.Carolina-2 and second, individuals from Louisiana were assigned to Western Mediterranean populations (Spain, Morocco, Canary Islands, Fig. 10). The hypothesis of unidirectional gene flow between Western and Eastern Mediterranean Q lineages is examined using the model Isolation with Migration (IMa2) in Chapter 3.

Our finding that introgression likely occurs between the two lineages in the invaded states Louisiana and N.Carolina-2, would also explain why we found much higher (although non-significant) genetic diversity in populations sampled from these states. Such cases, of elevated genetic variation through admixture in populations from different geographic sources was demonstrated in Cuban lizards (Kolbe et al. 2004), and has since been documented in studies of other taxa (Kolbe et al. 2007). We found that two geographically isolated lineages of a highly invasive biotype overlap in their distribution in the invaded range, with potential introgression from one lineage to the other. These findings have enormous implications for managing the spread of this biotype, particularly since the two lineages are known to possess different levels of insecticide resistance (Nauen et al. 2002). Other traits associated with high pest status and invasiveness, as well as increase in genetic variation through introgression from multiple geographical sources enhance the potential of this invasive biotype to become even more adapted and destructive. Further monitoring of these invaded populations with the use of genetic tools would be particularly important for tracking the spread of invasive genotypes across the introduced range.

Although this study did not address the hypothesis of that the US invasion originated *via* Guatemala or other Central America countries, our results point to the original sources of invasion in the Mediterranean native range, which likely acted as sources for all the New World invasions of biotype Q. Inclusion of samples from these areas that only recently reported the presence of this biotype (Bethke et al. 2009) in future studies will clarify the hypothesis that Central America was the intermediate source for the US invasion.

Identification of potential dispersal pathways of introduced populations in the US using Approximate Bayesian Computation

The ABC analysis suggested that, with California as an initial introduction point and source of subsequent invasions, the Q biotype colonized greenhouses and nurseries across the US in independent dispersal events. This scenario was supported over alternative scenarios of dispersal from California into each of the other states sequentially. This was true at least for populations introduced into Alabama, Michigan, N.Carolina 1, and N.York, although we cannot exclude the hypothesis that introduction in

the two latter states was serial since these populations had no differentiation. Our finding that California is the common source for multiple introductions across the US implicates the trade of poinsettias as a potential means of whitefly invasions. A high percentage of the US's seasonal poinsettias (over 80 percent) are produced and distributed at a national scale from California (http://urbanext.illinois.edu/poinsettia/facts.cfm). Thus, the Q biotype invasions across the US seem to have hitchhiked along on these plants and colonized multiple states. Alternative means of dispersal would be through natural colonization of several states, through wind-blown populations, or with human movement. However, the very low F_{ST} values between populations sampled across the US suggest these scenarios would be rather unlikely, since such natural dispersal would have given these populations some time for differentiation. Instead, we observed very low among population variation in AMOVA, suggesting that these populations spread across the US in a short period of time from a single source population, as suggested by the ABC analysis.

This study aimed to disentangle the sources of invasion and pathways of subsequent dispersal of the biotype Q in the US. Our findings showed that multiple independent, likely cryptic introductions from different sources contributed to the widespread distribution of this invasive biotype across the US, which has important implications for its monitoring, quarantine and management. First, it seems that stricter regulations at entry points are necessary if we are to avoid further invasions of this biotype. Second, although it seems that populations have been restricted to greenhouses, possible spread and colonization of field cultures would pose a major threat to the US agriculture. In addition, the presence of two distinct genetic lineages of this destructive biotype in its introduced range may have unpredictable implications for its future management, especially since they were shown to have differential insecticide resistance levels (Nauen et al. 2002). The two lineages have not yet been reported to be sympatric in the native range, although our results suggest that there may be admixture/and or introgression in Greece. However, their co-occurrence and range overlap in the invaded range provide opportunities for introgression in a novel environment, potentially bringing together combinations of genes associated with high resistance, virus transmission and invasiveness. Such phenomena of gene flow between genetically diverse populations from the native range, potentially resulting in the spread of invasive genotypes have also been suggested previously (Sakai et al. 2001). Our study highlights the significance of these possibilities and provides grounds for future research on the interactions of the two Q biotype lineages in its invaded range. Knowledge of the potential of this biotype for invasion success and the contributing characteristics such as insecticide resistance have enormous implications in the strategies implemented to manage this pest and for future policies to prevent further invasions.

References

Ahmed MZ, Shatters RG, Ren SX, Jin GH, Mandour NS, Qiu BL. 2009. Genetic distinctions among the Mediterranean and Chinese populations of *Bemisia tabaci* Q biotype and their endosymbiont *Wolbachia* populations. *J. Appl. Entomol.* 133: 733-741.

Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS. 2006. Automated binning of microsatellite alleles: problems and solutions. *Mol. Ecol. Notes* 7: 10-14.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16: 37-48.

Banks G, Green RL, Cerezo ER, Louro D, Markham PG. 1998. *Use of RAPD–PCR to characterise whitefly species in the Iberian peninsula*. Abstracts of the 2nd International Workshop on *Bemisia* and Geminiviral Diseases, San Juan, Puerto Rico.

Beaumont MA, Zhang WY, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162: 2025-2035.

Beitia F, Mayo I, Robles-Chillida EM, Guirao P, Cenis JL. 1997. Current status of *Bemisia tabaci* (Gennadius) in Spain: The presence of biotypes of this species. *IOBC/WPRS Bulletin* 20: 99–107.

Bethke JA, Byrne FJ, Hodges GS, McKenzie CL, Shatters RG. 2009. First record of the Q biotype of the sweetpotato whitefly, *Bemisia tabaci*, in Guatemala. *Phytoparasitica* 37: 61-64.

Boykin LM, Shatters Jr RG, Rosell RC, McKenzie CL, Bagnall RA, De Barro P, Frohlich DR. 2007. Global relationships of *Bemisia tabaci* (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Mol. Phylogen. Evol.* 44: 1306-1319.

Brown JK. 2007. The *Bemisia tabaci* complex: genetic and phenotypic variability drives begomovirus spread and virus diversification. Plant Disease Feature. http://www.apsnet.org/online/feature/btabaci Accessed: July 15, 2010.

Brown JK, Frohlich DR, Rosell RC. 1995. The sweetpotato or silverleaf whiteflies: Biotypes of *Bemisia tabaci* or a species complex? *Annu. Rev. Entomol.* 40: 511-534.

Chapuis MP, Estoup A. 2007. Microsatellite null alleles and estimation of population differentiation. *Mol. Biol. Evol.* 24: 621-631.

Chu D, Wan FH, Tao YL, Liu GX, Fan ZX, Bi YP. 2008. Genetic differentiation of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) biotype Q based on mitochondrial DNA markers. *Insect Sci.* 15: 115-123.

Chu D, Zhang YJ, Brown JK, Cong B, Xu BY, Wu QJ, Zhu GR. 2006. The introduction of the exotic Q biotype of *Bemisia tabaci* from the Mediterranean region into China on ornamental crops. *Fla. Entomol.* 89: 168-174.

Collins TM, Trexler JC, Nico LG, Rawlings TA. 2002. Genetic diversity in a morphologically conservative invasive taxon: Multiple introductions of swamp eels to the Southeastern United States. *Conserv. Biol.* 16: 1024-1035.

Cornuet J-M, Ravign V, Estoup A. 2010. Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*. In press. doi:10.1186/1471-2105-11-401

Cornuet J-M, Santos F, Beaumont MA, Robert CP, Marin J-M, Balding DJ, Guillemaud T, Estoup A. 2008. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* 24: 2713–2719.

Costa HS, Brown JK. 1991. Variation in biological characteristics and esterase patterns among populations of *Bemisia tabaci*, and the association of one population with silverleaf symptom induction. *Entomol. Exp. Appl.* 61: 211-219.

Dalmon A, Halkett F, Granier M, Delatte H, Peterschmitt M. 2008. Genetic structure of the invasive pest *Bemisia tabaci*: evidence of limited but persistent genetic differentiation in glasshouse populations. *Heredity* 100: 316-325.

Dalton R. 2006. The Christmas invasion. *Nature* 443: 898-900.

Davies N, Villablanca FX, Roderick GK. 1999. Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics. *Trends Ecol. Evol.* 14: 17-21.

De Barro PJ, Scott KD, Graham GC, Lange CL, Schutze MK. 2003. Isolation and characterization of microsatellite loci in *Bemisia tabaci*. *Mol. Ecol. Notes* 3: 40-43.

De la Rua P, Simon B, Cifuentes D, Martinez-Mora C, Cenis JL. 2006. New insights into the mitochondrial phylogeny of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in the Mediterranean Basin. *J. Zool. Syst. Evol. Res.* 44: 25-33.

Dennehy TJ, DeGain B, Harpold G, Brown JK, Byrne F, Fabrick JA, Morin S, Nichols R. 2005. New challenges to management of whitefly resistance to insecticides in Arizona. 2005 Vegetable Report, College of Agriculture and Life Sciences, The University of Arizona, Tucson. http://cals.arizona.edu/pubs/crops/az1382/contents.html Accessed: July 15, 2010.

Dennehy TJ, DeGain B, Harpold G, Brown JK, Byrne F, Morin S, Nichols R. 2006. First New World report of Q biotype of *Bemisia tabaci* (Gennadius) reveals high levels of resistance to insecticides. *Resistant Pest Management Newsletter* 15: 18-20.

Dlugosch KM, Parker IM. 2008. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Mol. Ecol.* 17: 431-449.

Earl DA. 2009. Structure Harvester v0.3. http://users.soe.ucsc.edu/~dearl/software/struct_harvest/

Estoup A, Beaumont M, Sennedot F, Moritz C, Cornuet JM. 2004. Genetic analysis of complex demographic scenarios: Spatially expanding populations of the cane toad, *Bufo marinus*. *Evolution* 58: 2021-2036.

Estoup A, Clegg SM. 2003. Bayesian inferences on the recent island colonization history by the bird *Zosterops lateralis lateralis*. *Mol. Ecol.* 12: 657-674.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1: 47-50.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164: 1567-1587.

Falush D, Stephens M, Pritchard JK. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* 7: 574-578.
Frohlich DR, Torres-Jerez I, Bedford I, Markham P, Brown J. 1999. A phylogeographic analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers. *Mol. Ecol.* 8: 1593-1602.
Gauthier N, Dalleau-Clouet C, Bouvret ME. 2008. Twelve new polymorphic microsatellite loci and PCR multiplexing in the whitefly, *Bemisia tabaci*. *Mol. Ecol. Res.* 8: 1004-1007.
Guillemaud T, Beaumont MA, Ciosi M, Cornuet JM, Estoup A. 2010. Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity* 104: 88-99.
Guirao P, Beitia F, Cenis JL. 1997. Biotype determination of Spanish populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Bull. Entomol. Res.* 87: 587-593.
Hamon AB, Salguero V. 1987. *Bemisia tabaci*, sweet potato whitefly, in Florida (Homoptera: Aleyrodidae: Aleyrodinae). *Entomology Circular, Division of Plant Industry, Florida Department of Agriculture and Consumer Services*: 2pp.
Herrmann M, Kienle S, Rochat J, Mayer WE, Sommer RJ. 2010. Haplotype diversity of the nematode *Pristionchus pacificus* on Reunion in the Indian Ocean suggests multiple independent invasions. *Biol. J. Linn. Soc.* 100: 170-179.
Hoffmann AA, Turelli M. 1997. Cytoplasmic incompatibility in insects. In *Influential Passengers: Inherited Microorganisms and Arthropod Reproduction*, ed. SL O'Neill, AA Hoffman, JH Werren, pp. 42-80. Oxford: Oxford University Press.
Horowitz AR, Denholm I, Gorman K, Cenis JL, Kontsedalov S, Ishaaya I. 2003. Biotype Q of *Bemisia tabaci* identified in Israel. *Phytoparasitica* 31: 94-98.
Hubisz MJ, Falush D, Stephens M, Pritchard JK. 2009. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Res.* 9: 1322-1332.
Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801-1806.
Kalinowski ST. 2005. HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. *Mol. Ecol. Notes* 5: 187-189.
Kolbe JJ, Glor RE, Schettino LR, Lara AC, Larson A, Losos JB. 2004. Genetic variation increases during biological invasion by a Cuban lizard. *Nature* 431: 177-181.
Kolbe JJ, Glor RE, Schettino LR, Lara AC, Larson A, Losos JB. 2007. Multiple sources, admixture, and genetic variation in introduced Anolis lizard populations. *Conserv. Biol.* 21: 1612-1625.
Leberg PL. 1992. Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. *Evolution* 46: 477-494.
Lee CE. 2002. Evolutionary genetics of invasive species. *Trends Ecol. Evol.* 17: 386-391.
Lee M, Uhm KB, Lee S, Park H, Choi HS, Park JW. 2007. Occurrence of *Bemisia tabaci* B-and Q-biotypes in Korea. Proceedings of the Fourth International *Bemisia* Workshop / International Whitefly Genomics Workshop. Organizers: PA Stansly, CL McKenzie. *J. of Insect Sci.* 8: 27.

Lombaert E, Guillemaud T, Cornuet JM, Malausa T, Facon B, Estoup A. 2010. Bridgehead effect in the worldwide invasion of the biocontrol *Harlequin* ladybird. *Plos One* 5

Maddison DR, Maddison WP. 2003. MacClade version 4: Analysis of phylogeny and character evolution. http://macclade.org/macclade.html

Martinez-Carrillo JL, Brown JK. 2007. First report of the Q biotype of *Bemisia tabaci* in southern Sonora, Mexico. *Phytoparasitica* 35: 282-284.

McKenzie CL, Hodges G, Osborne LS, Byrne FJ, Shatters RG. 2009. Distribution of *Bemisia tabaci* (Hemiptera: Aleyrodidae) biotypes in Florida - investigating the Q invasion. *J. Econ. Entomol.* 102: 670-676.

Miller N, Estoup A, Toepfer S, Bourguet D, Lapchin L, Derridj S, Kim KS, Reynaud P, Furlan L, Guillemaud T. 2005. Multiple transatlantic introductions of the western corn rootworm. *Science* 310: 992-992.

Moya A, Guirao P, Cifuentes D, Beitia F, Cenis JL. 2001. Genetic diversity of Iberian populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae) based on random amplified polymorphic DNA-polymerase chain reaction. *Mol. Ecol.* 10: 891-897.

Nauen R, Stumpf N, Elbert A. 2002. Toxicological and mechanistic studies on neonicotinoid cross resistance in Q-type *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Pest Manage. Sci.* 58: 868-875.

Nei M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Nei M. 2005. Bottlenecks, genetic polymorphism and speciation. *Genetics* 170: 1-4.

Nei M, Maruyama T, Chakraborty R. 1975. Bottleneck effect and genetic variability in populations. *Evolution* 29: 1-10.

Novak SJ, Mack RN. 1993. Genetic variation in *Bromus tectorum* (Poaceae) - comparison between native and introduced populations. *Heredity* 71: 167-176.

Paetkau D, Slade R, Burden M, Estoup A. 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Mol. Ecol.* 13: 55-65.

Pairon M, Petitpierre B, Campbell M, Guisan A, Broennimann O, Baret PV, Jacquemart AL, Besnard G. 2010. Multiple introductions boosted genetic diversity in the invasive range of black cherry (*Prunus serotina*: Rosaceae). *Ann. Bot.* 105: 881-890.

Pascual M, Chapuis MP, Mestres F, Balanya J, Huey RB, Gilchrist GW, Serra L, Estoup A. 2007. Introduction history of *Drosophila subobscura* in the New World: a microsatellite-based survey using ABC methods. *Mol. Ecol.* 16: 3069-3083.

Peakall R, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6: 288-295.

Perring TM. 2001. The *Bemisia tabaci* species complex. *Crop Protect.* 20: 725-737.

Perring TM, Cooper AD, Rodriguez RJ, Farrar CA, Bellows TS, Jr. 1993. Identification of a whitefly species by genomic and behavioral studies. *Science* 259: 74-77.

Pimentel D, Zuniga R, Morrison D. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol. Econ.* 52: 273-288.

Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A. 2004. GENECLASS2: A software for genetic assignment and first-generation migrant detection. *J. Hered.* 95: 536-539.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
R core development team. 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org
Rannala B, Mountain JL. 1997. Detecting immigration by using multilocus genotypes. *Proc. Natl. Acad. Sci. USA* 94: 9197-9201.
Raymond M, Rousset F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* 86: 248-249.
Rosenberg NA. 2004. DISTRUCT: a program for the graphical display of population structure. *Mol. Ecol. Notes* 4: 137-138.
Rousset F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Res.* 8: 103-106.
Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496-2497.
Sakai AK, Allendorf FW, Holt JS, Lodge DM, Molofsky J, With KA, Baughman S, Cabin RJ, Cohen JE, Ellstrand NC, McCauley DE, O'Neil P, Parker IM, Thompson JN, Weller SG. 2001. The population biology of invasive species. *Annu. Rev. Ecol. Syst.* 32: 305-332.
Scott IAW, Workman PJ, Drayton GM, Burnip GM. 2007. First record of *Bemisia tabaci* biotype Q in New Zealand. *N. Z. Plant Protect.* 60: 264-270.
Simon B, Cenis JL, De La Rua P. 2007. Distribution patterns of the Q and B biotypes of *Bemisia tabaci* in the Mediterranean Basin based on microsatellite variation. *Entomol. Exp. Appl.* 124: 327-336.
Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial-gene sequences and a compilation of conserved polymerase chain-reaction primers. *Ann. Entomol. Soc. Am.* 87: 651-701.
Stouthamer R, Breeuwer JAJ, Hurst GDD. 1999. *Wolbachia pipientis*: Microbial manipulator of arthropod reproduction. *Annu. Rev. Microbiol.* 53: 71-102.
Thomas ML, Becker K, Abbott K, Feldhaar H. 2010. Supercolony mosaics: two different invasions by the yellow crazy ant, *Anoplolepis gracilipes*, on Christmas Island, Indian Ocean. *Biol. Invasions* 12: 677-687.
Tsagkarakou A, Roditakis N. 2003. Isolation and characterization of microsatellite loci in *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Mol. Ecol. Notes* 3: 196-198.
Tsagkarakou A, Tsigenopoulos CS, Gorman K, Lagnel J, Bedford ID. 2007. Biotype status and genetic polymorphism of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Greece: mitochondrial DNA and microsatellites. *Bull. Entomol. Res.* 97: 29-40.
Ueda S, Brown JK. 2006. First report of the Q biotype of *Bemisia tabaci* in Japan by mitochondrial cytochrome oxidase I sequence analysis. *Phytoparasitica* 34: 405-411.
Weir BS. 1996. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data (2nd ed.)*. Sunderland, MA: Sinauer Associates, Inc.
Werren JH. 1997. Biology of *Wolbachia*. *Annu. Rev. Entomol.* 42: 587-609.

Tables

Table 1. Whitefly populations used in this study, number of individuals genotyped (microsatellites) and sequenced (mtCOI), host plant, collection date, number of loci out of HWE in each population, and annealing temperature (Ta) for sequencing of the mtCOI. Populations with the same number were collected from neighboring locations and were pooled together for analysis after exact tests of population differentiation showed non-significant differentiation ($P > 0.05$).

Pop. No	Pop. ID (collection site)	Number of individuals genotyped	Host Plant	Collection date	Number of individuals sequenced	Number of loci out of HWE	Ta
1	Alabama	31	Poinsettia	2005	5	1	51
2	Arizona	8	Host plant not available	2005	1	-	45
3	California_1	2	Greenhouse plants*	2005	2	-	51
3	California_2	6	Greenhouse plants	2005	2		51
3	California_3	3	Greenhouse plants	2005	2		51
3	California_4	3	Greenhouse plants	2005	2		51
4	Louisiana	41	*Hibiscus*	2005	12	3	52
5	Michigan	15	Host plant not available	2005	3	-	51
6	N. Carolina_1	5	Poinsettia	2005	2	-	42
7	N. Carolina_2	7	Poinsettia	2005	4	-	51
8	N. York	6	Poinsettia	2005	3	-	51
9	China	14	Host plant not available	2007	-	1	-
10	Cyprus	30	Gerbera	2008	27	-	51
11	Canary Islands_1	5	Tomato	2009	-	-	-
11	Canary Islands_2	6	Tomato	2009	-		-
11	Canary Islands_3	11	Tomato	2009	-		-
11	Canary Islands_4	10	*Solanum nigrum*	2009	-		-
12	France	25	Soybean	2007	8	-	52
13	Greece	17	Tomato	2006	5	2	40
14	Israel_Q	29	Cotton	1992 (colony until 2003)	12	1	48
15	Israel_2007	6	Sunflower	2007	3	-	51
15	Israel_2007	4	Cotton	2007	3		51
16	Morocco	9	Tomato	Unknown	-	1	-
17	Spain_2005	30	Tomato	2005 (colony until 2008)	3	-	51
18	Spain_2008	30	Zucchini	2008	18	2	48
18	Spain_2008	28	Pepper	2008	11		45
19	Turkey	29	Cotton	1994	3	1	48

* Host plant not available for greenhouse collections

Table 2. Properties of microsatellite loci and their descriptive statistics: mean expected (H_E) and observed (H_O) heterozygosities across populations (±SE), inbreeding coefficient (F_{IS}), and number of populations that deviated from HWE in each locus out of a total of 19 populations

Locus	Allele size range (bp) (*n* of alleles)	Fluorescent dye color[1]	Repeat motif	Mean H_O (±SE)	Mean H_E (±SE)	F_{IS}	Number of pop. out of HWE	Reference
WF2C01	141-237 (20)	PET	$(GTTT)_{11}$imp	0.351 (0.068)	0.370 (0.064)	0.051	-	Schwartz unpublished
WF2H06	166-188 (8)	NED	$(TTTG)_{11}$	0.259 (0.051)	0.257 (0.045)	-0.007	1	Schwartz unpublished
WF1B06	128-166 (7)	PET	$(ACTC)_8$	0.049 (0.023)	0.093 (0.039)	0.473	-	Schwartz unpublished
WF2E11	159-248 (15)	PET	$(GATT)_{27}$imp	0.476 (0.056)	0.458 (0.039)	-0.040	-	Schwartz unpublished
BEM6	160-222 (11)	6-FAM	$(CA)_8$imp	0.047 (0.017)	0.201 (0.058)	0.767	5	De Barro et al (2003)
BEM31	105-126 (5)	HEX	$(GCT)_4(GTT)_2$	0.475 (0.062)	0.370 (0.039)	-0.282	-	De Barro et al (2003)
BT-4	281-305 (11)	HEX	$(GT)_3(CA)_8$	0.403 (0.054)	0.411 (0.048)	0.019	-	Tsagkarakou and Roditakis (2003)
BT-b103	131-151 (8)	HEX	$(AC)_8(TC)_3$	0.244 (0.069)	0.316 (0.052)	0.228	3	Tsagkarakou and Roditakis (2003)
BT-e49	272-390 (28)	6-FAM	$(TTG)_{12}(TTC)_{11}$	0.634 (0.060)	0.706 (0.043)	0.102	-	Tsagkarakou et al (2007)
Bt*Is*1.2	273-367 (24)	6-FAM	$(CA)_{13}N_{14}(CA)_8$	0.460 (0.062)	0.596 (0.033)	0.228	1	Gauthier et al (2008)
MS145	173-181 (5)	PET	$(AC)_9$	0.401 (0.049)	0.452 (0.045)	0.112	-	Dalmon et al (2008)
MS177	241-265 (7)	6-FAM	$(CA)_7$	0.423 (0.037)	0.382 (0.028)	-0.108	-	Dalmon et al (2008)
MS11	173-187 (6)	VIC	$(GT)_9$	0.285 (0.057)	0.332 (0.064)	0.142	2	Delatte et al. (2006)

[1] 6-FAM, PET, NED, VIC (Applied Biosystems), 6-FAM, HEX (SIGMA-ALDRICH)

Table 3. Private alleles from native range populations detected in introduced populations and their respective frequencies

Locus	Private allele	Native population	Frequency	Introduced population detected in	Frequency
BT4	287	Spain_2005	0.017	Michigan	0.067
BTb103	133	France	0.020	Louisiana	0.012
Btls1.2	301	Canary Islands	0.031	Michigan	0.133
Btls1.2	349	Israel-Q	0.069	Louisiana	0.061
				N.Carolina_2	0.429
Btls1.2	359	Cyprus	0.017	Louisiana	0.012
WF2E11	159	Israel-Q	0.431	Louisiana	0.012
WF2E11	195	Spain_2008	0.027	Louisiana	0.110
				N.Carolina_2	0.071
WF2H06	170	Cyprus	0.083	Louisiana	0.024
				N.Carolina_2	0.071
WF2H06	184	Cyprus	0.033	Louisiana	0.390
				N.Carolina_2	0.214

Table 4. Global AMOVAs of microsatellite data in populations from the introduced and native range. "Groups" corresponded to populations with Eastern or Western Mediterranean origin. This level of hierarchy was determined after analysis in STRUCTURE identified two genetic clusters corresponding to geographic origin in the native range.

a) Introduced range populations

Source of variation	df	Sum of squares	Variance components	Percentage variation	F-statistics
Among groups F_{CT}	1	145.48	1.07	28.23	$F_{CT} = 0.29$, $P < 0.001$
Among populations within groups F_{SC}	6	52.12	0.26	6.9	$F_{SC} = 0.10$, $P < 0.001$
Within populations F_{ST}	246	603.03	2.46	64.88	$F_{ST} = 0.35$, $P < 0.001$

b) Native range populations

Source of variation	df	Sum of squares	Variance components	Percentage variation	*F-statistics*
Among groups F_{CT}	1	262.95	0.79	18.69	$F_{CT} = 0.19$ (P<0.001)
Among populations within groups F_{SC}	9	342.88	0.73	17.08	$F_{SC} = 0.21$ (P<0.001)
Within populations F_{ST}	555	1507.99	2.73	64.23	$F_{ST} = 0.36$ (P<0.001)

Table 5. F_{ST} values between pairs of populations introduced from the Western Mediterranean.

Population pair	$\boldsymbol{F_{ST}}$
N.Carolina1-N.York	0.000
California-N.Carolina1	0.015
Michigan-California	0.019
Alabama-California	0.020
California-N.York	0.042
Michigan-N.York	0.046
Alabama-Michigan	0.057*
Alabama-N.Carolina1	0.060*
Michigan-N.Carolina1	0.075
Alabama-N.York	0.111*
Alabama-Arizona	0.139*
California-Arizona	0.180*
Michigan-Arizona	0.190*
N.Carolina1-Arizona	0.250*
N.York-Arizona	0.320*

* $P<0.05$

Table 6. Descriptive statistics and variability of mitochondrial COI sequences in populations from the introduced and native range

Sample	Number of individuals	Number of haplotypes	Haplotype composition (as defined in haplotype network)	Haplotype diversity *h* (±SD)	Nucleotide diversity π
Alabama	5	1	Hap2	0.000 (0.000)	0.000
Arizona	1	1	Hap2	0.000 (0.000)	0.000
California	8	1	Hap2	0.000 (0.000)	0.000
Louisiana	12	2	Hap1,2 (1 ind)	0.167 (0.134)	0.001
Michigan	3	1	Hap2	0.000 (0.000)	0.000
N.Carolina_1	2	1	Hap2	0.000 (0.000)	0.000
N.Carolina_2	4	1	Hap1	0.000 (0.000)	0.000
N.York	3	1	Hap2	0.000 (0.000)	0.000
Cyprus	27	1	Hap1	0.000 (0.000)	0.000
France	8	1	Hap1	0.000 (0.000)	0.000
Greece	5	4	Hap1,2,4,6	0.900 (0.161)	0.004
Israel-Q	12	1	Hap1	0.000 (0.000)	0.000
Israel_2007	6	3	Hap1,5,7	0.733 (0.155)	0.001
Spain_2005	3	1	Hap2	0.000 (0.000)	0.000
Spain_2008	29	6	Hap2,3,8,9,10,11	0.645 (0.056)	0.001
Turkey	3	1	Hap1	0.000 (0.000)	0.000

Table 7. Assignment tests results obtained from GENECLASS 2. The maximum $L_{i\to s}$ given here expressed on a –log scale represents the highest mean individual assignment likelihood of each introduced population to a potential source population. The highest probability of assignment of an individual determined by the exclusion test with the simulation algorithm of Paetkau et al. (2004) is also given, with the source population it was assigned to.

Introduced population	Highest mean individual assignment likelihood $L_{i\to s}$ (± SE) by Rannala and Mountain (1997) assignment test	Source population associated with $L_{i\to s}$	Highest individual probability of assignment by Paetkau et al (2004) simulation algorithm (source population assigned to)
Alabama	11.33 (0.35)	Spain_2008	0.73 (Spain_2008)
Michigan	12.15 (0.59)	Spain_2008	0.66 (Spain_2008)
California	11.32 (0.58)	Spain_2008	0.66 (Spain_2008)
Louisiana	18.96 (0.57)	Israel_2007	0.87 (Spain_2008) 0.13 (Israel_2007)
N.Carolina_1	11.91 (0.66)	Spain_2008	0.32 (Spain_2008)
N.Carolina_2	15.65 (0.71)	Cyprus	0.01 (Cyprus)
N. York	12.26 (0.75)	Spain_2008	0.43 (Spain_2008)
Arizona	13.37 (0.43)	Spain_2008	0.21 (Spain_2008)
China	15.26 (0.86)	Spain_2008	0.21 (Spain_2008)

Figures

Fig. 1. Map of the US showing the scenario of independent dispersal of whitefly populations (in red) examined by the diyABC analysis against one example of sequential dispersal (in green). The other scenarios include all possible combinations of dispersal from California to each of the other states in sequence. The N.York and N.Carolina-1 populations were pooled because they were not differentiated and were examined as a single sampled population. States colored in bright yellow represent sampled populations from those states that were included in the diyABC analysis, states in light yellow those sampled but not included in this analysis, and in light blue are states with known presence of biotype Q but not sampled.

Pre-edited map source: http://edit.freemap.jp

Fig. 2. Mean observed H_O and expected H_E heterozygosity (±SE) of populations in the introduced and native range. Different letters above the error bars denote statistically significant differences in H_O between populations (P<0.05) as determined from Tukey's multiple comparison test. Differences in H_E were not statistically significant (P>0.05).

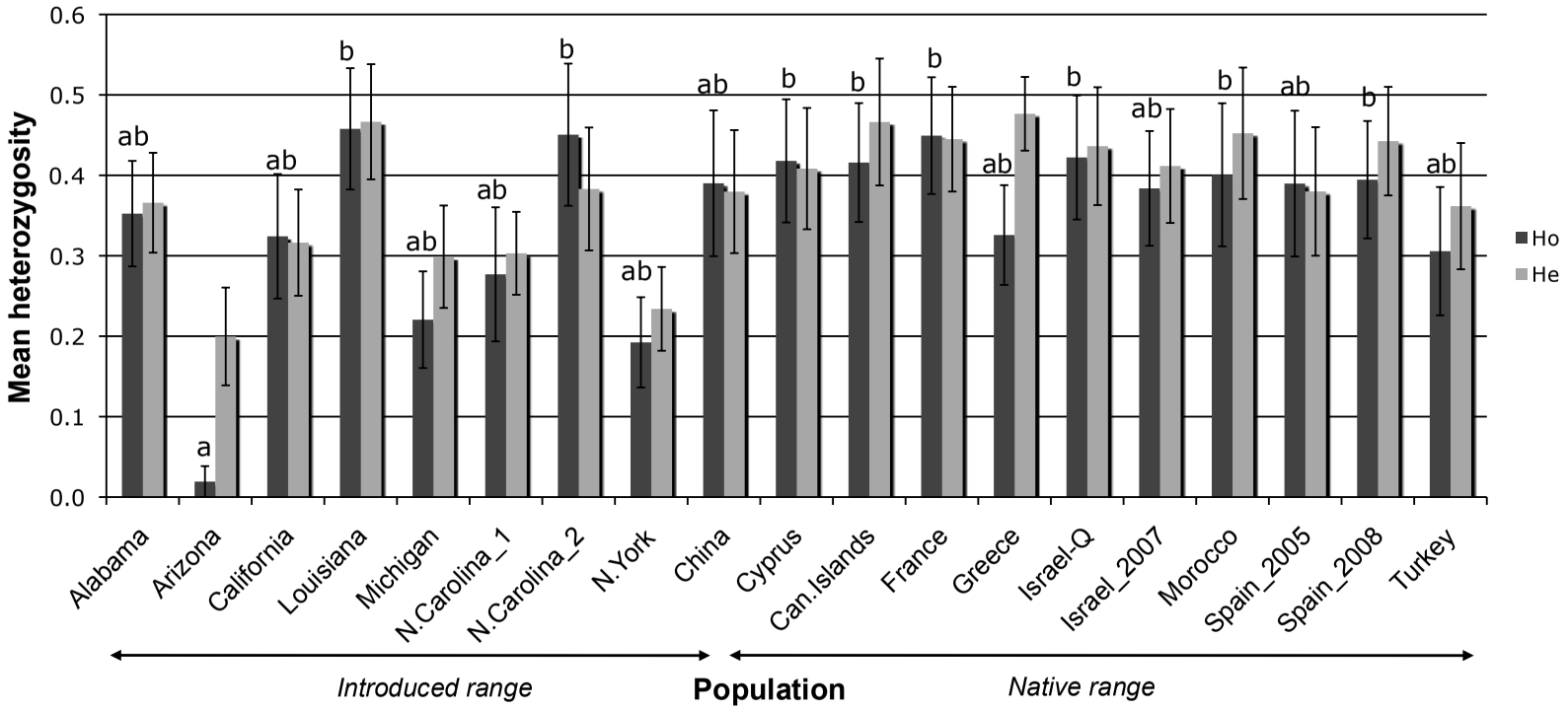


Fig. 3. Mean allelic richness in introduced and native range (±SE) corrected for sample size after rarefaction. Differences were not statistically significant (P>0.05).

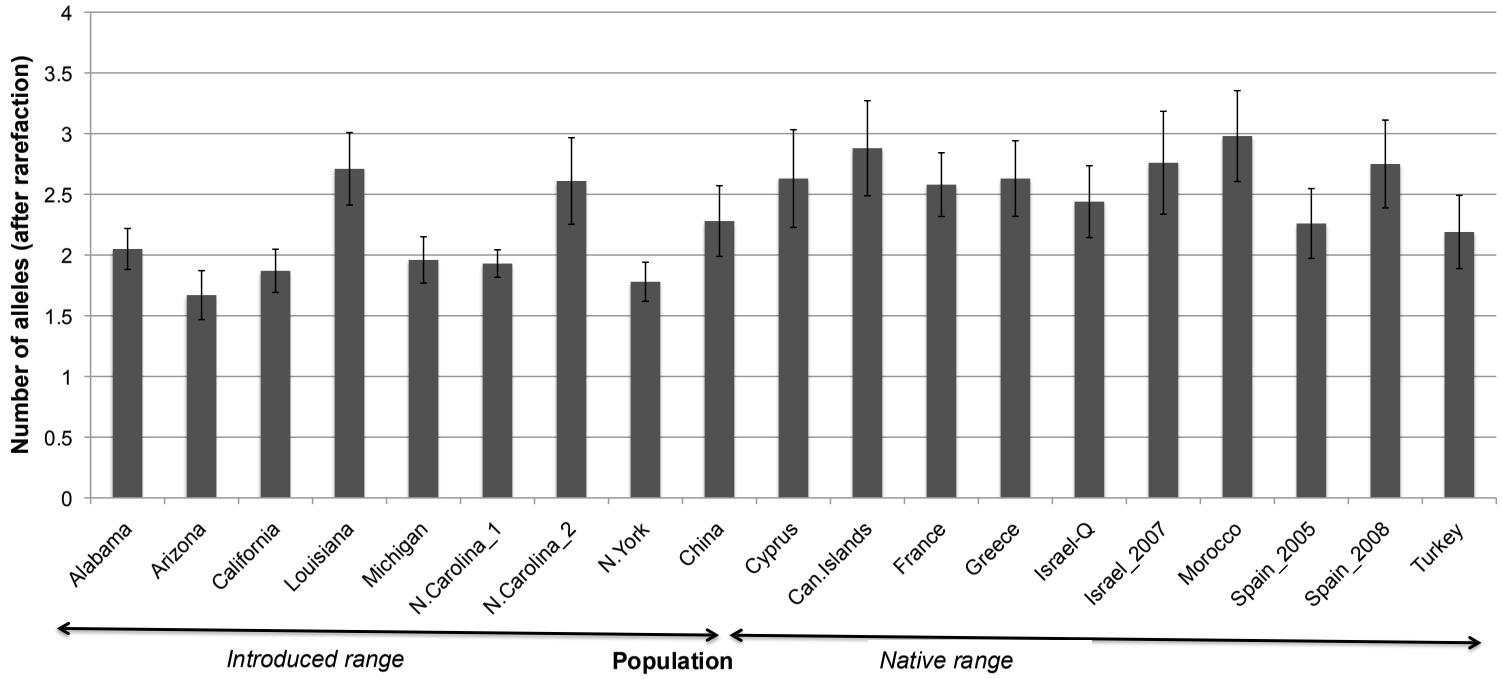


Fig. 4. Bayesian clustering analysis results of microsatellite data performed in STRUCTURE. Individuals are arranged on the x-axis, each represented by a thin vertical line and partitioned into each of two inferred clusters (*K*) with their estimated membership fractions on the y-axis. Labels below the plot separated by black vertical lines correspond to the sampled populations and groupings above and below the plot represent geographic sampling groups. Loci BEM6, BTb103, and MS11 were excluded from this analysis because of deviations from HWE in more than 1 population.

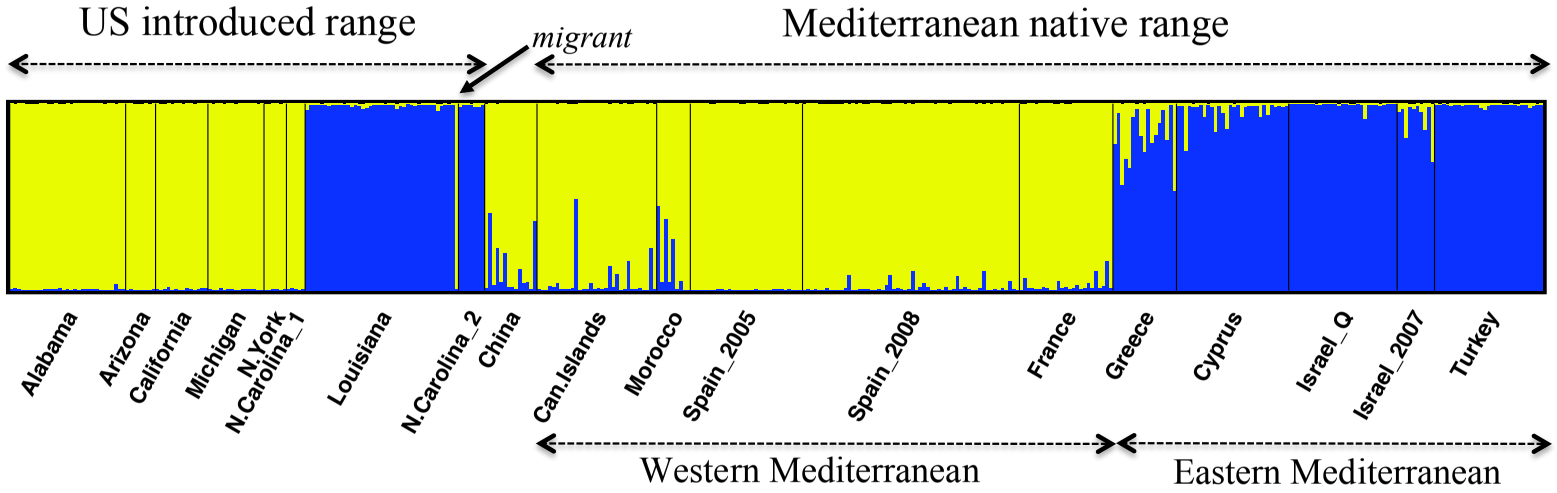
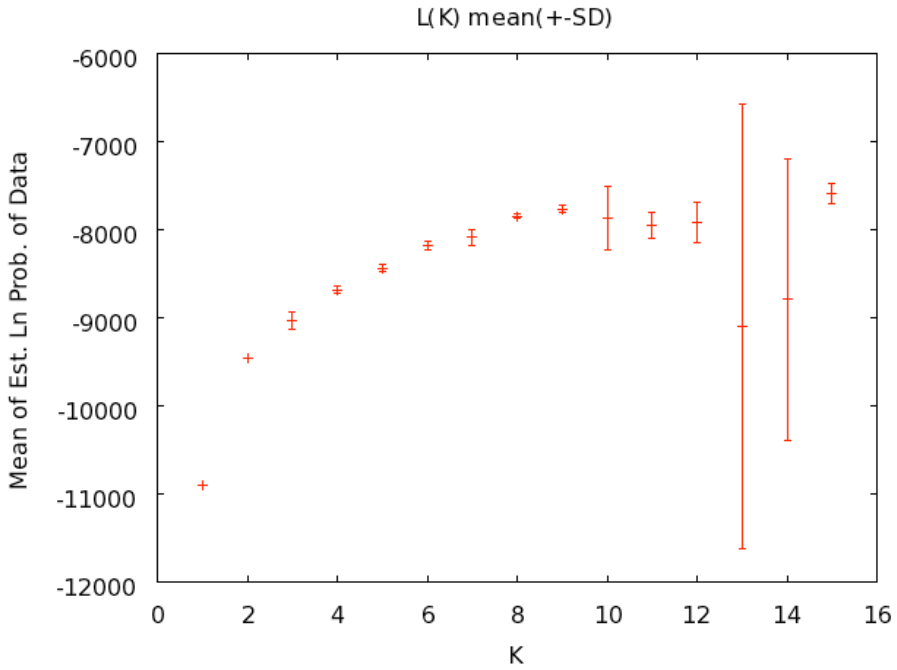


Fig. 5. Estimation of the number of clusters *K* that best explains the genetic structure in the microsatellite data with all populations analyzed together (introduced and native range) in STRUCTURE. Plots were produced using the program Harvester. Runs were done with *K*=1-15 and 5 replicates for each K.

a) Posterior probability of *K*, LnP(D), against *K* and SD (error bars)



b) Plot of the rate of change in the log probability of data between successive *K* values (Δ*K*, Evanno et al, 2005) against *K*.

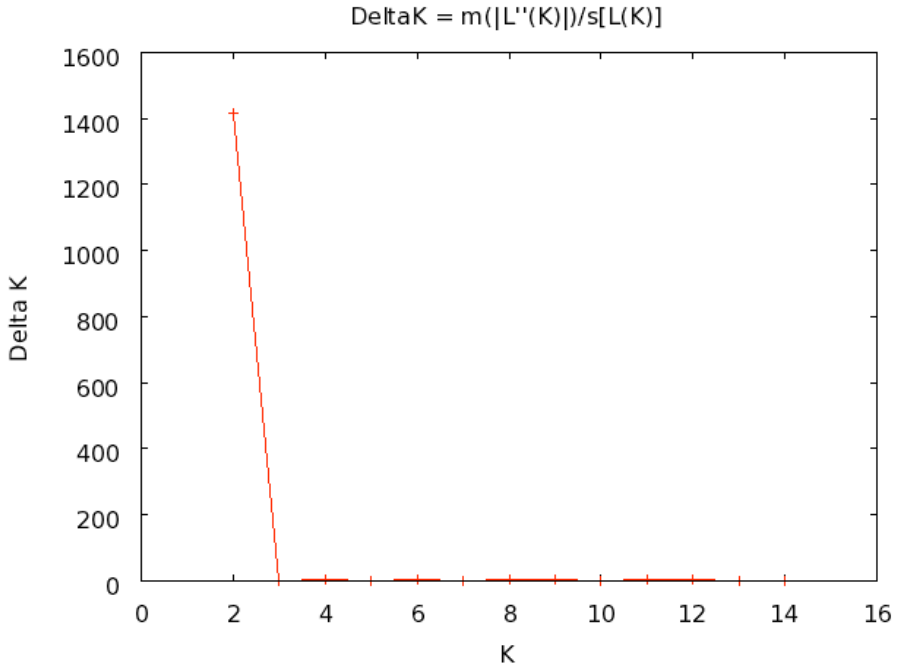
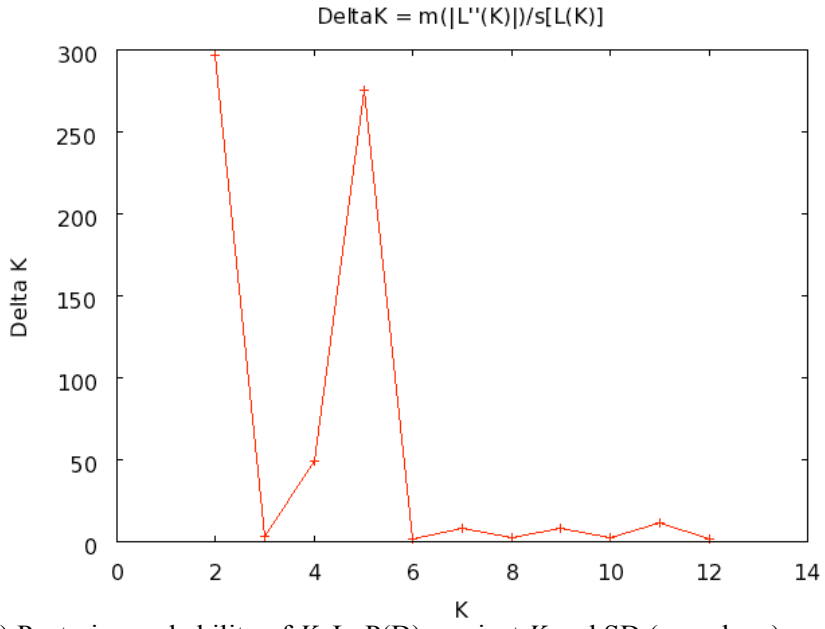


Fig. 6. Estimation of the number of clusters *K* that best explains the genetic structure in the microsatellite data of the Western Mediterranean cluster populations (native and introduced analyzed together) in STRUCTURE. Plots were produced using the program Harvester. Runs were done with *K*=1-13 and 5 replicates for each K.

a) Plot of the rate of change in the log probability of data between successive *K* values (ΔK, Evanno et al. 2005) against *K*.



b) Posterior probability of *K*, LnP(D), against *K* and SD (error bars)

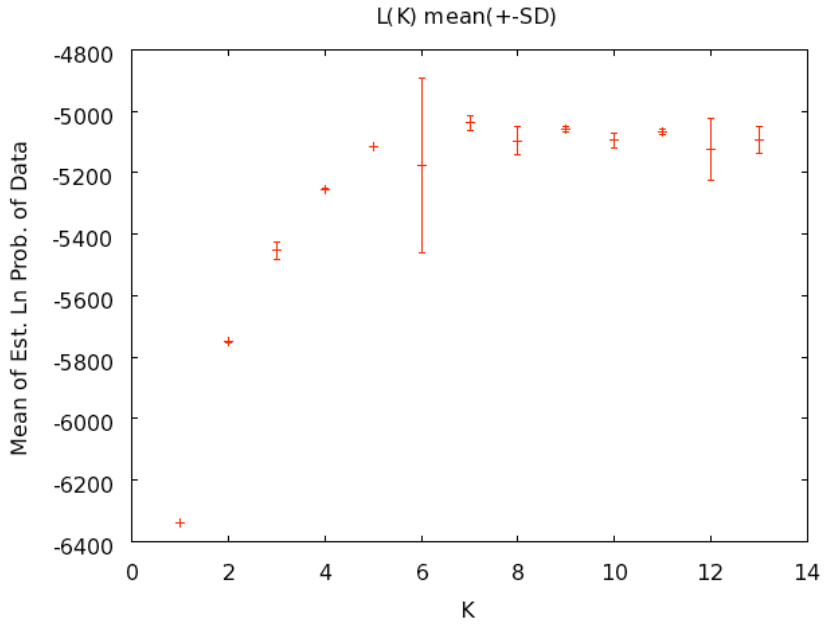


Fig. 7. Bayesian clustering analysis results of microsatellite data in STRUCTURE. Individuals are arranged on the x-axis, each represented by a thin vertical line and partitioned into each of inferred clusters (*K*) with their estimated membership fractions on the y-axis. Labels below the plot separated by black vertical lines correspond to the sampled populations and groupings above the plot represent geographic sampling groups. Locus BEM6 was excluded from both runs, and locus BTb103 was excluded only from the Eastern Mediterranean analysis because of deviations from HWE in more than 1 population.

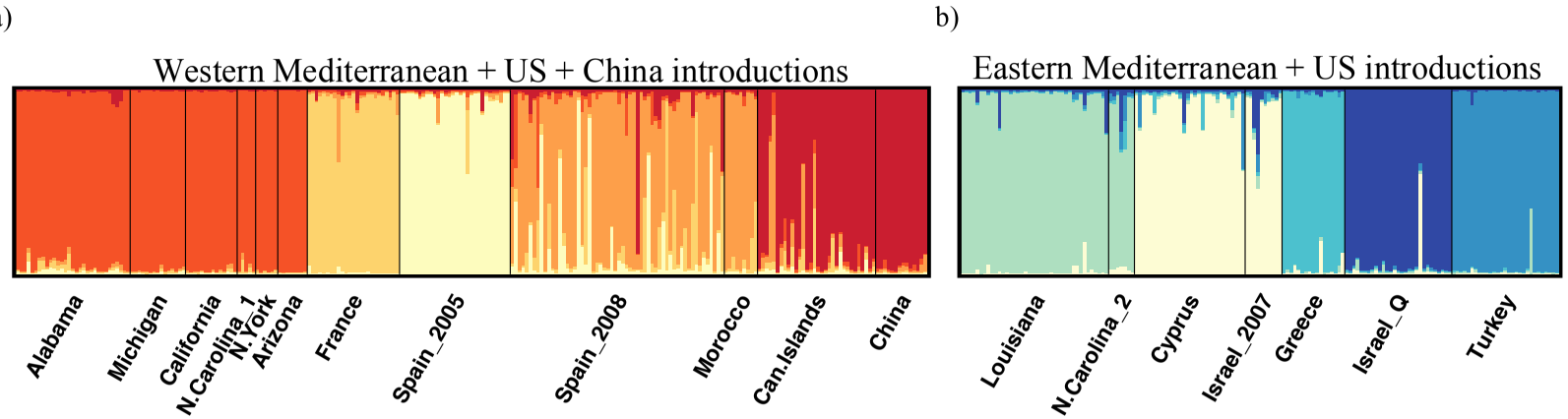
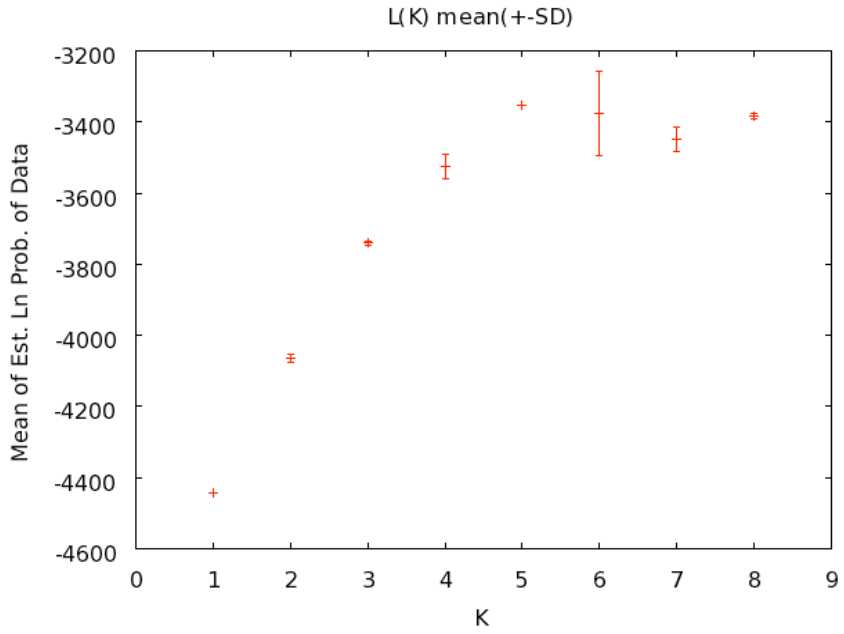


Fig. 8. Estimation of the number of clusters *K* that best explains the genetic structure in the microsatellite data of the Eastern Mediterranean cluster populations (native and introduced analyzed together) in STRUCTURE. Plots were produced using the program Harvester. Runs were done with *K*=1-8 and 5 replicates for each K.

a) Posterior probability of *K*, LnP(D), against *K* and SD (error bars)



b) Plot of Δ*K* (rate of change in the log probability of data between successive *K* values, Evanno et al, 2005) against *K*.

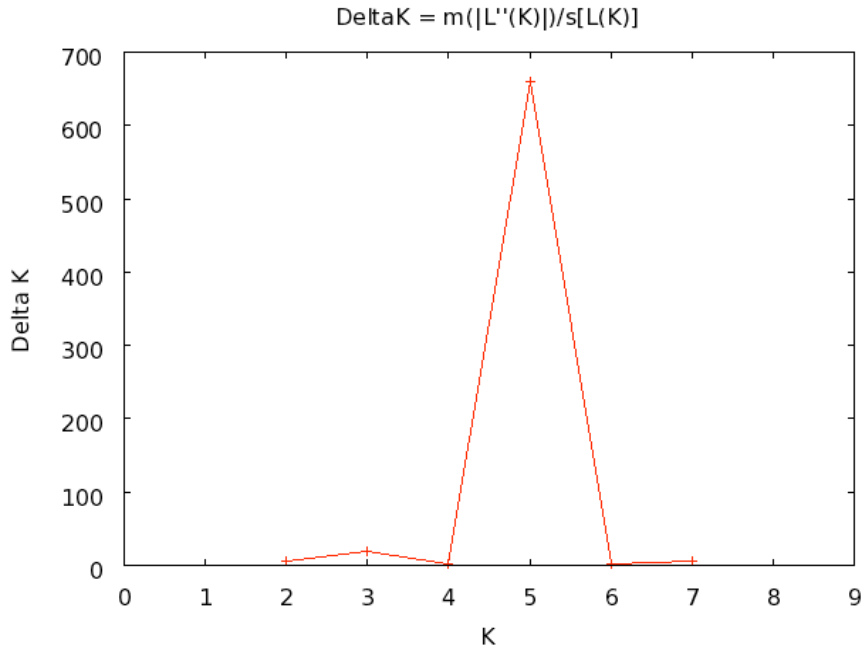


Fig. 9. MJ haplotype network of the mitochondrial COI sequences. The red numbers connecting branches represent substitutions in the mtCOI sequence. The colored pie segments of the haplotype correspond to the frequency of each sample (region) represented in the haplotype. Labels above the two major haplotype groupings are those determined from the microsatellite STRUCTURE analysis.

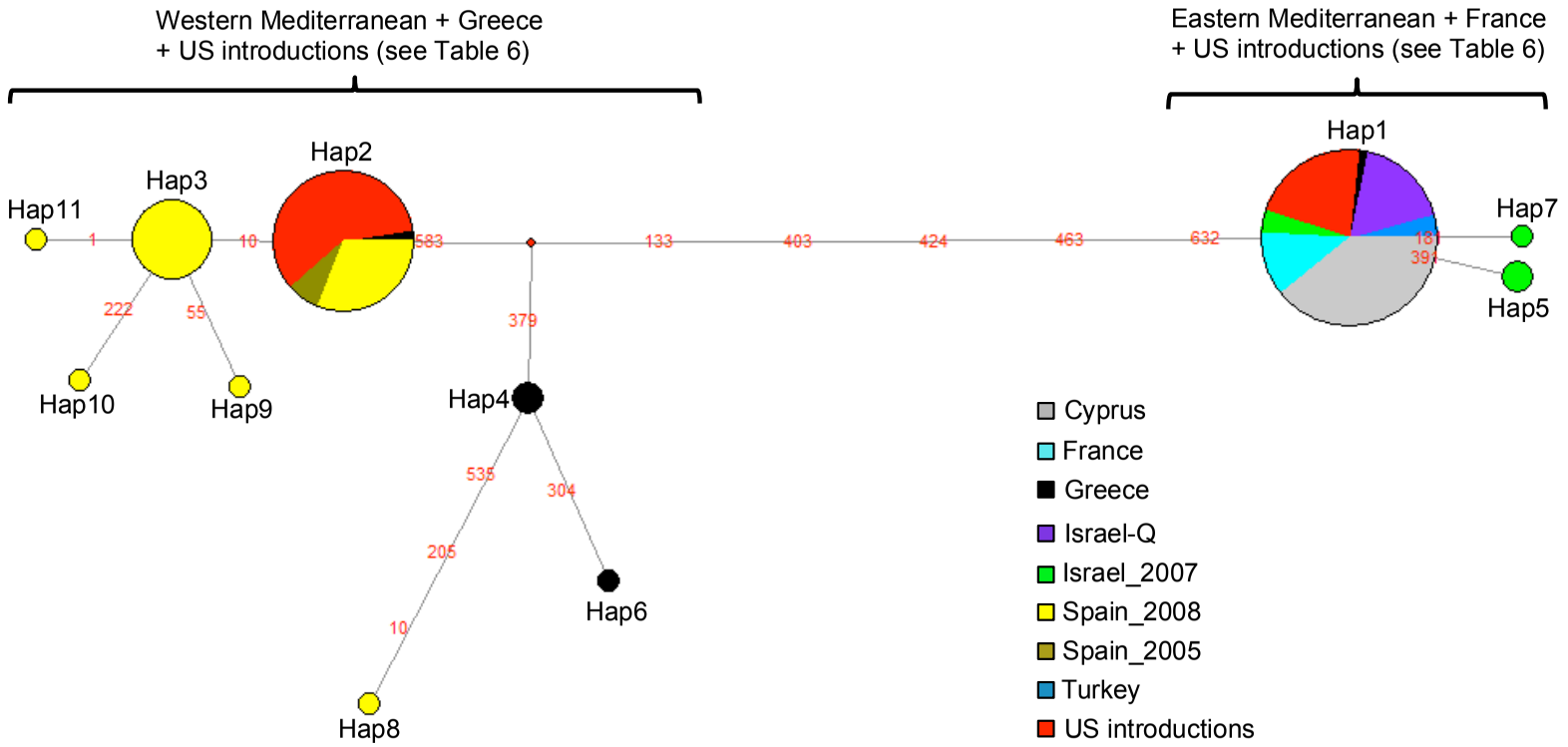


Fig. 10. Number of individuals from the introduced range assigned to each of the potential source populations from the Mediterranean native range. Assignment tests were done using the Bayesian approach by Rannala and Mountain (1997) in GENECLASS 2.

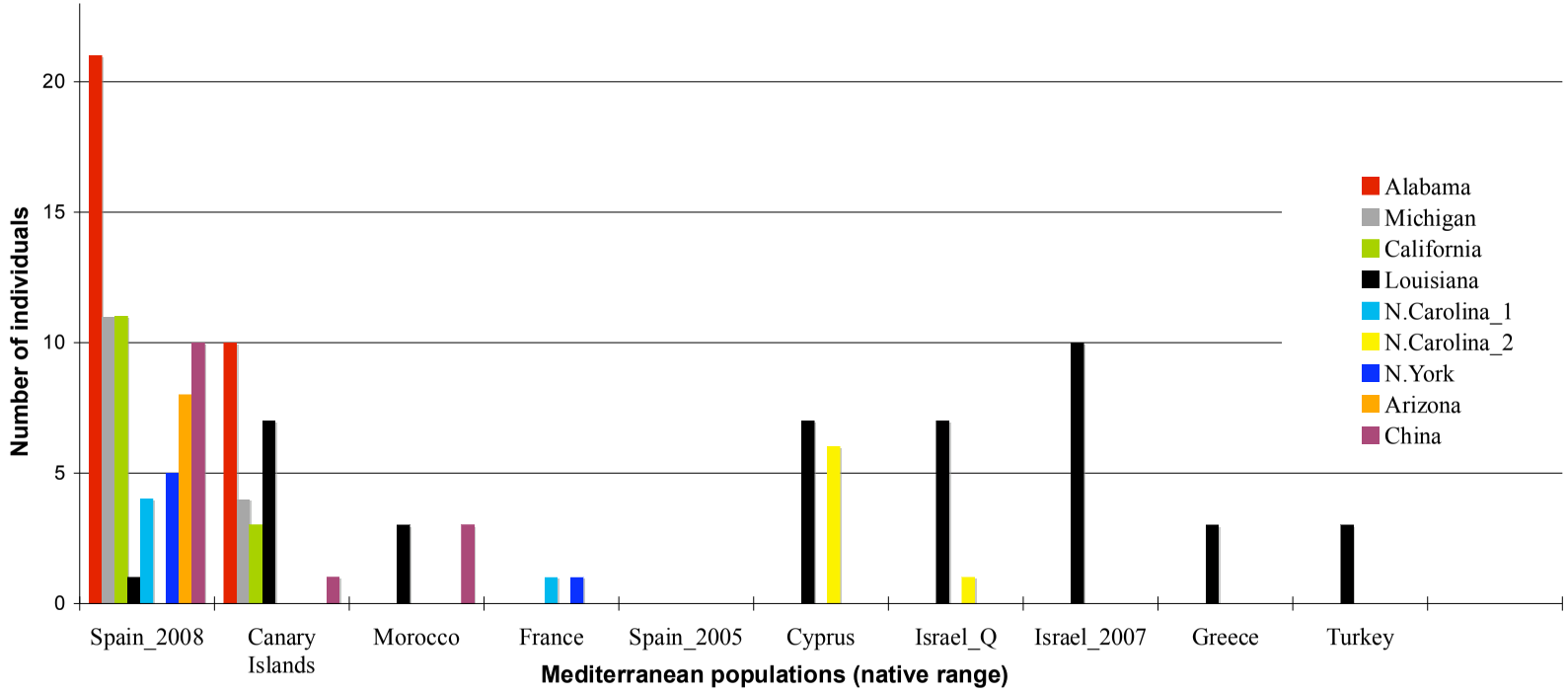
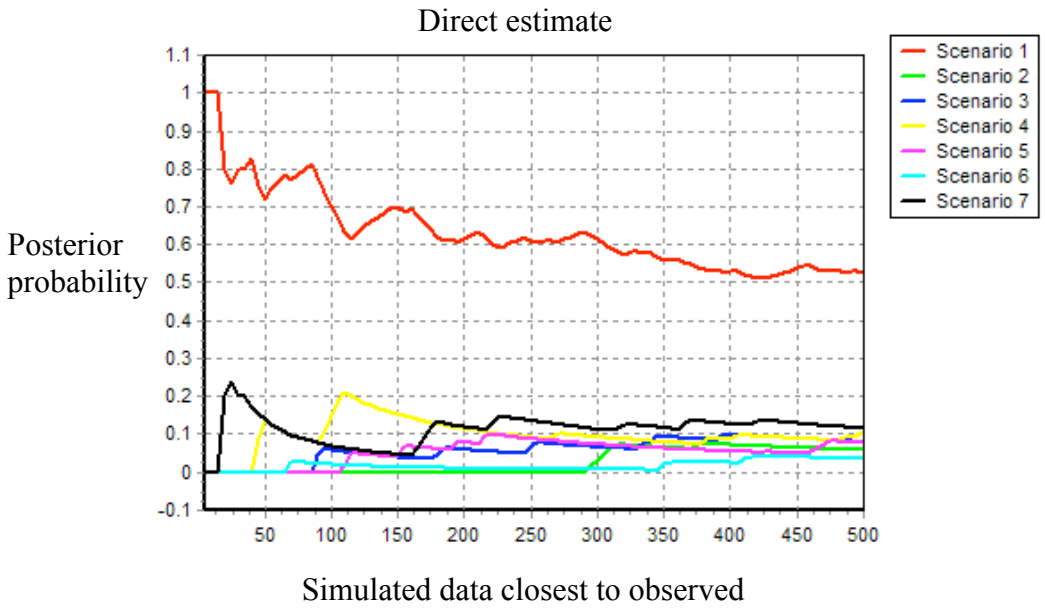
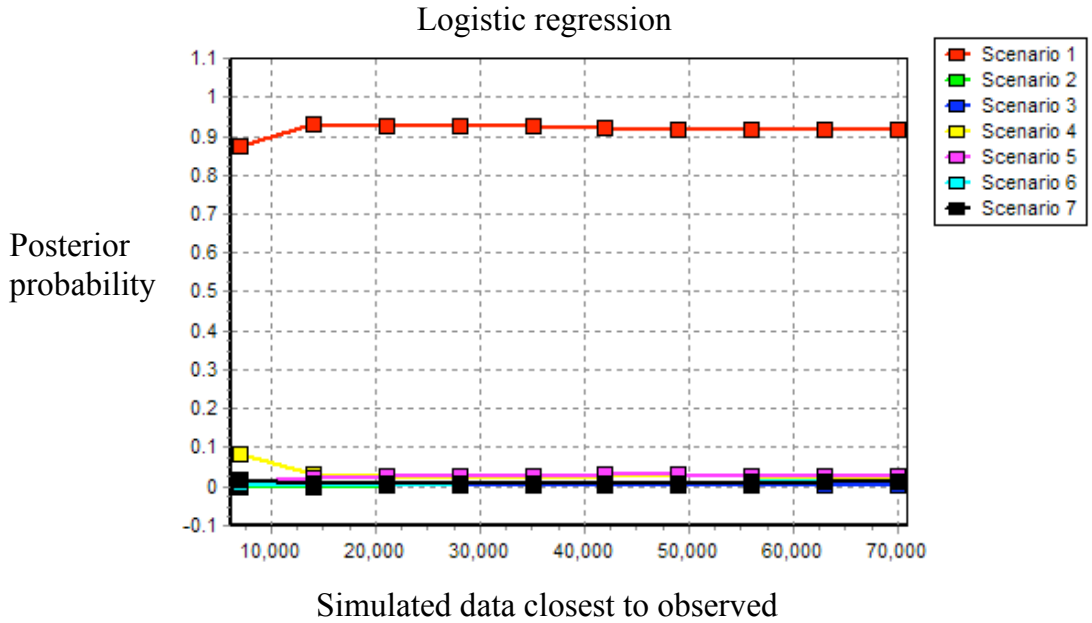


Fig. 11. DiyABC output graph, showing support for each of the seven competing scenarios (independent dispersal vs 2-6: sequential dispersal from California to the other states). The two different estimates, direct and logistic regression, show the posterior probability of each scenario in 500 or 70,000 simulated data respectively closest to the observed.

a)



b)



Conclusions

The whitefly *Bemisia tabaci* consists of more than 25 morphologically cryptic lineages, termed biotypes, that exhibit marked genotypic, phenotypic and physiological variability. The species therefore represents an excellent system for the study of multiple theoretical and applied questions. As a complex of biotypes, with various levels of reproductive isolation and genetic differentiation, as well as diverse ecological niches, *B. tabaci* provides ideal grounds to study population level demographic processes and evolution in the early stages of speciation. Furthermore, the presence of highly invasive pest biotypes in this complex offers unique opportunities to study ancestral and contemporary population histories contributing to invasiveness as well as pathways of invasion. This dissertation employs a combination of genetic approaches to examine multiple aspects of evolution and population genetics in the *B. tabaci* species complex, with studies that have both theoretical and applied implications.

In the first Chapter of this dissertation I reviewed the historical use of multi-locus molecular markers in *B. tabaci*, from the early application of esterases used to detect polymorphism in biotypes, to the more recent applications of microsatellite DNA. I summarized studies that addressed questions on biotype association with host or geography (e.g. De Barro 2005), insecticide resistance (e.g. Byrne et al. 2000), and possible links with virus epidemics (e.g. Legg et al. 2002), endosymbiont composition (e.g. Ahmed et al. 2009), and mating incompatibilities (e.g. Bedford et al. 1994). I concluded with suggestions for future work in this system, as the focus shifts from genetics to genomics, with research promising to elucidate the genetic basis for reproductive isolation and extreme phenotypes in biotypes, such as invasiveness, insecticide resistance, and polyphagy.

Previously, the genetic relationships among biotypes and cryptic lineages have been examined in a phylogenetic context with a limited number of genes: mitochondrial cytochrome oxidase I (mtCOI) (e.g. Frohlich et al. 1999), mitochondrial 16S ribosomal subunit (Frohlich et al. 1996), 18 rRNA (Campbell 1993), and ribosomal ITS1 (De Barro et al. 2000). Of these, only mtCOI showed sufficient resolution to distinguish populations and a several studies to date have used this gene to examine the relationships of biotypes at the global level (Boykin et al. 2007; Brown 2010; Dinsdale et al. 2010; Frohlich et al. 1996; 1999). Thus, our knowledge of worldwide genetic relationships in this complex to date has relied on the phylogeny of a single mitochondrial gene, which can be problematic due to its maternal mode of inheritance (Galtier et al. 2009). In Chapter 2, I used a population genetics approach to examine representative cryptic lineages at the global scale using multiple variable microsatellite markers. I showed that biotypes and other geographic lineages correspond to real genetic entities with strong geographic structure, and limited to no historical gene flow among them. Invasive biotypes did not exhibit lower levels of genetic diversity compared to non-invasives, suggesting both large founding populations with ancestral variation maintained despite the homogenizing effects of human-mediated gene flow among populations. Although the resulting genetic population clusters from this analysis are in general agreement with the mtCOI phylogeny, some cases of conflict in the two markers do exist, perhaps associated with the different modes of inheritance of nuclear and mitochondrial genomes. Thus, results from this work caution against the use of mtCO1 as a single marker for

species and subspecies delineation. Further phylogenetic studies that explore variation in additional genes evolving at various rates, will be necessary to circumscribe the boundaries of cryptic species and resolve the species complex conundrum. Possible examples of such genes are nuclear *wingless* (*wg*) gene, which at least in some insects evolves at a similar rate as mtDNA, but becomes saturated more slowly (Brower and DeSalle 1998), nuclear protein coding gene elongation factor-1a, and mitochondrial 12S rDNA. For population genetics studies, the development of multiple markers such as SNPs through next-generation sequencing approaches (Hughes et al. 2008; Metzker 2010; Pool et al. 2010), offers promising opportunities to examine demographic processes that have shaped the contemporary distribution of biotypes.

Earlier research on biotypes B and Q, currently two most prevalent, invasive biotypes in this complex, showed that although they are reproductively isolated and belong to sister clades in the mtCOI phylogeny (Brown 2007). Phylogeographic studies suggested that the closest relatives of biotype B are populations from the eastern African Sahel region and the Middle East/Arabian Peninsula, whereas the Q biotype relatives are located in North and West Africa (Brown 2007; 2010; Frohlich et al. 1999; Kirk et al. 2000). In Chapter 3, I investigated the divergence and demographic histories of these two biotypes and their closest relatives from these regions. Using an "Isolation with Migration" model in a Bayesian coalescent-based framework, I found evidence that biotype B diverged from its relative in Yemen around 3600 years ago. Biotype Q, consists of two genetically isolated lineages, one in the Western and one in the Eastern part of the Mediterranean (Chapter 4) that appear to have diverged from their common ancestor roughly 1400 years ago. Gene flow estimates suggest a direction of migration from Western to Eastern Mediterranean for biotype Q, and from Eastern Africa to the Middle East for biotype B. These results suggest that diversification of the two biotypes coincided with periods of extensive human movement and trade of agricultural goods in the Mediterranean, the Middle East, and Africa during the Iron and Bronze Ages and the Roman period. Further studies, focusing on additional populations from the hypothesized native range of these biotypes, complemented by phylogenetic dating of gene trees will help to elucidate further their history of divergence and pinpoint their exact origins. This knowledge will provide insights into why these two biotypes became such notable invasive pests and would also be useful in targeting potential sources for efficient natural enemies of this pest in areas where they co-evolved (Kirk et al. 2000).

The recent invasion of biotype Q in the USA caused great concerns about its potential damage, which could be as devastating as the previous 1980's invasion of biotype B (Dennehy et al. 2005). Nationwide surveys detected biotype Q in 25 states thus far, but the sources of this invasion were speculative; it has was proposed that whiteflies were introduced on cuttings of poinsettia from Central America, likely Guatemala (Dalton 2006). However, the invading populations must have originated more recently from their native range, the Mediterranean, since the presence of biotype Q in Central America was only recently reported (Bethke et al. 2009). In Chapter 4, I undertook a detailed analysis and investigation of this invasion and showed, in both microsatellite and mitochondrial DNA, that introduced populations originated from both the Western and Eastern Mediterranean. The results also suggest that this introduction comprised at least three independent cryptic invasions; likely from Spain for the Western Mediterranean introductions and Cyprus or Israel for the Eastern Mediterranean introductions. These

areas were most probably the original sources, even if Guatemala acted as an intermediate source. However a more comprehensive analysis of this invasion could be undertaken in future work with the inclusion of samples from Guatemala, which would allow to test this scenario. Using an approximate Bayesian computation (ABC) approach, I showed that if the Western Mediterranean invasions were first introduced to California as prior information has suggested (Dalton 2006), they most likely spread across the US directly in independent events and not sequentially from state to state. The application of such powerful methods could be extended with the inclusion of more samples from other states, as well as samples from Guatemala, to test more explicitly additional scenarios of this invasion - from potential sources of origin, to a potential intermediate source, to dispersal across the US. Such findings highlight the importance of better monitoring of whitefly (and other pest) invasions at points of entry and dispersal of associated host plants. Further, the work broadens our understanding of how insects travel on plant material through international corridors and spread from the initial sites of introduction.

This dissertation has contributed important insights to our knowledge of the history and biology of the *B. tabaci* complex, and indicates the potential of this species to serve as a model system to address intriguing applied and evolutionary questions pertaining to the origins, diversification, emergence, and spread of invasive species.

References

Ahmed MZ, Shatters RG, Ren SX, Jin GH, Mandour NS, Qiu BL. 2009. Genetic distinctions among the Mediterranean and Chinese populations of *Bemisia tabaci* Q biotype and their endosymbiont *Wolbachia* populations. *J. Appl. Entomol.* 133: 733-741.

Bedford ID, Briddon RW, Brown JK, Rosell RC, Markham PG. 1994. Geminivirus transmission and biological characterization of *Bemisia tabaci* (Gennadius) biotypes from different geographic regions. *Ann. Appl. Biol.* 125: 311-325.

Bethke JA, Byrne FJ, Hodges GS, McKenzie CL, Shatters RG. 2009. First record of the Q biotype of the sweetpotato whitefly, *Bemisia tabaci*, in Guatemala. *Phytoparasitica* 37: 61-64.

Boykin LM, Shatters RG, Rosell RC, McKenzie CL, Bagnall RA, De Barro P, Frohlich DR. 2007. Global relationships of *Bemisia tabaci* (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Mol. Phylogen. Evol.* 44: 1306-1319.

Brower AVZ, DeSalle R. 1998. Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: the utility of wingless as a source of characters for phylogenetic inference. *Insect Mol. Biol.* 7: 73-82.

Brown JK. 2007. The *Bemisia tabaci* complex: genetic and phenotypic variability drives begomovirus spread and virus diversification. Plant Disease Feature. http://www.apsnet.org/online/feature/btabaci Accessed: July 15, 2010.

Brown JK. 2010. Phylogenetic biology of the *Bemisia tabaci* sibling species group. In *Bemisia: Bionomics and Management of a Global Pest*, ed. PA Stansly, SE Naranjo. Dordrecht: Springer Science. In press.

Byrne FJ, Gorman KJ, Cahill M, Denholm I, Devonshire AL. 2000. The role of B-type esterases in conferring insecticide resistance in the tobacco whitefly, *Bemisia tabaci* (Genn). *Pest Manage. Sci.* 56: 867-874.

Campbell BC. 1993. Congruent evolution between whiteflies (Homoptera: Aleyrodidae) and their bacterial endosymbionts based on respective 18S and 16S rDNAs *Curr. Microbiol.* 26: 129-132.

Dalton R. 2006. The Christmas invasion. *Nature* 443: 898-900.

De Barro PJ. 2005. Genetic structure of the whitefly *Bemisia tabaci* in the Asia-Pacific region revealed using microsatellite markers. *Mol. Ecol.* 14: 3695-3718.

De Barro PJ, Driver F, Trueman JWH, Curran J. 2000. Phylogenetic relationships of world populations of *Bemisia tabaci* (Gennadius) using ribosomal ITS1. *Mol. Phylogen. Evol.* 16: 29-36.

Dennehy TJ, DeGain B, Harpold G, Brown JK, Byrne F, Fabrick JA, Morin S, Nichols R. 2005. New challenges to management of whitefly resistance to insecticides in Arizona. 2005 Vegetable Report, College of Agriculture and Life Sciences, The University of Arizona, Tucson. http://cals.arizona.edu/pubs/crops/az1382/contents.html Accessed: July 15, 2010.

Dinsdale A, Cook L, Riginos C, Buckley YM, De Barro P. 2010. Refined global analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. *Ann. Entomol. Soc. Am.* 103: 196-208.

Frohlich DR, Brown JK, Bedford I, Markham P. 1996. Mitochondrial 16S ribosomal subunit as a molecular marker in *Bemisia*, and implications for population variability. In *Bemisia: 1995. Taxonomy, Biology, Damage, Control and Management*, ed. D Gerling, RT Mayer, pp. 143-145. Wimbome: Intercept Publications.

Frohlich DR, Torres-Jerez I, Bedford I, Markham P, Brown J. 1999. A phylogeographic analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers. *Mol. Ecol.* 8: 1593-1602.

Galtier N, Nabholz B, Glemin S, Hurst GDD. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol. Ecol.* 18: 4541-4550.

Hughes AL, Welch R, Puri V, Matthews C, Haque K, Chanock SJ, Yeager M. 2008. Genome-wide SNP typing reveals signatures of population history. *Genomics* 92: 1-8.

Kirk AA, Lacey LA, Brown JK, Ciomperlik MA, Goolsby JA, Vacek DC, Wendel LE, Napompeth B. 2000. Variation in the *Bemisia tabaci* s.i. species complex (Hemiptera: Aleyrodidae) and its natural enemies leading to successful biological control of *Bemisia biotype* B in the USA. *Bull. Entomol. Res.* 90: 317-327.

Legg JP, French R, Rogan D, Okao-Okuja G, Brown JK. 2002. A distinct *Bemisia tabaci* (Gennadius) (Hemiptera: Sternorrhyncha: Aleyrodidae) genotype cluster is associated with the epidemic of severe cassava mosaic virus disease in Uganda. *Mol. Ecol.* 11: 1219-1229.

Metzker ML. 2010. Sequencing technologies — the next generation. *Nat. Rev. Genet.* 11: 31-46.

Pool JE, Hellmann I, Jensen JD, Nielsen R. 2010. Population genetic inference from genomic sequence variation. *Genome Res.* 20: 291-300.