Stony Brook University

Caribbean *Drosophila melanogaster*:

a model for incipient sexual isolation and admixture between genetically and phenotypically divergent populations in widespread species

A Dissertation Presented

By

Roman Yukilevich

To

The Graduate School

In Partial Fulfillment of the

Requirements

For the Degree of

Doctor of Philosophy

In

Ecology and Evolution

Stony Brook University

August 2008

Stony Brook University

The Graduate School

Roman Yukilevich

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

John R. True, Dissertation Advisor,
Associate Professor, Department of Ecology and Evolution

Walt Eanes, Chairperson of Defense,
Professor, Department of Ecology and Evolution

Robert R. Sokal, Committee member,
Distinguished Professor, Department of Ecology and Evolution

Christine RB Boake, Outside committee member,
Professor and Chair, Ecology and Evolutionary Biology,
University of Tennessee, Knoxville, TN

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation:

Caribbean *Drosophila melanogaster*:

a model for incipient sexual isolation and admixture between genetically

and phenotypically divergent populations in widespread species

by

Roman Yukilevich

Doctor of Philosophy

in

Ecology and Evolution

Stony Brook University

2008

Abstract: Understanding the beginning stages of speciation is of central importance in evolutionary biology. In this project I focused on the evolution of sexual isolation, which is one of the key mechanisms of speciation in nature. In this case populations genetically diverge for both their mating preferences and mating cues such that individuals reproductively discriminate against each other. It is of interest to ask whether widespread and human commensal species that have only recently expanded their geographical range can evolve such mechanisms. Here I study natural populations of the fruit fly *Drosophila melanogaster* in United States, Caribbean islands and Africa. The *D. melanogaster* African populations are ancestral and exhibit incipient sexual isolation from more temperate populations outside of Africa. Previous research suggests that Caribbean populations may be uniquely derived directly from Africa several hundred years ago via the human trans-Atlantic slave trades. Here I ask whether Caribbean populations harbor exceptional African traits in morphology, behavior and pheromones. My results show that Caribbean populations segregate many African traits and exhibit partial sexual isolation from some temperate US populations. I then study genome-wide patterns of differentiation between US and Caribbean populations. Caribbean genomes, despite sharing many sites in common with US populations, have also many important African genomic regions. This suggests that Caribbean flies are admixed populations between US and African flies and retain their African traits associated with sexual isolation despite gene flow from US. These results suggest that natural and sexual selection maintains African identity in the Caribbean. It implies that widespread and human commensal species may spread alleles for sexual isolation throughout the species range fairly rapidly. This facilitates the process of speciation within such species on an ecological time scale.

Table of Contents

List of Figures

Chapter I

Chapter II

Chapter III

List of Tables

Chapter I

Chapter II

Chapter III

Acknowledgements

I would like to thank my family and friends, especially my lovely wife Sri, for support and patience during my Ph.D. work. Without them I would simply not have made it. I also want to thank Abhi, my 1 year old son, who has given me a new happiness and perspective on life. I would also like to thank my Ph.D. advisor John True for his friendship and his inspiration. His encouragement to let me pursue my own questions and systems has allowed me to flourish as a scientist and made me realize my potential and creativity. He is truly an amazing advisor. I would also like to thank Shian-Ren for teaching me everything that I know about genetic work and for her great friendship. I thank C. Boake, W. Eanes, D. Futuyma, J. Lachance, S.R. Liou, P. Nosil, J. Rohlf, E. Rolán-Álvarez, R. Sokal and J. True for valuable comments and discussions on earlier versions of my dissertation. I thank P. Chang, L. Jung, and S.R. Liou for assistance with *desat*-2 locus genotyping, S. Fang for first assaying some of our lines for *desat-2* locus variation, J. Logan for assistance with courtship behavior assays, and J. Nocero, R. Oge and P. Chang for morphological measurements and B. Pollack for assistance with multiple-choice mating tests. I thank F. Aoki for writing the program for sparse Mantel tests and for other simulation work. I am also grateful to J. Pool and C. Aquadro for sending African isofemale lines and the Bahamas Agriculture Department in Nassau, Bahamas for permission and permit to collect isofemale lines. I thank NSF (DIG) and Stony Brook University for supporting this work.

Curriculum Vita (short)

Postdoctoral Fellow, University of Chicago (2008-
Department of Ecology and Evolution: Advisor Chung-I Wu.

Ph.D., Stony Brook University, 2008.
Department of Ecology and Evolution: Advisor John R. True.

BS, University of Illinois at Chicago, 2000
Major: Biology

Publications:

Yukilevich, R., and J. R. True (2008) African phenotypes and incipient sexual isolation between US and Caribbean *Drosophila melanogaster*. *Evolution.* In Press.

Yukilevich, R., J. Lachance, F. Aoki, and J. R. True (2008) Long-term adaptation of epistatic genetic networks. *Evolution*. In Press.

Nosil, P. and R. Yukilevich (2008) Mechanisms of reinforcement in natural and simulated polymorphic populations. *Biol. J. Linn. Soc*. In Press.

Yukilevich, R., and J. R. True (2008) Incipient sexual isolation among cosmopolitan *Drosophila melanogaster* populations. *Evolution*. doi:10.1111/j.1558-5646.2008.00427.x

Yukilevich, R., and J. R. True (2006) Divergent outcomes of reinforcement speciation: the relative importance of assortative mating and migration modification. *American Naturalist.* 167(55):638-654.

Yukilevich, R. (2005) Dispersal evolution in fragmented habitats: the interplay between the tendency and ability to disperse. *Evolutionary Ecology Research*. 7(7):973-992.

Previous Funding:

Postdoctoral Fellowship from University of Chicago (2008)

National Science Foundation: Dissertation Improvement Grant (2006)

CHAPTER I:
INCIPIENT SEXUAL ISOLATION AMONG COSMOPOLITAN *DROSOPHILA MELANOGASTER* POPULATIONS

Abstract: Understanding the biological conditions and the genetic basis of early stages of sexual isolation and speciation is an outstanding question in evolutionary biology. It is unclear how much genetic and phenotypic variation for mating preferences and their phenotypic cues is segregating within widespread and human commensal species in nature. A recent case of incipient sexual isolation between Zimbabwe and cosmopolitan populations of the human-commensal fruit fly *Drosophila melanogaster* indicates that such species may initiate the process of sexual isolation. However, it is still unknown whether other geographical populations have undergone evolution of mating preferences. In this study I present new data on multiple-choice mating tests revealing partial sexual isolation between US and Caribbean populations. I relate my findings to African populations, showing that Caribbean flies are partially sexually isolated from Zimbabwe flies, but mate randomly with West African flies, which also show partial sexual isolation from US and Zimbabwe flies. Thus, Caribbean and West African populations seem to exhibit distinct mating preferences relative to populations in the US and in Zimbabwe. These results suggest that widespread and human commensal species may harbor different types of mating preferences across their geographical ranges.

INTRODUCTION

The evolution of reproductive isolation and speciation has recently received a great deal of both theoretical and empirical attention (for reviews see Gavrilets 2004; Coyne and Orr 2004). One of the most pervasive conclusions of this body of research has been the idea that speciation becomes more probable in allopatry and is in general more likely with increasing geographical distance between populations (e.g. Mayr 1963; Coyne and Orr 2004). Geographically distant populations are more likely to experience both divergent ecological selective conditions as well as a reduction in homogenizing gene flow that is likely to be one of the greatest impediments to the evolution of isolating mechanisms (e.g. Dobzhansky 1937; Mayr 1942, 1963; Gavrilets 2004; Coyne and Orr 2004). However, many questions still remain about the exact biological conditions and the genetic basis of incipient reproductive isolation.

Of particular interest are geographically widespread species that have attained this status fairly recently either due to natural causes or with the aid of human transport (Dobzhansky 1965, 1973; Mayr 1965; Parsons 1983). It is well known that such widespread species often show clinal geographical differentiation, but what is less clear is the extent to which these species segregate genetic and phenotypic variation for incipient sexual and reproductive isolation (i.e. clinal sexual isolation; Mayr 1963; Endler 1977; Lande 1982; Doebeli and Dieckmann 2003). In principle, the very processes responsible for how these species become geographically widespread, namely high migration rates and rapid colonization of novel environments, should generally prevent populations of these species from diverging. On the other hand, since these species often span great geographical distances, often experiencing a wide range of environmental and ecological conditions, there is potential to diverge in phenotypic and behavioral traits that may lead to incipient sexual isolation.

It is still unclear which factors are generally more important in the evolution of widespread species. This question becomes especially pertinent when considering recently widespread human commensals, since I would expect that migration and gene flow is still ongoing and is relatively high in these species compared to species that spread due to natural processes (e.g. Dobzhansky 1965; Carson 1965; Parsons 1983; Coyne and Orr 2004). Thus it is interesting to ask whether variation can segregate for mating preference behaviors within such species and whether there is potential for early stages of sexual isolation as in other systems of recently diverged populations or species (e.g. Verrell and Arnold 1989; Tilley et al. 1990; Endler and Houde 1995; Johannesson et al. 1995; Noor 1995; Funk 1998; Tregenza et al. 2000; Jiggins et al. 2001; Tregenza 2002; Nosil et al. 2002; Pfennig and Simovich 2002; Jiggins et al. 2004; Boughman et al. 2005; Ortíz-Barrientos and Noor 2005).

One of the few and perhaps best known cases of incipient sexual isolation in a recently widespread human-commensal species is the fruit fly, Drosophila melanogaster (David and Capy 1988). This species has expanded from its ancestral African range to encompass most of the world in the last several thousand to hundred years with the aid of human transport (David and Capy 1988; Lachaise et al. 1988; Lachaise and Silvain 2004; Keller 2007). Despite earlier claims that this species mates randomly across its range (Henderson and Lambert 1982), recent evidence indicates that some African populations in and near Zimbabwe ("Z-type") have evolved incipient sexual isolation from

"cosmopolitan" ("M-type") populations (Wu et al. 1995; Hollocher et al. 1997a; Hollocher et al. 1997b; Greenberg et al. 2003).

Females from both Zimbabwe and cosmopolitan strains tend to preferentially mate with their own "Z-type" or "M-type" males, respectively, with some Zimbabwe strains showing very strong mating preferences (Wu et al. 1995; Hollocher et al. 1997a). The Zimbabwe-cosmopolitan sexual isolation also parallels substantial divergence in nuclear genes (Begun and Aquadro 1993), microsatellite loci (Kauer et al. 2002; Caracristi and Schlotterer 2003; Kauer and Schlotterer 2004), chromosomal inversions (Aulard et al. 2002) and various phenotypic traits, including body size, pigmentation, cuticular hydrocarbon composition and wing beat frequency, that are all known to be under genetic control (David and Capy 1988; Colegrave et al. 2000; Rouault et al. 2001; Takahashi et al. 2001). This system suggests that the evolution of reproductive isolation in recently widespread and human commensal species is possible.

However, it is still unclear whether other populations in this species may also show divergence in mating behaviors and preferences. With the possible exception of "microhabitat" isolation in Brazzaville populations of West Africa (Capy et al. 2000; Haerty et al. 2002, 2005), and segregation of weaker Zimbabwe-like mating preferences across South Africa (Hollocher et al. 1997a), no other large-scale geographic case of sexual isolation has been described in this species (Henderson and Lambert 1982). In general, the cosmopolitan populations of D. melanogaster are assumed to mate randomly with one another and are not known to segregate genetic and phenotypic variation for sexual isolation (Handerson and Lambert 1982; Hollocher et al. 1997a).

However, previous authors have noted that Caribbean island D. melanogaster populations show some peculiar phenotypic and pheromonal differences from other cosmopolitan populations (David and Capy 1988). Of special interest is the fact that Caribbean populations are the only populations outside of Africa known to harbor the African insertion allele at the desat2 locus (Takahashi et al. 2001). This locus solely determines a female cuticular hydrocarbon polymorphism between African and non-African flies and has been recently implicated in the sexual isolation between Zimbabwe and cosmopolitan strains (Greenberg et al. 2003; but see Coyne and Elwyn 2006). Further, Caribbean populations are known to segregate exceptional African-like morphologies, making them phenotypically distinct from other cosmopolitan populations in US and Europe (Capy et al. 1993, 1994). Recent microsatellite evidence also indicates that US flies are more genetically similar to African flies than are European flies, suggesting that African alleles may have introgressed into North America, possibly via the Caribbean islands (Caracristi and Schlotterer 2003, but see Capy et al. 1986).

Despite these intriguing phenotypic and genetic observations, the Caribbean populations remain largely unexplored. So far, mating preferences of only four isofemale lines have been studied. These lines were more similar to cosmopolitan than Zimbabwe lines (Fang et al. 2002). In the present work I remedy this situation by explicitly performing multiple-choice mating tests across southeastern US and Caribbean populations. My findings reveal that some Caribbean populations exhibit partial sexual isolation from US populations and that there is a substantial geographical variation in the presence and strength of these mating preferences among the islands. Further, I find that Caribbean flies mate randomly with West African flies and flies from both of these

regions show partial sexual isolation from US and Zimbabwe populations. These results suggest that D. melanogaster cosmopolitan populations are segregating mating preference behaviors that may be distinct from those of Zimbabwe lines. This implies that widespread and human commensal species may possibly evolve different types of mating preferences across their ranges.

MATERIALS AND METHODS

STUDY SYSTEM AND REARING CONDITIONS OF ISOFEMALE LINES

In the summer of 2004, R. Yukilevich collected wild females (established as isofemale lines) from 37 locations, 18 from the southeastern United States (Missisippi, Alabama, Georgia and Florida) and 19 from the islands of the Bahamas. Additional collections were undertaken in 2005 and 2006 in St. Lucia and Haiti (Fig. 1). African strains were acquired from J. Pool and C. Aquadro in 2005. These included 31 isofemale lines from West Cameroon (Mbalang-Djalingo) collected by J. Pool in 2004, 13 isofemale lines collected in 1990 from Lake Sengwa, Zimbabwe (Zim-s) that were previously studied by Wu et al. (1995) and Hollocher et al. (1997a), and 20 isofemale lines collected in 1994 from Lake Kariba, Zimbabwe (Zim-k). The two Zimbabwe locations are about 50 km apart. Prior to my experiments, all isofemale lines were maintained in the laboratory at Stony Brook University on instant Drosophila food (Carolina Biol. Supply Inc., Burlington) in a 25°C incubator with a 12h light:12h dark regime. Thus, other than for the Zimbabwe lines, my analyses were performed on lines that were less than two years old.

MULTIPLE-CHOICE MATING TESTS

All mating trials were done in the laboratory at Stony Brook University in 2005 and 2006. I followed the recommendation of Cesares et al. (1998) by concentrating on multiple-choice mating trials since mating propensity and preference are known to be confounded in no-choice male and female tests (see also Coyne et al. 2005). I also studied a single US Bahamas comparison under no-choice mating conditions that qualitatively confirmed my multiple-choice results. I randomly chose two locations at a time. For multiple-choice mating tests I used individuals from all available isofemale lines from each location (see Fig. 1). For each pairwise test I simultaneously set up all available isofemale lines of both locations by placing seven fertilized females into each experimental narrow glass vial (25 x 95 mm) and allowed them to lay eggs for four days under identical conditions. I then cleared the vials, inserted paper for pupation and waited about 15 days until eclosion of adults. Virgin males and females from each line were then collected using light CO_2 anesthesia and were aged separately in pooled sex-specific vials for four to five days. Individuals of each line were equally represented from each location. The day before the mating trial, I randomly picked equal numbers of females and males and placed them separately on instant Drosophila colored food for about 12 hours to mark their abdomens for identification (color has no effect on mating preferences; Wu et al. 1995; Boake et al. 2003).

I performed all mating trials within three to five hours of "lights on" in the laboratory. I used a common Plexiglas mating chamber (28_L x 17.5_H x 16_W cm) by placing standard corn meal/molasses/agar fly food on the floor and walls of the chamber

to simulate D. melanogaster mating in the wild, which typically takes place on or near food. I introduced 30 males from each location into the chamber without anesthesia by simply opening the vial into a hole in the chamber and letting the individuals fly out. These individuals were allowed to habituate for five minutes. I then introduced 30 females from each location using the same technique, for a total of 120 individuals per test. I let the mating trials run until 50% of possible copulations had occurred (as recommended by Cesares et al. 1998 to avoid bias) and then placed the chamber into a -20˚C or -80˚C freezer for about 20 minutes to kill the flies in the state of copulation (as in Boake et al. 2003). This procedure avoids disturbing copulating pairs and is effective at retrieving about 98-100% of copulations (Boake et al. 2003). I then scored all copulations based on abdomen color. I cleaned the chambers between tests with 95% ethanol, followed by soap and water to ensure all pheromones from previous tests had been eliminated. Note that all replicates used the same isofemale lines.

The approach of using all isofemale lines from each locality to perform mating tests has several advantages over strategies that focus on testing specific isofemale lines (as in Wu et al. 1995; Hollocher et al. 1997a). The first advantage is that I can survey the mating behavior of many isofemale lines simultaneously in a given experiment. This is more likely to mimic mating choices in the wild since it is very likely that individuals encounter potential mates with different genetic backgrounds during mating. Second, introducing all isofemale lines from a given population into the experimental mating cage avoids any biases that may come from "vial-effects" or individuals having unusual preferences for members of the same isofemale line. An individual is able to choose a mating partner from many isofemale lines of its native and foreign location. This approach also allows the important possibility of behavioral interactions among individuals of different isofemale lines. Therefore this design is a better representation of a "populational" measure of sexual isolation as opposed to an "isofemale" line measure that may or may not be representative of mating behavior of the population as a whole. Since I are interested in surveying many populations for sexual isolation, it would have been prohibitively difficult to perform many pairwise isofemale line tests for any sensible number of population pairs. The major disadvantage of this approach is that I cannot identify specific isofemale lines that show especially strong sexual isolation as has been done by Wu et al. (1995) and Hollocher et al. (1997a, 1997b). This is important for genetic dissection of mating behavior, which is not the focus of the present work.

We used the sexual isolation index I_{psi} of Rolán-Álvarez and Caballero (2000), which ranges from -1 to +1 with 0 = random mating. The standard joint sexual isolation index of Merrel (1950; see also Malagolowkin-Cohen et al. 1965) gave virtually identical results. Significance and P-values were derived by resampling 10,000 times in JMATING software (http://webs.uvigo.es/acraaj/JMsoft.htm - Carvajal-Rodríguez and Rolán-Álvarez 2006). I also performed a X^2 contingency test of independence, which tests if there is a significant mating interaction between individuals of two locations (Sokal and Rohlf 1995).

RESULTS

INCIPIENT SEXUAL ISOLATION BETWEEN US AND CARIBBEAN FLIES

Table 1 shows the results of pairwise multiple-choice tests of sexual isolation performed between different localities both within southeastern US and Bahamas and between these regions. First, among the 11 different pairwise tests within regions, none showed a significant excess of homotypic matings or deviation from random mating, as indicated by extremely low sexual isolation (I_{psi}) indexes and chi-square values. One test yielded a fairly high negative I_{psi} index between two neighboring US localities, but was not significant. There was no significant heterogeneity within either region among different pairwise tests ($G_{Sex.Isol.Heterog.(withinUS)} = 2.61$; df = 8; P = 0.95; $G_{Sex.Isol.Heterog.(withinBahamas)} = 0.56$; df = 10; P > 0.995). After a total of 1629 copulations the average sexual isolation indexes for within-US and within-Bahamas did not significantly deviate from random mating (average $I_{psi(US)} = 0.0019$ (st.dev. = 0.032), $X^2 = 0.379$, P = 0.544; average $I_{psi\ (Bahamas)} = -0.001$ (st.dev. = 0.029), $X^2 = 0.0006$, P = 0.973; Fig. 2a). Together these results show that there does not appear to be geographically based mating preference within either the US or the Bahamas.

However, tests performed between US and Bahamas populations gave strikingly different results. For a total of eighteen US and Bahamas pairwise tests performed, the average I_{psi} index was 0.0896 with a standard deviation of 0.017, which is significantly different from random mating (3785 copulations, $X^2 = 29.27$, $P < 10^{-5}$). This means on average 54% homotypic matings relative to 0.46% heterotypic matings. This index is also significantly different from average I_{psi} indexes of within-US and within-Bahamas comparisons (P < 0.017; Fig. 2a). I found variation in the presence and strength of sexual isolation among different US and Bahamas pairwise tests. Out of a total of 18 tests, seven had significant chi-square values with an average of 58% homotypic and 42% heterotypic matings and another test had marginal significance with 56% homotypic and 44% heterotypic matings. Thus, nearly half of the US-Bahamas pairwise tests revealed significant assortative mating. For these eight pairwise tests, the average I_{psi} index equals 0.165 with a standard deviation of 0.026 (1416 copulations, $X^2 = 37.99$, $P < 10^{-5}$). Note however that most between-region pairwise tests showed positive assortative mating, such that there was no significant heterogeneity among different pairwise tests ($G_{Sex.Isol.Heterog.(between)} = 18.31$; df = 34; P = 0.985).

When I analyzed all eighteen US-Bahamas mating tests further, I found that on average, homotypic US and Bahamas pairs comprised 25% and 29% of all matings, respectively, while the reciprocal heterotypic pairs made up only 23% and 22% of all matings (Fig. 2b). This result indicates that the strength of sexual isolation is asymmetrical with the Bahamas individuals showing stronger mating preferences compared to US individuals (Estimation of Asymmetry 'I_{apsi}' = 1.0119; st. dev. 0.0044; P = 0.0016; Rolán-Alvarez and Caballero 2000). Similar asymmetry is observed for the significant pairwise US-Bahamas tests (average: $27\%_{US}$ and $31\%_{Bahamas}$ homotypic matings, 21% heterotypic matings).

I further investigated whether the observed sexual isolation between US and Caribbean populations could be of the same type as that of the Zimbabwe case or if this represents a distinct case of sexual isolation. In addition to comparing my results to Zimbabwe isofemale lines, I also tested my populations against West African (Cameroon) isofemale lines. It has been hypothesized that West African populations may

have been the primary colonization source of Caribbean populations (e.g. David and Capy 1988; Caracristi and Schlotterer 2003).

RELATIONSHIP TO AFRICA

Table 2 gives the results of pairwise multiple-choice mating tests between North American and African populations. As expected, I found that US populations were significantly sexually isolated from Zimbabwe (particularly from the strongly sexually divergent Sengwa population that was used by Wu et al. 1995; see Table 2). This result confirms previous reports of Wu et al. (1995) and Hollocher et al. (1997a) using my "populational" multiple-choice mating approach and is strong evidence that this approach is able to capture assortative mating in D. melanogaster in general (see *Materials and Methods*). The I_{psi} indexes between US and Zimbabwe (Sengwa) were 0.189 and 0.292, with an average I_{psi} index of 0.24 (st. dev. 0.05), which is significantly higher than the average I_{psi} index for all eighteen US-Bahamas pairwise tests (0.089; st. dev. 0.0017; Kruskal-Wallis test: $P = 0.038$), and higher than, but not significantly different from, the average I_{psi} index between the eight significant/marginally significant US-Bahamas pairwise tests (0.164; st. dev. 0.02; Kruskal-Wallis test: $P = 0.09$). This suggests that US-Bahamas sexual isolation is weaker than the isolation between US and Zimbabwe. However, the Zimbabwe population from Lake Kariba (Zim-k) did not show significant sexual isolation with US and Bahamas flies (Table 2). This supports previous findings that mating preferences in Zimbabwe are not fixed within populations (Hollocher et al. 1997a).

Mating tests between Bahamas islands and the Zimbabwe (Sengwa) population resulted in the average I_{psi} index of 0.159 (st. dev. 0.05) and only one of the tests was significant (Table 2). This suggests that Zimbabwe-Caribbean populations have weaker mating discrimination compared to US-Zimbabwe populations. Using my Caribbean lines and the standard cosmopolitan French line FrV3-1, S. Fang (pers. comm.) also found that Caribbean females mate more readily with Zimbabwe males than with the French males. Despite relatively low sexual isolation between Bahamas and Zimbabwe lines, it is evident that these populations do not mate at random. Thus it is unlikely that sexual isolation between US-Caribbean populations is of the same type as between US-Zimbabwe populations.

I performed mating tests between North American populations and a West African population from Cameroon. Interestingly, I found that US-West African populations were significantly sexually isolated from one another with I_{psi} indexes of 0.202 and 0.17 for different replicates and an average I_{psi} index of 0.186 (Table 2). Even though this US-West African case of sexual isolation is novel, these results are consistent with previous reports showing that some West African lines show sexual discrimination against other cosmopolitan lines (Cohet and David 1980; Scott 1994; Capy et al. 2000; Haerty et al. 2005; see Discussion). Strikingly, when I performed mating tests between Bahamas and West African populations, I found complete random mating (Table 2). This suggests that Caribbean and West African populations have similar mating preferences and/or mating cues.

If Caribbean and West African populations share mating preferences, I would expect significant sexual isolation between West Africa and Zimbabwe since partial

sexual isolation was observed between Bahamas and Zimbabwe (see above). Interestingly, I found that the sexual isolation between West African and Zimbabwe (Sengwa) populations was the highest observed in my study with I_{psi} index of 0.282 (Table 2). This suggests that West African and Zimbabwe lines have diverged substantially in their mating preferences. I summarize my mating choice results in Fig. 3, which shows the average isolation indexes of relevant regional tests.

DISCUSSION

INCIPIENT SEXUAL ISOLATION BETWEEN COSMOPOLITAN POPULATIONS

Our results provide evidence of incipient sexual isolation occurring between cosmopolitan strains of D. melanogaster. Since cosmopolitan populations are generally assumed to not deviate from random mating, I discuss some of the potential reasons for my observations. I note that the original paper studying multiple-choice mating tests of world-wide D. melanogaster populations did not sample any lines from the Caribbean or West Africa (Henderson and Lambert 1982). Subsequent work by Wu et al. (1995) and Hollocher et al. (1997a) only looked at a single West African line (from Tai, Ivory Coast), which showed a significant level of sexual isolation from Zimbabwe lines and was therefore taken as a representative of cosmopolitan mating behavior. Even though my results are consistent with these observations, they also indicate that US and West African populations are partially sexually isolated.

Interestingly, the original paper that described divergent mating behavior between African and cosmopolitan flies discovered the phenomenon in an isofemale line from a West African population (Brazzaville, Congo) that is geographically adjacent to Cameroon (Cohet and David 1980; see also Capy et al. 2000; Haerty et al. 2005). Since the Congo line has not been directly tested against Zimbabwe lines, it is not clear whether it is also sexually isolated from Zimbabwe. Similar evidence comes from the West African Tai, Ivory Coast isofemale line, which was shown to be partially sexually isolated from the cosmopolitan Canton-S isofemale line (Scott 1994; Cobb and Ferveur 1996; Grillet et al. 2006; but see Coyne et al. 1999). This may mean that both the Congo and the Ivory Coast mating preferences may ultimately be of the same type as the Cameroon case described here. The partial sexual isolation between West African and Zimbabwe is consistent with evidence indicating that West and East (including Zimbabwe) African strains make up two genetically distinct groups (e.g. Bénassi and Veuille 1995; Aulard et al. 2002; Baudry et al. 2004; Pool and Aquadro 2006).

The only previous study to examine mating preferences of Caribbean populations was by Fang et al. (2002). These authors analyzed four isofemale lines (one from Cuba, one from Dominican Republic and two from St. Croix). These results showed that these Caribbean lines were more M-like (cosmopolitan) than Z-like in mating preferences, resembling lines from Central and West Africa (see Hollocher et al. 1997a). However, partial sexual isolation was described in a recent study of mating behavior of desaturase-2 pheromone locus transgenic lines (lines that carry the Zimbabwe insertion allele in an otherwise cosmopolitan background; Coyne and Elwyn 2006). These authors argued that the transgenic lines may represent Caribbean-like genotypes since some Caribbean populations also carry the insertion allele at this pheromone locus, but are otherwise

presumed to have a cosmopolitan genetic background (Takahashi et al. 2002; Fang et al. 2002; Coyne and Elwyn 2006). My results of partial sexual isolation between US and Caribbean populations are consistent with these findings. However since I observe geographical variation for incipient sexual isolation in this system, it is not surprising that Fang et al. (2002) found weak sexual isolation in their lines. Strongly sexually isolated isofemale lines in the Caribbean, if these exist at all, would have been difficult to find without prior geographical knowledge of variation in mating preferences.

HOW AFRICAN ARE CARIBBEAN FLIES?

Clearly my results raise new questions about the genetic and phenotypic makeup of Caribbean populations and their relationship to African strains. The finding that Caribbean and West African populations mate randomly with each other, but are sexually isolated from both US cosmopolitan and Zimbabwe strains suggests that this case of sexual isolation is likely to have a distinct genetic and phenotypic basis from that of Zimbabwe. This may mean that both Caribbean and West African flies prefer certain phenotypes that are absent in cosmopolitan and Zimbabwe strains or that these flies prefer intermediate trait values between extreme cosmopolitan and Zimbabwe phenotypes.

The observation that some Caribbean populations harbor exceptional African-like phenotypes and pheromones has led some to suggest that Caribbean flies may have had direct origins from Africa via historical trans-Atlantic slave trades (David and Capy 1988; Capy et al. 1993, 1994; Rouault et al. 2001; Takahashi et al. 2002; Caracristi and Schlotterer 2003; Kauer and Schlotterer 2004, but see Capy et al. 1986). My sexual isolation patterns are consistent with this historical scenario, but it is certainly possible that Caribbean populations may represent convergent evolution onto African-like tropical phenotypes and mating preferences. I have begun to characterize in detail phenotypic and behavioral variation associated with incipient sexual isolation between US, Caribbean and African populations in order to determine whether these cases of incipient sexual isolation are associated with similar or different trait divergence (Yukilevich and True, 2008). Future work should also focus on phylogenetic relationships between Caribbean, US and African populations using genetic markers and sequences (see also Caracristi and Schlotterer 2003). Finding a genetic signature of African (especially West African) ancestry in the Caribbean would clearly support the hypothesis that Caribbean populations are recent, direct descendents of African populations and that the mating preferences in the Caribbean are likely to be the same as those in West Africa.

IMPLICATIONS FOR SEXUAL ISOLATION IN WIDESPREAD AND HUMAN COMMENSAL SPECIES

These results extend our understanding of the evolution of incipient sexual isolation within recently expanding widespread and human commensal species. Previous findings of incipient sexual isolation between Zimbabwe and cosmopolitan populations indicated that species with expanding geographical ranges and those associated with humans are capable of rapid divergence in sexual mating preferences and phenotypic cues (Wu et al. 1995; Hollocher et al. 1997a). However, it was generally assumed that populations from

the rest of the species range mate randomly and do not segregate mating preference behaviors.

My results indicate otherwise, by showing that populations in Caribbean and West Africa exhibit incipient sexual isolation with other cosmopolitan populations in US and that these mating preferences are likely to be distinct from those of Zimbabwe. This suggests that more than one type of mating preference has evolved in D. melanogaster. These results contribute to a rapidly changing view of the evolutionary history and biology of D. melanogaster, which was originally thought to be a genetically undifferentiated and panmictic unit (see David and Capy 1988 for details). These findings imply that widespread and human commensal species can segregate genetic and phenotypic variation for mating preferences and initiate the process of sexual isolation in different parts of the species range.

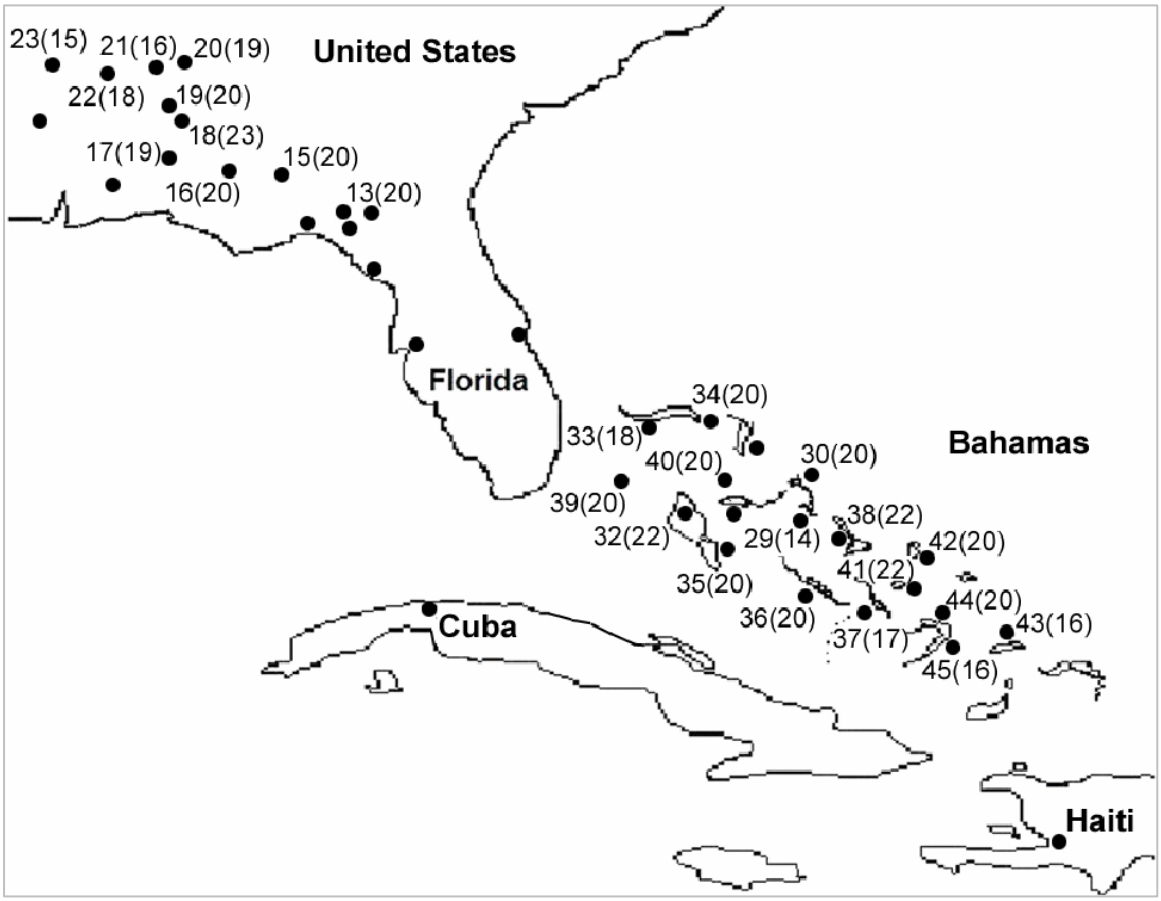


Figure 1. Map of localities in southern US and Bahamas used in multiple-choice mating tests of sexual isolation. Each locality is represented by an identification number, followed by the number of isofemale lines (in parentheses) from each location used in mating tests. Other localities where lines are available, but were not tested for mate choice are shown as isolated dots.

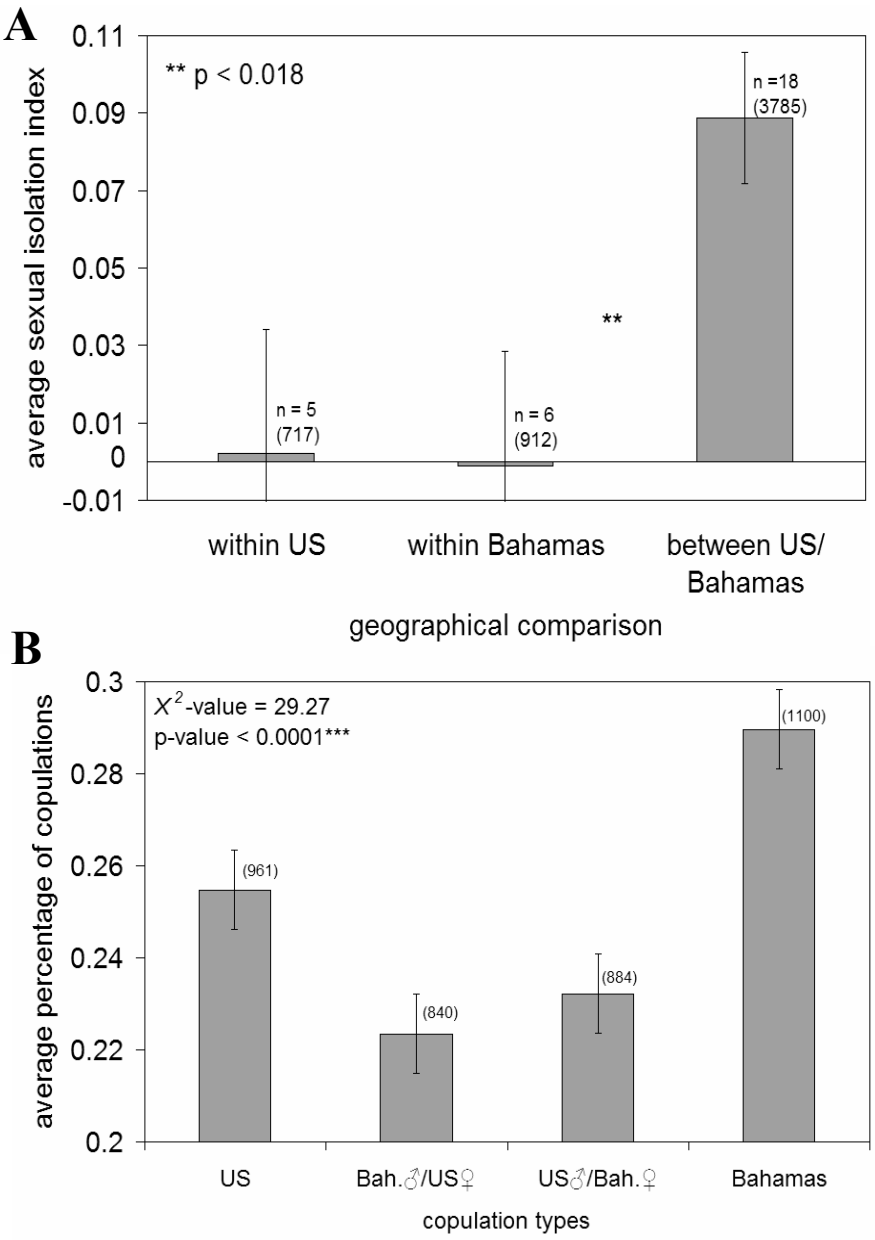


Figure 2. Average sexual isolation I_{psi} index for multiple-choice mating tests within US, within Bahamas, and between US and Bahamas localities (A). Average percent of copulations of homotypic and heterotypic pairs for between US-Bahamas comparisons (B). The I_{psi} index in (A) was determined using the software JMATING (range: –1 to +1 with 0 = random mating; see text). For each geographic comparison, n represents the number of pairwise mating tests (see Table 1) and values in parentheses represent the total number of copulations accumulated. Error bars represent standard deviations.

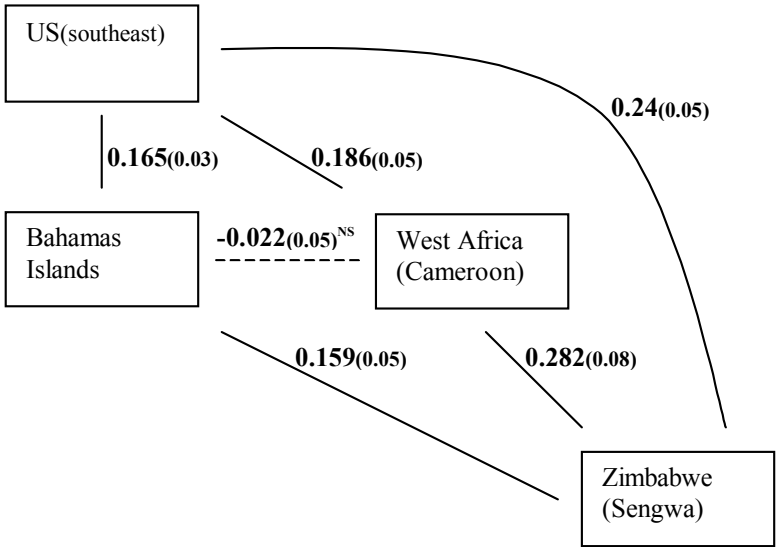


Figure 3. Average pairwise sexual isolation (I_{psi}) indices from multiple-choice mating tests between populations of US, Bahamas, West Africa and Zimbabwe in Tables 1 and 2. Standard deviation of the sexual isolation index is shown in parentheses (all indices except Bahamas-Cameroon, are significant at $P < 0.0001$). The averages are based on the following number of pairwise comparisons (and number of copulations): US-Bah: 7(1416), US-Zim: 2(371), US-Cam: 2(403), Bah-Zim: 2(352), Bah-Cam: 2(371), Cam-Zim: 1(156). Note the average I_{psi} index of 0.164 is only for the eight significant and marginally significant pairwise tests between US and Bahamas (see text). If all eighteen pairwise tests were used, the average I_{psi} index would be 0.089 (0.02) for US-Bah. Similarly, only pairwise tests with Zimbabwe (Sengwa) were used, not Zimbabwe (Kariba) since my interest is to compare my results to the strong sexual isolation of Zimbabwe (Sengwa) described by Wu et al. 1995 and Hollocher et al. 1997a. If Zim(k) is also included, the average I_{psi} index would be 0.206 (0.04) for US-Zim, and average I_{psi} index = 0.099 (0.04) for Bah-Zim. Cam-Zim(k) isolation was not tested.

Table 1. Pairwise multiple-choice mating tests of sexual isolation within and between US and Bahamas regions. Numbers of copulations for each type are shown from left to right locations as (n_{aa}, n_{ab}, n_{ba}, n_{bb}), with the left subscript designating females and the right subscript males. Abbreviations are for number of copulations, number of replicates, and percent homotypic matings.

id#'s	# copul.	# repl.	(n_{aa}, n_{ab}, n_{ba}, n_{bb})	%homot.	*Ipsi* index	st.dev.	GI or X^2 values	*P*-value
Within US								
15-19	190	6	(57, 50, 43, 40)	0.511	0.021	0.073	0.040	0.841
13-17	104	4	(25, 21, 30, 28)	0.510	0.019	0.098	0.071	0.8
21-22	68	2	(10, 22, 17, 19)	0.426	-0.147	0.120	1.820	0.151
22-23	169	5	(42, 41, 40, 46)	0.521	0.041	0.077	0.283	0.582
13-22	186	5	(60, 45, 41, 40)	0.538	0.075	0.073	0.785	0.363
Within Bahamas								
36-38	128	4	(31, 29, 36, 32)	0.492	-0.016	0.088	0.021	0.962
30-32	67	2	(20, 20, 13, 14)	0.507	0.015	0.122	0.022	0.878
29-39	185	6	(57, 45, 46, 37)	0.508	0.016	0.074	0.004	0.926
35-41	227	7	(53, 54, 58, 62)	0.507	0.013	0.066	0.033	0.844
33-37	127	4	(44, 39, 26, 18)	0.488	-0.024	0.089	0.430	0.522
33-45	178	5	(36, 43, 47, 52)	0.494	-0.011	0.075	0.064	0.782
Between US and Bahamas								
19-34	246	7	(77, 48, 58, 63)	0.569	0.138	0.063	4.650	**0.021**
13-36	244	8	(69, 44, 60, 71)	0.574	0.154	0.063	5.700	**0.013**
22-33	176	6	(34, 47, 32, 63)	0.551	0.088	0.075	1.282	0.268
20-40	181	6	(58, 38, 35, 50)	0.597	0.193	0.073	6.720	**0.009**
15-37	238	7	(47, 49, 59, 83)	0.546	0.075	0.065	1.273	0.275
21-32	288	9	(81, 60, 77, 70)	0.524	0.052	0.059	0.746	0.3876
22-35	172	6	(44, 33, 37, 58)	0.593	0.184	0.075	5.680	**0.016**
17-38	152	6	(41, 30, 42, 39)	0.526	0.059	0.081	0.530	0.4632
22-30	320	10	(74, 71, 71, 104)	0.556	0.106	0.056	3.503	**0.056**
16-45	175	5	(43, 37, 34, 61)	0.594	0.182	0.074	5.686	**0.014**
21-42	127	4	(32, 36, 27, 32)	0.504	0.013	0.089	0.021	0.885
20-37	129	5	(30, 37, 29, 33)	0.488	-0.020	0.088	0.052	0.8174
18-44	224	7	(50, 44, 63, 67)	0.522	0.047	0.067	0.488	0.4906
23-36	328	9	(80, 87, 73, 88)	0.512	0.027	0.055	0.216	0.623
23-41	177	5	(51, 42, 32, 52)	0.582	0.169	0.074	5.000	**0.02**
22-45	180	6	(48, 33, 40, 59)	0.594	0.190	0.074	6.370	**0.011**
23-44	256	8	(57, 65, 66, 68)	0.488	-0.023	0.062	0.164	0.661
21-43	131	4	(36, 28, 39, 28)	0.489	-0.020	0.087	0.051	0.822

Note: Bold values designate significant ($P < 0.05$) or marginally significant *P*-values. All P-values were determined using JMATING software by bootstrapping 10,000 times (see text).

Table 2. Pairwise multiple-choice mating tests of sexual isolation between US and Bahamas and Zimbabwe and West African populations.

id#'s	# copul.	# repl.	(n_{aa}, n_{ab}, n_{ba}, n_{bb})	%homot.	*Ipsi* index	st.dev.	GI or X^2 values	*P*-value
Between Zimbabwe								
and US:								
21-Zim(s)	217	7	(78, 53, 35, 51)	0.594	0.189	0.067	7.420	**0.003**
19-Zim(s)	154	5	(65, 33, 21, 35)	0.649	0.292	0.081	12.009	**0.0006**
16-Zim(k)	184	6	(67, 31, 48, 38)	0.571	0.136	0.076	3.080	0.076
and Bahamas:								
45-Zim(s)	202	7	(79, 43, 38, 42)	0.599	0.177	0.072	5.902	**0.016**
35-Zim(s)	150	5	(51, 35, 29, 35)	0.573	0.141	0.082	2.885	**0.0864**
41-Zim(k)	193	7	(56, 37, 62, 38)	0.487	-0.018	0.072	0.065	0.8122
Between West Africa (Cameroon)								
and US:								
22-Cam	198	6	(67, 36, 43, 52)	0.601	0.202	0.070	7.835	**0.004**
19-Cam	205	6	(59, 35, 51, 60)	0.580	0.170	0.070	5.791	**0.011**
and Bahamas:								
45-Cam	179	6	(57, 43, 47, 32)	0.497	-0.026	0.077	0.113	0.747
35-Cam	192	6	(57, 44, 53, 38)	0.495	-0.019	0.075	0.064	0.804
Between Zimbabwe and West Africa (Cameroon)								
Cam-Zim(s)	156	6	(56, 39, 17, 44)	0.641	0.282	0.077	14.411	**0.0009**

Note: See description in Table 1.

CHAPTER II:
AFRICAN PHENOTYPES AND INCIPIENT SEXUAL ISOLATION BETWEEN U.S. AND CARIBBEAN *DROSOPHILA MELANOGASTER*

Abstract: Understanding the beginning stage of incipient sexual isolation and speciation is an important pursuit in evolutionary biology. The fruit fly *Drosophila melanogaster* is a useful model to address questions about the early stages of sexual isolation occurring within widespread species. This species exhibits sexual isolation between cosmopolitan and African flies, especially from Zimbabwe populations. In addition, I have recently described another example of partial sexual isolation between some US and Caribbean populations. This and other phenotypic data suggests that Caribbean flies might be segregating African-like traits that may be responsible for US-Caribbean sexual isolation. In the present work I investigate this question. I study the geographical variation at the pheromone locus *desaturase*-2, as well as morphology and courtship behavior across the US-Caribbean region. I find that US and Caribbean populations show sharp geographical clines in all phenotypes and demonstrate that Caribbean traits are more similar to those of Africa than to US populations. Further, I find that African-like traits in the Caribbean are geographically associated with sexual isolation and best explain variation in sexual isolation when all traits are considered together. Thus my results suggest that Caribbean mating preferences are likely to be based on African-like phenotypes and that even at such early stages of sexual isolation individuals may already cue in on several traits simultaneously during mate choice.

INTRODUCTION

Sexual or behavioral isolation is common in nature and plays a central role in maintaining genetic and phenotypic differences between sibling species (Dobzhansky 1937; Mayr 1942; 1963; Coyne and Orr 2004). Understanding how sexual isolation evolves requires that I capture the process before it has reached completion. Of particular interest are natural populations that occupy different environments and that have diverged in traits involved in assortative mating (Dobzhansky and Streisinger 1944; Koref-Santibáñez 1972; Verrell and Arnold 1989; Tilley et al. 1990; Endler and Houde 1995; Johannesson et al. 1995; Noor 1995; Funk 1998; Tregenza et al. 2000; Tregenza 2002; Jiggins et al. 2000; Nosil et al. 2002; Pfennig and Simovich 2002; Jiggins et al. 2004; Boughman et al. 2005; Ortiz-Barrientos and Noor 2005; Nosil et al. 2007). These and other examples of incipient sexual isolation have helped to elucidate the geographical, ecological and historical context of early stages of speciation.

Recent evidence suggests that incipient sexual isolation may occur multiple times in various parts of the species range, especially among geographically widespread species (e.g. Rundle et al. 2000; Nosil et al. 2002; Tregenza 2002; Boughman et al. 2005). It thus becomes important to understand the similarities and differences between multiple cases of sexual isolation within species. For instance, how much genetic and phenotypic differentiation accompanies each case of incipient sexual isolation, which phenotypic traits typically diverge, and what is the genetic and phenotypic basis of assortative mating in each case (e.g. Panhuis et al. 2000; Hendry 2001; Boughman et al. 2005; Nosil 2005; Hendry et al. 2006).

If multiple cases of sexual isolation were to show unique genetic and phenotypic bases, it would indicate that species typically segregate different types of variation associated with assortative mating in nature. On the other hand, if different cases of sexual isolation share common phenotypes and occupy similar environments, it would imply that there might be adaptive constraints on the conditions and the type of variation underlying these processes. For instance, it is likely that assortative mating may evolve as a byproduct of adaptation to similar environments (see Schluter and Nagel 1995; Rundle and Schluter 2004; also see artificial selection evidence: Killas et al. 1980; Dodd 1989, but see Rundle 2003). This is known as “parallel speciation” and has been recently documented in benthic versus limnetic forms of sticklebacks (e.g. Rundle et al. 2000; McKinnon et al. 2004; Boughman et al. 2005; Vines and Schluter 2006), host-associated races of phytophagous insects (e.g. Funk 1998; Nosil et al. 2002) and possibly in color morphs of cichlids (Allender et al. 2003). In these examples, parallel cases of sexual isolation within species have been argued to evolve independently of one another since populations with parallel phenotypes are typically genetically distant (e.g. Taylor and McPhail 1999; 2000; Rundle and Schluter 2004; but see Coyne and Orr 2004 pg. 406-7). More direct tests of independent genetic basis of parallel evolution of assortative mating behaviors and cues have yet to be performed.

On the other hand, if the same genetic basis is involved in multiple cases of sexual isolation within species it is most likely that this is a result of common genetic history (e.g. Rundle and Schluter 2004). Migration of alleles responsible for parallel cases of adaptation has recently been described in several species (Colosimo et al. 2004, 2005; also see Schluter et al. 2004). Similarly, it is possible for assortative mating behaviors

and associated traits that initially caused sexual isolation in one geographical region to be introduced into other populations, recreating the phenomenon in whole or in part. This scenario may also involve ecological adaptation. For instance, if assortative mating originally evolved as a byproduct of ecological adaptation, it is likely that these behaviors would establish in places where they are preadapted to similar environments. Thus, assortative mating behaviors and associated traits may migrate in a non-random way throughout the species range.

This "historical migration" scenario would imply that once certain population pairs evolve incipient sexual isolation, assortative mating could spread into other populations and thus accelerate the process of speciation across the species range. This would likely lead to multiple cases of incipient sexual isolation more quickly, compared to independent or parallel speciation since migration of relevant alleles is likely to occur on ecological rather than evolutionary time scales. Presently, it is unclear whether there are any biological candidates of this scenario in nature.

The fruit fly *D. melanogaster* is a powerful genetic model species to address questions about the phenotypic basis of incipient sexual isolation in nature. In this species, matings between Zimbabwe females and males from outside this region (known as "cosmopolitan") are typically very rare (Wu et al. 1995; Hollocher et al. 1997a). Cosmopolitan females also tend to prefer to mate with their own males, but in this case the preferences are often weaker (Wu et al. 1995). It has been postulated that sexual isolation may have originated as a byproduct of adaptation to temperate versus tropical environments of cosmopolitan and Zimbabwe populations, respectively (Greenburg et al. 2003; but see Coyne and Elwyn 2006).

This species also segregates for mating preferences in other parts of the world, such as the case of "microhabitat" sexual isolation in the West African Brazzaville populations (Capy et al. 2000; Haerty et al. 2002, 2005). Moreover, I have recently discovered that Caribbean and West African populations mate randomly with each other, but show partial sexual isolation against US cosmopolitan flies and against Zimbabwe flies (Yukilevich and True, submitted). It is also noteworthy that crosses between Caribbean and Zimbabwe populations show weaker sexual isolation compared to crosses between US and Zimbabwe populations. These results are consistent with previous findings that some Caribbean populations have African-like morphology and pheromones (David and Capy 1988; Capy et al. 1993, 1994; Rouault et al. 2001;Takahashi et al. 2002; but see Fang et al. 2002; Coyne and Elwyn 2006). This has led some to suggest that Caribbean populations may ultimately be of African, rather than European origin, perhaps having come directly via the Trans-Atlantic slave trade from various parts of Africa, especially from West Africa (e.g. David and Capy 1988; Caracristi and Schlotterer 2003).

The Caribbean island populations of *D. melanogaster* have been little studied. Not much is known about pheromonal, phenotypic and behavioral variation across US and Caribbean populations and the extent to which African-like phenotypes segregate in the region. Also, it is unknown what geographical relationships exist between sexual isolation and phenotypic variation that may be involved in mate choice in this region (e.g. Bastock and Manning 1955; Bastock 1956; Ewing 1961, 1964; Partridge et al. 1987; Taylor and Kekic 1988; Pitnick 1991; Grillet et al. 2006; Coyne and Elwyn 2006).

In the present paper I study variation in several traits that are of particular interest for sexual isolation in *D. melanogaster.* Thus I address whether the case of incipient sexual isolation between US and Caribbean flies shares phenotypic similarity to that found between cosmopolitan and African flies. First, I focus on the *desaturase*-2 (*desat*-2) locus that is solely responsible for the difference in a major female cuticular hydrocarbon between African and non-African flies (Coyne et al. 1999; Dallarec et al. 2000; Takahashi et al. 2001; Greenburg et al. 2003). It has been previously reported that non-African populations carry a derived deletion allele at this gene, which results in a high 5,9-heptacosadiene (HD) female phenotype, whereas African populations have retained the ancestral allele, which confers much less 5,9-HD, but high levels of 7,11-HD (see e.g. Takahashi et al. 2002; Coyne and Elwyn 2006). The only non-African populations known to segregate the African ancestral allele are Caribbean populations (Takahashi et al. 2002).

This locus has recently been proposed to play a major role in local adaptation to tropical versus temperate environments and as a byproduct to cause sexual isolation between cosmopolitan and Zimbabwe strains (Greenburg et al. 2003). However, the view that *desat*-2 contributes to environmental adaptation has recently been challenged (Coyne and Elwyn 2006). Nevertheless, Coyne and Elwyn still provided some support for the claim that *desat-2* may be partially responsible for sexual isolation. Thus it is of interest to characterize geographical variation at this gene across the US and Caribbean region and to test for any associations with sexual isolation.

Second, I study male morphology. Previous research on a few isofemale lines from the Lesser Antilles revealed that these individuals have smaller body size than those from temperate populations of Europe and US and are thus most similar to African flies, which are also relatively small (Capy et al. 1993, 1994). Male body size is known to be associated with male mating success in both laboratory and field conditions, with females typically preferring to mate with large males (Ewing 1961, 1964; Partridge et al. 1987; Taylor and Kekic 1988; Pitnick 1991). Here I present geographical patterns of male morphology and its relationship to sexual isolation.

Finally, I investigate male courtship behavior since it is known to be under direct sexual selection in *D. melanogaster* (e.g. Bastock and Manning 1955; Bastock 1956). In particular, these studies have demonstrated that mutants with lower frequencies of certain courtship elements have reduced mating success compared to wildtype males. Even though little is known about variation in courtship behaviors in *D. melanogaster*, including courtship behavior of African males, I have characterized this variation because of its potential role in mate choice.

Our results show that most Caribbean populations are divergent from US populations at the *desat-2* pheromone locus and in morphology and courtship behaviors and that these traits in the Caribbean populations are more similar to both West Africa and Zimbabwe than to US populations. I also demonstrate that these African-like traits are positively geographically correlated with the strength of sexual isolation, indicating that incipient sexual isolation between US and Caribbean flies and between cosmopolitan and African flies are associated with similar phenotypic trait differences. Thus my results reveal many phenotypic parallels between these different cases of incipient sexual

isolation and raise further questions about the genetic makeup of Caribbean populations relative to West Africa and Zimbabwe.

MATERIALS AND METHODS

STUDY SYSTEM

In the summers of 2004 and 2005, R. Yukilevich collected and established isofemale lines from 39 different locations in the southeastern US and various Caribbean islands including the Bahamas Archipelago, Cuba, Haiti, and St. Lucia in the Lesser Antilles (see Fig. 1 for map and Appendix Table A1 for details). African isofemale lines were acquired from J. Pool and C. Aquadro in 2005, which included West African populations as well as southeastern African lines from Zimbabwe (Appendix Table A1). All lines have been maintained at Stony Brook University laboratory on instant *Drosophila* food (Carolina Biol. Supply Inc., Burlington, NC) in a 25°C temperature incubator with a 12h light:12h dark regime. My analyses were performed on lines that were generally less than two years old.

The sampled populations of *D. melanogaster* represent a 10-degree change in latitude and a 15-degree change in longitude, with about 2,500 km separating the most distant populations (Fig. 1). Preliminary collections in 2003 by R. Yukilevich suggest that these populations are fairly stable within and between years, although episodic regional events, such as droughts and hurricanes, may potentially affect the demography of the populations. Climate differs across the localities. Cooler and more humid conditions prevail on the mainland, where pines are the dominant vegetation, and hotter and dryer conditions occur on the islands, which are dominated by mangrove and scrub (Correll and Correll 1982; Baker 2001). The number of *Drosophila* species also varies across locations ranging from about fifteen sympatric species in the U.S. to zero sympatric species on many of the far-flung islands. Human population size declines from mainland to islands with a few thousand to a few hundred people on the islands (Baker 2001). *D. melanogaster* abundance tends to parallel this decline.

DESATURASE-2 LOCUS VARIATION

The *desaturase-2* locus is responsible for variation in a major female cuticular hydrocarbon (CH) in *D. melanogaster* (Dallarec et al. 2000; Takahashi et al. 2001). A single 16-bp indel at site 12,552 in the 13-kb region of *desat-2* shows complete concordance with CH profiles, with the 16-bp *insertion* allele conferring the ancestral (African) phenotype of high 5,9/7,11 CH ratio and the 16-bp *deletion* allele conferring the derived (Cosmopolitan) phenotype of low 5,9/7,11 CH ratio (Takahashi et al. 2001). I developed a 200-bp marker with primers flanking the 16-bp 12,552 site indel (forward primer sequence: 5' gct cgc cag cta tct acg ac 3'; reverse primer sequence: 5' ata caa tcc ggc agc ttt tg 3'). I then amplified the sequence region using polymerase chain reaction (PCR) (annealing temp. 59°C, 50_{mm}MgCl concentr.1/20 ml., 34 cycles) and ran the PCR products on 1% agarose gels for genotyping. I genotyped a single first generation (F1) lab individual from each isofemale line (see Appendix Table 1 for number of isofemales sampled per locality). I then calculated the percentage of *desat-2* locus homozygotes and heterozygotes and overall frequency of *insertion* and *deletion* alleles in each population

and calculated the inbreeding coefficient (*Fis*) as 1 - (observed heterozygote freq./expected heterozygote freq. based on HWE).

MORPHOLOGICAL MEASUREMENTS AND UPGMA CLUSTERING

All morphological measurements were performed on individuals from single isofemale lines maintained in the laboratory for less than two years. Earlier measures of thorax length and pigmentation of males maintained for only one year showed virtually identical patterns. I used a randomized block experimental design, where each block represented a single day for a total of thirteen day-blocks. On each of the thirteen days, I setup three well-maintained lines (defined as having over 50 adults) from each available location. For each line, I transferred approximately 50 parental adults into experimental narrow glass vials (25 x 95 mm) with standard corn meal/molasses/agar fly food and with a piece of paper for pupation and let females lay eggs for about two hours, at which time the adults were cleared. This ensured low egg laying density in experimental vials (about 30 eggs per vial) so as to create ideal growing conditions for the next generation. I then placed the experimental vials into a 25°C chamber until adult emergence (approximately 12-13 days).

For each block, I aged the emerging adults to 4-5 days before imaging. I randomly chose on average 56 isofemale lines to image on each of the thirteen-day blocks (on average 1 line per locality). For each isofemale line, I placed live adults onto a CO_2 pad under a LEICA MZ7 stereomicroscope with a fiberoptic light source at fixed light level and randomly picked one male and one female to image. Images were taken using a Zeiss AxioCam video camera attached to the microscope with AxioVision (Rel. 4.3) software on adults standing upright, with the dorsal view facing the camera. This approach increases the speed of imaging and avoids any surgical manipulation of the fly. I repeated this procedure for each experimental block for 13 days.

Images of flies were measured (with the worker blind to the line of origin of each fly) using the ImageJ software (http://rsb.info.nih.gov/ij/). I measured: 1) thorax length (measured along the dorsal midline from the most anterior point of the thorax to the most posterior point of scutellum), 2) thorax width (measured between posterior supraalar macrochaetae), 3) head width (between the left and right postvertical bristles), 4) wing length (from the intersection of the L4 vein and the anterior crossvein to the distal intersection of the L1 and L3 veins), 5) wing width (from the distal tip of the axillary vein to the intersection of the posterior crossvein and the L5 vein), and 6) thorax pigmentation (measured as the luminosity of digital pixels of the total thorax area using ImageJ software; presence and absence of thorax trident pigmentation was also scored with qualitatively similar results; data not shown). Luminosity is inversely related to the degree of pigmentation. In total, 1,463 individuals were measured (on average 17 isofemale lines per locality; see Appendix Table 1). Location means for each variable were used for spatial analyses and UPGMA clustering (ANOVAs are available upon request).

To determine how population structure was associated with morphological variation, a standardized matrix of location means for each of the six variables was created (NTSYSpc 2.20; Rohlf 2004). I then calculated a distance matrix based on all six

variables (see below). Finally, the morphological distance matrix was used to create a phenogram using the UPGMA clustering method (NTSYSpc 2.20; Rohlf 2004).

NO-CHOICE COURTSHIP BEHAVIOR ASSAYS

Male courtship behavior with standard Canton-S isofemale line

Male courtship behavior was first studied using no-choice tests with females from the standard isofemale line Canton-S (obtained from W. Eanes). Since it is possible that females may influence the courtship behavior of males, it was important to standardize the tester female (Cook 1979; Tompkins et al. 1981; O'Dell 2003, but see below). The assays were performed from January to May of 2005 using a randomized block experimental design. Each experimental block was initiated by picking two stock lines from each represented locality, totaling about 32 lines per block for a total of 10 blocks. These lines were then set up simultaneously using standard cornmeal/molasses/agar media in a 25°C incubator with a 12h light: 12h dark cycle). For each line I placed seven fertilized females into an experimental narrow glass vial (see above) and let them lay eggs for four days. I then cleared the vials, inserted paper for pupation and waited about 12-13 days until eclosion of next generation adults. I then collected 10 virgin males from each line using light CO_2 and aged them separately from females for 4-5 days.

I randomly picked an isofemale line and aspirated a random male from that line into a plastic Petri-dish (35 x 10 mm) containing a small piece of *Drosophila* food (since individuals typically mate on food in the wild) through a hole on its side that was then closed using a piece of cotton. The Petri-dish was then placed under a stereomicroscope. My tests were always performed within 1-5 hours of "lights on" at 25°C to 27°C. I let the male habituate for one min. and then aspirated a random virgin female, 4-5 days old, from the Canton-S isofemale line into the chamber. Upon introducing the female into the mating chamber I began timing. *Courtship latency* was measured as the time from introduction of the female to the initiation of male courtship behaviors. I studied standard male courtship behavior elements of *D. melanogaster* (as defined by Spieth 1952; Bastock and Manning 1955; Bastock 1956; Welbergen et al. 1987; Greenspan and Ferveur 2000; O'Dell 2003). These included: 1) orienting toward the female by either following her or standing facing her, 2) wing vibration and extension, 3) wing scissoring, 4) licking female ovipositor with proboscis, and 5) attempted copulation. I also measured acts of no courtship, defined when a male ignored the female by not orienting towards her (running away or standing on the opposite side of the chamber). Variation in some of these courtship elements has a known genetic basis. For instance the heritability of wing vibration frequencies in *D. melanogaster* is around 7% in the lab (e.g. McDonald 1979). Mutations at many well-known genes affect these behaviors (e.g. Bastock 1956; Greenspan and Ferveur 2000; Billeter et al. 2002; Drapeu et al. 2003).

These behaviors were recorded using voice-recognition software IBM Via Voice that entered one of the male behavioral acts every 1.5 seconds into a Microsoft word document in a sequential manner (video recordings are available upon request). I studied each male-female pair until successful copulation or when 10 minutes passed from the initiation of male courtship. I then determined the frequency of each element (as in Bastock 1956; Welbergen et al. 1987). The frequency of each element is simply the number of instances of each element (e.g. wing vibration) out of the total number of male

courtship acts during the courtship period. In total, I were able to study 327 mating pairs (on average 9 isofemale lines per locality).

Courtship in homotypic and heterotypic matings of US and Bahamas flies

I also studied courtship behavior of males and females from populations in Columbus, Mississippi, US (#23) and from Port Nelson, Rum Cay Island, Bahamas (#41). The setup of isofemale lines and the protocol for studying courtship behavior were identical to the above procedures. On each day of the experiment, I studied an equal number of homotypic and heterotypic matings from both US and Bahamas isofemale lines. In this study I focused on the rate of a particular behavior during courtship, determined by counting all acts of a particular behavior that occurred during the courtship period (defined above), then dividing this number by the total time of courtship in seconds. Video recordings (available upon request) were used to assay courtship.

SPATIAL STATISTICAL ANALYSES

Distance matrices

For each measured variable, I created a n x n distance matrix, where n is the number of data points. Each value in the matrix, d_{ij}, is a measure of the distance between the ith and jth data values (x_i and x_j). I used standard Euclidean distances to calculate d_{ij}. For a single variable (such as geography or *desat*-2 locus allele frequency), $d_{ij} = \text{abs}\,(x_i - x_j)$. For distance matrices that include multiple variables, such as for morphological distances, courtship distances, and combined trait distances $d_{ij} = \sqrt{\sum (x_{iv} - x_{jv})^2}$, where v represents each included variable. Before distances with multiple variables were calculated I standardized the variables using NTSYS software (NTSYSpc 2.20; Rohlf 2004). All distance matrices were created using PASSAGE software (Rosenberg 2002).

Spatial Correlograms

I used a correlogram to describe the average spatial autocorrelation between pairwise localities as a function of their geographical distance (e.g. Sokal and Oden 1978a; 1978b; Epperson 2003). I grouped location pairs into different distance classes and measured the average correlation for each distance class using PASSAGE (Rosenberg 2004). I used Moran's *I* coefficient which ranges from -1 to +1. I created ten distance classes based on the geographical distance matrix by assuming an equal number of location pairs per distance class. My connection matrix assumed that all location pairs within a certain distance range are connected and all others are not (binary weight matrix; a Gabriel connection gave similar results). I used PASSAGE (Rosenberg 2004) to determine the significance of individual autocorrelation coefficients and the entire correlogram. Linear clines are revealed when short distance classes show positive Moran's *I* and long distance classes show negative Moran's *I* (Sokal and Oden 1978a, 1978b; Sokal 1979a; Sokal et al. 1987; Barbujani 2000; see Epperson 2003 for further details).

Design Matrices

I created design matrices to test explicit hypotheses about the relationship between geography and phenotypic variation. A design matrix describes the relative distances among populations *expected* under a particular geographical hypothesis (Waddle et al.

1998). A significant correlation between the design matrix and the actual phenotypic distance matrix indicates that phenotypic differentiation is described well by a particular geographical clustering of localities in the design matrix. To construct a design matrix I use a binary assignment of distances of 0's and 1's, where 0 means relatively short distance and 1 means relatively long distance (for an example, see Appendix Fig. A1). I developed six design matrices with specific geographical delimitations across the US-Caribbean region (Appendix Fig. A2). This determines which geographical regions show more phenotypic differentiation across the US-Caribbean localities.

Mantel tests

Correlations between different distance matrices were calculated using the standard Mantel test of matrix correlation using PASSAGE (Mantel 1967; Sokal 1979b; see Epperson 2003). I tested the significance of the correlation by permuting the rows and columns of one of the matrices 499 times and comparing observed to permuted correlations (Rosenberg 2004). I also used a partial Mantel test, which tests the correlation between two matrices, while holding the third constant (Smouse et al. 1986; Epperson 2003; but see Castellano and Balletto 2002; Rousset 2002).

Finally, I developed a 'sparse' Mantel test to test the significance of geographical associations between matrices where one of the matrices was sparse (i.e. not all pairwise comparisons have data available, such as for my sexual isolation matrix; see below). As in the standard Mantel test, I permuted the rows and columns of one of the matrices, but only calculated the product-moment correlation for values that existed in both matrices. This is valid as long as only one of the matrices is sparse (program available upon request).

RESULTS

DESATURASE-2 LOCUS

Fig. 1 shows geographical variation at *desat*-2 locus across the US-Caribbean system. I found that the African insertion allele at this locus is widespread throughout the Bahamas and exhibits a sharp geographical cline in the area as indicated by location-specific pie charts and the corresponding spatial correlogram (Fig. 1). This generates a significant *Fst* value of 0.221 across the region. US populations north of Alabama and Mississippi are nearly fixed for the deletion allele, while St. Lucia in the Lesser Antilles shows a similar insertion allele frequency ($freq_{\cdot insertion} = 0.55$) to the southern Bahamas islands. Compared to African populations in Cameroon (West Africa) and Zimbabwe (Sengwa and Kariba), which are either fixed or nearly fixed for the insertion allele (see inset in Fig. 1), the insertion allele in the Caribbean is intermediate in frequency between US and African populations. These data suggest that the CH pheromone frequencies determined by *desat*-2 locus will follow the same pattern.

I also found that this gene shows a very different pattern of observed to expected heterozygosity as determined by the inbreeding coefficient (*Fis*) across Bahamas versus US populations (Fig. 2). Across the islands and southern Florida, there is a strong clinal pattern with the southern islands showing a positive *Fis* or deficit in heterozygotes and northern islands showing a negative *Fis* (heterozygote excess). This pattern may be a result of either: 1) general inbreeding, 2) within-island assortative mating or 3) selection

against heterozygotes at this locus (Hedrick 2000). The excess of heterozygotes in the northern islands and Florida indicates either outbreeding in general or specifically based on this locus. No clinal pattern exists across US localities as typically very small deviations in heterozygote frequencies from expectations were seen (Fig. 2). In general, my data indicate that most Bahamas islands are more similar to African populations at this locus than to US populations.

MORPHOLOGY

To investigate how morphology varies across the region, I first performed a Principal Component Analysis (PCA) on five morphological variables (head width, thorax length, thorax width, and wing length and wing width; Table 1). The PCA revealed that the first Principal Component (PC1) explained 61.61% of total variance. All traits loaded positively on PC1, indicating that it is associated with general body size. PC2 explained 13.84% of total variance, which primarily corresponded to a tradeoff between: a) head width and b) wing length and width (Table 1). Despite large variation among individuals, there is a clear difference between US and Caribbean flies with respect to these first two Principal components (Fig. 3a). US flies tend to be typically larger (positive PC1), but have more narrow heads with longer and wider wings (negative PC2).

To see how this individual variation translates to location-specific patterns, I pooled individuals based on their location. I find that both location means of PC1 and PC2 show steep clines across the US-Caribbean region with PC1 having a much steeper cline compared to PC2, as revealed by spatial correlograms (Fig. 3B). Similarly, PC1 means are highly correlated with both longitude and latitude (Fig. 3C). Virtually all US localities exhibit positive PC1 means and all Bahamas localities exhibit negative PC1 means. Further, US and Caribbean flies significantly differ in thorax pigmentation, with dark flies in US and lighter flies in the Caribbean (see clinal differentiation in Figs. 3B and 3D). Similar R^2 values as in Fig. 3 were observed among females. This extends previous results of clines in morphology among eastern US populations (e.g. Coyne and Beecham 1987).

To illustrate how these morphological patterns relate to African populations, Fig. 4 shows regional means of morphological traits of US and Caribbean flies relative to West and Southeast African flies. Caribbean populations exhibited an intermediate body size between US and African flies, but showed smaller wing size and lighter pigmentation than even in West or Southeast African populations. The UPGMA cluster analysis on all morphological traits resulted in two major phenotypic clusters, one with only US populations and the other containing all Caribbean and African populations with a few US populations from Florida and southern Alabama (Fig. 5). For example, West African populations from Niger and Guinea were morphologically most similar to Haiti and Long Island in the southern Bahamas. Note also that Caribbean islands from northern and southern parts of the archipelago cluster in different groups. In general, my results indicate that Caribbean flies are morphologically more similar to African populations than to US populations.

MALE COURTSHIP BEHAVIOR

Below I describe results from no-choice mating tests with a single male and a tester-female from Canton-S isofemale line. First, the initiation of male courtship or latency was not significantly different between US and Caribbean populations (avg. $latency_{US}$ = 81.2 secs. n = 159; avg. $latency_{Carib.}$ = 75.9 secs. n = 168; Z-score = 0.332; $P = 0.74$; Kruskal-Wallis test). However, upon initiation of courtship, males differed in the frequency of their courtship behavior elements towards the tester female. Table 2 summarizes a PCA of all courtship elements. PC1 explained 42.28% of the variance and corresponded to a tradeoff between: a) wing vibrations/extensions and ovipositor licking behavior and b) following/orienting toward the female and ignoring the female during courtship. This indicates that the most variable male behavior was the degree to which males were active in attempting to stimulate females using courtship song and licking behavior. PC2 explained 20.32% of total variance, which primarily corresponded to a trade-off between: a) ignoring the female during courtship and b) following/orienting toward the female. PC3 explained 16.17% of the variance and primarily corresponded to a trade-off between: a) wing scissoring and attempting copulation and b) wing vibrations/extensions and ovipositor licking. Wing scissoring is rare in *D. melanogaster*, but more common in the sibling species *D. simulans* (e.g. Welbergen et al. 1987). This analysis shows that males exhibit different mating strategies during courtship across this region.

To see if male courtship varies with respect to geographical location, I calculated the average frequencies of courtship elements across all isofemale lines studied in each population. I found significant clines in the frequency of most common courtship elements (Fig. 6). PC1 cline indicated that the proportion of wing vibrations and ovipositor licking behaviors relative to ignoring and simply following the female was significantly higher among US males relative to Caribbean males (Fig. 6). Thus Caribbean males predominantly followed or oriented towards females during courtship instead of performing wing vibrations and ovipositor licking. PC2, which is related to ignoring versus following the female during courtship, was not significantly different between US and Caribbean males, indicating that both types of males pursued females with equal perseverance. Finally, PC3 also showed a significant but weaker cline, with US males typically stimulating the female with wing vibrations and ovipositor licking, while Caribbean males displayed more wing scissoring followed by attempted copulations. Thus US and Bahamas males seem to attract females using different behavioral strategies.

Interestingly, the Cuban population was more similar to US populations in courtship behavior than it was to Bahamas populations (Fig. 6). Recent evidence indicates that Zimbabwe males also tend to exhibit lower wing vibration frequencies relative to cosmopolitan males (C.-T. Ting, unpublished data; pers. comm.). This suggests that Caribbean and African courtship behaviors are likely to be more similar to each other compared to cosmopolitan populations.

GEOGRAPHICAL PATTERNS OF DIFFERENTIATION

It is of interest to extend the above clinal analyses to explicitly describe geographical patterns of differentiation for each variable. This would allow us to determine if different African-like traits show similar or different distributions across the US and Caribbean

populations. To address this question, I created six design matrices, each representing a particular geographical hypothesis (see Table 3, Appendix Fig. A2, and Materials and Methods).

The *desat*-2 locus was significantly correlated with design matrix #5 (Divergent Family Island Hypothesis), which considers locations south of Nassau, New Providence Island to be different from all northern localities (Table 4). The finding that *desat*-2 was not significantly correlated with any other geographical hypothesis shows that most differentiation occurs between the northern and middle Bahamas islands.

On the other hand, the morphological distance matrix (based on all six traits) was most significantly correlated with design matrix #1 (Divergent Northern US Hypothesis) and with design matrix #2 (Florida-Islands Connection Hypothesis; Table 4). These hypotheses group Florida and southern parts of Georgia and Alabama with Caribbean island populations rather than other US mainland localities. This indicates that individuals from most northern parts of US localities in the study system have different morphology from all southern localities and suggests that African-like phenotypes are spread well across the Florida peninsula.

Finally, I found that the courtship behavior distance matrix (based on all six courtship elements in Table 2) was most significantly correlated with design matrix #3 (Mainland and Cuba Hypothesis) and to a lesser degree with design matrix #5 (Divergent Family Islands Hypothesis; Table 4). This suggests that most differentiation in courtship behavior occurs between mainland and island populations with the exception that Cuba is included with US. Thus although all traits show parallel clines across the region, the specific geographical patterns of differentiation vary between traits. This indicates that African-like traits differ in their distribution across the US-Caribbean region.

CORRELATIONS BETWEEN VARIABLES

Using the pairwise Mantel test of matrix correlations I found that *desat*-2 locus alleles, morphology and courtship behavior traits were positively correlated with geographical distance between locations and with each other (see Table 5 and Appendix Table A2). The morphological distance matrix was best correlated with geographical distance, followed by *desat-2* locus distance and least by courtship behavior distance. Since variables could be correlated simply because all are correlated with geographical distance, I also used the partial Mantel test to determine whether residual variances remain correlated when holding geographical distance constant (see Materials and Methods).

Table 5 shows that male morphology and *desat*-2 locus distance matrices are still significantly correlated when geographical distance is accounted for. However, the courtship behavior distance matrix is no longer significantly correlated with either the morphology or *desat*-2 locus matrices (Table 5). Therefore the positive association between courtship behavior and the other two variables comes only from their common association with geography. Thus variation in courtship behavior between equally distant populations is not predicted from either morphology or *desat-2* allelic frequency.

PATTERN OF SEXUAL ISOLATION

Recently, I performed multiple-choice mating tests between various populations across the US and Caribbean region (Yukilevich and True, submitted). This included 11 mating tests between populations either within US or within Bahamas regions and 18 mating tests between US and Bahamas populations. Only between-region tests revealed significant deviation from random mating with substantial variation in the presence and strength of sexual isolation. Here I would like to test whether these sexual isolation indices are geographically associated with divergence at the *desat*-2 locus, morphology, or courtship behavior. First I included all available pairwise mating tests for my analysis, which considers both within and between-region tests. I found that sexual isolation was significantly correlated with geographical distance between localities (Fig. 7A). This result occurred because tests between US and Caribbean populations have higher sexual isolation indices than tests within regions (compare gray and black points in Fig. 7). Given that all variables were positively associated with geographical distance (see Table 5 above), I found that these were also significantly correlated with sexual isolation, with *desat*-2 locus distance having the highest correlation, followed by morphology and then by courtship behavior distances (see Fig. 7B-D). However, upon removing the effect of geographical distance on sexual isolation using the partial Mantel test, none of the variables remained significantly associated with sexual isolation. Therefore, none of the phenotypic variables explained variation in sexual isolation beyond what has already been explained by geographical distance.

Since individuals across the region varied in all of these traits (see above), I asked how these variables combine together to influence the association with sexual isolation. I created a combined distance matrix based on standardized values of all thirteen variables in my dataset (see Materials and Methods). Interestingly, the combined distance matrix had the highest significant correlation with sexual isolation ($R^2 = 0.388$; $P = 0.001$; Fig. 8), which was substantially higher than any other R^2 seen previously (see Figs. 7). Moreover, the relationship between the combined distance matrix and sexual isolation remained significant even when geographical distance was held constant (Partial Mantel test: $P = 0.014$). Combining variance across multiple variables could have possibly resulted in a worse or a similar fit to sexual isolation data compared to individual variables. Since I found the fit to be better, this result suggests either that individuals utilize multiple traits for mate choice or that the combined matrix from several variables is best correlated to other unmeasured traits that best explain sexual isolation (see Discussion).

Finally, I focused exclusively on between-region tests (i.e. between US and Bahamas locations) with the highest sexual isolation indexes (black points in Figs. 7 and 8). Interestingly, I found that sexual isolation between US and Caribbean populations was not significantly associated with geographical distances ($R^2 = 0.049$; $P = 0.71$), *desat*-2 locus allele frequency distances ($R^2 = 0.051$; $P = 0.16$) or morphological distances ($R^2 = 0.0002$; $P = 0.28$). Thus sexual isolation between US and Caribbean populations exhibits more of a mosaic geographical pattern. It also supports the above conclusion that relationships of the *desat*-2 locus and morphological variables with sexual isolation were significant because each variable was correlated with geography.

However, I did find that sexual isolation is marginally significantly correlated with differentiation in courtship behavior between US and Caribbean flies ($R^2 = 0.146$; $P = 0.08$; see black points in Fig.7D). This suggests that US and Caribbean females may directly cue in on courtship behavior differences of males when choosing mates. Finally, sexual isolation between US and Caribbean flies was significantly correlated with the combined distance matrix of *desat*-2 locus alleles, morphology and courtship behavior ($R^2 = 0.161$; $P = 0.03$; see black points in Fig.8). Since the significant correlation occurred regardless of whether all comparisons or only between US and Caribbean comparisons were considered, I conclude that variation in sexual isolation is best explained when multiple traits are considered together. It is thus likely that US and Caribbean flies utilize multiple phenotypic variables when choosing mates.

NO-CHOICE MATING TESTS BETWEEN US AND BAHAMAS FLIES

To further understand how male courtship behavior influences mating preferences in US and Bahamas females, I performed homotypic and heterotypic mating tests with US individuals from Columbus, Mississippi (#23) and Bahamas individuals from Rum Cay Island (#41). These populations showed significant partial sexual isolation in multiple-choice mating tests (Yukilevich and True submitted). First, I found that US males had significantly higher rates of wing vibrations compared to Bahamas males towards wildtype females from both populations (Fig. 9A). This is consistent with my previous no-choice results using Canton-S females (see Fig. 6). Second, male wing vibrations were associated with male mating success (Fig. 9B). However, upon comparing matings with US versus Bahamas females, I observed that only US females mated successfully with males that had significantly higher wing vibration rates (Fig. 9C). Bahamas females mated randomly with respect to male wing vibration rates. These results suggest that Bahamas females have weaker mating preferences for male wing vibrations than US females, consistent with the lower wing vibration rates of Bahamas males (see Figs.6 and 9A). Despite weaker mating preferences for wing vibration rates in males, I found that Rum Cay females actively rejected US males significantly more often than their own local males by decamping during courtship (Fig. 9D). I did not see the reciprocal rejection behavior in US females (Fig. 9D). These results suggest that Bahamas females have weaker mating preferences for male wing vibrations, but actively reject US males presumably based on other phenotypic cues.

DISCUSSION

Previously I found a mosaic pattern of sexual isolation between US and Caribbean populations. Some Caribbean populations exhibited partial mating discrimination against US populations, while pairing with other islands did not deviate from random mating (Yukilevich and True submitted). Further mating tests revealed that West African flies from Cameroon mated randomly with Caribbean flies, but also showed mating discrimination against US strains. Also, Caribbean and West African populations were both partially sexually isolated from the behavioral race of Zimbabwe, although Caribbean flies showed weaker sexual isolation. This suggested that Caribbean flies might share mating preferences and other traits involved in mate choice with those of African flies, especially those from West Africa. In the present paper I study variation in

several traits that are of particular interest for sexual isolation in *D. melanogaster.* Thus I address whether the case of incipient sexual isolation between US and Caribbean flies shares phenotypic similarity to that found between cosmopolitan and African flies.

I have found that Caribbean island populations are divergent from cosmopolitan US populations in *desat*-2 pheromone locus allele frequencies and in various morphological and courtship behavior traits. These traits exhibit steep parallel clines across the US and Caribbean region, with Caribbean flies being more similar to African flies than to US flies. Therefore, this incipient sexual isolation is accompanied by pheromonal, phenotypic, and behavioral differentiation. Further, I found that despite parallel clines across the region, these African-like traits exhibit different geographical patterns. Some African-like traits were predominantly restricted to southern islands of the Bahamas, while others were distributed well into the Florida peninsula. This indicates that African-like traits are distributed unevenly across US and Caribbean populations.

Finally, I revealed that these African-like traits are geographically associated with incipient sexual isolation between US and Caribbean flies. All traits and sexual isolation were positively correlated with geographical distance. Thus all traits were correlated with sexual isolation, with the *desat*-2 locus having the highest correlation, followed by morphology and lastly by courtship. However, when only comparisons between the US and the Caribbean were considered, courtship was then best correlated with sexual isolation. These results illustrate that different phenotypes vary in the extent to which they may predict variation in sexual isolation in this region.

Interestingly I found that when all traits were considered together, they explained variation in sexual isolation much better than either geographical distance or any individual trait. This correlation between sexual isolation and combined trait divergence remained significant when I controlled for geography. Using no-choice mating tests I showed that only US females preferred males with significantly higher vibration rates. However, Bahamas females actively rejected US males presumably based on some other trait(s). Therefore, even though courtship behavior is likely to play an important role in sexual isolation, my total results suggest that multiple variables are likely to fully explain female mating preferences. This may occur because preferences are weak when based on each particular trait, but are stronger when based on several traits together. Future experiments are necessary to test these predictions.

RELATIONSHIP TO SEXUAL ISOLATION BETWEEN COSMOPOLITAN-ZIMBABWE POPULATIONS

My results relate to recent findings that the *desat*-2 locus may be responsible for climatic adaptation and sexual isolation between cosmopolitan and Zimbabwe strains (Fang et al. 2001; Greenburg et al. 2003; Coyne and Elwyn 2006). I have demonstrated that *desat*-2 exhibits a steep geographical cline across the US and Caribbean region. This suggests that this allele was introduced from Africa in the past and may perhaps be maintained presently by natural and/or sexual selection in the Caribbean. Even though a genetic signature of selection near the *desat*-2 locus has been identified (Takahashi et al. 2002; Greenberg et al. 2006), it remains to be seen if this locus is under direct selection across the US and the Caribbean.

The observation that *desat*-2 contributes to explaining sexual isolation in the Caribbean supports previous results that this locus is at least partially involved in sexual isolation (Fang et al. 2001; Coyne and Elwyn 2006). I found that most southern Bahamas islands exhibit a deficit in heterozygotes at this locus. Aside from other potential explanations, such as inbreeding, one interesting possibility may be that southern Bahamas islands carry genetically distinct subpopulations that partially discriminate against each other based on this locus. Further, as Coyne and Elwyn (2006) point out, this locus may epistatically interact with other loci of African genetic background in causing sexual isolation. This view is supported by my observation that sexual isolation between US and Caribbean flies is best explained when the African *desat*-2 allele is combined with African-like morphology and courtship behavior.

In general, these findings support the idea that the US-Caribbean case of sexual isolation may also be partially determined by *desat*-2 locus variation. However, I emphasize that because my multiple-choice mating tests indicated that both Caribbean and West African populations are significantly sexually isolated from Zimbabwe populations, *desat-2* cannot explain sexual isolation fully since all of these populations harbor the African insertion allele. It is thus likely that the basis of mating preferences in Caribbean and West Africa flies may turn out to be somewhat different from that of Zimbabwe flies. The focus of the present work is on the differences between US flies versus both Caribbean and African flies. Future research should address how Caribbean and West African mating preferences and their associated traits differ from those of Zimbabwe populations.

THE ROLE OF MALE MORPHOLOGY AND COURTSHIP BEHAVIOR IN SEXUAL ISOLATION

The observation that both male morphology and courtship behavior elements contributed to explaining patterns of sexual isolation is of particular interest. Cosmopolitan females typically discriminate against males with smaller body and wing sizes as well as against males with lower frequencies/rates of wing vibration and ovipositor licking behaviors (Bastock and Manning 1955; Bastock 1956; Ewing 1961, 1964; Partridge et al. 1987; Taylor and Kekic 1988; Pitnick 1991). It is fascinating that these are the very traits that are widespread throughout the Caribbean islands. This suggests that US and Caribbean females differ in mating preferences for these traits, which was partially confirmed by my no-choice mating experiments (see above). Mating preferences for male body size and pigmentation among US and Caribbean females have not been analyzed.

Our study does not exclude other potentially divergent traits between US and Caribbean populations from being involved in sexual isolation. Indeed, it has been previously noted that West African and Caribbean *D. melanogaster* populations harbor unique 7-tricosene and 7-pentacosene male pheromones (Rouault et al. 2001). I would thus predict that these pheromones may differ between US and Caribbean males and may be potential targets for female mate choice.

PARALLEL EVOLUTION VERSUS HISTORICAL PREADAPTIVE MIGRATION OF SEXUAL ISOLATION ALLELES

My major finding is that Caribbean populations harbor both West African and Zimbabwe-like phenotypes and behaviors, and that these traits are geographically associated with incipient sexual isolation between US and Caribbean flies. It is possible that Caribbean populations, independently of Africa, have converged onto various traits associated with adaptation, for instance, to similar tropical conditions and that this has led to partial sexual isolation from US cosmopolitan and more temperate flies. Artificial selection experiments in *Drosophila* suggest that this scenario is plausible (Kilias et al. 1980; Dudd 1989). I would then expect that the genetic basis of each case of incipient sexual isolation would be different and that Caribbean flies would be genetically divergent from African flies. This would then be a very incipient case of independent parallel sexual isolation in the New and Old World populations of this species, similar to recent examples in *Timmema* walking sticks (e.g. Nosil et al. 2002) and in sticklebacks (e.g. Schluter and Nigel 1995; Rundle et al. 2000).

However, it is also possible, and perhaps most likely, that these tropical traits first evolved in Africa and subsequently spread into the Caribbean islands with the historical human slave trades several hundred years ago (e.g. David and Capy 1988). This "historical migration" scenario is supported by the observation that Caribbean and West African flies mate randomly with each other (Yukilevich and True 2008), and by the fact that the same African insertion allele at *desat*-2 locus is segregating in the Caribbean and is associated with sexual isolation.

The few genetic studies of Caribbean populations in relation to Non-African and African populations have so far provided mixed results. First, using ten microsatellite loci, Schlotterer et al. (1997) had shown that Lesser Antilles Caribbean populations were genetically closer to European flies than to African flies (US populations were not included). However, more recent work with 48 microsatellite loci revealed that US populations are genetically more similar to Africa than are European populations and segregate many putative African alleles, suggesting recent admixture of African alleles in the New World (Caracristi and Schlotterer 2003). Further genetic analyses of Caribbean populations are necessary to determine which of these two historical scenarios are more likely.

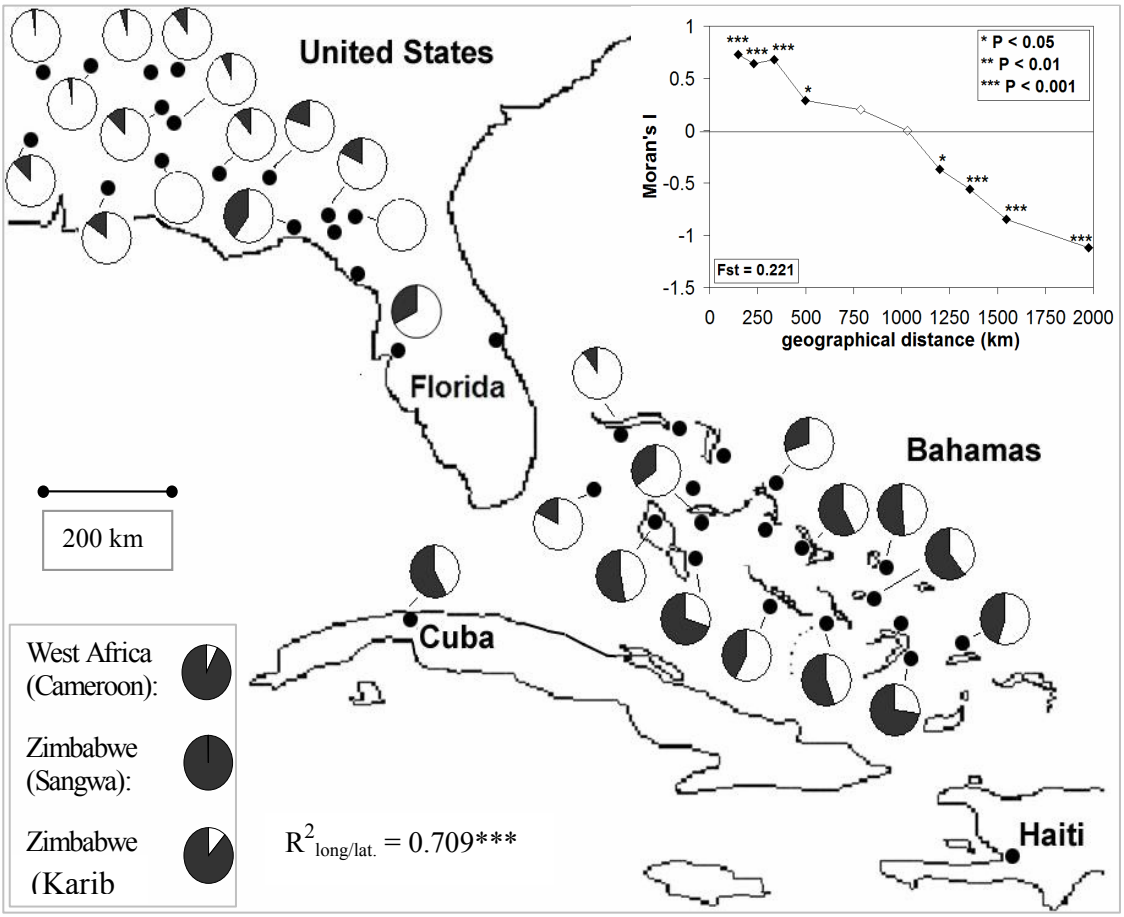


Figure 1. Map of 38 collection cites (shown as black crosses) of *Drosophila melanogaster* iso-female lines and *desaturase-2* locus allelic frequency data (per sampled locality) and its geographic correlogram across the system (top-right corner). See Appendix Table 1 for number of isofemale lines used to determine allelic frequency per location. *Desat2* allelic frequencies are described as pie charts across the region, with black representing the African insertion allele frequency and white representing the cosmopolitan deletion allele frequency. Also shown at left bottom corner are allelic frequencies of African locations at *desat2* locus for comparison. St. Lucia allelic frequency is not shown (insertion allele = 55%). Multiple regression of longitude and latitude on *desat2* allelic frequency is shown. Partial $R^2_{long.}$ = 0.007 and $R^2_{lat.}$ = 0.057 (P = 0.605 and P = 0.003, respectively). The *desat2* locus correlogram shows how Moran's I coefficient of correlation changes as the distance between paired populations increases in km from US to Bahamas. The overall correlogram is significant at P-value < 0.05 and a Moran's I value for each distance class is designated either as significant (black with a given significance value) or not significant (white). The Fst for *desat2* locus across US - North Caribbean region is 0.221.

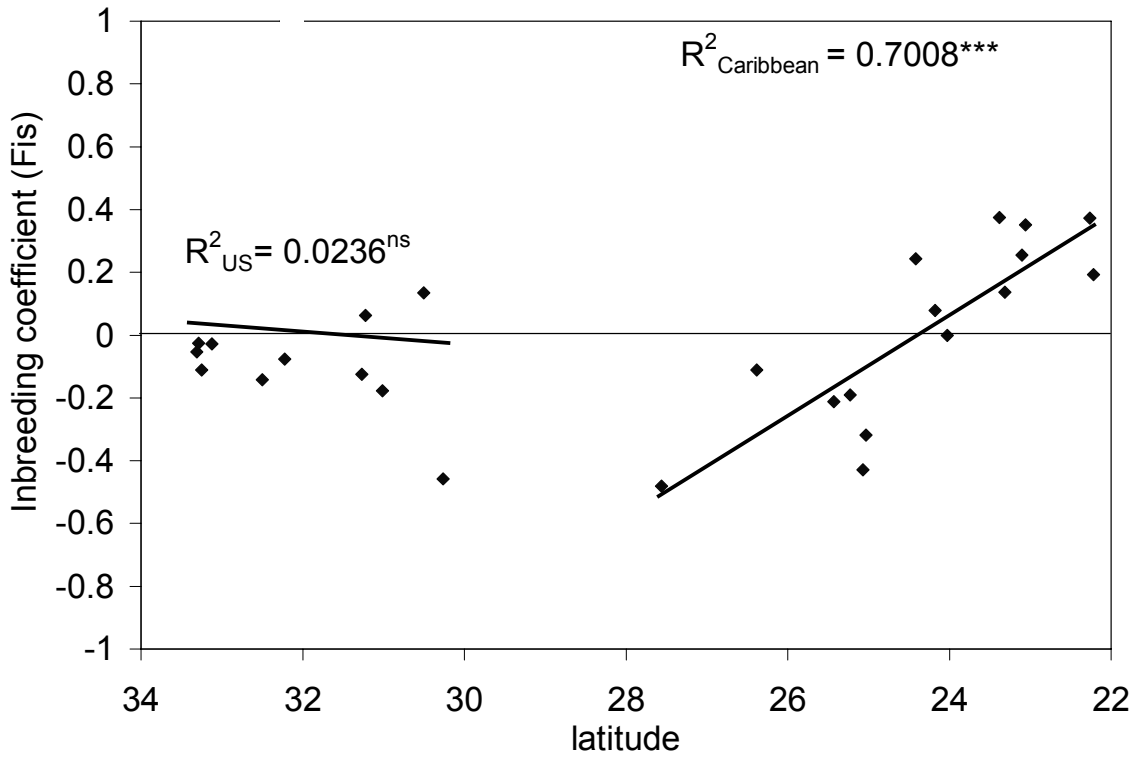


Figure. 2. Relationship between observed and expected frequency of heterozygotes at the *Desaturase-2* described by the inbreeding coefficient (Fis) across US and Caribbean (Bahamas and Cuba) island populations as a function of latitude. R^2 values were determined independently for each regional test. Non-significant R^2 in US corresponds to F = 0.242, $P = 0.633$. Asterisks for Caribbean populations signify F = 30.45, $P < 0.0001$. Note that only for this analysis the Tampa Bay Florida population (latitude 27.56) was included in the Caribbean regional test since this locality is geographically distant from other studied US populations and is more consistent with the pattern across northern Bahamas islands. Including Tampa Bay population in US still results in non-significant $R^2 = 0.14$, F = 1.8, $P = 0.21$ across US and its exclusion from Caribbean still results in a significant $R^2 = 0.617$, F = 19.29, $P < 0.0009$ across Caribbean populations.

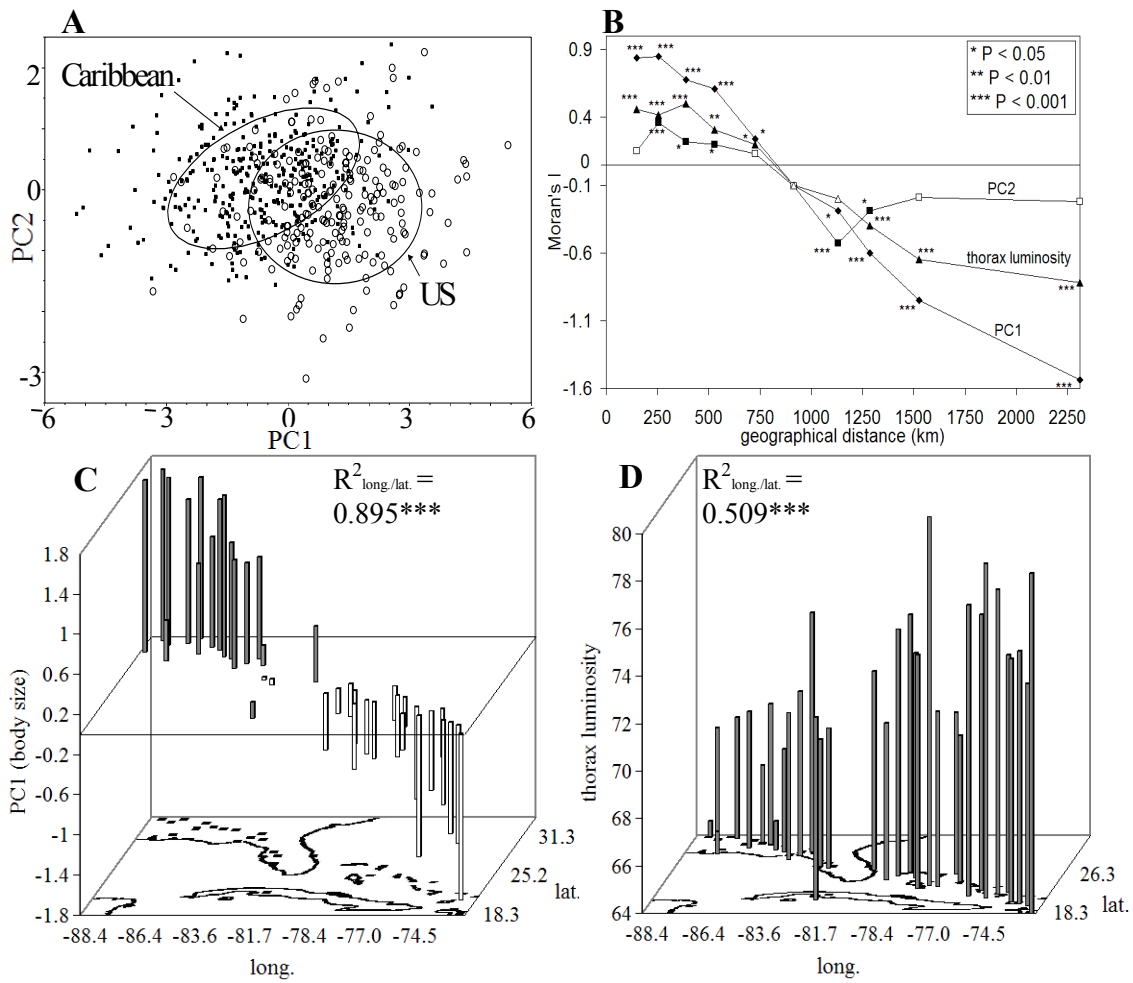


Figure 3. Geographical differentiation in morphological traits across US and Caribbean populations (only male data shown). A) Principal Component values (PC1 and PC2) for US and Caribbean individuals (shown as circles and squares, respectively) from Table 1. Binary normal ellipses for US and Caribbean individuals are shown, representing 65% variation in each region. B) Geographic correlograms for location-means of PC1 and PC2 and thorax luminosity (performed in PASSAGE 1.1; M. Rosenberg 2004). All 3 correlograms are significant at $P < 0.05$. Black symbols represent significant Moran's I values (each value for each distance class) and white symbols represent non-significant Moran's I values. C) Clinal differentiation of location means of PC1 from Table 1. Multiple regression of longitude and latitude on PC1 shown. Partial $R^2_{long.} = 0.089$ and $R^2_{lat.} = 0.125$ are both significant at $P = 0.009$ and $P = 0.004$, respectively. Black bars represent positive values and white bars represent negative values (standard errors not shown). D) Clinal differentiation of thorax luminosity (inverse of pigmentation; see Materials and Methods). Partial $R^2_{long.} = 0.214$ and $R^2_{lat.} = 0.440$ are not significant ($P = 0.433$ and $P = 0.205$, respectively).

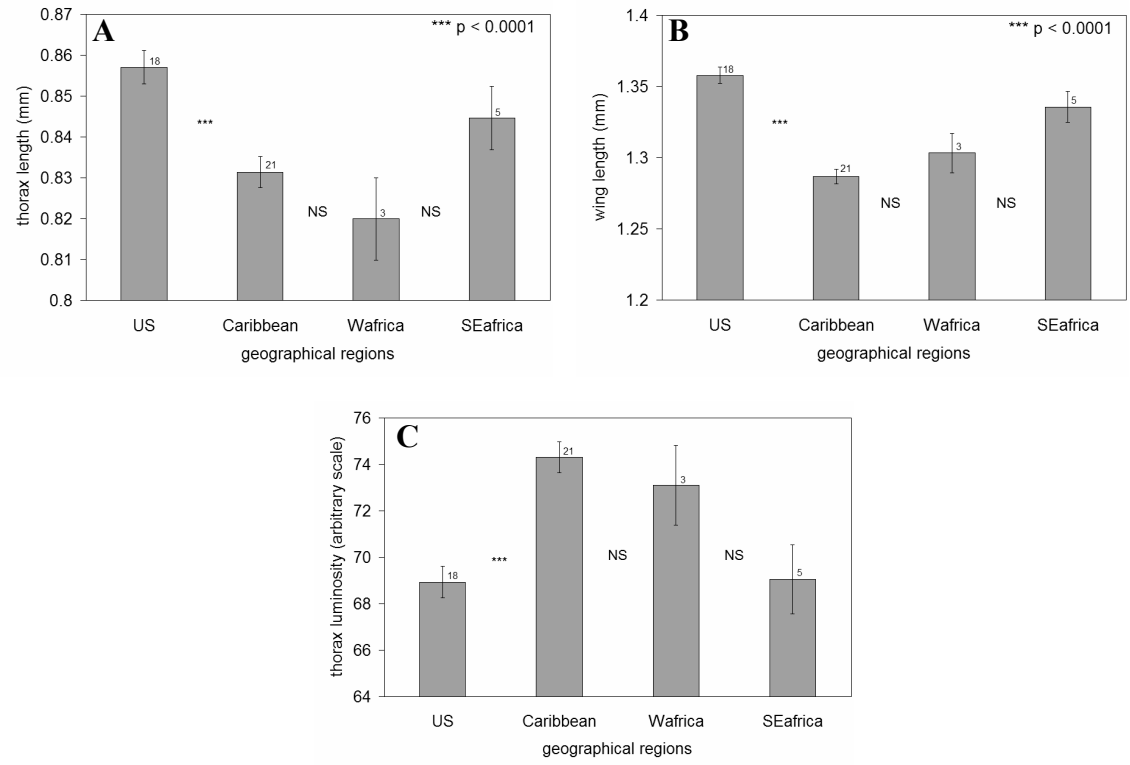


Figure 4. Morphological differentiation between US, Caribbean, West African and Southeast African populations in thorax length (A), wing length (B), and thorax luminosity (C) (only male data is shown). Numbers above bars represent the number of locations being averaged and error bars designate standard error. Significance was determined using the Kruskal-Wallis test.

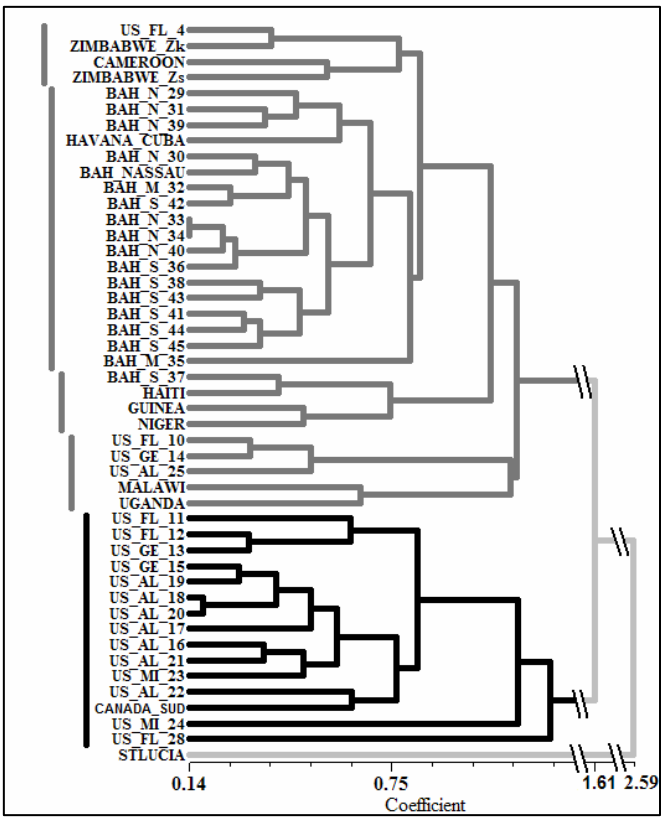


Figure 5. UPGMA Cluster Analysis of males based on six standardized morphological variables: thorax length and width, head width, wing length and width and thorax pigmentation. Analysis performed using NTSYS (Rohlf 2004). Labels consist of the regional name, location area in the region, and id # (REGION_LOCATION_ID#). In United States, FL is for Florida, GE is for Georgia, AL is for Alabama, and MI is for Mississippi. In Bahamas, N is for north, M is for middle and S is for south Bahamas islands (see Appendix Table A1 for location details).

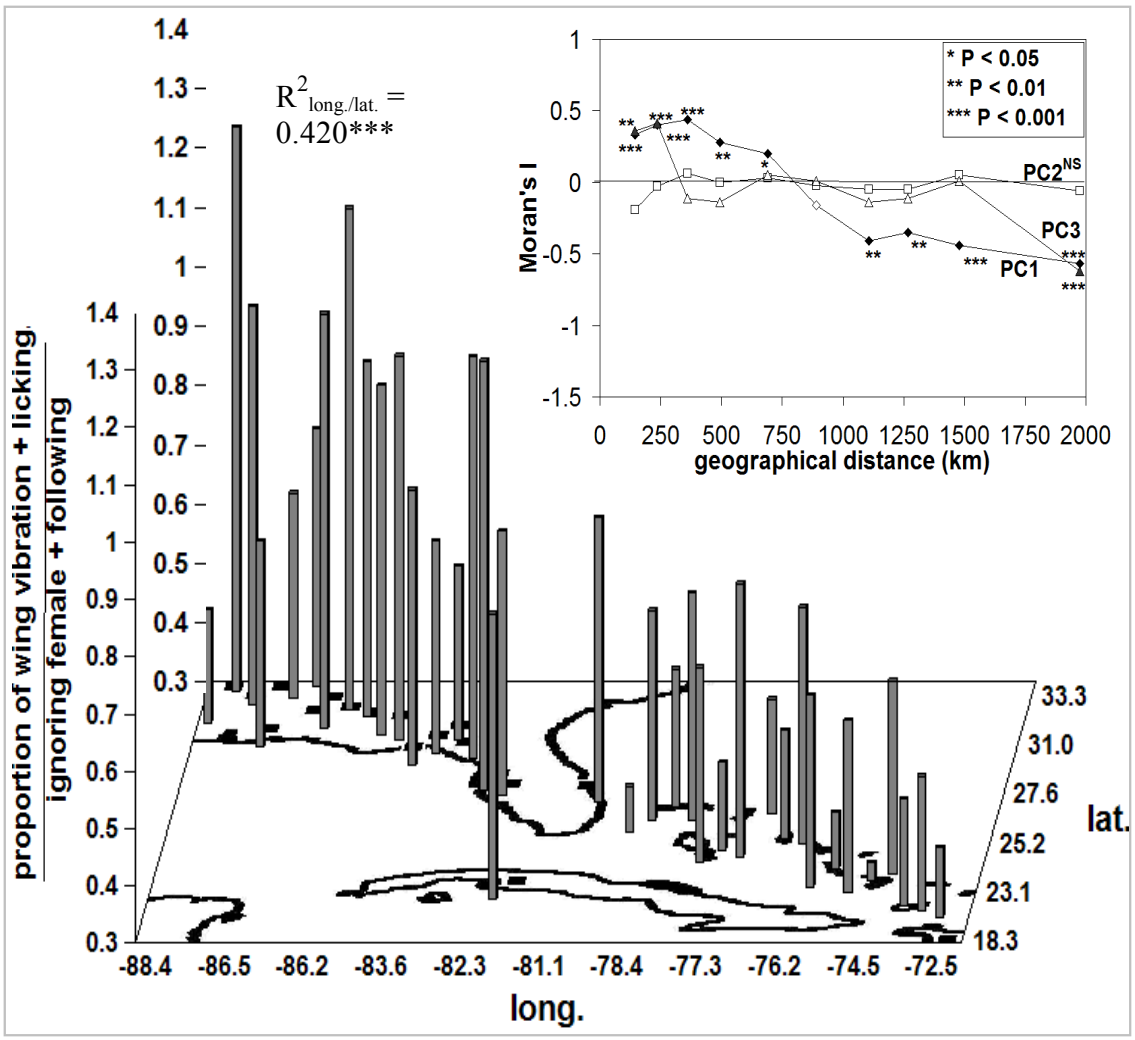


Figure 6. Male courtship behavior across US and Bahamas locations. Bar graph represents a relationship described by first Principal component in Table 2, which is predominantly a tradeoff between the proportion of events a male performs wing vibrations and ovipositor licking versus ignoring and simply following the female during courtship. See text for number of isofemale lines studied per location. Inset graph shows a geographical correlogram of the first three Principal components of courtship behavior (see Table 2 for detail; PASSAGE 1.1; M. Rosenberg 2004). PC1 and PC3 are significant correlograms at $P < 0.05$ while PC2 does not deviate from random spatial distribution. Black symbols represent significant Moran's I values at the designated significance level (each value for each distance class) and white symbols represent non-significant Moran's I values. Multiple regression of longitude and latitude on proportion data is shown. Partial $R^2_{long.} = 0.022$ and $R^2_{lat.} = 0.011$ are not significant ($P = 0.22$ and $P = 0.63$, respectively).

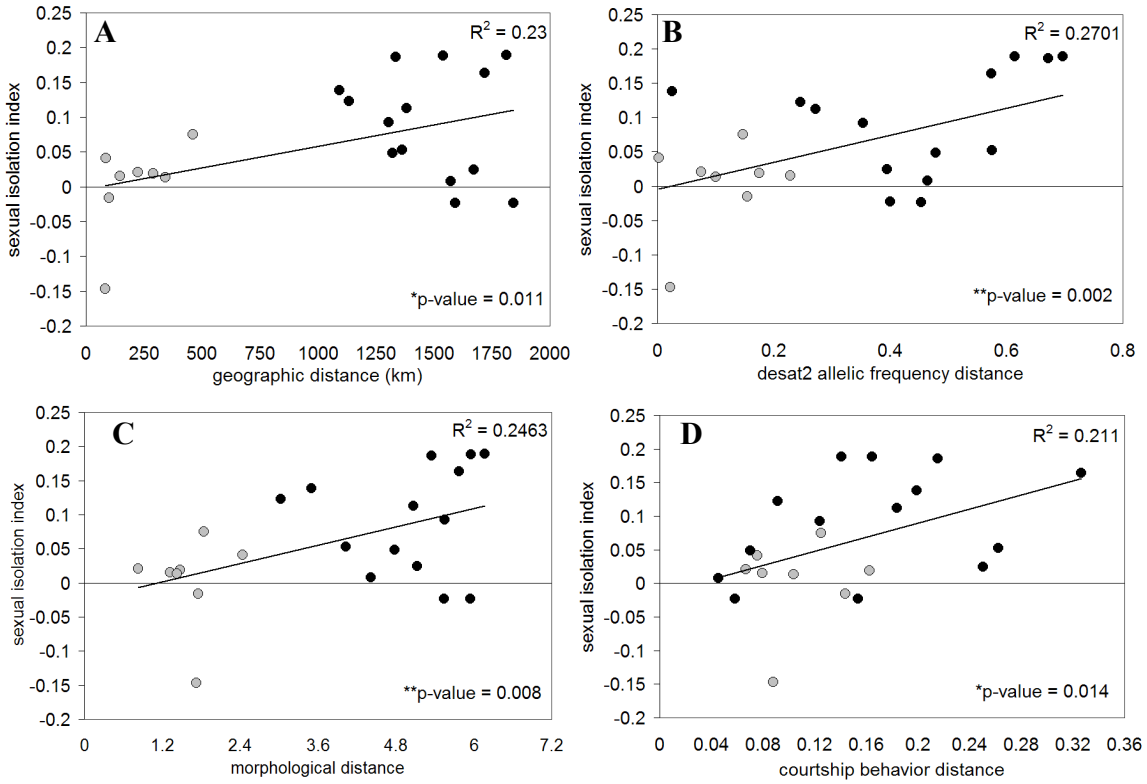


Figure 7. Geographical relationship between sexual isolation and distances based on geography or measured variables between all pairwise populations that were tested for mate choice. Gray symbols represent within-region comparisons (either within US or within Caribbean) and black symbols represent between-region comparisons. See Yukilevich and True (2008) for standard deviation of individual sexual isolation indexes. Since my sexual isolation matrix is incomplete, I determined *P*-values and significance using the 'sparse' Mantel test (see Materials and Methods).

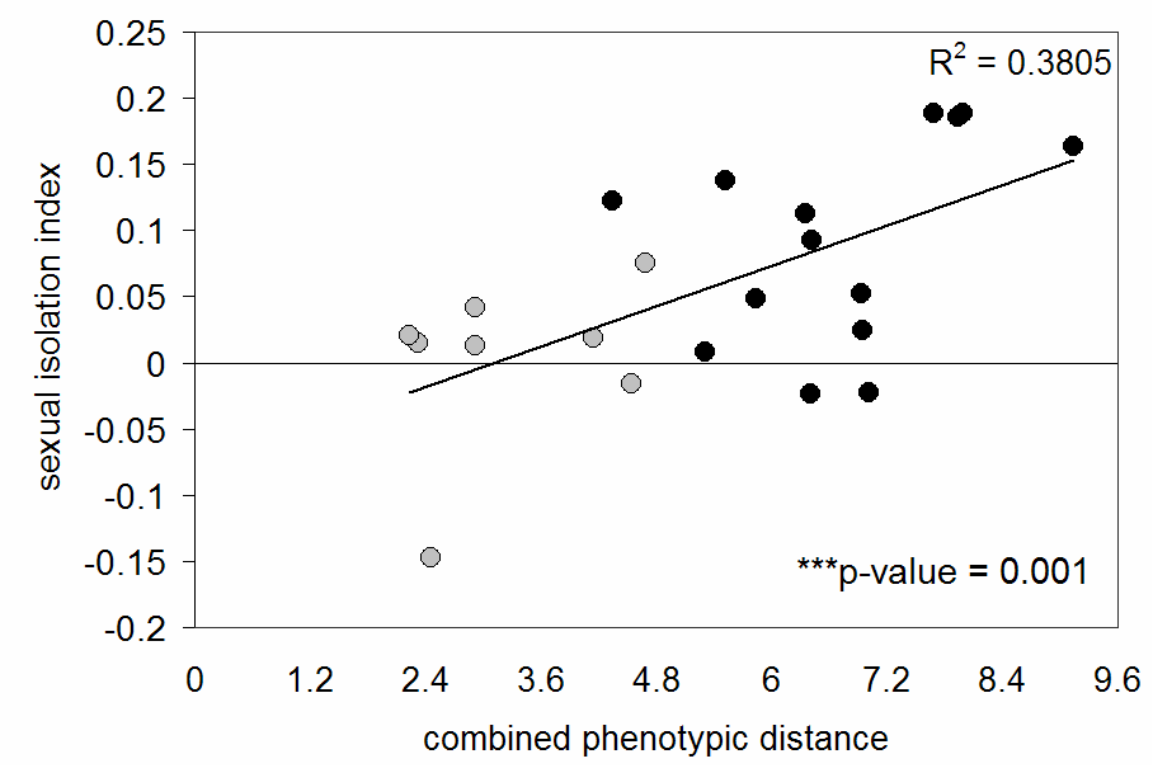


Figure 8. Geographical relationship between a combined distance matrix based on standardized values of *desat*-2 locus allele frequency, six morphological traits and six courtship behavior elements and sexual isolation index. Grey symbols represent within-region comparisons (within US or within Bahamas) and black symbols represent between-region comparisons (see Fig. 7 for further details). For statistical significance, I used the 'sparse' Mantel test (see Materials and Methods).

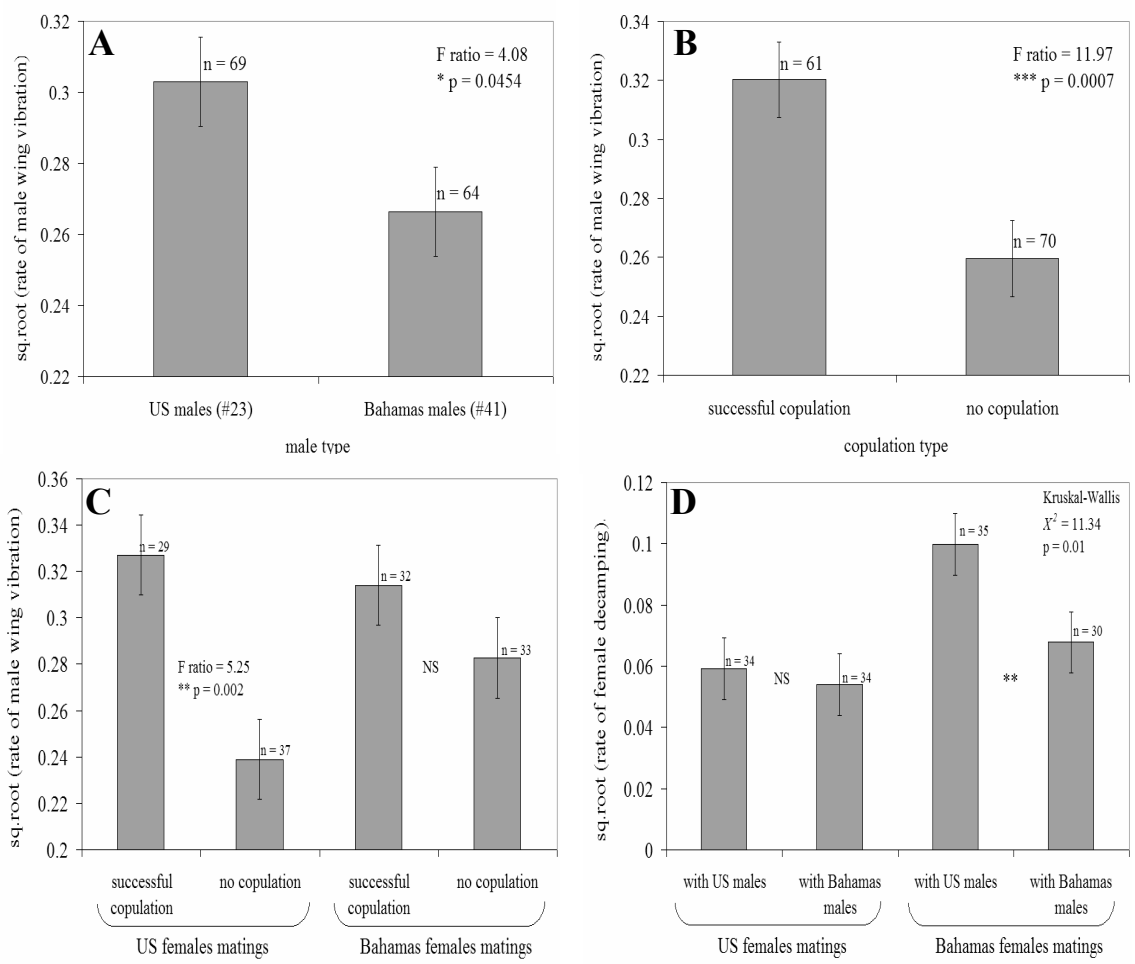


Figure 9. No-choice mating tests results between a US population (Columbus, Mississippi #23) and a Bahamas island population (Port Nelson, Rum Cay #41). All rates were determined by the number of acts divided by the total courtship time (in seconds) from initiation of male courtship to copulation or 10 mins. (note the rate of male wing vibration was measured differently than in Fig. 6; see Materials and Methods). Wing vibration rates were then square-root transformed to achieve normality. A) The average rate of wing vibrations displayed by males during courtship: A) from US location #23 and Bahamas location #41 towards females from both locations, B) in all successful matings versus unsuccessful matings, and C) in successful versus unsuccessful matings by US or Bahamas females. D) The average rate of decamping by females from courting males from US location #23 and Bahamas location #41 in homotypic and heterotypic mating tests. Note that the square-root transformation of decamping rates did not achieve normality. Kruskal-Wallis test was used to test significance. Error bars designate standard errors and 'n' is the sample size per category.

Table 1. Principal Component Analysis based on five morphological traits

PC's	1	2	3	4	5
EigenValue	3.08	0.692	0.581	0.377	0.27
Percent	61.61	13.84	11.61	7.54	5.40
Cum Percent	61.61	75.45	87.06	94.60	100.00
Eigenvectors (loadings)					
thorax length	0.437	0.011	-0.729	0.526	-0.045
thorax width	0.467	0.227	-0.297	-0.801	0.036
head width	0.385	0.761	0.441	0.277	0.023
wing length	0.472	-0.409	0.326	0.007	-0.710
wing width	0.469	-0.449	0.283	0.073	0.702

Note: Analyses were performed in JMP software and all eigenvalues were based on correlation matrices of variables that standardize their variation. Distribution of variables were not necessarily normalized since multiple populations are included in analyses.

Table 2. Principal Component Analysis of Male Courtship Behavior

Note: Analyses were performed in JMP software and all eigenvalues were based on correlation matrices of variables that standardize their variation. Distribution of variables were not necessarily normalized since multiple populations are included in analyses.

PC's	**1**	**2**	**3**	**4**	**5**	**6**
EigenValue	2.537	1.220	0.970	0.861	0.408	0.005
Percent	42.284	20.328	16.173	14.341	6.793	0.081
Cum Percent	42.284	62.612	78.785	93.126	99.919	100.000
Eigenvectors (loadings)						
attemp copulation	0.282	-0.058	0.493	0.806	-0.016	0.159
ovipositor licking	0.496	-0.163	-0.310	0.140	0.252	0.528
wing scissoring	0.240	0.174	0.748	-0.007	0.783	0.134
vibrations+extensions	0.553	0.067	-0.288	-0.555	0.150	0.149
ignore female	-0.369	0.702	-0.096	-0.079	0.061	0.603
following+standing	-0.417	-0.666	0.097	-0.130	-0.545	0.541

Table 3. Description of different geographical design matrices used to test explicit geographical patterns of differentiation in phenotype space

Design Matrices	Hypotheses	Regional Splits*
Design1	Divergent Northern US	Northern US vs. All
Design2	Florida-Islands Connection	Northern+Middle US vs. Florida+Cuba+Bahamas+Haiti
Design3	Mainland and Cuba vs Islands	US+Cuba vs. Bahamas+Haiti
Design4	Mainland vs All Islands	US vs. Cuba+Bahamas+Haiti
Design5	Divergent Family Islands	US+Cuba+Northern Bahamas vs. Middle+South Bahamas
Design6	Divergent South Bahamas and Haiti	US+Cuba+Northern+Middle Bahamas vs. South Bahamas+Haiti

*see Appendix Fig. A2 for exact geographical locality delimitations of regional splits

Table 4. Partial matrix correlations among six geographic design matrices and phenotypic distance matrices (holding geographical distance constant)

	Design 1	*Design 2*	*Design 3*	*Design 4*	*Design 5*	*Design 6*
desat2	-0.1087^{ns}	-0.0444^{ns}	-0.2782^{ns}	-0.1041^{ns}	**0.2252****	0.1281^{ns}
morphology	**0.2456***	0.1785*	0.039^{ns}	-0.0344^{ns}	-0.1102^{ns}	0.0209^{ns}
courtship	-0.1259^{ns}	-0.0532^{ns}	**0.2266****	0.1109^{ns}	0.1538*	0.0191^{ns}

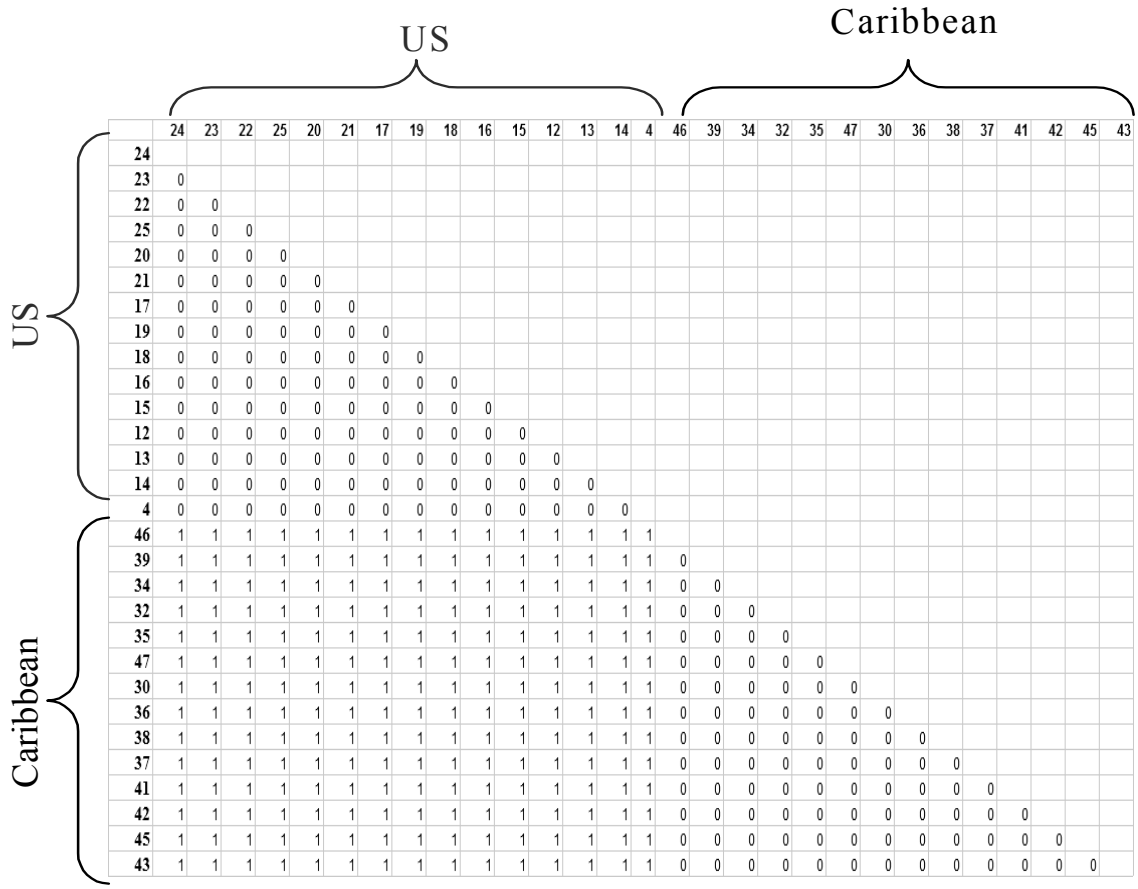
* $p < 0.05$
**$p < 0.01$
***$p < 0.001$

Table 5. Matrix correlations of *desat*-2 locus, morphology and courtship behavior to geographical distance and partial matrix correlations to each other (holding geographical distance constant)

	geography	*desat2*	*morphology*
desat2	0.6229***		
morphology	0.8472***	0.1991**	
courtship	0.3177***	0.0660^{NS}	0.1126^{NS}

**p < 0.01
***p < 0.001

Appendix



	24	23	22	25	20	21	17	19	18	16	15	12	13	14	4	46	39	34	32	35	47	30	36	38	37	41	42	45	43
24																													
23	0																												
22	0	0																											
25	0	0	0																										
20	0	0	0	0																									
21	0	0	0	0	0																								
17	0	0	0	0	0	0																							
19	0	0	0	0	0	0	0																						
18	0	0	0	0	0	0	0	0																					
16	0	0	0	0	0	0	0	0	0																				
15	0	0	0	0	0	0	0	0	0	0																			
12	0	0	0	0	0	0	0	0	0	0	0																		
13	0	0	0	0	0	0	0	0	0	0	0	0																	
14	0	0	0	0	0	0	0	0	0	0	0	0	0																
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0															
46	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1														
39	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0													
34	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0												
32	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0											
35	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0										
47	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0									
30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0								
36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0							
38	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0						
37	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0					
41	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0				
42	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0			
45	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0		
43	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	

Figure A1. Example of a binary design matrix #4 (Mainland vs. All Islands) where all localities that are grouped together (in this case either within US region or within Caribbean region) are assigned a value of 0 (‘less differentiation’) while all localities between groups (in this case between region locality pairs) are assigned a value of 1 (‘more differentiation’). Other design matrices in Table 3 vary with respect to which locality pairs are assigned values of 1 and 0, depending on the geographical hypothesis.

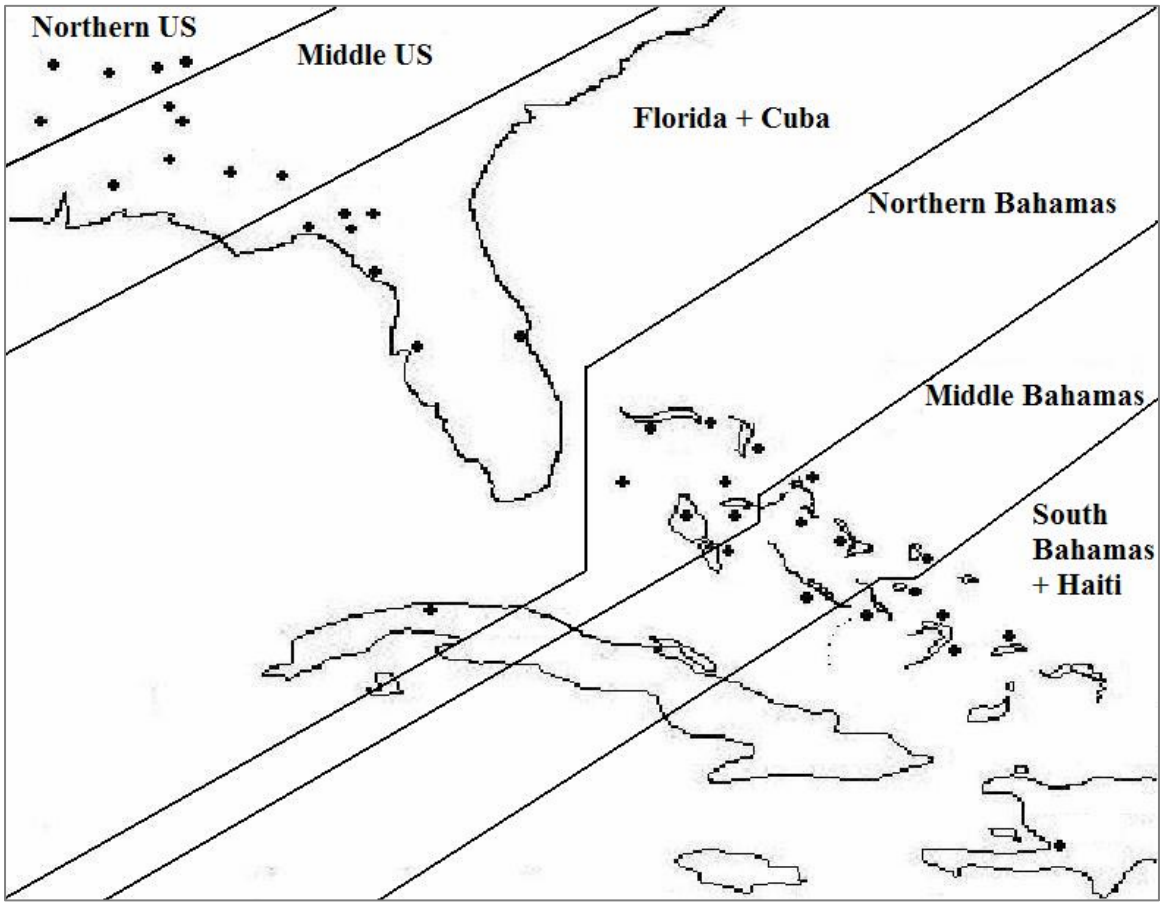


Figure A2. Map of arbitrary regional splits used to create geographical design matrices in Table 3 to test for specific spatial hypotheses about *desat*-2 locus, morphology and courtship behavior variation.

Table A1. Location information and the number of isofemale lines collected and maintained in the lab.

ID #	State/Island, Town	# of Lines	Date Collected	Longitude	Latitude
4	Florida, Tampa Bay (downtown)	13	Jun-04	-82.27	27.56
9,10	Florida, Crosscity-Chiefland	6	Jun-04	-83.07	29.38
11	Florida, Lamont	6	Jun-04	-83.48	30.22
12	Florida, Tallahassee (downtown)	15	Jun-04	-84.16	30.26
13	Georgia, Thomasville	20	Jun-04	-83.58	30.5
14	Georgia, Valdosta	16	Jun-04	-83.16	30.49
15	Geogria, Blakeley	20	Jun-04	-84.56	31.22
16	Alabama, Ozark	20	Jun-04	-85.38	31.27
17	Alabama, Greenville	19	Jun-04	-86.38	31.49
18	Alabama, Montgomery (downtown)	23	Jun-04	-86.18	32.22
19	Alabama, Clanton	20	Jun-04	-86.37	32.5
20	Alabama, Selva	19	Jun-04	-86.53	33.25
21	Alabama, Birmingham (downtown)	16	Jun-04	-86.48	33.31
22	Alabama, Tuscaloosa	18	Jun-04	-87.34	33.12
23	Mississippi, Columbus	15	Jun-04	-88.25	33.29
24	Mississippi, Meridian	7	Jun-04	-88.42	32.21
25	Alabama, Atmore	14	Jun-04	-87.29	31.01
28	Florida, Sabastian	6	Jun-04	-80.28	27.48
29	Eluthera (south), Governor's Harbor	14	Jul-04	-76.18	25.15
30	Eluthera (north), Gregory Town	20	Jul-04	-76.33	25.23
31	Abaco, Marsh Harbor	20	Jul-04	-77.03	26.32
32	Andros (north), Andros Town	22	Jul-04	-77.46	24.42
33	Grand Bahamas (west), Freeport	18	Jul-04	-78.38	26.3
34	Grand Bahamas (east), McLean's Town	20	Jul-04	-77.56	26.38
35	Andros (south), High Rock	20	Jul-04	-77.33	25.07
36	Exumas, George Town	20	Jul-04	-75.47	23.31
37	Long Island, Deadman's Cay	17	Jul-04	-75.06	23.1
38	Cat Island, New Bight	22	Jul-04	-75.25	24.18
39	Bimini Island, Alice Town	20	Jul-04	-79.17	25.43
40	Berry Islands, Bullock's Harbor	20	Jul-04	-77.51	25.45
41	Rum Cay, Port Nelson	22	Jul-04	-74.5	23.38
42	San Salvador, Cockburn Town	20	Jul-04	-74.31	24.03
43	Mayaguana, Mayaguana	16	Jul-04	-72.54	22.22
44	Crooked Island, Colonel Hill	20	Jul-04	-74.1	22.44
45	Acklins Island, Spring Point	16	Jul-04	-73.59	22.26
46	Cuba, Havana	17	Jul-04	-82.23	23.06
47	New Providence, Nassau	22	Jul-04	-77.2	25.03
48	Lesser Antelles, St. Lucia (Soufriere)	23	Dec-05	-60.58	13.54
52	Haiti, Port-Au-Prince (near downtown)		Jun-06	-72.2	18.32
49	West Africa, West Cammeroon (Mbalang-Djalingo)	31	2004	10.05	-5.23
50	Southeast Africa, Zimbabwe (Sengwa)	13	1990	28.34	-16.5
51	Southeast Africa, Zimbabwe (Lake Kariba)	20	1994	27.59	-17
53	Guinea	12	2004	10.52	-10.47
54	Niger	20	2004	7.12	-15.23
55	Uganda	20	2004	32.12	-0.48
56	Malawi	20	2004	33.35	-13.19

Note: All lines were collected by R. Yukilevich, except lines from Haiti (collected by S. Tumuluri) and from Africa (obtained from J. Pool and C. Acquadro).

Table A2. Full correlations between phenotypic distance measures.

	desat2	morphology
desat2		
morphology	0.5139***	
courtship	0.2180**	0.3259***

**p < 0.01
***p < 0.001

CHAPTER III:
A GENOME-WIDE ANALYSIS OF HISTORICAL ADMIXTURE IN THE CARIBBEAN *DROSOPHILA MELANOGASTER* REVEALS GENOMIC 'ISLANDS' OF AFRICAN ANCESTRY

Abstract: Understanding the genetic basis of phenotypes and behaviors involved in local adaptation and sexual isolation between incipiently diverging populations has recently received a great deal of attention in evolutionary biology. This question gains particular interest when two divergent populations hybridize with one another in a zone of secondary contact. Microarray techniques now allow us to assess such a hybridization event on the scale of the whole genome to determine the extent to which the two source populations dominate and partition the genome of hybrid populations and to isolate genomic regions that may explain the maintenance of phenotypic and behavioral differences. Recently, I described a possible hybridization event between US populations of European origin and Caribbean populations of African origin in the fruit fly *Drosophila melanogaster*. The Caribbean populations exhibit exceptional African-like traits that show parallel clinal differentiation from US populations. Further, Caribbean populations mate randomly with West African flies, but both populations have partial/incipient sexual isolation from US and from the known Zimbabwe 'race' in South Africa. In the present paper I use Affymetrix tiling arrays to study genome-wide DNA sequence differentiation between US, Caribbean, West African, and Zimbabwe populations to determine their overall relative level of sequence divergence as well as the extent to which Caribbean genomes have been shaped by contributions from US versus African populations. My findings reveal that US populations have predominantly shaped Caribbean genomes. However, I also find that Caribbean populations are consistently closer genetically to African populations than are US populations. Of the genomic regions that have significantly diverged between US and African populations, the Caribbean genome maintains an African genetic signature at about 20% to 40% of the differentiated sites. Consistent with the slave trade colonization hypothesis of Caribbean populations, the Caribbean genome is on average 10% closer to West Africa than it is to Zimbabwe at differentiated sites. I identify many unusually African-like regions in the Caribbean genome which generate hypotheses concerning candidate genes associated with particular African-like phenotypes and behaviors across the islands.

INTRODUCTION

Natural admixture or hybridization between ecologically divergent populations and incipient species is common in nature (Barton and Hewitt 1985; Coyne and Orr 2004; Mallet 2005). These intraspecific hybridizing populations typically show intermediate trait values between the two source populations and reveal parallel clines in the zone of secondary contact (Endler 1977; Barton 1979; Barton and Hewitt 1985; Howard 1993; Jiggins and Mallet 2000). Much interest involves the extent to which gene flow and natural selection shapes the genomes of hybridizing populations and which of the two source populations dominate their phenotypic and genetic makeup (Barton and Hewitt 1985; Mallet et al. 1990; Dasmahapatra et al. 2002; Mallet 2005; Alipaz et al. 2005; Mallet et al. 2007). These questions gain particular relevance when the initial divergent source populations are partially sexually isolated from each other. In such cases, both natural and sexual selection may reduce the effect of gene flow at certain regions of the genome, while the rest of the genome may become homogenized in the secondary contact zone (Mallet 2005). These isolated regions of the genome become excellent candidates for understanding the genetic basis of incipient sexual isolation and local adaptation to ecological environments in nature (Nosil et al. 2008).

So far admixture events in nature have been studied using approaches that focus on selectively neutral markers in conjunction with phenotypic and behavioral assays of hybrid populations to assess the contribution of gene flow and selection to clinal differentiation (Mallet and Barton 1989; Mallet et al. 1990; Dasmahapatra et al. 2002). Traits that have diverged between source populations have also been studied using QTL approaches that identify candidate loci involved in the genetic basis of differentiated traits (e.g. True et al. 1997; Hawthorne and Via 2001; Noor et al. 2001; Hedrick 2006; Noor and Feder 2006; Stinchcombe and Hoekstra 2007; Albert et al. 2008). More recent advances in molecular technology now allow diverging populations to be studied at the level of the whole genome with extremely fine-scale genetic resolution of only several base pairs. Such 'genome scans' have been recently applied to the study of the genetic basis of incipient speciation and local adaptation between divergent populations (e.g. Vines et al. 2003; Campbell and Bernatchez 2004; Turner et al. 2005; Stinchombe and Hoekstra 2007; Bonin et al. 2006; Nosil et al. 2008). These studies are beginning to shed light on the extent to which genomes are differentiated, the patterns of differentiation across the genome, and the identification of candidate loci and regions for relevant phenotypes, behaviors, and local adaptation between populations.

The use of whole genome scans to study natural intraspecific hybridization between distinct geographic populations has thus far received less empirical attention (Emelianov et al. 2003; Murray and Hare 2006; Yatabe et al. 2007). Recently, I have described a case of a possible historical admixture in the Caribbean *Drosophila melanogaster* fruit fly between partially sexually isolated 'races' of this species in the US (cosmopolitan) and Africa. First, Caribbean populations are known to harbor exceptional African-like morphology, pheromones, and courtship behaviors that differ from other 'cosmopolitan' populations outside of Africa (David and Capy 1988; Capy et al. 1993, 1994; Colegrave et al. 2000; Rouault et al. 2001; Takahashi et al 2001; Yukilevich and

True 2008b). Most of these traits in the Caribbean show strong parallel clinal differentiation from northern US locations and are either intermediate between US and Africa or similar to Africa (Yukilevich and True 2008b). These data suggest that Caribbean populations may be experiencing a different suite of adaptations to tropical environments compared to the more temperate climates of US populations (e.g. David and Capy 1988; Rouault et al. 2001; Greenberg et al. 2003). I have also recently shown that Caribbean and West African populations mate randomly with each other, but are both partially sexually isolated from US populations and from the Zimbabwe 'race' in South Africa (Yukilevich and True 2008a). It is also evident that Caribbean flies show weaker sexual isolation from Zimbabwe populations compared to West Africa-Zimbabwe crosses.

These results suggest that Caribbean fruit flies are direct descendents from Africa, especially West Africa, and were colonized during the historical human slave trades into the region several hundred years ago (David and Capy 1988; Caracristi and Schlotterer 2003). The fact that this species is a human commensal that is particularly attracted to molasses and alcohol transported on the slave ships makes this scenario plausible. This historical colonization combined with possible secondary gene flow from US populations into the region would be consistent with the Caribbean phenotypic patterns. Microsatellite genetic evidence suggests that European and US populations are very closely related to each other (Caracristi and Schlotterer 2003; Schlotterer et al. 2006). These populations are also very similar in phenotypes and pheromones (e.g. David and Capy 1988; Capy et al. 1993). This makes it highly probable that US *D. melanogaster* populations originally came from Europe and then encountered African-derived Caribbean flies secondarily in the New World. In essence, the Caribbean populations may represent a widespread secondary contact zone between US and African populations. Thus this system offers an unprecedented opportunity to study a possible natural hybridization event at the level of the whole genome. Given the wealth of functional information related to candidate loci associated with specific phenotypes in the *Drosophila melanogaster* genome, a search for genomic outliers gains particular power in this model organism (e.g. Turner et al. 2008).

In the study reported here I used Affymetrix tiling arrays, which have a 25 bp probe (marker) for approximately every 40 bp of the genome to study large-scale, genome-wide differentiation between US, Caribbean, and African *D. melanogaster* populations. Affymetrix tiling arrays have recently been used to identify significant differentiated regions or 'outliers' in the genome between divergent populations or incipient species (Borevitz et al. 2003; Turner et al. 2005; Turner et al. 2008). This approach allows us to study the presumed hybridization event between US and African populations in the Caribbean with an unprecedented fine-scale genomic resolution. Thus I can address questions concerning the relative genome-wide relationship of Caribbean populations to US and Africa and perform a detailed scan of Caribbean genomes for probes that are highly differentiated between US and African populations to determine the degree to which Caribbean genomes have been shaped by US versus African genetic variation. This generates a list of candidate regions that may be involved in maintaining African-like traits and adaptations in the Caribbean populations. Further, because it is still unclear which parts of Africa have predominantly contributed toward Caribbean genetic

and phenotypic variation, I analyzed both a West African (Cameroon) and a South African (Zimbabwe 'race'; Hollocher et al. 1997a) population that have been previously studied for phenotypic and behavioral traits (see above).

My findings reveal that Caribbean populations generally show fewer significantly differentiated probes against US than against African populations, indicating that contributions from US populations have predominantly shaped Caribbean genomes. However, despite this general result, I also find that Caribbean populations are consistently closer genetically to African populations than are US populations. Caribbean and African populations share many genome-region differences in common against US. A genome-wide scan of Caribbean populations further revealed that among highly divergent US-African genomic regions, Caribbean populations are closer to US populations at about 60% to 80% of the differentiated sites. At the remaining 20% to 40% of US-African differentiated sites, depending on the chromosome and particular Caribbean island, the Caribbean genome is closer to African populations. Consistent with the slave trade colonization hypothesis, the Caribbean genome is on average 10% closer to West Africa than it is to Zimbabwe at differentiated sites. I identify many unusually African-like regions in the Caribbean genome that generate hypotheses concerning candidate genes associated with particular African-like phenotypes and behaviors across the islands. In general, my results suggest that African-like traits in the Caribbean are maintained by the effects of relatively few genomic regions despite the domination of rest of the Caribbean genome by US genetic variation.

MATERIALS AND METHODS

ISOFEMALE LINES

In the summer of 2004 R. Yukilevich collected and established isofemale lines from southeastern US (#22: Alabama, Tuscaloosa - 18 lines and #23: Mississippi, Columbus - 15 lines) and from Bahamas islands in the Caribbean (#35: South Andros Island, High Rock - 20 lines, #41: Rum Cay, Port Nelson - 22 lines, #45: Acklins Island, Spring Point - 16 lines). For further details about the Caribbean populations see Yukilevich and True (In press; submitted). African isofemale lines were acquired from J. Pool and C. Aquadro in 2005, including a population from West Africa (West Cameroon, Mbalang-Djalingo - 31 lines collected in 2004) and a population from Southeast Africa (Zimbabwe, Sengwa - 13 lines collected in 1990). All lines have been maintained at Stony Brook University laboratory on instant Drosophila food (Carolina Biol. Supply Inc., Burlington, NC) in a 25°C temperature incubator with a 12h light:12h dark regime.

DNA EXTRACTION AND PURIFICATION

DNA was extracted from pooled individuals of multiple isofemale lines for each of the seven locations described above. First, I randomly collected an equal number of males and females from each isofemale line of a given location to equal 300 individuals per location. I randomly divided the 300 individuals into 3 replicates of 100 individuals and froze the individuals. This yielded 21 samples (3 replicates * 7 locations). For each 100 fly sample, I used a phenol:chloroform extraction to purify the initial DNA extract. I then used ethanol precipitation and resuspended the DNA in 38 ul of H_2O. To eliminate RNA I added 1 ul of RNAse. To check the concentration of DNA sample (ng/ul), I used

lambda DNA (350 ng/ul) and ran it parallel to all 21 diluted samples on 1% Agarose gels. Before DNA fragmentation the DNA concentration of all 21 samples was standardized to 7.8ug/100 fly sample.

DNA FRAGMENTATION AND LABELING

DNA samples of volume 39 ul were fragmented with a mix of 4ul of 10X One-Phor-All buffer (Amersham Biosci. #27-0901-02), 0.14 ul of Acetylated BSA (Invitrogen: 15561-020) and 0.64 ul of DNase1 (Promega M6101-RQ1) (total mix = 4.78 ul) per sample. Fragmentation of all 21 samples was done simultaneously in a PCR thermocycler at 37˚C for 16 min., 99˚C for 15 min., and 12˚C for 15 min. and then the DNA was stored at 0˚C. Fragmention of DNA was assessed by running 3 ul of DNA fragment on 2% Agarose gel. Fragment sizes of all samples were about 35 bp. with standardized intensity and variance. Labeling was done with 2 ul of Biotin-N6-ddATP (Enzo: 42809) and 3 ul of RTdT enzyme (Promega: M1875) mix added to each sample. RTdT was first diluted from 30 U/ul to 15 U/ul enzyme by mixing a ratio of 5:1:4 of RTdT enzyme, RTdT 5X buffer and H_2O. PCR conditions for labeling were 37˚C for 90 min., 99˚C for 15 min., and 12˚C for 5 min. and then the DNA was stored at -20˚C. Labeling was done simultaneously on all 21 samples using the same master mix.

AFFYMETRIX TILING ARRAY HYBRIDIZATION AND DATA EXTRACTION.

To measure genome-wide differentiation in my populations, I used Affymetrix tiling arrays that have a 25 bp marker for approximately every 40 bp of the genome. It has been established that hybridization intensity of DNA to a microarray depends on its sequence similarity (Winzeler et al. 1998; Borevitz et al. 2003; Gresham et al. 2006). Differentiated sites in the genome can be identified when different DNA samples hybridize to an array with different affinities (Borevitz et al. 2003; Turner et al. 2005). Limitations of this technique may include variable sensitivity of hybridization intensity across the genome and a possible non-linear relationship between DNA sequence divergence and hybridization intensity (Zhang et al. 2003). Several approaches are used to minimize these effects (see below).

All 21 samples were hybridized to Affymetrix tiling arrays that contained a haploid reference genome of *D. melanogaster*. Hybridization was done at the UC Davis Genome Center (Affymetrix facility). Approximately 7.8 ug of DNA per sample were hybridized to each tiling array chip at the same time. The 2X Hybridization mix included BSA, Herring Sperm and MES. I used a protocol for measuring whole transcript double stranded target as described in Turner et al. (In Press). The Affymetrix core facility generates a binary file from scanned arrays with intensities for each probe and converts this file to text format. A NCBI megablast identified single perfect matched probes using 4.3 *D. melanogaster* reference genome. After eliminating mismatched and control oligos, I retained 3,015,075 randomly distributed probes throughout the genome, including 2,950,143 probes on the major chromosomal arms, 24,726 probes on the 'dot' 4th chromosome, and 32,256 probes in heterochromatic regions of chromosomes X, 2, and 3.

DATA NORMALIZATION

The purpose of normalizing the data is to control for heterogeneous and spatially non-random patterns of signal intensities on chips and to remove the correlation between mean and variance of signal intensity for each probe. Briefly, I divided each array into 1600 subarrays of 64 x 64 probes and log transformed raw intensity values. I then divided the intensity of each oligo by median intensity of unique probes on each local 64 x 64 probe subarray (see Turner et al. 2008 for details). I then normalized the data using quantile normalization to standardize the distribution of intensities of probes among the geographical locations. This was done to ensure that average signal intensity was not stronger in any particular population. After quantile normalization the average difference in the means of intensities of any two populations was less than $4x10^{-10}$ (nearly 50% of the probes were more intense in each location; see also Turner et al. 2008).

PAIRWISE *T*-TESTS

I performed two-tailed pairwise *t*-tests on normalized signal intensities for all combinations of individual locations and pooled locations within geographical regions (see below for details). Location-specific pairwise *t*-tests were based on three replicates of each location, while *t*-tests between regional comparisons were based on pooled replicates of locations within each region (e.g. six replicates in Africa). The resulting *P*-values indicate the degree to which the signal intensities of probes have diverged between populations (see below). Signal intensity differences are well correlated with DNA sequence-level divergence at particular probes (see Turner et al. 2008). The *P*-values for each pairwise comparison were also converted into q-values, which estimate the false discovery rates of each *P*-value or FDR (Storey and Tibishirani 2003; Turner et al. 2008). The FDR of each *P*-value was calculated by obtaining the Bonferoni adjusted *P*-value, which takes into account the number of tests performed (*P*-value * 3,015,075 probes), and dividing the adjusted *P*-values by their rank order.

RELATIVE SIGNAL INTENSITIES OF PROBES

Following *t*-tests, I focused on the most significantly differentiated probes ('outliers') between US and each African population at the threshold of $P < 0.0001$ (FDR around 16-19%). I also studied 259 significantly differentiated probes between pooled US and African populations at the FDR < 0.05. These probes are the most divergent between US and African populations and thus become most informative as to whether Caribbean populations at these differentiated probes are closer to US or African populations. Non-differentiated probes between US and African populations are not informative because the Caribbean populations are expected to randomly deviate around 50% with respect to being closer to US or African populations in their signal intensities at these probes (see Data normalization above). I analyzed average signal intensities of pooled Caribbean samples as well as signal intensities of each Caribbean island population separately.

For each significantly differentiated probe between US and African populations I asked whether the Caribbean probe is more similar to US or to African populations in relative signal intensity, which is an indication of the relative allelic frequency of that probe. Because the scale of signal intensities among probes is slightly different, I first standardized the signal intensities across all probes to a scale from 0 to 1, where 0 is the

relative signal intensity of US probes and 1 is the relative signal intensity of African probes. I then determined the relative signal intensity difference (or 'distance') of the Caribbean probe on a standardized scale. The formula for the standardization is: (Carib. – US)/(Africa – US). For probes where signal intensity of African populations was lower than that of US populations, I standardized the relative distance and then reversed the signal intensities to retain the relative distances among populations.

RESULTS

OVERALL POPULATION GENOMIC PATTERNS

Geographical differentiation among localities

To understand genome-wide population structure using tiling array probes I performed 21 pairwise population t-tests among the seven locations in this study (e.g. three replicates from US location #23 versus three replicates from Caribbean island #35). I focused on the most significantly differentiated probes or 'outliers' for each pairwise comparison at different levels of significance ranging from highly significant outliers ($P < 0.0001$ to less significant outliers $P < 0.01$. I also studied much less significant cutoffs such as $P < 0.05$ and $P < 0.15$. Significant probe outliers represent candidates for sequence divergence between populations and indicate the most differentiated localized regions in the genome. Here I investigate gross scale differences by focusing on the *number* of significantly differentiated probes between a pair of localities as an estimate of overall genetic divergence between populations. I assume that population pairs with relatively few outliers at some significance threshold are less differentiated than population pairs with more outliers at the same threshold (also see *Genome Scan* results below for qualitatively similar results using another approach). I created a genetic distance matrix based on the number of significantly differentiated outliers per pairwise test and clustered populations with the Neighbor-Joining Algorithm (Table 1 and Figure 1; NTSYSpc; Rohlf 2004).

The clustering of the seven localities paralleled their general geographical relationships, providing high confidence in the reliability of this genetic data to reveal genome-wide population structure (Figure 1). The two US localities always clustered together and showed the fewest number of significantly differentiated probes between each other regardless of the significance cutoff value (Figures 1a,b and Table 1a,b). When considering the most significant outlier probes, the two African populations (Cameroon and Zimbabwe) clustered together (Figure 1a and Table 1a), but when I included probes with a less stringent significance cutoff value ($P < 0.05$), Cameroon clustered together with the Caribbean-US populations and the Zimbabwe population became a geographical outgroup (Figure 1b and Table 1b). This indicates that West Africa has more differences from New World populations relative to Zimbabwe at the most extreme outliers, but has more differences from Zimbabwe relative to New World populations at less extreme outliers (i.e. subtly differentiated probes).

The three Caribbean island populations were always clustered between US and African populations, with Caribbean islands #41 (Rum Cay) and #45 (Acklins) showing the fewest significantly differentiated probes between each other and Caribbean island #35 (South Andros) showing less differentiation from Caribbean island #45 than from Caribbean island #41 (Table 1). In general, all three islands belonged to the US cluster (Figure 1a,b). However, African populations had consistently fewer significantly

differentiated probes from Caribbean flies compared to US flies, regardless of the significance cutoff value (Table 1a,b). This indicates that African populations are genetically closer to Caribbean populations than to US populations. Also, Caribbean populations were generally genetically closer to West Africa than to Zimbabwe (Table 1a,b).

Upon closer examination it can be seen that Caribbean islands exhibit variation in their relative genetic relationships with US and African populations. Caribbean island #41 was consistently found to be genetically closer to US relative to African populations. However, Caribbean island #45 was equally genetically close to US and Africa, while Caribbean island #35 was genetically closer to West African than to US populations. In sum, this genome-wide data is consistent with the hypothesis that Caribbean populations represent a mixture of US and African genes and tends to support the slave trade colonization hypothesis from West Africa (David and Capy 1988; Caracristi and Schlotterer 2003; Yukilevich and True 2008a).

Shared differences between regional comparisons

To understand the distribution of significant outlier probes in greater detail across various geographical regions, I pooled individual localities based on their geography (as suggested by the above clustering analysis). This allowed us to investigate how many significantly differentiated probes differ across geographical regions and to study shared differences between these regional comparisons. For each pooled pairwise regional test, two locations were included. Because the Caribbean region contains three localities in this study, while US and Africa each contain two localities, in order to keep statistical power constant across different pairwise tests, I performed three tests, each with a different combination of Caribbean locations against US and African regions (Figure 2). However, regardless of which Caribbean populations were included in the tests, the results were similar in terms of relative number of differentiated probes. For this analysis I used the significance threshold $P < 0.0005$ (FDR between 18.1% to 37.1%; see below). My conclusions remain the same for different significant threshold values.

Using these regional comparisons, I confirmed the above clustering results that were based on individual localities. US and Caribbean regions are genetically most similar (2,129-4,059 differentiated probes at $P < 0.0005$; FDR = 37.1%), followed by Caribbean and African regions are (5,306-5,470 differentiated probes at $P < 0.0005$; FDR = 28.1%), and finally the most differentiation occurs between US and African regions (8,337 differentiated probes at $P < 0.0005$; FDR = 18.1%; Figure 2). Notice that the false discovery rate also drops in this order, implying that US and African populations have more true discoveries of differentiated probes than either of the other two regional comparisons.

This approach also allowed us to investigate shared differences between regional comparisons. First, I found that US and Caribbean populations shared on the order of 500 differentiated probes in common against African populations at $P < 0.0005$ for a combined FDR of 5.1% (28.1% * 18.1%; see Figure 2). This indicates that most genome-wide differentiation is associated with North American versus African differences (see below). Strikingly, probes that differentiate African from Non-African flies are overwhelmingly found on the X chromosome, with a dramatic 59.9% of all significant

shared probes while all other chromosome arms carry only 7%-12% of significant probes (X^2-test; $P << 0.0001$). This X-effect is much larger compared to that of significant probes between individual North American and African population comparisons (Yukilevich et al. ms in prep). This implies that the great majority of genomic differentiation out of Africa has occurred on the X chromosome. Thus I expect US and Caribbean populations to be genetically most similar on the X chromosome relative to autosomes (see below for an explicit test of this prediction).

Second, I found that Caribbean and African populations shared on the order of 50 differentiated probes in common against US populations at $P < 0.0005$ for a combined FDR of 6.7% (37.1% * 18.1%). These probes could either be associated with common genetic history of Caribbean and African populations and/or involved in adaptation to similar tropical environments of these regions relative to the more temperate environments of US populations (see below). Here, the X chromosome contains only 11.5% of significant shared probes and the distribution among chromosomes does not deviate from random expectations (X^2 test; $P = 0.16$). Because these US versus Caribbean and African probes are an order of magnitude fewer compared to North American versus African probes, it suggests that the genome of Caribbean individuals is predominantly US in ancestry, but with a few isolated regions of African ancestry. An explicit test of this prediction and an identification of genomic regions of African signature in the Caribbean are described below.

Further, US and African populations shared only about 5 to 8 differentiated probes in common against Caribbean populations at a combined FDR of 10.4% (37.1% * 28.1%; see Figure 2). This indicates that relative to US and African populations, the Caribbean populations exhibit very few unique differences, consistent with these populations being a mixture of US and African flies (see below). Finally, one to two differentiated probes are shared between all three regional comparisons for a combined FDR of 1.9%. These include probes found within the following annotated genes: intron of *Neuropeptide Y receptor* (NepYr), intron of *frizzled3* (fz3), exon of *CG11380*, and an intron of a *Synaptosomal-associated protein 25* (Snap25).

SCAN OF CARIBBEAN GENOMES FOR US VERSUS AFRICAN GENETIC SIGNATURE

I now explicitly study the extent to which Caribbean genomes have been shaped by US versus African genetic variation. Because I have identified genomic regions of significant differentiation between US and African populations, including both West Africa and Zimbabwe, I can now ask how Caribbean genomes compare in their signal intensities on array chips and their corresponding *t*-test differentiation at those highly differentiated probes between US and Africa. This would determine the degree of US versus African influence on the Caribbean genome as well as identify African-like genomic regions within the Caribbean populations.

The use of average signal intensities of probes in determining the relative genetic similarity between populations at these probes is well justified. This is because signal intensity 'distance' of a probe between two populations is exponentially negatively correlated with the significance (*P*–value) of that probe (Figure 3). In Figure 3 I show the relationship between average signal intensity of Caribbean populations and intensities of

both US and African populations. When the signal intensity distance between Caribbean and other populations is around 0.5 and above, over 95% of the probes are significantly differentiated at $P < 0.05$ or less with an average $P = 0.0099$ between such probes. At a distance of 0.8 or above, the average *P*-value drops to 0.0039 (Figure 3).

US and African populations differentiation

I study the most significantly differentiated probes between US and African populations at $P < 0.0001$ since these provide the most informative regions for determining whether the Caribbean signal intensity is closer to one or the other population (see Materials and Methods). For the US-West African test, 1566 probes were significantly differentiated at $P < 0.0001$ (FDR = 19.25%). For the US-Zimbabwe test, 1807 probes were significantly differentiated at $P < 0.0001$ (FDR = 16.67%). This includes probes on all four major chromosomal arms as well as in heterochromatin regions of each arm and on the 4th 'dot' chromosome. In total, these regions represent about 0.0519% and 0.0599% of the genome between US-West Africa and US-Zimbabwe, respectively, indicating that over 99% of the genome is undifferentiated or differentiated to a much lesser degree between these populations. For the following analysis, I averaged the signal intensities of the three Caribbean populations. However, below I also describe variation between different Caribbean islands with respect to relative US-African genomic makeup.

US versus West African contribution to the Caribbean genome

Figure 4 shows the relative average signal intensity of the Caribbean genome within probes that were significantly differentiated between US and West Africa. Signal intensities of Caribbean probes above 0.5 are closer to West African than to US signal intensities of 1 and 0, respectively, and vice versa. Also notice that a few Caribbean probes have signal intensities beyond those of US or West Africa (Figure 4). These probes are likely to have even more extreme allelic frequencies in the Caribbean compared to US or West Africa. Because these types of probes are not very common, this result supports the hypothesis that Caribbean populations generally represent a mixture of US and African populations.

First, I found that among the 1566 highly differentiated probes between US and West Africa, the Caribbean genome has only 491 probes (or 31.4%) that are closer to West Africa than to US signal intensities (i.e. above 0.5 in Figure 4). I confirmed that these are also significantly differentiated between US and Caribbean populations. For instance, among the 491 probes, 224 probes are also found to be significantly differentiated between US and Caribbean populations at $P < 0.0015$ (FDR = 33.4%) for a combined FDR of only 6.4% (19.25% * 33.4%). This is highly unlikely to result from chance alone; less than one probe is expected to be shared among the two comparisons at the combined significance level of $P = 1.5*10^{-7}$ ($1.5*10^{-7}$*3,015,075 probes). This result indicates that these differentiated probes between US and West African populations within the Caribbean genome are truly West African-like.

Second, my results indicate that there is a substantial variation in the Caribbean among different chromosomal arms for West African-like signature. The X chromosome is least African-like with only 20.2% of probes more similar to West Africa than to US, while chromosome arm 3R is most West African-like, with 40.9% of probes closer to

West Africa than to US (Figure 4). The other chromosome arms have West African signatures of 29.7% (2R), 33.2% (3L), and 39% (2L). The observation that the X chromosome is predominantly US-like is in agreement with my earlier finding that most North American-West African differences are situated on the X chromosome (see above; Yukilevich et al. ms in prep). In addition to the heterogeneity between different chromosome arms, there is also heterogeneity within arms. In particular, West African-like regions occur at the right end of chromosome 3L, left and extreme right ends of 3R, left and right ends of 2L, and left ends of chromosomes 2R and X. A majority of probes within these regions have a signal intensity of 0.5 or above (Figure 4).

These data also reveal the extent to which Caribbean probes are close to West African signal intensity. Figure 5 shows the frequency of probes with signal intensities below 0.333 (more US-like), between 0.333 and 0.666 (intermediate between US and West Africa) and above 0.666 (more West African-like). Only 13.5% (211 probes) out of the total 1566 differentiated probes are substantially West African-like in the Caribbean (signal above 0.666). US-like and 'intermediate' probes in the Caribbean are similarly frequent at 45% and 41.5%, respectively (Figure 5). If one were to consider a more extreme West African-like signal, such as for instance 0.8 and above, the Caribbean genome only has 96 such probes or 6.1% (96/1566 probes). These probes are shown as isolated peaks or genomic 'islands' of West African signature in Figure 4. These are ideal candidate regions for determining the genetic basis of West African-like traits in the Caribbean populations (e.g. David and Capy 1988; Takahashi et al. 2001; Yukilevich and True 2008a; Yukilevich and True 2008b; see below). Even though this is not a comprehensive description of the differentiated genomic regions between US and West Africa (since the tiling array chips represent 25 bp out of every 40 bp in the genome), my results imply that the genomic signature of West African ancestry in the Caribbean is likely represented by a small number of genomic regions in an otherwise US-dominated genome.

US versus Zimbabwe contribution to the Caribbean genome

I would also like to understand the extent to which Caribbean genomes are shaped by other African populations since it is possible that South African populations may have also historically contributed to Caribbean genomes. Indeed, even though my evidence suggests that Caribbean populations are genomically and reproductively closer to West Africa than to Zimbabwe (see above; Yukilevich and True 2008a), I still do not understand the relative magnitude of the genomic contribution of these different African populations in the Caribbean. Hence I performed a similar analysis on significantly differentiated probes between US and Zimbabwe populations in the Caribbean (Figure 6). First, I found that only 409 out of 1807 (22.4%) probes in the Caribbean are closer to Zimbabwe than to US signal intensities (i.e. above 0.5 in Figure 6). This is about 10% less than the contribution from West Africa to the Caribbean genome (see above). Of the 409 probes that are Zimbabwe-like, 178 probes were also significantly differentiated at $P < 0.0015$ between US and Caribbean populations for a combined FDR of 5.6% (16.67% * 33.4%). Figure 5 illustrates that only 8.6% (155 probes) out of the total 1807 differentiated probes are substantially Zimbabwe-like in the Caribbean (signal above 0.666). US-like probes now make a considerable majority at 60.5% and 'intermediate'

probes are at 30.9%. The number of Zimbabwe-like probes drops substantially if I consider more extreme Zimbabwe-like signal intensities (e.g. probes with signal intensity of 0.8 and above make up only 3.4% or 61 out of 1807 probes). These are excellent candidate regions for Zimbabwe-like traits in the Caribbean (see below).

Interestingly, once again I found that the X chromosome is the least African, with only 12.9% of the differentiated probes closer to Zimbabwe than to US, while chromosome 3R is again the most African-like with up to 31.8% of probes being closer to Zimbabwe than to US. African signatures for other arms were 22.2% (3L), 24.8% (2R) and 26.6% (2L). Finally, heterogeneity within chromosomal arms in terms of the relative contribution of US and Zimbabwe signal intensities was less apparent (Figure 6). US-like probes now dominate each chromosomal arm fairly evenly, with the possible exception of regions on the left and right ends of chromosome 3L, 3R and 2R and left end of chromosome X where signal intensity increases somewhat toward Zimbabwe.

Overlapping differentiated probes between US-West African and US-Zimbabwe tests

Among the significantly differentiated probes between US-West African and US-Zimbabwe comparisons at $P < 0.0001$, I found that 77 probes (2.3%) overlap for a combined FDR of 3.1% (19.25% * 16.67%). These presumably represent general US-African differences. These probes allow us to understand whether the Caribbean signal intensities correlate between US-West African and US-Zimbabwe tests. First, I observe that Caribbean populations are closer in their signal intensities to US than to Africa in the majority (68.8%) of US-African probes (Figure 7). This is consistent with what I have described above. The correlation in Caribbean signal intensities between US-West African tests and US-Zimbabwe tests is excellent for these US-like probes. This means that if a probe in the Caribbean is more similar to US than to African populations, it will be equally different against West Africa and Zimbabwe.

About 23.4% of the US-African differentiated probes in the Caribbean are closer to both African than to US populations (Figure 7). For these probes, I find that the Caribbean genome exhibits more variation in its relative signal intensities across West African and Zimbabwe populations. Thus, 'African-like' probes in the Caribbean may sometimes be more similar to West Africa and sometimes be more similar to Zimbabwe, with equal frequencies (Figure 7). This 'flaring out' of the signal intensities above 0.5 suggests that West Africa and Zimbabwe populations differ somewhat in their own signal intensities at these probes.

Regional US-African differentiated probes

Despite some variation among African signal intensities at probes that differentiate US from African populations, I can gain significant power by pooling samples from West Africa and Zimbabwe populations. Doing so should increase the number of differentiated probes between US and Africa at very low FDR levels. Indeed, my pooled US-African *t*-test resulted in 259 significantly differentiated probes at the FDR < 5% with up to 50 probes (19.3%) that are closer to Africa than to US. By pooling African populations, I more than doubled the number of identified differentiated US-African regions compared to above. These genomic regions are described in detail below. Recall that specific US-West African and US-Zimbabwe comparisons identified 31.4% and 22.4% of Caribbean

probes, respectively, as being closer to African populations than to US. The present analysis suggests that the African signature in pooled (19.2%) and overlapping (23.4%) US-African differentiated probes is limited by the US-Zimbabwe differentiated probes in the Caribbean showing African signature.

Variation between Caribbean islands in African genetic influence

The above analysis averaged signal intensities over the three hybridized Caribbean island populations, including South Andros Island (#35), Rum Cay (#41), and Acklins Island (#45; see *Methods*). I now ask whether these islands vary in their African genetic influence throughout the genome. I studied the signal intensity of each Caribbean island within the same significantly differentiated probes between US-West African and US-Zimbabwe pairwise tests as above. My results indicate that there is substantial geographical variation in the degree to which African-like probes have contributed to genetic variation within different islands (Table 2). Consistently, South Andros Island (#35) is the most African-like with up to 40.2% of the differentiated US-African probes being closer to West Africa than to US and up to 27.5% of the differentiated US-Zimbabwe probes being closer to Zimbabwe than to US. Rum Cay was consistently found to be least African-like, in both West African (26.2%) and Zimbabwe probes (20.5%). Acklins Island had intermediate degree of African influence from both West Africa (34.8%) and from Zimbabwe (25%).

In agreement with results based on average Caribbean signal intensities, West Africa has repeatedly contributed more genetic variation to all three islands compared to Zimbabwe. Importantly, these results are also consistent with the overall phylogenetic relationships based on the relative *number* of significantly differentiated probes between different populations, providing further confidence in both results (see Figure 1 above). Interestingly, the geographical distance between these Caribbean islands and Africa is not a clear predictor of African ancestry since South Andros Island is geographically farthest from Africa, but shows the greatest African signature. Geographical variation among Caribbean islands in their African signature parallels their substantial variation in phenotypic traits and supports the idea that the Caribbean displays a mosaic of US and African genetic influences (Yukilevich and True 2008b).

Relative African genetic influence in US versus Caribbean populations

In addition to understanding how different Caribbean islands have been influenced by African relative to US genetic variation, it is also important to know to what extent Caribbean populations are more African-like than US populations in their probe signal intensities. To answer this question, I pooled African populations once again to reduce false discovery rates of differentiated probes. I then identified significantly differentiated probes between US populations and Africa, this time using one US population at a time. This generated 1474 (FDR = 20.4%) and 1507 (FDR =20%) probes at $P < 0.0001$ between Tuscaloosa, Mississippi (US#23) and Africa and between Columbus, Alabama, (US#22) and Africa, respectively. I then determined the proportion of these probes in the other US population that are closer to US (signal < 0.5) versus African signal intensities (signal > 0.5). I then used all three Caribbean populations to determine their proportion of probes that are closer to the signal intensities of the US population versus Africa. I

performed the same test for the two US-African comparisons above. My results indicate that Caribbean islands have about 13% to 21% more probes with African-like signal intensities than either of the two US populations. Relative to Mississippi population (US#23), South Andros Island (#35) has 21.2% more African-like probes, Rum Cay (#41) has 13.3% more, and Acklins Island (#45) has 17.5% more African-like probes. Relative to Alabama population (US#22), the Caribbean islands have 21.8%, 13.5% and 16.6% more African-like probes, respectively. Once again I see that South Andros is the most African-like and Rum Cay is the most US-like. These consistent results reconfirm the relative phylogenetic relationships of US and Caribbean populations to Africa and indicate that Caribbean islands harbor substantially more African-like genetic variation than US populations.

CANDIDATE GENOMIC 'ISLANDS' FOR AFRICAN-LIKE TRAITS IN CARIBBEAN

My results revealed that among the highly differentiated genomic regions between US and African populations, the Caribbean genome harbors many sites of African-like signature (see Figures 4 and 6 above). These become excellent candidate genomic regions for African-like phenotypes, behaviors, and pheromones that segregate in the Caribbean islands (Yukilevich and True 2008b). In Table 3 I list these candidate loci and their specific genomic regions. It is clear that these hypotheses need to be further validated. This requires first testing for false positives using genomic sequencing and second to perform functional work to test for involvement of these regions in the traits of interest (e.g. Turner et al. 2008; Greenberg et al. 2003). Also, because my primary interest is to identify genomic regions that may possibly correlate with African-like phenotypes known to segregate in the Caribbean islands, I do not detail genomic regions in the Caribbean that show a US-like genetic signature. A full list of annotated genes that show probes with both US-like and West African and Zimbabwe-like signatures within the Caribbean genome is available by request from the corresponding author.

I found that many of the highly differentiated genomic sites between US and African populations harbor annotated loci that are known to be associated with various phenotypic, behavioral, pheromonal and other potentially adaptive functions (Table 3). Note that many of these candidate loci are highly pleiotropic and are also associated with other biological functions. In Table 3 I list the 25 bp probes located in these candidate genes, denote whether the probes are differentiated between US-West Africa, US-Zimbabwe, or both comparisons, and show the relative Caribbean signal intensity at these probes (we only list those probes that are more African-like in the Caribbean, i.e. signal > 0.5). As in Figures 4 and 6, Caribbean signal intensities closer to 0.5 are likely to be associated with intermediate allelic frequencies between US and African populations. I also describe whether these probes are situated within introns or protein-coding exons and the degree of sequence conservation among fifteen different *Drosophila* genomes (www.genome.ucsc.edu).

Consistent with known phenotypic divergence between US and Caribbean populations, I identified several unusually African-like genomic regions in the Caribbean within genes known to be associated with mating and courtship behaviors, perception of smell and olfactory behaviors, and morphological features such as growth and

pigmentation (Table 3). The majority of these regions are within non-coding introns, while some are in highly conserved protein-coding exons of genes. For instance, I found three highly significantly differentiated introns between US and African populations within the gene *fruitless,* which is regulates male courtship and mating behavior. Caribbean populations tend to show African-like signal intensities of 0.66 to 0.76 within these probes. Another probe is situated within a highly conserved exon region of the gene *Shaker*, which is known to be associated with mating behavior as well as other relevant functions such as visual and taste perception and flight behavior. Caribbean populations are nearly identical in their signal intensities to those of West African populations in this region. Some probes are differentiated between US and both African populations such as the probe within the gene *lingerer*, which is associated with male copulation. Caribbean flies seem to be intermediate between US and African populations at this probe.

Even though this approach has identified genes that are associated with phenotypes that are also differentiated between US versus Caribbean and African populations, many regions may not be identified since the tiling array does not cover the whole genome. For instance, US and African populations are known to be highly differentiated at the pheromone (olfactory) gene *desaturase-2*, with a 16 bp indel differentiating populations (Takahashi et al 2001). This gene is also strongly clinally differentiated across the US-Caribbean populations, with many Caribbean islands, including those in this study, segregating the African insertion allele at high frequencies (Yukilevich and True 2008b). However, the tiling array analysis did not identify this region since the available probes do not overlap with the 16 bp indel. For this reason, many other potentially divergent sites will also not be identified with this approach.

In addition to a priori biological functions based on known trait differences between US and Caribbean populations, I have also uncovered many other genomic regions with African-like genetic signature within the Caribbean that have not been previously described at the phenotypic level. Among these are regions within annotated genes associated with circadian rhythm, such as the *Serotonin receptor 1A (5-HT1A),* visual perception of light, such as the gene *rdgA*, adult and larval locomotion, such as an exon region in the gene *pickpocket* (*ppk*), many regions within environmental stress and response genes such as the gene *Myd88,* which is divergent between US and both African populations, and genes associated with perception of taste, and flight behavior such as the gene *flightless I*. Thus, these genomic differences generate new hypotheses concerning more African-like behaviors and phenotypes that may potentially segregate in the Caribbean populations that have not yet been described at the phenotypic level.
Table 4 lists 48 probes within annotated genes that are differentiated between pooled or overlapping US-African comparisons and which also show African-like signature within the Caribbean genome (see above). A few of these regions have already been described either in the three-way regional shared differences above (US-Caribbean-African comparisons) or in Table 3 where US are differentiated from both Caribbean as well as from the two African populations (e.g. *Myd88*). Many other regions are newly identified because of increased statistical power from pooling samples in Africa. Note that many of these probes are situated in protein-coding regions that are highly conserved. A list of probes within annotated genes that show a US-like signature within the Caribbean genome is available upon request from corresponding author.

DISCUSSION

In this study, I used Affymetrix tiling arrays to investigate genome-wide differentiation between US, Caribbean and African populations of *Drosophila melanogaster* that have been previously characterized in their morphological, behavioral, pheromonal and reproductive traits. Phenotypic work suggests that Caribbean populations represent a mixture of US and African populations that have previously diverged in many phenotypic traits likely involved with adaptation to their temperate versus tropical environments (David and Capy 1988; Capy et al. 1993, 1994; Hollocher et al. 1997a; Colgave et al. 2000; Takahashi et al. 2001; Greenberg et al. 2003; Coyne and Elwyn 2006). Caribbean populations harbor many African-like morphology, courtship, behavior, and pheromone traits that are rarely found outside of Africa (David and Capy 1988; Capy et al. 1993; Takahashi et al. 2001; Yukilevich and True 2008b). These findings strongly suggest that Caribbean populations were colonized directly from Africa during the historical human slave trades into the region several hundred years ago (David and Capy 1988; Caracrsti and Schlotterrer 2003). The observation that Caribbean and West African populations mate randomly with each other, but both show incipient sexual isolation from US populations further strengthens the plausibility of this scenario (Yukilevich and True 2008a). Thus it is likely that Caribbean populations represent a hybridization event between US and African populations.

However, despite these phenotypic observations, very little was known previously about genetic differentiation of Caribbean populations relative to US or Africa. In general, I still do not know the genetic basis of much of the phenotypic differentiation that occurs between these populations. Further, it was unknown the extent to which Caribbean population genomes are shaped by US versus African genetic variation. The Affymetrix tiling arrays allow us to begin to describe genome-wide patterns of differentiation among these populations, determine the extent of US versus African genetic influence on the Caribbean populations and identify unusually African-like genomic regions in the Caribbean genome that are excellent candidates for phenotypic and behavioral traits across the islands. As a result this study offers an unprecedented look at a natural hybridization event at the scale of the whole genome.

My findings indicate that despite being phenotypically and behaviorally closer to African populations, the Caribbean populations are overall closer to US than to African populations at the level of the genome. This conclusion is based on several lines of evidence. First, I found that Caribbean populations have fewer significantly differentiated probes or outliers in comparison with the US than with African populations and Caribbean populations cluster together with US populations by a Neighbor Joining analysis. Second, Caribbean populations share more differences in common with the US against Africa than they share in common with Africa against the US. Third, among the highly differentiated probes between US and African populations, 60% to 80% in the Caribbean genome are closer in their probe signal intensity to US than to Africa. The fact that I observe the same genetic patterns using these different approaches provides strong confidence in this general result and validates the use of tiling array-based genomic data to infer such population genomic patterns. Perhaps the most likely explanation for this genome-wide pattern is that gene flow from US into the Caribbean islands has been

substantial throughout the genome. These results are in general agreement with previous findings based on ten neutral microsatellite loci in two Lesser Antilles Caribbean populations that showed closer genetic relationship to non-African populations (Schlotterer et al. 1997).

However, more recent work with more microsatellite loci in Europe, US and Africa has indicated that US populations harbor more African-based microsatellite alleles compared to European populations, suggesting that African alleles have been introduced into the New World either through the Caribbean or directly from Africa in a secondary colonization wave (Caracristi and Schlotterer 2003). This is consistent with phenotypic observations in Caribbean populations, leading to the possibility that Caribbean populations represent an admixture of US and African populations, with African variation likely being maintained by natural or sexual selection despite gene flow from US (David and Capy 1988; Caracristi and Schlotterer 2003; Yukilevich and True 2008a,b). Additional work is necessary to determine the extent of gene flow between US and Caribbean populations at neutral genetic markers.

My results have provided evidence for this latter hypothesis on a genome-wide scale. I found that African populations are clearly genetically closer to Caribbean than to US populations. This is based on the finding that fewer probes are significantly differentiated between Caribbean and African populations compared to between US and African populations, and on the observation that Caribbean and African populations share substantially more differentiated probes in common against US than US shares with Africa against the Caribbean. Finally, my genome scans of Caribbean populations revealed on the order of 20% to 40% of differentiated US-African probes with African-like signatures, with West Africa being more represented than Zimbabwe in the Caribbean genome. This suggests that the Caribbean genome is highly structured or "porous" with respect to US versus African genetic variation. It also implies that different genes within Caribbean populations will show different evolutionary histories. Thus, depending on which markers are used, inferences concerning further evolutionary/demographic relationships of these populations will strongly vary. These results support the notion that admixture events in nature could maintain few strongly selected regions or "genomic islands" in an otherwise non-differentiated genome (Via 2001; Wu 2001; Vines et al. 2003; Turner et al. 2005).

I also found that Caribbean populations tend to have about 13% to 21% more African-like probes genome-wide than US populations. This is consistent with results of Caracristi and Schlotterer (2003) who found that US populations also harbor African alleles in their genomes. Future work should determine the relative importance of European genomic influences versus Caribbean/African influences on the genomes of US populations. I would expect the Caribbean/African signature to show a clinal North-South pattern along the Eastern US coast.

It is also noteworthy that Caribbean islands consistently show different degrees of African genomic influence. For instance, South Andros island is consistently genetically closer to West Africa than it is to US populations based on having fewer significantly differentiated outliers with Africa and having the highest percentage of probes with African-like signal intensities than any other Caribbean population. The opposite result is seen for the Caribbean island Rum Cay, which is consistently genomically closest to US

populations, while Acklins Island tends to be intermediate in terms of African genomic influence. These results suggest that gene flow from the US into the Caribbean has been uneven across different islands with some being less influenced by US genetic variation than others.

Interestingly, I also observed variation in the African signature among different chromosomes in the Caribbean, with the X chromosome consistently most US-like (around 10% to 20%), while chromosome 3R is the most African-like (around 30% to 40%). Other autosomes also show much more African-like signatures compared to the X chromosome. These results are consistent with the finding that Africa is predominantly differentiated from US and Caribbean populations on the X chromosome. Nearly 60% of shared differences between US and Caribbean populations against Africa are situated on the X. One likely possibility for this striking unevenness among chromosomes in African signature in the Caribbean genome is that introgression of US genetic variation into the Caribbean was non-random with respect to chromosomes. It seems that US gene flow into the Caribbean has primarily occurred on the X chromosome, but has been less effective on the autosomes, probably due to natural or sexual selection on the autosomes that favor more African-like traits across the Caribbean islands.

This data is consistent with the predominantly autosomal genetic basis of Zimbabwe female mating discrimination (Hollocher et al. 1997b; Ting et al. 2001). If autosomes were also primarily responsible for the West African mating discrimination, this could provide a clear explanation for why US gene flow could introgress easily into the X, but not into autosomes. Similarly, autosomes may be primarily responsible for ecologically important adaptations to tropical environments, which could then explain why Caribbean genomes retain these tropical adaptations. Indeed, I find that the great majority of African-like probes in the Caribbean are situated in autosomal candidate genes (see Tables 3 and 4). Interestingly, my results are completely opposite to those of Kauer et al. (2003) who studied 151 neutral microsatellite loci in a recent admixture event between reintroduced European *D. melanogaster* into Zimbabwe populations. They found that the X chromosome retained the greatest African ancestry, while the third chromosome became most European-like at these microsatellite loci. Even though they point out that this chromosomal pattern was inconsistent with the genetic basis of Zimbabwe mating behaviors, it is consistent with the X being strongly divergent between non-African and African populations (see above; Yukilevich et al. ms in prep). Perhaps it is adaptive to retain the African X within Africa, but not outside of the continent due to specific ecological adaptations that are only favorable in Africa.

In general I have identified many strongly African-like regions in the Caribbean genome in an otherwise US-dominated genome. Future work should test whether similar signal intensities in Caribbean and African probes are due to the same allelic sequences. At present, these become ideal candidate regions for genetic investigations of phenotypic differences between US and Caribbean populations. Many of these regions are situated in annotated genes known to be associated with relevant phenotypic functions ranging from mating and courtship behavior, to morphology and pheromones. Further I identified other genomic regions of exceptional African-like signature in the Caribbean within genes that are associated with other potentially adaptive functions whose variation has not yet been described in these populations. These include flight ability, circadian rhythm,

environmental stress responses, perception of taste and light, and locomotor activity. These generate excellent hypotheses concerning other African-like phenotypes that may be segregating in the Caribbean.

In conclusion, I suggest that Caribbean populations may be particularly unusual among worldwide *Drosophila melanogaster* in that they have acquired many African-like traits and adaptations to tropical environments and tend to exhibit partial sexual isolation from US populations (Yukilevich and True 2008a,b). It is likely that these populations have partially retained their ancestral African phenotypic makeup even in the presence of substantial gene flow from US populations. The observation that West African populations also show partial sexual isolation from US populations (Yukilevich and True 2008a) and that Caribbean genomes have many strongly West African-like genomic regions suggests that the two cases of incipient sexual isolation from US populations are not independent, but are instead related through shared genetic history (also see Yukilevich and True 2008b). This is in contrast to other recent multiple cases of incipient sexual isolation in nature that show independent or parallel evolution such as in the case of sticklebacks (Rundle et al. 2000; Rundle and Schluter 2004; Boughman et al. 2005) and phytophogous insects (Funk 1998; Nosil et al. 2002; Nosil 2005; Nosil et al. 2007). This study has generated a list of genomic regions that may be responsible for these cases of incipient sexual isolation between US-Caribbean and US-African populations. Future research should determine if both cases of incipient sexual isolation have the same genetic basis.

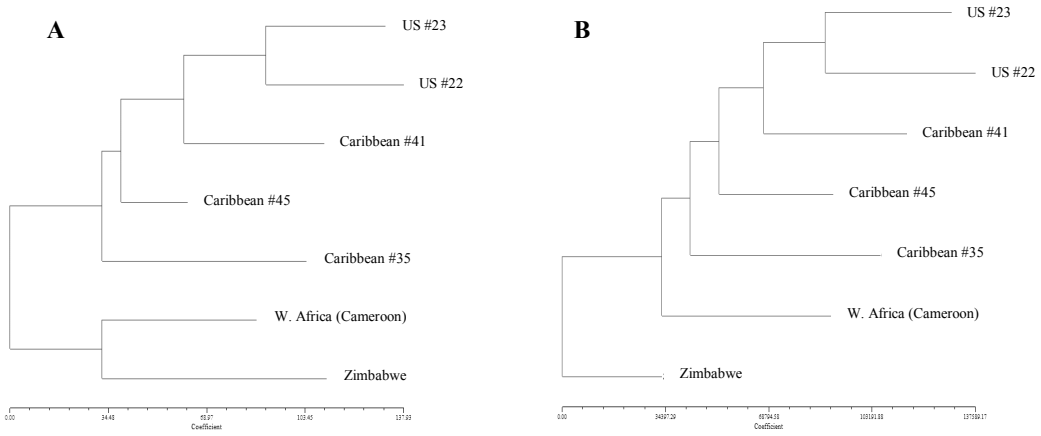


Figure 1. Neighbor-Joining clustering of geographical localities based on number of significantly differentiated probes between each pair of localities at significance level of $P < 0.0001$ (A) and $P < 0.05$ (B). Results for $P < 0.01$ gave qualitative identical results to those of panel A (data not shown). Significance was determined using pairwise t-test (see Materials and Methods). See Table 1 for distance matrices used in clustering. Clustering was performed using NTSYSpc (Rohlf 2004).

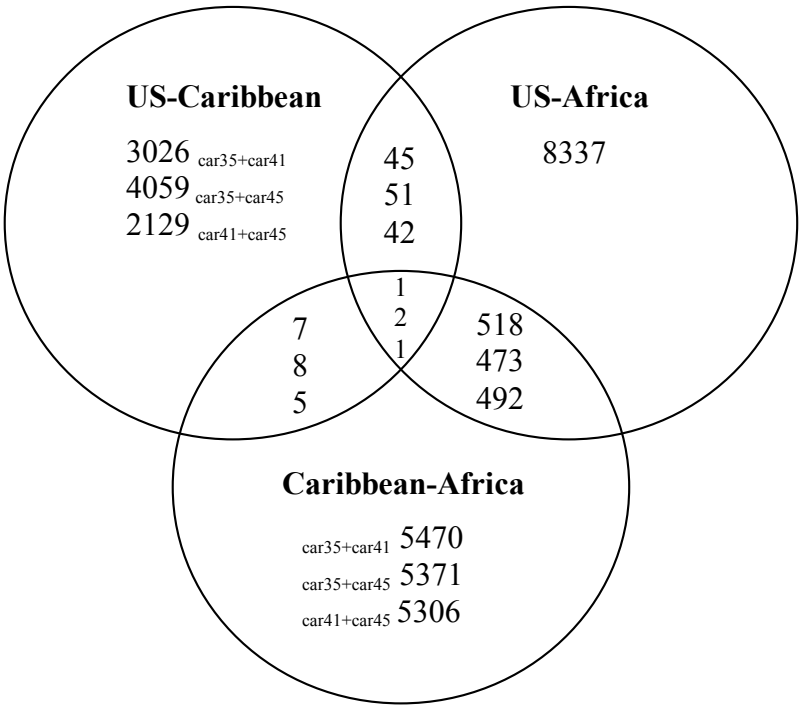


Figure 2. Venn diagram of number of significantly differentiated probes at significance level of $P < 0.0005$ between regional comparisons (in non-overlapping part of each circle) and number of shared probes significantly differentiated between different regional comparisons (in overlapping parts of circles). Statistical significance was determined using pairwise t-tests (see *Methods*). The US region contains US#23 and US#22 locations, Africa contains West Africa (Cameroon) and Zimbabwe locations, and Caribbean contains car#35, car#41, car#45 islands (see text). I studied two locations per region for each pairwise *t*-test to keep statistical power constant across tests. Since the Caribbean region has three locations, I performed three different independent tests against US and African regions (see above). Results for shared differences follow the same order of Caribbean comparisons as for pairwise tests.

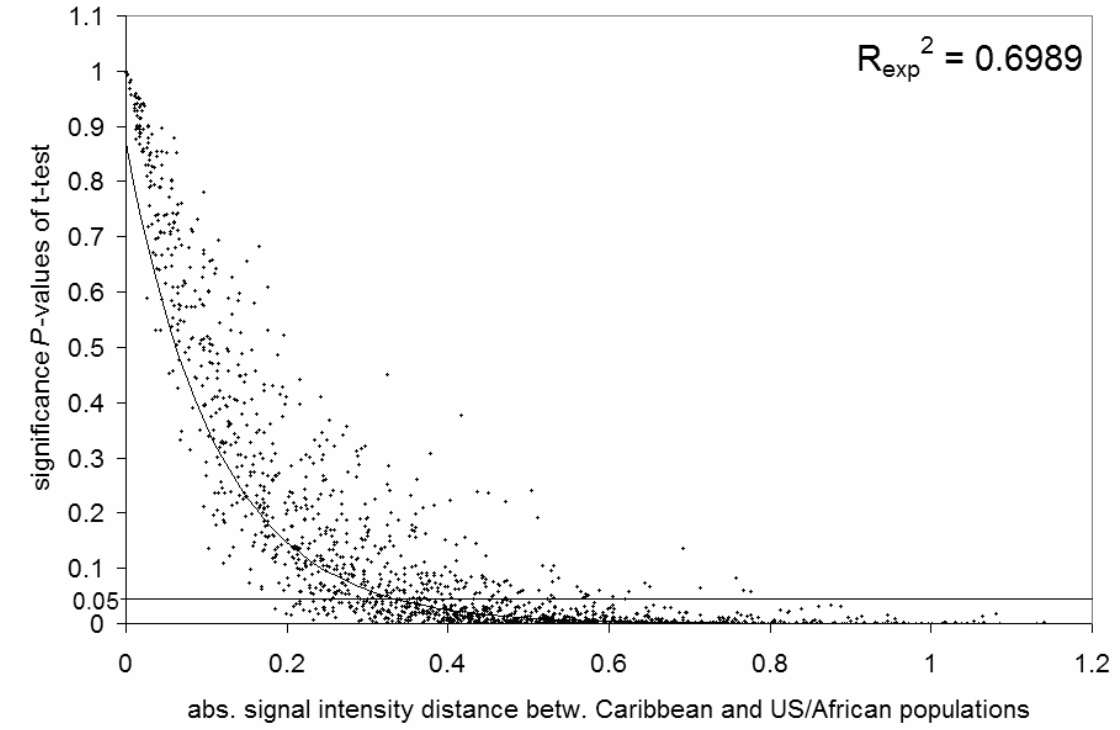


Figure 3. Correlation between absolute signal intensity distance between Caribbean and US or African populations and its associated significance *P*-value using the t-test. The relationship is identical for US-Caribbean and Caribbean-African comparisons. Horizontal line represents *P*-value = 0.05.

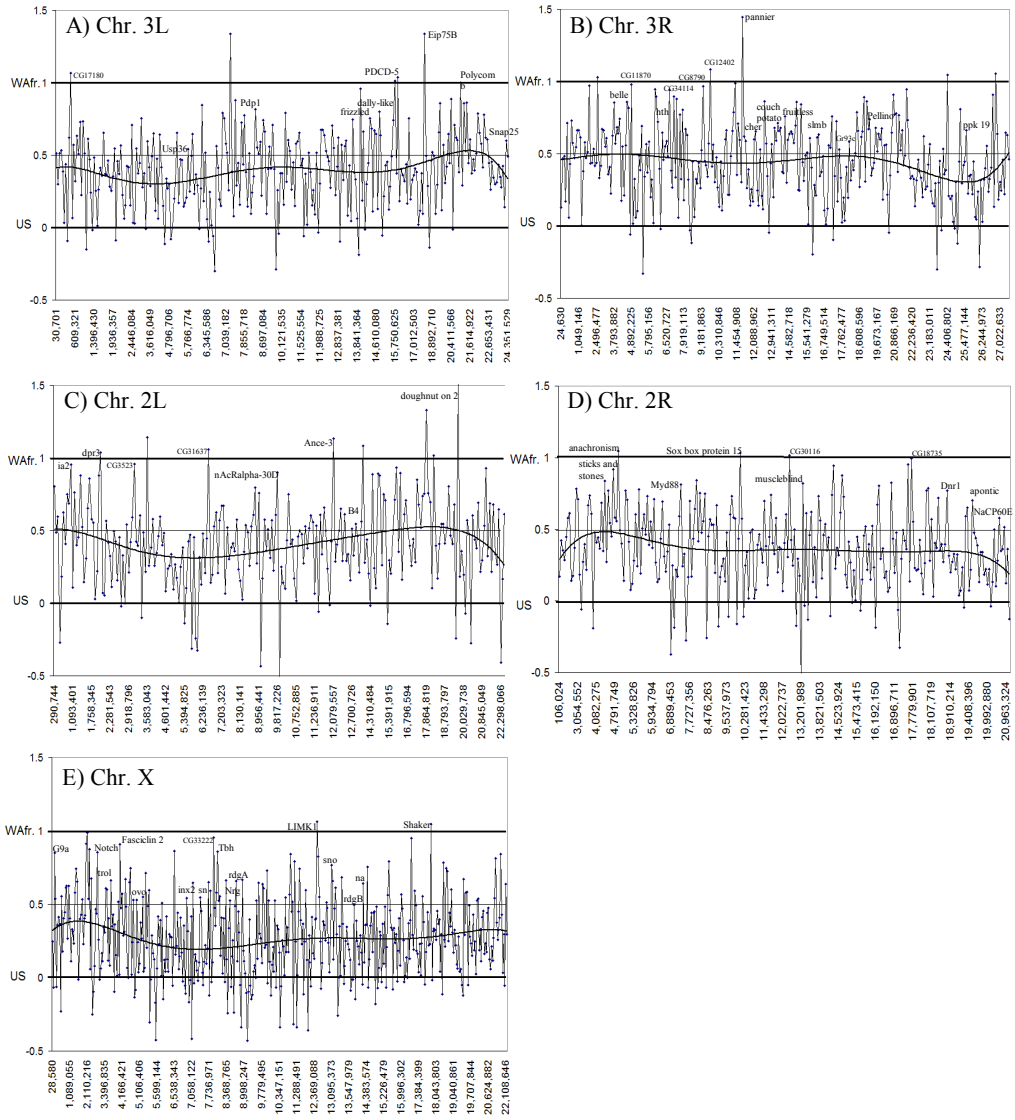


Figure 4. Genome-wide scan in Caribbean islands of relative signal intensities (y-axis) in significantly differentiated probes at $P < 0.0001$ between US and West African (Cameroon) populations. The x-axis represents chromosomal position in base pairs. A) Chrom. arm 3L, B) Chrom. arm 3R, C) Chrom. arm 2L, D) Chrom. arm 2R, E) X Chrom. Horizontal lines at 1 and 0 are the standardized signal intensities of Zimbabwe and US populations, respectively. Horizontal line at 0.5 is the standardized 50% intermediate intensity between West Africa and US populations. Signal intensities above 0.5 are closer to West Africa than to US and vice versa for signal intensities below 0.5. Standardization across all probes was obtained using the relation Signal = (Car.-US)/(W.Afr. –US). For probes where signal intensity of US was higher than West Africa, all data (including Caribbean) was reversed so as to always have West Africa signal at 1. Points show actual probe intensities across the chromosomes and the connecting line is to facilitate visualization of the pattern. Also shown is a 6-order polynomial trendline in each plot. Probes associated with genes in Table 3 and 4 or above signal intensity 0.9 are labeled with gene name.

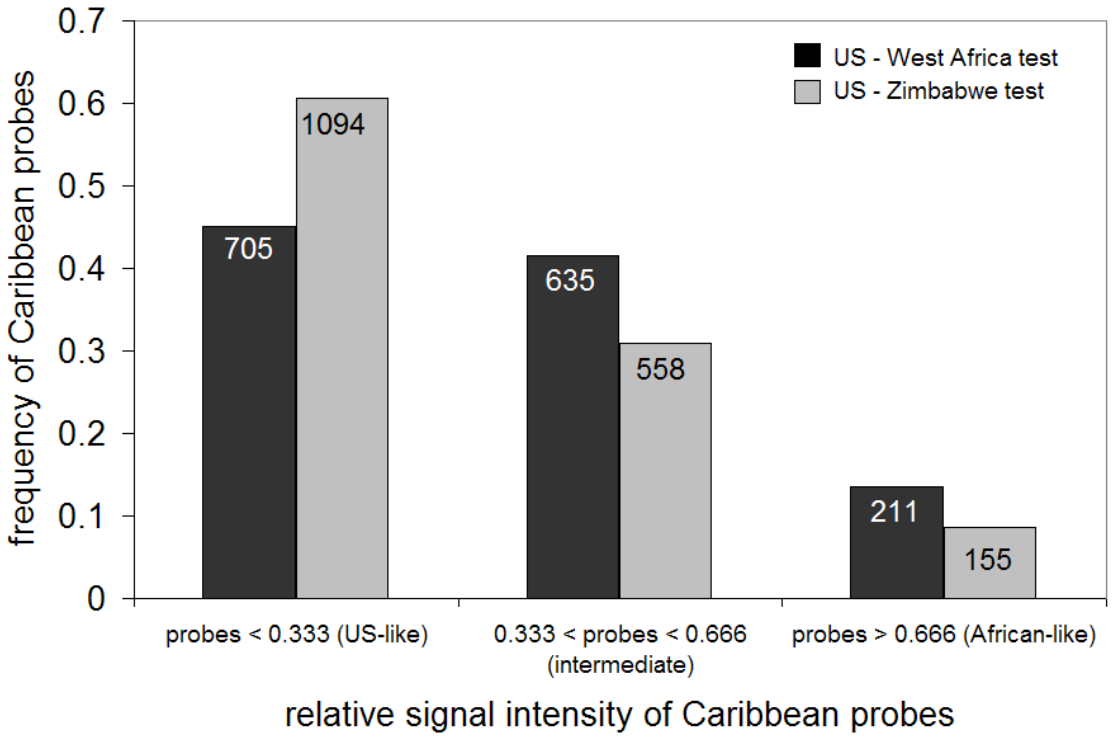


Figure 5. Overall frequency of Caribbean signal intensities of probes that are differentiated between US and African populations. Probe intensities range from very US-like (signal < 0.3333), intermediate intensities between US and Africa (0.333 < signal < 0.666) and intensities that are very African-like (signal > 0.666).

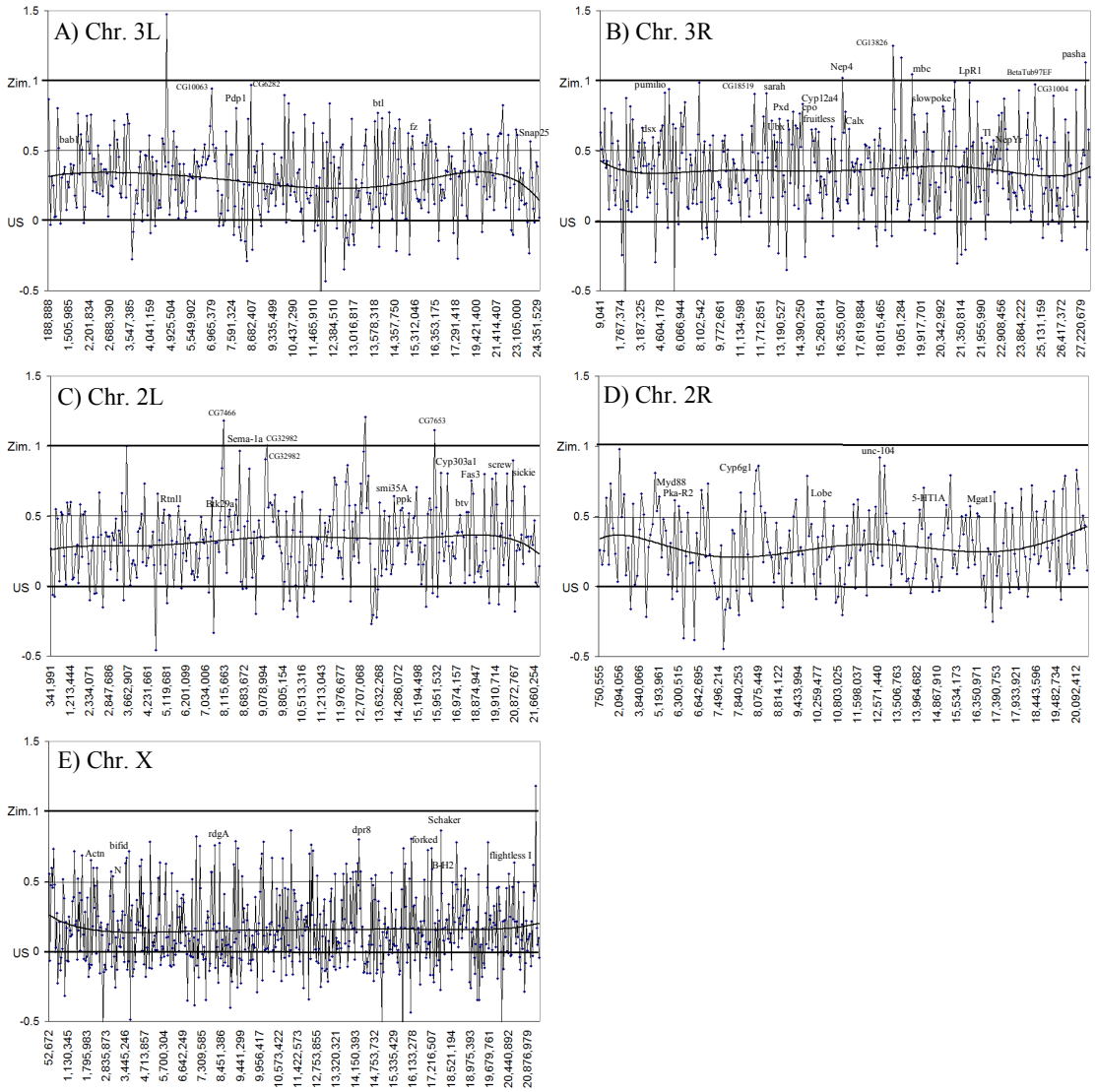


Figure 6. Genome-wide scan in Caribbean islands of relative signal intensities (y-axis) within significantly differentiated probes at $P < 0.0001$ between US and Zimbabwe populations. The x-axis represents chromosomal position in base pairs. A) Chrom. arm 3L, B) Chrom. arm 3R, C) Chrom. arm 2L, D) Chrom. arm 2R, E) X Chrom. Horizontal lines at 1 and 0 are the standardized signal intensities of Zimbabwe and US populations, respectively. Horizontal line at 0.5 is the standardized 50% intermediate intensity between Zimbabwe and US populations. Signal intensities above 0.5 are closer to Zimbabwe than to US and vice versa for signal intensities below 0.5. Standardization across all probes was obtained using the relation Signal = (Car.-US)/(Zimb. –US). For probes where signal intensity of US was higher than of Zimbabwe, I reversed all the data (including Caribbean) so as to always have Zimbabwe signal at 1. Points show actual probe intensities across the chromosomes and the connecting line is to facilitate visualization of pattern. Also shown is a 6-order polynomial trendline in each plot. Probes associated with genes in Table 3 and 4 or above signal intensity 0.9 are labeled with gene name.

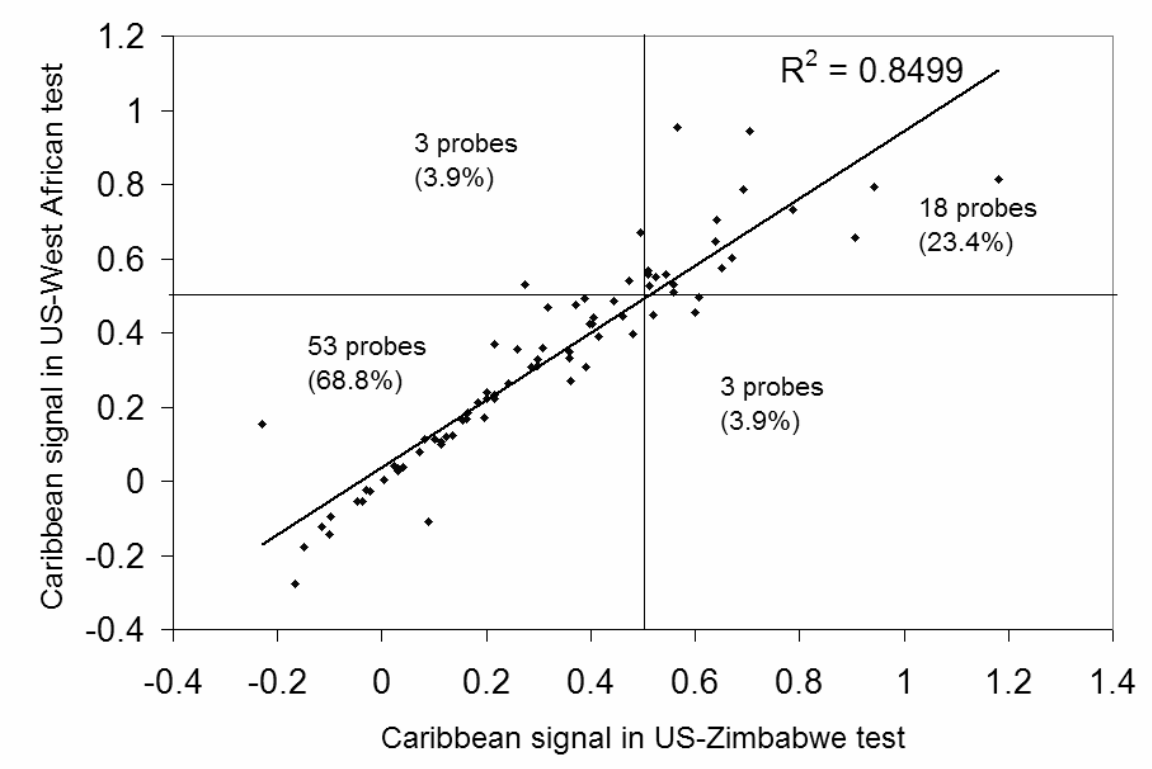


Figure 7. Caribbean signal intensity in 77 overlapping differentiated probes between US-West African test and US-Zimbabwe test (i.e. general US-African differences). Horizontal and vertical lines represent 0.5 signal intensity, which is the half-way point between US and African signal intensities.

Table 1a. Number of significantly differentiated probes between each pair of populations at $p<0.0001$.

	US23	US22	Carib35	Carib41	Carib45	Cameroon
US23						
US22	90					
Carib35	166	197				
Carib41	121	125	140			
Carib45	123	114	103	95		
Cameroon	226	213	173	218	142	
Zimbabwe	228	240	232	218	177	133

Table 1b. Number of significantly differentiated probes between each pair of populations at $p<0.05$.

	US23	US22	Carib35	Carib41	Carib45	Cameroon
US23						
US22	92,034					
Carib35	159,780	170,916				
Carib41	113,843	115,022	137,762			
Carib45	114,587	120,402	104,630	102,380		
Cameroon	152,134	167,215	125,072	141,127	117,579	
Zimbabwe	151,432	158,021	142,934	140,789	125,469	122,629

Table 2. African genetic influence relative to US on the genomes of different Caribbean populations.

	South Andros (#35)	Acklins (#45)	Rum Cay (#41)
West African contribution	40.2%	34.8%	26.2%
Zimbabwe contribution	27.5%	25.0%	20.5%

Table 3. Candidate genes and specific genomic regions for African-like traits in the Caribbean islands

Short Name	Full Name	Chromos.	Start Probe	End Probe	Carib. signal (1 = Afr.)	Differentiation	Intron/Exon	Seq. Conservation
Mating and courtship behavior								
Btk29A	Btk family kinase at 2	Chr2L	8,271,429	8,271,453	0.55	U-Z, U-W	I	Low
btv	beethoven	Chr2L	17,972,568	17,972,592	0.53	U-Z	I	High
dsx	doublesex	Chr3R	3,781,078	3,781,102	0.59	U-Z	I	Medium
fru	fruitless	Chr3R	14,279,909	14,279,933	0.68	U-Z	I	Low
fru	fruitless	Chr3R	14,321,371	14,321,395	0.66	U-Z	I	High
fru	fruitless	Chr3R	14,320,821	14,320,845	0.76	U-W	I	Medium
lig	lingerer	Chr2R	3,959,855	3,959,879	0.61	U-Z, U-W	I/E	Low/High
Sh	Shaker	ChrX	17,856,075	17,856,099	0.86	U-Z	I	Low
Sh	Shaker	ChrX	17,847,304	17,847,328	1.04	U-W	E	High
slo	slowpoke	Chr3R	20,524,443	20,524,467	0.82	U-Z	I	High/Low
sra	sarah	Chr3R	12,013,051	12,013,075	0.91	U-Z	I	High/Low
Olfaction Behavior/Learning*								
cher	cheerio	Chr3R	12,941,311	12,941,335	0.57	U-W	I	Low
cpo	couch potato (diapause)	Chr3R	13,770,279	13,770,303	0.65	U-Z	I	Low/High
cpo	couch potato (diapause)	Chr3R	13,762,146	13,762,170	0.71	U-W	I	Low/High
Fas2	Fasciclin 2	ChrX	4,086,080	4,086,056	0.91	U-W	I	Low
Fas3	Fasciclin 3	Chr2L	18,363,822	18,363,846	0.75	U-Z	I	High
inx2	innexin 2	ChrX	6,895,613	6,895,589	0.54	U-W	I	Low/High
NaCP60E	-	Chr2R	20,789,349	20,789,373	0.52	U-W	I	Medium
rdgB	retinal degeneration B	ChrX	13,666,065	13,666,041	0.50	U-W	E	High
Rtnl1	Rtnl1	Chr2L	5,006,528	5,006,552	0.55	U-Z	E	Low
slmb	supernumerary limbs	Chr3R	16,949,947	16,949,971	0.56	U-W	I	Low
smi35A	smell impaired 35A	Chr2L	14,194,951	14,194,975	0.64	U-Z	I	Low
Circadian rhythm/Sleep/Wake Cycle								
5-HT1A	Serotonin receptor 1A	Chr2R	14,996,882	14,996,906	0.58	U-Z	I	Low
B4	-	Chr2L	13,507,443	13,507,467	0.54	U-W	I	Medium
na	narrow abdomen	ChrX	14,168,136	14,168,112	0.64	U-W	I	Low
Pdp1	PAR-domain protein 1	Chr3L	7,855,718	7,855,742	0.78	U-W	I	Low
Pdp1	PAR-domain protein 1	Chr3L	7,810,625	7,810,649	0.80	U-Z	E	High
Pka-R2	cAMP-dependent protein kinase R2	Chr2R	5,896,458	5,896,482	0.61	U-Z	I	Low/High
Growth/larval development*								
bel	belle	Chr3R	4,485,682	4,485,706	0.74	U-W	E	High
G9a	G9a	ChrX	141,910	141,934	0.85	U-W	I	Low
E2f	E2f transcription factor	Chr3R	17461311	17461287	0.55	U-Z, U-W	I	Low/High
L	Lobe (imaginal disc)	Chr2R	10,374,295	10,374,319	0.61	U-Z	I	Low/High
scw	screw	Chr2L	19,699,845	19,699,869	0.76	U-Z	E	High
Ubx	Ultrabithorax (imaginal disc)	Chr3R	12,504,341	12,504,365	0.56	U-Z	I	High
Usp36	Ubiquitin specific protease 36	Chr3L	5,766,774	5,766,798	0.54	U-W	E	High
Pigmentation								
bab1	bric a brac 1	Chr3L	1,042,931	1,042,955	0.52	U-Z	I	Low
bi	bifid	ChrX	4,306,918	4,306,942	0.66	U-Z	E	High
ovo	ovo	ChrX	4,946,152	4,946,128	0.53	U-W	I	High
Visual Stimulus/Sensory to light/Gravitaxis*								
B-H2	BarH2	ChrX	17,216,507	17,216,531	0.59	U-Z	I	Medium
dlp	dally-like	Chr3L	14,639,470	14,639,494	0.80	U-W	I	High
nmo	nemo	Chr3L	7,992,200	7,992,224	0.51	U-W	I	Low/High
rdgA	retinal degeneration A	ChrX	8,810,266	8,810,242	0.67	U-W	I	Medium
rdgA	retinal degeneration A	ChrX	8,820,826	8,820,850	0.62	U-Z	I	Low
rdgA	retinal degeneration A	ChrX	8,915,522	8,915,546	0.79	U-Z	E	High
Locomotion/Larval Movement*								
ppk	pickpocket	Chr2L	14,380,206	14,380,230	0.56	U-Z	E	High
Tbh	Tyramine beta hydroxylase	ChrX	7,891,867	7,891,843	0.86	U-W	I	Medium
na	narrow abdomen	ChrX	14,168,136	14,168,112	0.64	U-W	I	Low
Pka-R2	cAMP-dependent protein kinase R2	Chr2R	5,896,458	5,896,482	0.61	U-Z	I	Low/High
unc-104	unc-104	Chr2R	12,644,843	12,644,867	0.92	U-Z	E	High
Mgat1	UDP-GlcNAc	Chr2R	16,448,171	16,448,195	0.50	U-Z	E	High
Environmental Stress/Response/Resistance								
Cyp12a4	Cyp12a4	Chr3R	14,962,322	14,962,346	0.65	U-Z	E	High
Cyp6g1	Cyp6g1	Chr2R	8,075,449	8,075,473	0.86	U-Z	E	High
Dnr1	Defense repressor 1	Chr2R	18,477,381	18,477,405	0.72	U-W	I	Low
Myd88	-	Chr2R	5,193,961	5,193,985	0.71	U-Z, U-W	I	Low/High
nAcRalpha-30D	nicotinic Acetylcholine Receptor	Chr2L	9,817,226	9,817,250	0.90	U-W	I	Low/Medium
nAcRalpha-30D	nicotinic Acetylcholine Receptor	Chr2L	9,882,780	9,882,804	0.53	U-Z	I	Low/High
Pli	Pellino	Chr3R	19,715,364	19,715,340	0.64	U-W	I	Medium/Low
Pxd	Peroxidase	Chr3R	12,847,742	12,847,766	0.92	U-Z, U-W	I	High
sick	sickie	Chr2L	19,910,714	19,910,738	0.81	U-Z	I	Low
Tbh	Tyramine beta hydroxylase	ChrX	7,891,867	7,891,843	0.86	U-W	I	Low/Medium
Tl	Toll	Chr3R	22,630,851	22,630,875	0.53	U-Z	I	Low
Flight Behavior/Ability*								
Actn	alpha actinin	ChrX	1,923,232	1,923,256	0.65	U-Z	I	Medium
fliI	flightless I	ChrX	21,215,446	21,215,470	0.61	U-Z	E	Low
Perception of Taste*								
Gr93c	Gustatory Receptor 93	Chr3R	17,670,183	17,670,207	0.53	U-W	E	High
ppk19	pickpocket 19	Chr3R	25,499,104	25,499,080	0.66	U-W	E	High/Low

*Note: Candidate loci *Shaker* and *sarah*, *slmb* and *Pdp1* are associated with multiple functions in the table, but only shown once. *Shaker* is associated with mating, flight behavior, visual and taste perception. *sarah* is associated with mating, and olfactory behavior. *slmb* is associated with olfactory behavior, circadian rhythm and locomotion. *Pdp1* is associated with circadian rhythm and growth. Sequence conservation is based on 15 *Drosophila* species genomes in www.genome.ucsc.edu.

Table 4. Candidate genes and genomic regions for general African-like traits in the Caribbean islands

Short Name	Full Name	Chromos.	Start Probe	End Probe	Carib. signal (1 = Afr.)	Intron/Exon	Seq. Conservation
PDCD-5	PDCD-5	Chr3L	16,224,448	16,224,472	1.04	E	High
Pxd	Peroxidase	Chr3R	12,847,742	12,847,766	0.92	I	High
CG31004	CG31004	Chr3R	27,018,451	27,018,475	0.91	I	Low
CG13176	CG13176	Chr2R	8,068,376	8,068,400	0.91	E	High
CG32982	CG32982	Chr2L	9,134,366	9,134,390	0.88	I	Low
CG8602	CG8602	Chr3L	7,321,463	7,321,487	0.88	E	Low
CG10063	CG10063	Chr3L	6,965,379	6,965,403	0.87	I/E	Low/High
tna	tonalli	Chr3L	10,852,806	10,852,830	0.87	I	Low/Medium
CG33967	CG33967	Chr3R	10,533,795	10,533,819	0.86	I	Low/High
CG17600	CG17600	ChrX	21,920,591	21,920,615	0.84	I	Medium
CG30116	CG30116	Chr2R	14,345,517	14,345,541	0.82	I	High
Aats-ala-m	-	Chr3L	5,140,508	5,140,532	0.80	I	Low/High
CG11791	CG11791	Chr3R	20,877,886	20,877,910	0.78	I	Low/High
CG33222	CG33222	ChrX	7,793,355	7,793,379	0.76	I	Low
stj	straightjacket	Chr2R	9,706,236	9,706,260	0.76	E	High
fz	frizzled	Chr3L	14,345,598	14,345,622	0.76	I	Low/High
yellow-b	yellow-b	Chr2L	16,757,165	16,757,189	0.75	E	High
Vrp1	Verprolin 1	Chr2R	18,008,483	18,008,507	0.74	I	Low
hoe2	hoepel2	Chr2L	4,934,423	4,934,447	0.72	E	High
Nos	Nitric oxide synthase	Chr2L	10,836,798	10,836,822	0.71	I	Low
rno	rhinoceros	Chr3L	188,888	188,912	0.71	E	High
Myd88	Myd88	Chr2R	5,193,961	5,193,985	0.71	I	High
CG14606	CG14606	Chr3R	3,041,226	3,041,250	0.67	E	Low/High
CG2604	CG2604	Chr3R	1,049,146	1,049,170	0.66	E	Medium
CG18519	CG18519	Chr3R	11,365,705	11,365,729	0.66	E	Low/High
NmNat	-	Chr3R	20,772,979	20,773,003	0.64	I/E	Low/High
CG15685	CG15685	Chr3R	16,169,224	16,169,248	0.62	E	Low/High
lig	lingerer	Chr2R	3,959,855	3,959,879	0.61	I/E	Low/High
Oat	Ornithine aminotransferase prec.	Chr3L	19,641,680	19,641,704	0.61	I	Low
Snap25	Synapse protein 25	Chr3L	24,072,065	24,072,041	0.60	I	Low
bor	belphegor	Chr3R	12,088,962	12,088,986	0.60	E	High
Nep3	Neprilysin 3	ChrX	19,857,145	19,857,169	0.59	I	Low/Medium
CG7597	CG7597	Chr3L	21,510,532	21,510,556	0.59	E	Low
CG31149	CG31149	Chr3R	14,801,020	14,801,044	0.58	I	Low
fz3	frizzled 3	ChrX	668,982	669,006	0.56	I	High
CG10960	CG10960	Chr3L	12,825,643	12,825,667	0.55	I	High
E2f	E2F transcription factor	Chr3R	17,461,287	17,461,311	0.55	I	Low/High
CG6296	CG6296	Chr3R	22,830,400	22,830,424	0.54	E	Low/High
CG10632	CG10632	Chr3L	12,556,354	12,556,378	0.54	I	Low
tai	taiman	Chr2L	9,200,148	9,200,172	0.53	I	Low/High
kal-1	kal-1	Chr3R	19,971,841	19,971,865	0.53	E	Low/High
CG7722	CG7722	Chr2R	6,828,214	6,828,238	0.53	E	Low/High
CG30084	CG30084	Chr2R	11,650,486	11,650,510	0.51	E	High
Btk29A	Btk family kinase at 29A	Chr2L	8,271,429	8,271,453	0.50	I	Low
L	Lobe	Chr2R	10,374,295	10,374,319	0.50	I	Low/Medium
NepYr	Neuropeptide Y receptor-like	Chr3R	22,880,818	22,880,794	0.44*	I	Low

*Note: *NepYr* was included in the list even though its more US-like in the Caribbean because it was shown to be significantly differentiated between US, Caribbean and African populations (see text of Figure 2).

LITERATURE CITED

Albert, A. Y. K., S. Sawaya, T. H. Vines, A. K. Knecht, C. T. Miller, B. R. Summers, S. Balabhadra, D. M. Kingsley, and D. Schluter. 2008. The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. Evolution 62: 76–85.

Alipaz, J. A., S. Fang, N. Osada, and C.-I Wu. 2005. Evolution of sexual isolation in laboratory populations. Am. Nat. 165:420–428.

Allender, C.J., Seehausen, O., Knight, M.E., Turner, G.F. & Maclean, N. 2003. Divergent selection during speciation of Lake Malawi cichlid fishes inferred from parallel radiations in nuptial coloration. Proc. Natl. Acad. Sci. USA 100, 14074-14079.

Aulard, S., J. R. David, and F. Lemeunier. 2002. Chromosomal inversion polymorphism in Afrotropical populations of Drosophila melanogaster. Genet. Res. 79:49-63.

Barbujani, G. 2000. Geographic patterns: how to identify them and why. Human Biol. 72(1):133-153.

Barton, N. H. 1979. The dynamics of hybrid zones. Heredity 43:341-359

Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. Ann. Rev. Ecol. Syst. 16: 113–148.

Barton, N. H., and K. S. Gale. 1993. Genetic analysis of hybrid zones. Pp. 13-45 in Harrison, R. G., ed. Hybrid zones and the evolutionary process. Oxford Univ. Press, Oxford.

Barton, N.H., Hewitt, G. M. 1985. Analysis of Hybrid Zones. Ann. Rev. Ecol. Syst. 16:113-148.

Bastock, M. 1956. A gene mutation which changes a behavior pattern. Evolution 10: 421-39.

Bastock, M. and Manning, A. 1955. The courtship of Drosophila melanogaster. Behavior 8: 85-111.

Baudry, E., B. Viginier, and M. Veuille. 2004. Non-African populations of Drosophila melanogaster have a unique origin. Mol. Biol. Evol. 21:1482-1491.

Begun, D., and C. Aquadro. 1993. African and North American populations of Drosophila melanogaster are very different at the DNA level. Nature. 365:548-550.

Bénassi, V., and M. Veuille. 1995. Comparative population structuring of molecular and allozyme variation of Drosophila melanogaster Adh between Europe, west Africa and east Africa. Genet. Res. 65:95-103.

Billeter J. C., Goodwin, S. F., and O'Dell, K. M. C. 2002. Genes mediating sex-specific behaviors in Drosophila. Advances in Genetics, 47: 87-116.

Boake, C. R. B., K. McDonald, S. Maitra, and R. Ganguly. 2003. Forty years of solitude: life-history divergence and behavioral isolation between laboratory lines of D. melanogaster. J. Evol. Biol. 16:83-90.

Bonin, A., P. Taberlet, C. Miaud, and F. Pompanon. 2006. Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (Rana temporaria). Mol. Biol. Evol. 23:773–783.

Borevitz, JO, Liang, D, Plouffe, D, Chang, HS, Zhu, T, Weigel, D, Berry, CC, Winzeler, and E. Chory, J. 2003. Large scale identification of single feature polymorphisms in complex genomes. Genome Res. 13:513–523.

Boughman, J. W., H. D. Rundle and D. Schluter. 2005. Parallel evolution of sexual isolation in sticklebacks. Evolution 59:361–373.
Campbell, D., and L. Bernatchez. 2004. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. Mol. Biol. Evol. 21:945–956.
Capy, P. J., David, R., Allemand, Y. Carton, G., Febvay, and A. Kermarec. 1986. Genetic analysis of D. melanogaster in the French West Indies and comparison with populations from other parts of the world. Genetica 69:167-176.
Capy, P., E. Pla, and J. R. David. 1993. Phenotypic and genetic variability of morphometrical traits in natural populations of D. melanogaster and D. simulans. I. Genet. Sel. Evol. 25:517-536.
Capy, P., E. Pla, and J. R. David. 1994. Phenotypic and genetic variability of morphometrical traits in natural populations of D. melanogaster and D. simulans. II. Genet. Sel. Evol. 26:15-28.
Capy, P., M. Veuille, M. Paillette, J.-M. Jallon, J. Vouidibio, and J. R. David. 2000. Sexual isolation of genetically differentiated sympatric populations of Drosophila melanogaster in Brazzaville, Congo: the first step towards speciation? Heredity 84 (4):468–475.
Caracristi, G., and C. Schlotterer. 2003. Genetic differentiation between American and European D. melanogaster populations could be attributed to African alleles. Mol. Biol. Evol. 20:792-799.
Carson, H. L. 1965. Chromosomal morphism in geographically widespread species of Drosophila. In: Baker HG & GL Stebbins (eds) The genetics of colonizing species: 533-546. New York: Academic Press.
Carvajal-Rodríguez, A. and E. Rolán-Álvarez. 2006. JMATING: a software for the analysis of sexual selection and sexual isolation effects from mating frequency data. BMC Evol. Biol. 6:40.
Casares, P., M. C. Carracedo, B. Rio, R. Pineiro, L. Garcia-Florez, and A.R. Barros. 1998. Disentangling the effects of mating propensity and mating choice in Drosophila. Evolution. 52(1):126-133.
Castellano, S., and E. Balletto. 2002. Is the partial Mantel test inadequate? Evolution 56: 1871-1873.
Cobb, M. and J. F. Ferveur. 1996. Female mate discrimination or male responses to female stimulation? Evolution. 50:1719–1720.
Cohet, Y. and J. R. David. 1980. Geographical divergence and sexual behavior: comparisons of mating systems in French and Afrotropical populations of Drosophila melanogaster. Genetica 54:161-165.
Colegrave, N., H. Hollocher, K. Hinton, and M. G. Ritchie. 2000. The courtship song of African Drosophila melanogaster. J. Evol. Biol. 13:143-150.
Colosimo, P. F., C. L. Peichel, K. Nereng, B. K. Blackman, M. D. Shapiro, D. Schluter, and D. M. Kingsley. 2004. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLOS Biology 2: 635–641.
Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarrea, Jr., M. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley. 2005.

Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science 5717: 1928–1933.
Cook, R. 1979. The courtship tracking of Drosophila melanogaster. Biol. Cybernet. 34:91-106.
Coyne, J. A. and H. A. Orr. 2004. Speciation. Sinauer, Sunderland, MA.
Coyne, J. A., and S. Elwyn. 2006. Desaturase-2, environmental adaptation, and sexual isolation in Drosophila melanogaster. Evolution 60:626-627.
Coyne, J. A., C. Wicker-Thomas, and J.-M. Jallon. 1999. Localization of a gene responsible for a pheromonal polymorphism in Drosophila melanogaster. Genet. Res. 73:189-203.
Coyne, J. A., S. Elwyn, and E. Rolán-Álvarez. 2005. Sexual isolation between Drosophila yakuba and D. santomea: effects of environment and experimental design. Evolution 59: 2588-2601.
Coyne, J. and E. Beecham. 1987. Heritability of two morphological characters within and among natural populations of Drosophila melanogaster. Genetics. 117 (4): 727-737.
Dallerac, R., C. Labeur, J. M. Jallon, D. C. Knipple, W. L. Roelofs, C. Wicker-Thomas. 2000. A delta 9 desaturase gene with a different substrate specificity is responsible for the cuticular diene hydrocarbon polymorphism in Drosophila melanogaster. Proc Natl Acad Sci U S A. 97 (17):9449-54
Dasmahapatra, K.K., Blum, M.J., Aiello, A., Hackwell, S., Davies, N., Bermingham, E.P. & Mallet, J. 2002. Inferences from a rapidly moving hybrid zone. Evolution 56: 741-753.
David J.R. and P. Capy. 1988. Genetic variation of Drosophila melanogaster natural populations. Trends Genet. 4(4):106-11.
Dobzhansky T. 1937. Genetics and the origin of species. Columbia University Press, New York.
Dobzhansky T. 1965. "Wild" and "domestic" species of Drosophila. In: Baker HG & GL Stebbins (eds) The genetics of colonizing species: 533-546. New York: Academic Press.
Dobzhansky, T. 1973. Active dispersal and passive transport in Drosophila. Evolution. 27:565–575.
Dobzhansky, T., and Streisinger, G. 1944. Experiments on Sexual Isolation in Drosophila: II. Proc Natl Acad Sci U S A. 30(11):340–345.
Dodd, D.M.B. 1989. Reproductive isolation as a consequence of adaptive divergence in Drosophila pseudoobscura. Evolution 43:1308–1311
Doebeli, M., and U. Dieckmann. 2003. Speciation along environmental gradients. Nature 421 (6920):259-264.
Drapeau, M.D., Radovic, A., Wittkopp, P.J., Long, A.D. 2003. A gene necessary for normal male courtship, yellow, acts downstream of fruitless in the Drosophila melanogaster larval brain. J. Neurobio. 55(1):53-72.
Emelianov, I., F. Marec, and J. Mallet. 2003. Genomic evidence for divergence with gene flow in host races of the larch budmoth. Proc. R. Soc. Lond. B 71:97–105.
Endler, J. A. 1977. Geographic variation, speciation, and clines. Princeton Univ. Press. Princeton, NJ.

Endler, J. A., and A. E. Houde. 1995. Geographic variation in female preferences for male traits in Poecilia reticulata. Evolution 49:456-468.
Epperson, B. K. 2003. Geographical Genetics. Princeton Univ. Press.
Ewing, A. W. 1961. Body size and courtship behavior in Drosophila melanogaster. Anim. Behav. 9:93-99.
Ewing. A. W. 1964. The influence of wing area on the courtship behavior of Drosophila melanogaster. Anim. Behav. 12:316-320.
Fang, S., A. Takahashi and C. I. Wu. 2002. A mutation in the promoter of Desaturase-2 is correlated with sexual isolation between Drosophila behavioral races. Genetics 162:781-784.
Funk, D.J. 1998. Isolating a role for natural selection in speciation: host adaptation and sexual isolation in Neochlamisus bebbianae leaf beetles. Evolution 52: 1744-1759.
Gavrilets, S. 2004. Fitness landscapes and the origin of species. Princeton Univ. Press, Princeton, NJ.
Greenberg, A.J., J.R. Moran, J.A. Coyne, and C.I. Wu. 2003. Ecological adaptation during incipient speciation revealed by precise gene replacement. Science. 302(5651):1754 – 1757.
Greenberg, A. J., J. R. Moran, S. Fang, and C.-I Wu. 2006. Adaptive loss of an old duplicated gene during incipient speciation. Mol. Biol. Evol. 23:401-410.
Greenspan, R.J., and Ferveur, J.F. 2000. Courtship in Drosophila. Ann. Rev. Genet. 34: 205--232.
Gresham, D., Ruderfer D, Pratt, S, Schacherer J, Dunham M, Botstein D, Kruglyak L, et al. 2006. Genome-wide detection of polymorphisms at nucleotide resolution with a single DNA microarray. Science 311(5769): 1932-1936.
Grillet, M., L. Dartevelle, and J.-F. Ferveur. 2006. A Drosophila male pheromone affects female sexual receptivity. Proc Biol Sci. 273(1584):315–323.
Grillet, M., L. Dartevelle, and J.-F. Ferveur. 2006. A Drosophila male pheromone affects female sexual receptivity. Proc Biol Sci. 273(1584): 315–323.
Haerty, W., J. M. Jallon, J. Rouault, C. Bazin, and P. Capy. 2002. Reproductive isolation in natural populations of Drosophila melanogaster from Brazzaville (Congo). Genetica. 116 (2-3):215-24.
Haerty, W., M. Lesbats, and P. Capy. 2005. Pre-reproductive isolation as a consequence of allopatric differentiation between populations of Drosophila melanogaster. Mol Ecol. 14(12):3801-7.
Hawthorne, D. J., and S.Via. 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. Nature 412:904–907.
Hedrick, P. W. 2000. Genetics of Populations. Jones and Bartlett Publ. Sudbury. Mass.
Hedrick, P.W. 2006. Genetic polymorphism in heterogeneous environments: The age of genomics. Annu. Rev. Ecol. Syst. 37:67–93.
Henderson, N. R. and D. M. Lambert. 1982. No significant deviation from random mating of worldwide populations of Drosophila melanogaster. Nature 300:437-440.

Hendry, A. P. 2001. Adaptive divergence and the evolution of reproductive isolation in the wild: an empirical demonstration using introduced sockeye salmon. Genetica 112-113:515-534.
Hendry, A.P., P. Nosil, and L.H. Rieseberg. 2007. The speed of ecological speciation. Functional Ecology 21:455-464
Hollocher, H., C.-T. Ting, F. Pollack, and C. I. Wu. 1997a. Incipient speciation by sexual isolation in Drosophila melanogaster: variation in mating preference and correlation between sexes. Evolution 51(4):1175-1181.
Hollocher, H., C. T. Ting, M. L. Wu, and C. I. Wu. 1997b. Incipient speciation by sexual isolation in Drosophila melanogaster: extensive genetic divergence without reinforcement. Genetics. 147: 1191-1201.
Howard, D. J. 1993. Reinforcement: origin, dynamics, and fate of an evolutionary hypothesis, in Hybrid zones and the evolutionary process (Harrison, R. G., ed.), pp.46-69, New York: Oxford University Press.
Jiggins, C.D., Mallet, J. 2000. Bimodal hybrid zones and speciation. Trends in Ecological Evolution. 15(6):250-255.
Jiggins, C. D., R. E. Naisbit, R. L. Coe, and J. Mallet. 2001. Reproductive isolation caused by colour pattern mimicry. Nature 411:302-305.
Jiggins, C. D., C. Estrada, and A. Rodrigues. 2004. Mimicry and the evolution of premating isolation in Heliconius melpomene Linnaeus J. Evol. Biol. 17:680-691.
Johannesson, K., Rol´an-Alvarez, E. and A. Ekendahl. 1995. Incipient reproductive isolation between two sympatric morphs of the intertidal snail Littorina saxatilis. Evolution 49: 1180-1190.
Kauer, M., B. Zangerl, D. Dieringer, and C. Schlötterer. 2002. Chromosomal patterns of microsatellite variability contrast sharply in African and non-African populations of Drosophila melanogaster. Genetics 160:247-256.
Kauer, M., and C. Schlötterer. 2004. An analysis of genetic differentiation among assortatively mating Drosophila melanogaster in Zimbabwe. J. Evol. Biol. 17: 493-500.
Keller, A. 2007. Drosophila melanogaster's history as a human commensal. Current Biology. 17:R77-R81.
Kilias, G., Alahiotus, S.N. and Pelecanos, M. 1980. A multifactorial genetic investigation of speciation theory using Drosophila melanogaster. Evolution 34:730–737
Koref-Santibáñez, S. 1972. Courtship behavior in the semispecies of the superspecies Drosophila paulistorura. Evolution. 26;108-15.
Lachaise, D., Cariou, M.L., David, J.R., Lemeunier, F., Tsacas, L. and M. Ashburner. 1988. Historical biogeography of the Drosophila melanogaster species subgroup. J. Evol. Biol. 22:159–225.
Lachaise, D., and J. F. Silvain. 2004. How two Afrotropical endemics made two cosmopolitan human commensals: the Drosophila melanogaster-D. simulans palaeogeographic riddle. Genetica. 120:17-39.
Lande, R. 1982. Rapid origin of sexual isolation and character divergence in a cline. Evolution 36:213-223.
Mallet, J. & Barton, N.H. 1989. Strong natural selection in a warning-color hybrid zone. Evolution 43:421-431.

Mallet, J. 2005. Hybridization as an invasion of the genome. Trends in Ecology and Evolution 20 (5): 229-237.
Mallet, J., Barton, N., Lamas M., G., Santisteban C., J., Muedas M., M. & Eeley, H. 1990. Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in Heliconius hybrid zones. Genetics 124:921-936.
Mallet, J., Beltrán, M., Neukirchen, W., & Linares, M. 2007. Natural hybridization in heliconiine butterflies: the species boundary as a continuum. BMC Evolutionary Biology 7:28.
Malogolowkin-Cohen, C., Simmons, A. S., and H., Levene. 1965. A study of sexual isolation between certain strains of Drosophila paulistorum. Evolution 19:95-103.
Mayr, E. 1942. Systematics and the origin of species. Columbia Univ. Press, New York.
Mayr, E. 1963. Animal species and evolution. Harvard University Press. Cambridge, MA.
Mayr, E. 1965. The nature of colonizations in birds. In The Genetics of Colonizing Species (H.G.Baker and G. L. Stebbins, eds.), pp. 29–47. New York: Academic Press.
McDonald, J. 1979. Genetic analysis of lines selected for wing vibration in Drosophila melanogaster. Behav. Genet. 9:579-584.
McKinnon, J. S., S. Mori, B. Blackman, L. David, D. M. Kingsley, L. Jamieson, J. Chou, and D. Schluter. 2004. Evidence for ecology's role in speciation. Nature 429: 294-298.
Merrel, D. J. 1950. Measurements of sexual isolation and selective mating. Evolution 4:326-331.
Murray, M. C., and M. P. Hare. 2006. Agenomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, Crassostrea virginica. Mol. Ecol. 15:4229–4242.
Noor, M. A. 1995. Speciation driven by natural selection in Drosophila. Nature. 375(6533):674–675
Noor, M. A. F., K. L. Grams, L. A. Bertucci, and J. Reiland. 2001. Chromosomal inversions and the reproductive isolation of species. Proc. Natl. Acad. Sci. USA 98:12084–12088.
Noor, M. A. F., and J. L. Feder. 2006. Speciation genetics: evolving approaches. Nat. Rev. Genet. 7:851–861.
Nosil P, Crespi B.J., and C. P. Sandoval. 2002. Host-plant adaptation drives the parallel evolution of reproductive isolation. Nature 417: 441-443.
Nosil P, Crespi, B.J., Gries, R., and G. Gries. 2007. Natural selection and divergence in mate preference during speciation. Genetica 128: 309-327.
Nosil, P. 2005. Divergent host plant adaptation and reproductive isolation between ecotypes of Timema cristinae walking sticks. Am Nat. 169(2):151-62.
Nosil, P., S. P. Egan and D. J. Funk. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: 'isolation by adaptation' and multiple roles for divergent selection. Evolution 62: 316–336.
O'Dell, K. M. C. 2003. The voyeurs' guide to Drosophila melanogaster courtship. Behavioral Processes 64:211-223.

Ortiz-Barrientos, D., and M. F. Noor. 2005. Evidence for a one-allele assortative mating locus. Science, 310: 1467.
Panhuis, T. M., R. Butlin, M. Zuk, and T. Tregenza. 2001. Sexual selection and speciation. TREE 16:364-371.
Parsons, P. A. 1983. The Evolutionary Biology of Colonizing Species. Cambridge Univ. Press. NY NY.
Partridge, L., Ewing, A. and A. Chandler. 1987. Male size and mating success in Drosophila melanogaster: the roles of male and female behavior. Anim. Behav. 35:555-562.
Pfennig, K. S. and M. A. Simovich. 2002. Differential selection to avoid hybridization in two toad species. Evolution 56:1840-1848.
Pitnick, S. 1991. Male size influences mate fecundity and remating interval in Drosophila melanogaster. Anim. Behav. 41:735-745.
Pool, J. E. and C. F. Aquadro. 2006. History and Structure of Sub-Saharan Populations of Drosophila melanogaster Genetics 174(2):915 - 929.
Rohlf, J. F. 2004. NTSYSpc. 2.20. Exeter Software © by Applied Biostatistics, Inc.
Rolán-Álvarez, E. and A. Caballero. 2000. Estimating sexual selection and sexual isolation effects from mating frequencies. Evolution 54:30-36.
Rosenberg, M.S. 2004. PASSAGE. Pattern Analysis, Spatial Statistics, and Geographic Exegesis, Version 1.0. Tempe: Dept. of Biology, Arizona State University.
Rouault, J., P. Capy and J. M. Jallon. 2001. Variations of male cuticular hydrocarbons with geoclimatic variables: an adaptative mechanism in Drosophila melanogaster? Genetica 110:117-130.
Rousset, F. 2002. Partial Mantel tests: Reply to Castellano and Balletto. Evolution 56: 1874-1875.
Rundle, H. D., L. Nagel, J. W. Boughman, and D. Schluter. 2000. Natural selection and parallel speciation in sympatric sticklebacks. Science 287: 306–308.
Rundle, H.D. 2003. Divergent environments and population bottlenecks fail to generate premating isolation in Drosophila pseudoobscura. Evolution 57: 2557-2565.
Rundle, H.D. and Schluter, D. 2004. Natural Selection and Ecological Speciation in Sticklebacks. In: Adaptive Speciation, eds. Dieckmann U, Doebeli M, Metz JAJ & Tautz D, pp. 192–209. Cambridge University Press. International Institute for Applied Systems Analysis.
Schlötterer, C., Vogl ,C., and D. Tautz. 1997. Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural Drosophila melanogaster populations. Genetics. 146:309-320.
Schlötterer, C., H. Neumeier, C. Sousa and V. Nolte. 2006. Highly structured Asian Drosophila melanogaster populations: A new tool for hitchhiking mapping? Genetics. 172 (1):287–292.
Schluter, D. and L. Nagel. 1995. Parallel speciation by natural selection. American Naturalist 146: 292–301.
Schluter, D., E. A. Clifford, M. Nemethy, and J. S. McKinnon. 2004. Parallel evolution and inheritance of quantitative traits. American Naturalist 163: 809–822.
Scott, D. 1994. Genetic variation for female mate discrimination in Drosophila melanogaster. Evolution. 48:112–121.

Smouse, P.E., J.C. Long, and R.R. Sokal. 1986. Multiple regression and correlation extensions of the Mentel Test of matrix correspondence. Syst. Zool. 35:627-632.
Sokal, R. R. 1979a. Ecological parameters inferred from spatial correlograms. In G. P. Patil and M. L. Rosenzweig, eds., Contemporary Quantitative Ecology and Related Econometrics. 167-196. Fairland, MD: International Cooperative Publishing House.
Sokal, R.R. 1979b. Testing statistical significance of geographic variation patterns. Syst. Zool. 28(2):227-232.
Sokal, R. R. and N. L. Oden. 1978a. Spatial autocorrelation in biology. 1. Methodology. Biol. J. Linnean Soc. 10: 199-228.
Sokal, R.R., and N.L. Oden. 1978b. Spatial autocorrelation in biology. 2. Some biological implications and four applications of evolutionary and ecological interest. Biol. J. Linn. Soc. 10:229-249.
Sokal, R., Oden, N.L., and J. S. F. Barker. 1987. Spatial structure in Drosophhila buzzatii populations, simple and directional spatial autocorrelation. Am. Nat. 129:122-142.
Sokal, R. R. and F. J. Rohlf. 1995. Biometry. W. H. Freeman and Company.
Spieth, H.T. 1952. Mating behaviour within the genus Drosophila (Diptera). Bull Am Mus Nat Hist 99: 401-474.
Stinchcombe, J. R., and H. E. Hoekstra. 2007. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity. In press.
Storey JD and Tibshirani R. 2003. Statistical significance for genome-wide studies. Proc. Natl. Acad. Sci. USA. 100: 9440-9445
Takahashi, A., S. C. Tsaur, J. A. Coyne, and C. I. Wu. 2001. The nucleotide changes governing cuticular hydrocarbon variation and their evolution in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA 98:3920-3925.
Taylor, C. E., and V. Kekic. 1988. Sexual Selection in a Natural Population of Drosophila melanogaster. Evolution, 42 (1):197-199.
Taylor, E.B. and J. D. McPhail. 1999. Evolutionary history of an adaptive radiation in species pairs of threespine sticklebacks (Gasterosteus). Biol J Linn Soc 66:271–291.
Tilley, S. G., P. A. Verrell, and S. J. Arnold. 1990. Correspondence between sexual isolation and allozyme differentiation: a test in the salamander Desmognathus ochrophaeus. Proc. Natl. Acad. Sci. USA 87:2715-2719.
Ting, C. T., A. Takahashi and C.-I Wu. 2001. Incipient speciation by sexual selection: concurrent evolution at multiple loci. Proc. Natl. Acad. Sci. 98: 6709-6713.
Tompkins, L., Gross, A.C., Hall, J.C., Gailey, D.A., and R. W. Siegel. 1982. The role of female movement in the sexual behavior of Drosophila melanogaster. Behav. Genet. 12(3): 295-307.
Tregenza, T. 2002. Divergence and reproductive isolation in the early stages of speciation. Genetica 116:291-300.
Tregenza, T., V. L. Pritchard, and R. K. Butlin. 2000. The origins of premating reproductive isolation: testing hypotheses in the grasshopper Chorthippus parallelus. Evolution 54:1687-1698.

True, J.R. J. Liu, L.F. Stam, Z.B. Zeng, and C.C. Laurie 1997 Quantitative genetic analysis of divergence in male secondary sexual traits between Drosophila simulans and Drosophila mauritiana. Evolution 51:816-832.

Turner, T. L., M. W. Hahn, S. V. Nuzhdin. 2005. Genomic islands of speciation in Anopheles gambiae. PloS. Biol. 3(9):1572-1574.

Turner, T. L., M. T. Levine, and D. J. Begun. 2008 Genomic analysis of adaptive differentiation in Drosophila melanogaster. Genetics. 179:455-473.

Verrell, P. A. and S. J. Arnold. 1989. Behavioral observations of sexual isolation among allopatric populations of the mountain dusky salamander Desmognathus ochrophaeus. Evolution 43:745-755.

Via, S. 2001. Sympatric speciation in animals: the ugly duckling grows up. TREE. 16:381-390.

Vines, T. H. and D. Schluter. 2006. Strong assortative mating between allopatric sticklebacks as a by-product of adaptation to different environments. Proceedings of the Royal Society of London Series B, Biological Sciences 273: 911–916.

Vines, T. H., S. C. K¨ohler, M. Thiel, I. Ghira, T. R. Sands, C. J. MacCallum, N. H. Barton, and B. N¨urnberger. 2003. The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads Bombina bombina and B. variegata. Evolution 57:1876–1888.

Waddle, D. M., R. R. Sokal, and P. Rudan. 1998. Factors affecting population variation in Eastern Adriatic Isolates (Croatia). Human Biology. 70(5):845-864.

Welbergen, P. van Dijken F.R. and Scharloo W. 1987. Collation of the courtship behaviour of the sympatric species Drosophila melanogaster and D. simulans. Behaviour 101: 253-274.

Winzeler, EA, Richards, DR, Conway, AR, Goldstein, AL, Kalman, S, McCullough, McCusker, M. J, J.H, Stevens, D.A, Wodicka, L and Lockhart, DJ, et al. 1998. Direct allelic variation scanning of the yeast genome. Science 281:1194–1197.

Wu, C. I., H. Hollocher, D.J. Begun, C.F. Aquadro, Y. Xu, and M. L. Wu. 1995. Sexual isolation in D. melanogaster: A possible case of incipient speciation. Proc. Natl. Acad. Sci. USA 92:2519-2523.

Wu, C.-I. 2001. The genic view of the process of speciation. J. Evol. Biol. 14: 851-865.

Yatabe, Y., N. C. Kane, C. Scotti-Saintagne, and L. H. Rieseberg. 2007. Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, Helianthus annuus and H. petiolaris. Genetics 175:1883–1893.

Yukilevich, R. and J.R. True. 2008a. Incipient sexual isolation among cosmopolitan Drosophila melanogaster. Evolution. doi:10.1111/j.1558-5646.2008.00427.x

Yukilevich, R. and J. R. True. 2008b. African phenotypes and incipient sexual isolation between US and Caribbean Drosophila melanogaster. Evolution.

Yukilevich, R., T. L. Turner, J. R. True, and S. Nuzhdin. ms. Geographical genomics in North American and African Drosophila melanogaster.

Zhang L, Miles MF, Aldape KD. 2003. A model of molecular interactions on short oligonucleotide microarrays. Nature Biotechnology 21(7):818-821.