

Copyright

by

Rafael Felipe Guerrero

2013

**The Dissertation Committee for Rafael Felipe Guerrero Certifies that this is the
approved version of the following dissertation:**

Models and Analyses of Chromosome Evolution

Committee:

Mark A. Kirkpatrick, Supervisor

Thomas E. Juenger

Claus O. Wilke

James J. Bull

Sara L. Sawyer

Models and Analyses of Chromosome Evolution

by

Rafael Felipe Guerrero, B.S, M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2013

Dedication

A Nubia Farias García, Rafael Guerrero Lozano, y María Lucía Guerrero Farías, porque
uno nunca sabe si esta es la última vez que los hago sentir orgullosos.

“Mamá, estoy triunfando!”

Noel Petro, el burro mocho

Acknowledgements

This work was vastly improved by the input and guidance of my supervisor, Mark Kirkpatrick. Mark has been a great mentor and personal friend. This work would have been impossible without my collaborators Nicolas Perrin, François Rousset, and Diego Ayala. My committee, Tom Juenger, Sara Sawyer, Jim Bill and Claus Wilke, made the process impossibly painless, being supportive and trusting at every step.

Robin Hopkins and Sam Scarpino, aside from becoming the closest of friends, contributed to this work by reading drafts, enduring practice talks, and taking part in countless discussions, rants and coffee breaks. Work in the sixth floor of Patterson has been a pleasure thanks to everyone in the lab: Will Stutz, Lisa Snowberg, Kate Behrman, Stephan Peischl, as well as illustrious visitors such as Sally Otto, Mike Whitlock and Claudia Bank.

Life in Austin has been spectacular, and I can't imagine having to go through graduate school anywhere else. Here's an incomplete list of people who massively contributed to my happiness at some point: Dave Des Marais, Mike Singer, Barrett Klein, Carlos Guarnizo, Sofía Rodríguez, Alejandro Puyana, Manuel Chávez, Karin Akre, Emily McTavish, Genevieve Smith, Eben Gering, Pam Willis, Thomas Keller, Jeremy Brown, Will Harcombe and María Jose La Rota. I was very lucky to be part of the awfully-named Competitive Excluders and Cliffdancers groups; soccer and climbing kept me sane.

Thank you, Mendy Black.

Lastly, I thank the reader, because if you're reading this text you're either my friend or family, or someone that needs to check how to properly format a dissertation and considers me a good source.

Models and Analyses of Chromosome Evolution

Rafael Felipe Guerrero, Ph.D.

The University of Texas at Austin, 2013

Supervisor: Mark Kirkpatrick

At the core of evolutionary biology stands the study of divergence between populations and the formation of new species. This dissertation applies a diverse array of theoretical and statistical approaches to study how chromosomes evolve. In the first chapter, I build models that predict the amount of neutral genetic variation in chromosomal inversions involved in local adaptation, providing a foundation for future studies on the role of these rearrangements in population divergence. In the second chapter, I use a large dataset of the geographic variation in frequency of a chromosomal inversion to infer natural selection and non-random mating, revealing that this inversion could be implicated in strong reproductive isolation between subpopulations of a single species. In the third chapter, I use coalescent models for recombining sex chromosomes coupled with approximate Bayesian computation to estimate the recombination rate between X and Y chromosomes in European tree frogs. This novel approach allows me to infer a rate so low that would have been hard to detect with empirical methods. In the fourth chapter, I study the theoretical conditions that favor the evolution of a chromosome fusion that reduces recombination between locally adapted alleles.

Table of Contents

Chapter 1: Coalescence Patterns for Chromosomal Inversions in Divergent Populations	
Introduction	1
Models and Results	5
Old inversions: Locally Adapted Breakpoints	8
Old inversions: Locally adapted alleles	10
Young inversions: Drift	12
Young inversions: Locally adapted breakpoints and alleles	13
Discussion	14
Chapter 2: Reproductive Isolation And Local Adaptation Quantified For A Chromosome Inversion In A Malaria Mosquito	
Introduction	19
Methods.....	21
Results.....	26
Discussion	32
Chapter 3: Cryptic recombination in the ever-young sex chromosomes of Hyliid frogs	
Introduction	43
Methods.....	46
Results	50
Discussion	52
Chapter 4: The Evolution of Chromosome Fusions by Local Adaptation	
Introduction	56
Models and Results	59
Continent-Island model.....	60
Two-deme model	62
Discussion	64

Tables	69
Figures.....	71
Bibliography	97
Vita	110

Chapter 1: Coalescence Patterns for Chromosomal Inversions in Divergent Populations*

INTRODUCTION

Chromosomal inversions have been important to evolutionary biology for decades (reviewed in (Hoffmann & Rieseberg, 2008a; Kirkpatrick, 2010)). As genetic markers, inversions have served the study of balanced polymorphisms (Dobzhansky, 1951), geographic clines (White, 1973), and meiotic drive (Sandler & Hiraizumi, 1960). Perhaps the most important property of chromosomal inversions is their potential to create reproductive isolation. The observation of polymorphisms within species and fixed differences between closely related species led to the suggestion that chromosomal rearrangements play a causative role in speciation (Dobzhansky, 1951; White, 1973). Heterokaryotypic individuals (chromosomal heterozygotes) of some species have problems in meiosis that could generate a reproductive barrier between populations fixed for different arrangements. Based on this feature, White (White, 1969; 1978) proposed his “stasipatric” mode of speciation in which inversions drive reproductive isolation. That hypothesis, however, relies on strong assumptions about population structure, genetic drift, and/or alternative forces (such as meiotic drive) to account for the spread of underdominant inversions (Barton & Charlesworth, 1984; Coyne & Orr, 2004). Moreover, not all inversions are structurally underdominant (Coyne *et al.*, 1993), rendering them useless as reproductive barriers.

Alternatively, inversions can facilitate speciation because they are potent recombination modifiers (Noor *et al.*, 2001; Rieseberg, 2001; Ortiz-Barrientos *et al.*,

* Significant portions of this chapter have been previously published as Guerrero, Rousset & Kirkpatrick (2012) *Philosophical Transactions of the Royal Society B*. 367:430-438. Author contributions: Francois Rousset developed a complementary method to check my results (not included in this dissertation); Mark Kirkpatrick, supervisor.

2002). Recombination in heterokaryotypes is severely reduced, causing associations between sets of alleles inside of the inversion (Sturtevant & Beadle, 1936; Navarro & Barton, 2003). For this reason, alternative chromosomal arrangements can play a key role in facilitating speciation. When alternative arrangements have been established in diverging populations, they may protect the inverted region from introgression and allow the accumulation of alleles that contribute to reproductive isolation (Noor *et al.*, 2001; Rieseberg, 2001; Feder & Nosil, 2009). Their effects on recombination may also be key to how inversions become established in the first place. Dobzhansky (Dobzhansky, 1951) suggested that inversions evolve because they reduce recombination between genes with epistatic interactions. Even without epistasis, local adaptation can favor reduced recombination between alleles adapted to the same habitat or genetic background (Charlesworth & Charlesworth, 1979; Kirkpatrick & Barton, 2006). Migration or hybridization is essential in this process: inverted chromosomes gain a fitness advantage because they keep locally adapted alleles together. By reducing recombination, inversions preserve the divergence between populations in the presence of gene flow, which can set the stage for speciation. The hypothesis that reduced recombination in inversions plays a role in speciation is consistent with some lines of empirical evidence (Brown *et al.*, 2004; Livingstone & Rieseberg, 2004; Noor *et al.*, 2007; Yatabe *et al.*, 2007).

Intuition suggests that different mechanisms that might drive the evolution of inversions will leave different signatures in the DNA. For example, if an inversion spread because it carried favorable alleles, two key patterns might be expected. First, neutral genetic diversity within chromosomal arrangements is likely to be low when compared to the divergence between them. Second, marked peaks of divergence might be found around the breakpoints and at sites close to the selected loci. Patterns consistent with these predictions have been found in *Drosophila pseudoobscura* (Schaeffer *et al.*, 2003)

and *Anopheles gambiae* (White *et al.*, 2007). Evidence for selection on loci inside the inversion has also been inferred from strong linkage disequilibria between the inversion and genetic markers putatively linked to selected genes (*e.g.* in *Drosophila melanogaster* (Kennington *et al.*, 2006)).

Selection could also target inversions directly, rather than alleles that they carry. The breakpoints of an inversion alter the DNA sequence and may be targets of selection by changing reading frames or expression patterns (Matzkin *et al.*, 2005; Tsuchimatsu *et al.*, 2010). It is expected that direct selection on the breakpoints will produce divergence between arrangements but no additional peaks of divergence inside of the rearranged region. This is the pattern seen in the inversion polymorphism O_{3+4}/O_{ST} in *D. subobscura*, which has strong geographic clines suggestive of selection and high divergence between arrangements throughout the length of the rearrangement (Munté *et al.*, 2005). In other cases, however, no apparent signatures of selection have been found (*e.g.* in *Anopheles funestus* (Cohuet *et al.*, 2004) and *A. gambiae* (White *et al.*, 2009)). These previous observations suggest that patterns can be found in linked neutral genetic data, but also highlight the fact that there are no quantitative predictions for those patterns that would enable tests of alternative hypotheses.

When are inversions expected to hold neutral divergence between chromosome arrangements? Reduced recombination in heterokaryotypes decreases the rate at which a gene lying within an inversion moves onto a standard chromosome, and vice versa. This gene flux between chromosome arrangements (a result of double recombination events and gene conversion (Navarro *et al.*, 1997; Schaeffer & Anderson, 2005)) has genetic consequences that are similar to migration between populations. Consequently, inversions are expected to show some of the same patterns of neutral genetic diversity seen in subdivided populations. Using a coalescent approach, Navarro *et al.* (Navarro *et*

al., 2000) found that an inversion maintained by selection as a balanced polymorphism in one population will show reduced diversity for a substantial period ($< N$ generations) after it becomes established. Divergence between the two arrangements is also expected to accumulate with time, especially near the breakpoints where the gene flux rate is very strongly reduced. Although this result suggests that inversions under direct divergent selection would show similar patterns, no expectations have been obtained for inversions that carry loci under divergent selection. I also lack predictions for a basic null model of neutral inversions.

Motivated by these gaps in the theory, and with the aim of developing expectations for the patterns observed in the data, here I develop coalescent models of chromosomal inversions. I consider two populations that exchange migrants or hybridize; these could be diverging populations of a single species or two hybridizing species (in either primary or secondary contact). I assume that this genetic exchange has been going on for a long period. These results therefore do not apply to cases of recent secondary contact, although the models could be adapted to study that situation.

I consider three alternative scenarios. The first I refer to as *locally adapted breakpoints*. Here the chromosomal lesion caused by an inversion is under selection with opposing direction in two populations that exchange migrants. Local adaptation might result from different environmental conditions experienced by the two populations or hybridizing species. Alternatively, it could result from interactions with genetic differences between the populations at other parts of the genome. The second model is of *locally adapted alleles* segregating at loci within an inversion. Motivated by the model of Kirkpatrick & Barton (Kirkpatrick & Barton, 2006), this scenario considers an inversion that has spread in one of two populations because it captures two alleles that are locally adapted. Both of these first two models are relevant to discussions about the roles the

inversions play in genetic isolation between species. The third model is of a selectively *neutral inversion* that has spread by random genetic drift. This model provides an appropriate null model for comparison with the first two models. Perhaps surprisingly, I find that drift can generate patterns that resemble those resulting from the first two models.

My aim is to provide intuition for the patterns of neutral diversity around inversions in divergent populations. I therefore focus on the expected coalescence time at a selectively neutral site for a pair chromosomes; this quantity is proportional to the expected neutral genetic diversity or divergence (Wakeley, 2009).

MODELS AND RESULTS

I am interested in the coalescent patterns of neutral sites (*e.g.* a nucleotide or a microsatellite) in a chromosome region that is polymorphic for an inversion. There are two chromosome arrangements that are referred to as standard (*S*) and inverted (*I*). In homokaryotypes, the rate of recombination between the breakpoints (that is, the map length of the inversion) is r . Gene flux, denoted ϕ , is defined as probability that a gene in a heterokaryotype recombines from a standard chromosome to an inversion or vice versa (by double recombination or gene conversion). Consider two populations of equal and constant size N . Mating within each population is random, and deviations from Hardy-Weinberg equilibrium are assumed negligible. Migration (or hybridization) between the populations occurs at rate m , and that rate has been constant for much longer than N generations. Selection favors arrangement *S* in population 1, and arrangement *I* in population 2. The polymorphism is maintained such that the rare (disfavored) arrangement is at frequency q in each population. Begin by assuming the inversion is infinitely old, so patterns of neutral genetic variation reflect a migration-selection-

recombination equilibrium. This assumption gives good approximations for inversions much older than N generations. In a later section I consider inversions that have invaded more recently.

I focus on obtaining the expected coalescent time (\bar{T} , in generations) for genes sampled from two chromosomes (which may or may not have the same arrangement). Moving backwards in time, three types of events can occur. First, a gene can move from one population to the other as the result of migration (hybridization). Because selection maintains differences in karyotype frequencies between populations, migration is not conservative (i.e. the karyotype frequencies of migrants into and out of a population are not equal (Nagylaki, 1980)). Second, a gene can recombine from one genetic background to another. Genes move between standard and inverted chromosomes (in heterokaryotypes), and also between chromosomes with the same arrangement but different selected alleles (in homokaryotypes). Third, coalescence occurs if two genes shared a common ancestor in the previous generation. This event can only occur if the genes are present in the same population and are carried on the same chromosomal arrangement. In my model of locally adapted alleles, coalescence further requires that the genes share the same genetic background of selected alleles. With the exception of the case of a neutral inversion, the models produce results for arbitrary values of N . In the cases where results for specific values of N are presented, it is done so solely for ease of presentation and comparison to empirical data.

I analyzed the models using two approaches. First, I derived analytical expressions for the expected coalescence times based on the structured coalescent (reviewed in (Wakeley, 2009), Ch.5). The calculations are complicated, so I developed programs in *Mathematica* (Wolfram Research, 2008) that perform the algebra. The details are given in Appendix A, and the *Mathematica* code is available on request from

the authors. Second, I developed stochastic simulations of coalescent processes. The simulations generate realizations of the coalescent process for sites linked to an inversion using algorithms similar to previous studies (Kaplan *et al.*, 1989; Kim & Stephan, 2002). In addition to verifying results, these simulations allow us to study the cases in which the inversion appeared recently. Simulations were implemented in C++, and the code is available on request.

The rate of gene flux between standard and inverted chromosomes is lowest near the breakpoints and increases towards the midpoint of the inversion (Navarro *et al.*, 1997). To model this effect, assume that ϕ declines linearly from a maximum at the midpoint of the inversion to 0 at the breakpoints. More complex models of gene flux (*e.g.* (Navarro *et al.*, 1997)) allow for a non-linear decline of ϕ . While those alternative assumptions will affect the quantitative patterns of coalescence times along the chromosome, the qualitative conclusions are insensitive to those details. Recombination rates in regions flanking an inversion are also reduced in heterokaryotypes (*e.g.* (Machado *et al.*, 2007)), but the scale and pattern of this effect are not well characterized. For simplicity, assume recombination is unaffected outside of the rearranged region, and include results for those regions to offer a contrast to the patterns inside of the inversion.

I now describe results for three situations. The first two are for locally adapted breakpoints and locally adapted alleles for old inversion polymorphisms. The third situation is of a selectively neutral inversion that is drifting through a single population. Neutral inversions that are still polymorphic will typically be young. I therefore end by revisiting the cases of locally adapted breakpoints and locally adapted alleles when selection has established the inversion polymorphism recently, and compare those results with the case of neutral drift.

Old inversions: Locally Adapted Breakpoints

In the first model, the inversion polymorphism is maintained as a result of local adaptation of the breakpoints. Arrangement S is favored by selection in Population 1. The viabilities of IS and II are $(1 - hs)$ and $(1 - s)$ relative to SS (set to unity). In Population 2, the I arrangement is favored with symmetric selection coefficients. Selection affects the results only through q , which I calculate numerically. Additional details are described in Appendix A. The analytical results are impractical to show here, so I present numerical evaluations for specific parameter values.

As a general result, the effect of the inversion on \bar{T} is smaller for larger values of N , q and ϕ . Divergence between arrangements is particularly high when the product of $Nq\phi < 1$. This result is consistent with previous analytical approximations for local adaptation by Nordborg (Eq.47 in (Nordborg, 1997)). Here I illustrate scenarios and compare my results to existing empirical observations using a small set of plausible parameter values. Different results will be obtained using other parameter values, of course, and my methods can be adapted to study those situations.

The left half of Figure 1.1 shows results for a large inversion of size $r = 10\text{cM}$ when each population is of size $N = 10^5$. Gene flux is $\phi = 10^{-5}$ at the inversion midpoint, consistent with data from *Drosophila melanogaster* (Payne, 1924; Chovnick, 1973), but other estimates vary from $\phi = 10^{-2}$ to 10^{-8} (Navarro *et al.*, 1997). For the case shown by the black curves, the migration (hybridization) rate is $m = 0.001$, and the selection parameters are $s = 0.02$ and $h = 0.5$. At selection-migration equilibrium, the frequency of the locally rare arrangement is $q = 0.09$. I show expected coalescence times of three types of samples: both genes are sampled from inverted chromosomes (\bar{T}_{II}), both from standard (\bar{T}_{SS}), and one from standard and the other from inverted (\bar{T}_{IS}). All samples are random with respect to population. The grey line shows \bar{T}_{IS} for the case where $s = 0.05$ leaving

other parameters unchanged, so that $q = 0.02$. The results from the two approaches are in good agreement (Figure 1.1).

Basic coalescence theory for structured populations (Kaplan *et al.*, 1991; Nordborg, 1997; Wakeley, 2009) shows that, in the absence of the inversion, \bar{T} for genes sampled from the same population is $4N = 4 \times 10^5$ generations for the case shown, and that between populations \bar{T} converges to this value as the migration increases. When the product Nm is much greater than 1, little population structure is expected (Wright, 1951). In the presence of an inversion, Figure 1.1 shows that the expected coalescence time for pairs of genes sampled from the same arrangement (from either population) is near to the neutral expectation from the structured coalescent (Wakeley, 2009) in the absence of an inversion. This is because although the migration rate is small ($m = 0.001$), Nm is large and the high exchange of migrants prevents the divergence between populations.

The expected amount of neutral polymorphism within each arrangement is very similar inside and outside of inverted regions when the inversions are old (Figure 1.1). When one gene is sampled from an inverted and the other from a standard chromosome, however, \bar{T}_{IS} is substantially longer. The difference grows moving closer to the breakpoints. Because of the simplifying assumption that the inversion is infinitely old, \bar{T}_{IS} becomes infinite at the breakpoints. Stronger selection also increases \bar{T}_{IS} by reducing q , hence reducing the opportunity for gene flux. At the center of the inversion, for these sets of parameters \bar{T}_{IS} is several times larger than the expectation under the standard neutral model. The implication is that I expect substantially greater divergence between arrangements than diversity within arrangements near the breakpoints, even for very old inversions.

Figure 1.2 shows the effect of migration rate on coalescence times within and between chromosomal arrangements. I quantify these effects using two statistics. The

first is related to the amount of neutral diversity expected between these two populations: $F_{ST} = 1 - (\bar{T}_S / \bar{T}_T)$, where \bar{T}_T is the expected coalescence time for a pair of genes sampled from the total (combined) population and \bar{T}_S is for a pair sampled from the same subpopulation (see eq. 22 in (Slatkin, 1991)). A similar statistic relates to the expected divergence in neutral genetic variation between chromosome arrangements: $F_{AT} = 1 - (\bar{T}_A / \bar{T}_T)$, where \bar{T}_A is the expected coalescence time for pairs of the same arrangement sampled randomly from the total population. I varied m while leaving other parameters constant, which alters q . Varying s while leaving the other parameters unchanged has similar effects.

The fraction of diversity found between populations (F_{ST}) goes down as migration increases. However, high levels of population structure are obtained under much stronger migration than the expected in the absence of the inversion ($Nm > 1$; see (Wright, 1951)), provided that the inversion reduces gene flux considerably ($\phi < 10^{-3}$). At low migration rates the arrangements are close to fixation in the populations where they are beneficial, reducing the effective gene flux rate. As migration increases, diversity between populations declines due to the reduced differences in the frequencies of the arrangements. The genetic structure between chromosome arrangements is less sensitive to migration, remaining essentially unaffected for $\phi = 10^{-8}$. This result illustrates the role of reduced gene flux in the maintenance of differences between populations and arrangements.

Old inversions: Locally adapted alleles

The second model considers the situation in which two loci are polymorphic, with alternative alleles at each locus adapted in the two populations (or species). Theory shows that this situation can maintain alternative chromosomal rearrangements even if the

breakpoints of the inversion are themselves selectively neutral (Kirkpatrick & Barton, 2006).

Denote the two loci as A (with alleles A_1 and A_2) and B (with alleles B_1 and B_2). Alleles A_1 and B_1 are favored in Population 1, while alleles A_2 and B_2 are favored in Population 2. I assume symmetric selection with no dominance: in Population 1, each copy of alleles A_2 and B_2 decreases fitness by $s/2$, while in Population 2 alleles A_1 and B_1 have that effect. I assume that the polymorphisms at these loci are infinitely old. The inversion captured alleles A_2 and B_2 , which caused it to invade population 2. The positions of the selected loci are assumed symmetric so that the distance between the left breakpoint and locus A is equal to the distance between locus B and the right breakpoint. Adding these selected loci to the model substantially increases its complexity (Appendix A). Consequently, I again only present numerical evaluations of the analytic expressions here.

Results for expected coalescence times are shown on the right side of Figure 1.1. An arrow indicates the position of one selected locus. (The left half of the inversion, including the second locus, has symmetric patterns and is not shown). Parameter values are as for the first model of locally adapted breakpoints. For pairs of genes sampled from the same arrangement, coalescence times are very similar to the first model (at left in the figure) and to the neutral expectation from the structured populations (dashed line).

When one gene is sampled from an inverted and one from a standard chromosome, however, a dramatic difference is apparent. Locally adapted alleles produce peaks in \bar{T} , a pattern consistent with classical balancing selection (Hudson & Kaplan, 1988; Kaplan *et al.*, 1988). Divergence decreases as we move away from the selected loci. The force behind this pattern is recombination in homokaryotypes, which is considerably large ($Nr = 10^4$). High recombination causes the region affected by selection

on loci to be small. The effect of locally adapted alleles is expected to be wider in inversions when the product of Nqr is smaller, for example as a result of strong selection (Figure 1.1, grey lines) or small population size. The implication of these potentially narrow peaks is that we may require genetic markers very tightly linked to a selected locus to detect a signature particular of this model.

Young inversions: Drift

To this point I have focused on inversion polymorphisms that have been maintained by selection for long periods of time. Results from *Drosophila*, however, suggest that some inversions are relatively young (Andolfatto *et al.*, 2001). This raises the question of how the age of the inversion affects the patterns described.

I begin the study of young inversions with a null model in which a polymorphic inversion has drifted to its current frequency in a single population. (Old inversions that evolved by drift will either be fixed or lost.) This situation provides a useful null model for comparison in the following sections in which young inversions initially spread by selection. I studied this case by simulation. In a first step, a stochastic trajectory is simulated to give the frequency of the inversion from its origin at a (random) time in the past to its current frequency x_0 (Przeworski *et al.*, 2005). In a second step, the backward coalescent process conditional on this trajectory is simulated (Kaplan *et al.*, 1989). Appendix B describes the algorithm in detail. I simulated 10^6 realizations of the process.

Figure 1.3a shows \bar{T} for a neutral inversion that is currently at a frequency of $x_0 = 0.5$ in a population of size $N = 2 \times 10^5$. The average age of the inversion is $T^* = 2.8 \times 10^5$ generations, which is consistent with the expected age of a neutral allele currently at frequency 0.5 (Kimura & Ohta, 1971). Inside the inversion, \bar{T} within both arrangement types is reduced with respect to the Standard Neutral Model (SNM, shown by a dashed

line). This is because only chromosomes with the same rearrangement can coalesce, and the number of chromosomes with a given arrangement is smaller than $2N$.

There is also an increase in coalescence times between S and I arrangements compared to the SNM, and a marked peak at the breakpoints. Going backwards in time, if a gene is completely linked to the breakpoint of an inverted chromosome, it cannot coalesce with a gene on a standard chromosome until before the origin of the inversion. At this point, the single ancestral I chromosome mutates into an S chromosome, and can coalesce with other S chromosomes. Outside of the inversion, coalescent times converge to the SNM as we move along the chromosome away from breakpoint. This contrast between inside and outside of the rearranged segment illustrates the effect of reduced recombination.

Qualitatively, these coalescent patterns are similar to what is seen at sites linked to a weak selective sweep (Barton, 2000; Gillespie, 2000). That point will become relevant shortly when comparing these results with those for inversions established by selection.

Young inversions: Locally adapted breakpoints and alleles

I now consider an inversion that appeared at time T^* in the past and was then established by selection. For this purpose, I obtain the frequency trajectories from deterministic forward-time models. In the coalescent simulations, I follow these trajectories backwards in time to frequency $1/2N$. I present the case of a very young inversion, where $T^* = 10^4$ generations and all other parameters are as for the old inversions described above.

Figures 1.3b and 1.3c show T^- for young inversions under both models of local adaptation. The inverted arrangement has extremely reduced diversity, product of the

recent selective sweep. On the other hand, there is little effect of this sweep on \bar{T}_{IS} and \bar{T}_{SS} , which remain close to the expectation given by the neutral structured coalescent. Both models have similar diversity, except again when tightly linked to the selected sites. Given the assumption of ancient polymorphism at the selected loci, \bar{T} between different selected alleles is infinite.

These coalescent patterns are qualitatively similar for inversions of about $T^* < N$ generations, age after which the values of \bar{T} start to increase towards those expected for old inversions (Figure 1.4). There is a period of time (say, $N < T^* < 4N$) during which \bar{T}_{II} is not considerably reduced and \bar{T}_{IS} is not dramatically increased. The coalescent patterns during this period are qualitatively similar to those for a neutral inversion. The implication is that inversions of intermediate ages are expected to present patterns of neutral diversity with little evidence of selection. Nevertheless, throughout this period, levels of F_{AT} (between 0.4 - 0.5) remain higher than those observed in drifting inversions ($F_{AT} = 0.3$). This difference may be a useful diagnostic for inversions of intermediate age.

DISCUSSION

Chromosomal inversions under local selection affect neutral diversity in two ways. For a period of about N generations after their origin, inversions cause decreased diversity within chromosome types. After this period, the effect of the partial sweep disappears and a second pattern emerges as the chromosome types diverge. These predictions are encouraging for the search of signatures of selection in chromosomal inversions, and they suggest that inversions may hold neutral divergence that can be detected under some conditions. Divergence between arrangements is particularly strong when $m \ll s$ and $Nq\phi < 1$. However, high divergence between chromosome types is not

always expected. That is the case for inversions of intermediate age (when T^* is about N generations), inversions that do not reduce gene flux considerably, and populations in which selection against migrants is not strong enough (say, when $m/hs > 0.1$). The patterns of divergence predicted by my models appear consistent with data from some inversion systems (Schaeffer *et al.*, 2003; Kennington *et al.*, 2006), and also offer explanations for why other systems do not show a significant departure from neutrality (Cohuet *et al.*, 2004; White *et al.*, 2010).

Reduced gene flux between chromosome arrangements is the key driver of the patterns observed. The neutral model shows that young inversions will have diversity patterns that differ from SNM expectations if they suppress gene flux. Additionally, reduced levels of ϕ under local adaptation maintain neutral divergence between arrangements even in situations where high migration has eliminated any signal of divergence between populations elsewhere in the genome. In my models, values of $N\phi < 10$ are necessary to obtain high \bar{T}_{IS} , a result consistent with Navarro *et al.* (Navarro *et al.*, 2000). Given the estimated range of ϕ in nature ($10^{-2} - 10^{-8}$, (Navarro *et al.*, 1997; Schaeffer & Anderson, 2005; Stump *et al.*, 2007); assuming N values of about 10^4 to 10^6 and $m \ll s$) many inversions - but not all - have the potential to harbor increased divergence between arrangements.

When an inversion polymorphism is established by selection, it is expected to show reduced diversity for a period of time on the order of N generations. Old inversions recover diversity and accumulate considerable divergence from standard chromosomes. In this continuum between recent and ancient polymorphism there is a period during which levels of diversity within and between arrangements will be very similar to neutral expectations (between about N and $4N$ generations; see Figure 1.4). This is a substantial period that spans the estimated ages of some inversion polymorphisms (*e.g.* (Andolfatto

et al., 2001)). For example, some inversions in *Anopheles gambiae* (0.4 to 1.7 N generations old) fall in this age range, and they show little divergence between standard and inverted arrangements (White *et al.*, 2009). Here the lack of signal does not imply that the inversions are neutral or uninvolved in divergence between populations.

Inversions maintain increased coalescence times between populations even in high levels of migration. As migration (or hybridization) increases, genetic divergence declines in regions of the genome with normal recombination. The genetic structure between chromosome arrangements can persist in these situations, however, particularly at very low values of ϕ . For gene flux observed around the breakpoints of some inversions (*e.g.* $\phi = 10^{-8}$, (Schaeffer & Anderson, 2005)), increased \bar{T}_{IS} remains unaffected by migration, even as the populations begin to resemble a panmictic population. High levels of migration are frequent in situations of local adaptation, and studies of polymorphic inversions have observed large values of Nm (2.2 – 16) (Onyabe & Conn, 2001; Schaeffer *et al.*, 2003). Populations with this amount of gene flow will show no appreciable differentiation at neutral sites in genomic regions that are not rearranged. For this reason, it may be more informative to compare diversity between arrangements, as it allows us to detect the elevated diversity still present.

The models predict different patterns of coalescence times and therefore neutral genetic diversity. There are, however, three hurdles that may complicate the process of distinguishing between evolutionary processes. First, the reduction of diversity within chromosome arrangements caused by the initial sweep is similar to the pattern produced by genetic drift. Relative measures of divergence (*i.e.* F_{AT}) may provide better evidence of inversions under selection, but ruling out drift may require evidence from other sources (such as stable geographic clines (Hoffmann & Rieseberg, 2008a)). Second, the both models of local adaptation predict the same patterns of divergence near the

breakpoints. Finding additional peaks of divergence within the inversion would suggest the presence of locally adapted alleles. Third, extremely reduced gene flux will cause increased divergence throughout the rearrangement, ‘swamping’ potential peaks of divergence around locally adapted alleles (see (Feder & Nosil, 2010; Via)).

In essence, these models confirm the intuition that coalescent patterns in inversions behave as special cases of selective sweeps and local selection in which a large portion of the chromosome is tightly linked to the selected sites, thus hitchhiking and diverging with them. The results are consistent with several previous theoretical results. The model of locally adapted breakpoints yields results equivalent to Navarro *et al.* (Navarro *et al.*, 2000) for the patterns within populations when migration is zero and frequencies are kept constant by balancing selection (results not shown). The models are also consistent with theoretical work on balancing selection (Hudson & Kaplan, 1988; Kaplan *et al.*, 1991), selective sweeps (Kaplan *et al.*, 1989; Kim & Stephan, 2002), local selection (Charlesworth *et al.*, 1997; Nordborg, 1997; Feder & Nosil, 2010) and divergence hitchhiking [56].

I present these models to provide quantitative predictions of the patterns expected. I consider this approach necessary, as with increasing complexity and number of interacting parameters intuition is not enough to understand the patterns observed in nature. Some studies have reported little differentiation between inverted and standard arrangements (Cohuet *et al.*, 2004; White *et al.*, 2009). The models predict that pattern when the inversion is very young or when conditions allow for high gene flux between the arrangements. Other studies have found high and uniform divergence between inverted and standard arrangements (*e.g.* (Munté *et al.*, 2005; Machado *et al.*, 2007; Lawniczak *et al.*, 2010)), which is consistent with my models when there is reduced gene flux and/or multiple locally adapted loci within the inversion (as suggested in [56]). Other

data show peaks of divergence within an inverted region (*e.g.* (Schaeffer *et al.*, 2003; Schaeffer & Anderson, 2005; Kennington *et al.*, 2006; Kolaczowski *et al.*, 2011; McGaugh *et al.*)), a pattern consistent with my locally adapted alleles model. Fine-scale scans may prove essential to distinguish between these and other hypotheses; comparisons of large sections of the rearranged chromosome (Yatabe *et al.*, 2007; Strasburg *et al.*, 2009) will typically not be sufficient. High-resolution data are increasingly available (*e.g.* (White *et al.*, 2009; Neafsey *et al.*, 2010; Kolaczowski *et al.*, 2011; McGaugh *et al.*)), and they show trends that lend themselves for further speculation. It is not possible, however, to justify any conclusions without quantitative data analyses for which these models are only the foundation.

How will data ultimately be linked with models to give quantitative conclusions for how inversions evolve? Even biologically simple models for inversions are sufficiently complex to make standard statistical approaches infeasible. One way forward is by combining coalescent simulations with analysis techniques akin to approximate Bayesian computation (reviewed in (Beaumont, 2010; Csillery *et al.*, 2010)). Results from models presented here will be important for identifying which summary statistics to use in that approach. (For example, F_{AT} is promising in some situations.) Developing expectations for other statistics that may be informative (such as linkage disequilibrium and long distance associations, *e.g.* (Schaeffer *et al.*, 2003; Kennington *et al.*, 2006)), and extending these models to other scenarios of interest, will require future theoretical work.

Chapter 2: Reproductive Isolation And Local Adaptation Quantified For A Chromosome Inversion In A Malaria Mosquito*

INTRODUCTION

Chromosome inversions have a rich history in evolutionary genetics, much of it focused on the roles they may play in speciation (White, 1973; King, 1993; Hoffmann & Rieseberg, 2008a; Kirkpatrick, 2010). Two genetic properties of inversions make them particularly favorable for the evolution of reproductive isolation (Butlin, 2005). First, large inversions cover hundred or even thousands of loci, and so could easily span genes involved both prezygotic and postzygotic isolation. Second, inversions dramatically reduce recombination when heterozygous, increasing linkage between any genes present that contribute to the two types of isolation. Consider the situation when prezygotic isolation results from assortative mating of inversion homozygotes and postzygotic isolation results from selection against inversion heterozygotes. Then pre- and postzygotic isolation will reinforce each other, a situation that greatly facilitates speciation (Rice, 1987; Kirkpatrick & Ravigné, 2002; Gavrillets, 2004). Inversions have been implicated in pre- and postzygotic isolation in *Rhagoletis* flies (Feder *et al.*, 2003; Michel *et al.*, 2010) and monkeyflowers (Lowry & Willis, 2010). Quantitative measures for the overall strength of reproductive isolation generated by an inversion, however, are as yet still lacking.

The mosquito *Anopheles funestus* offers an unusually compelling opportunity to study inversions and speciation. Widely distributed across the sub-Saharan Africa, *A. funestus* is one of the most proficient vectors of human malaria, and in some places its transmission rate surpasses *A. gambiae*, its better-studied congener (Coetzee &

* Significant portions of this work have been previously published as Ayala, Guerrero & Kirkpatrick (2013) *Evolution*. 67:946-958. Author contributions: Diego Ayala collected the karyotype data I analyze here; Mark Kirkpatrick, supervisor.

Fontenille, 2004). The species is highly polymorphic for inversions, but a clear picture of their evolutionary significance remains elusive because of conflicting results from different geographical regions and from different genetic markers (Costantini *et al.*, 1999; Kamau *et al.*, 2003; Boccolini *et al.*, 2005; Cohuet *et al.*, 2005; Michel *et al.*, 2006).

Inversion 3Ra in *A. funestus* is a prime candidate for detailed investigation. It is large, spanning roughly 30% of the right arm of Chromosome 3, and 7% of the entire genome (which has 3 pairs of chromosomes). Its frequencies in Africa are correlated with humidity (Costantini *et al.*, 1999; Guelbeogo *et al.*, 2005; Michel *et al.*, 2006; Ayala *et al.*, 2011a), suggesting that this inversion is involved in local adaptation. Other phenotypes that have been associated with 3Ra are resting behavior (Costantini *et al.*, 1999), host preference (Costantini *et al.*, 1999), and wing shape (Ayala *et al.*, 2011b). Previous studies have also suggested that this inversion contributes to the genetic isolation between populations in Burkina Faso (Costantini *et al.*, 1999; Guelbeogo *et al.*, 2005).

Chromosome inversions in *Anopheles* mosquitoes have important implications for human health (Sharakhov *et al.*, 2002; Besansky *et al.*, 2003; White *et al.*, 2007). Malaria is a severe problem in Africa, responsible for over one million deaths per year and more than 250 million people infected (Murray *et al.*, 2012). Inversion polymorphisms in *Anopheles* are critical to the epidemiology of malaria: they affect habitat preferences (Coluzzi *et al.*, 1979; Costantini *et al.*, 1999), feeding behavior (Petrarca & Beier, 1992; Lochouart *et al.*, 1998), and they may have enabled large range expansions (Toure *et al.*, 1994; Besansky *et al.*, 2003; Cohuet *et al.*, 2005; Manoukis *et al.*, 2008b; Ayala *et al.*, 2011b). Further, inversions affect mosquito control because they limit the potential for insecticide resistance alleles to introgress between populations and impact the possibility

of using genetically-modified mosquitoes for disease control (Alphey *et al.*, 2002; Boete & Koella, 2003; Tripet *et al.*, 2007; Enayati & Hemingway, 2010).

In this study I estimate the effects of inversion 3Ra in *Anopheles funestus* on viability, local adaptation, and both pre- and postzygotic reproductive isolation. Mosquitoes were sampled along a transect through diverse habitats in Cameroon. I constructed a series of population genetic models, and then fit the models to the data. The best-fit models suggest that the inversion is subject to strong viability selection that varies between the zones, and to strong assortative mating. The results have implications for how inversions evolve, and provide hypotheses for further experimental studies of their genetics, physiology, and ecology.

METHODS

Sampling: Sampled 751 mosquitoes from 105 villages that lie along a 1421 km transect in Cameroon that follows the only highway that runs the length of the country. The sampled transect, shown in Figure 1.1, traverses three markedly distinct ecological zones. It begins in the sub-Saharan savannah, crosses a highland region with mountains that reach 4000 m, and then descends into the lowland tropical rainforest. *Anopheles funestus* is largely commensal on humans. Most of its population is thought to be strongly associated with human presence, and the transect follows the region of suitable habitat that runs north to south in Cameroon (Ayala *et al.*, 2009).

Adult mosquitoes were collected between August and December (the wet season) in 2005 and 2006 from inside homes (Ayala *et al.*, 2009). Sampling dates for each zone were timed so that they were in comparable phases of the wet season. Species identification was confirmed using PCR (Cohuet *et al.*, 2003). Karyotypes were determined with a phase contrast microscope following the *Anopheles funestus*

cytological map (Sharakhov *et al.*, 2004). Sixty percent of the individuals were independently karyotyped a second time by different technician, and no discrepancies were found. Coordinates, sample sizes, and karyotype count for all of the localities are given in Supporting Information Table S1.

The models: I fit a series of models to the data in order to estimate the strengths of evolutionary forces acting on the inversion. The models differ in their assumptions about selection, mating, and migration. I implemented the models in simulations, and then used likelihood to determine the model and parameter values that best fit the data.

The life cycle, based on the natural history of *Anopheles*, assumes that viability selection occurs at the larval stage (Clements, 1999). Viability selection is followed by mating, which may be assortative. Mated females disperse, and then lay their eggs. Generations are not overlapping.

The viabilities of the three genotypes depend on the ecological zone in which a deme lies. I allowed there to be one, two, or three zones, and assumed that selection is uniform within a zone. With two and three zones, I assumed the zone boundaries run east-west, and allowed the locations of the boundaries to be variables in the model. Viability selection is soft, and so the number of surviving adults in a deme is independent of the fitnesses and genotype frequencies. Heterozygotes are assigned a relative viability of 1, and I denote the viabilities of standard and inverted homozygotes as W_{ss} and W_{II} , respectively. These parameters can be smaller or greater than one, which allows for directional, overdominant, or underdominant selection.

Other inversions in *Anopheles* are thought to be involved in assortative mating (Lehmann & Diabate, 2008; Perevozkin *et al.*, 2012). I therefore allowed for this possibility by considering four models for mating. The first assumes mating is random. In

the second, assortative mating occurs when a fraction F of individuals to mate only with others sharing the same karyotype, while the remainder mate at random. In the third model for mating, the strength of assortative mating takes on a different value in each ecological zone. In this case there are as many assortment parameters as there are ecological zones (F_1, F_2 , etc.). Last, I allowed the karyotypes to differ in the strength of assortment. This model includes, for example, the case in which only homozygotes have a mating preference. In this case there are three assortment parameters (F_{SS}, F_{SI} , and F_{II}).

There is limited information about how mosquitoes are distributed in space and how they move. I therefore considered two very different models for population structure and migration that hopefully bracket the actual situation. The first is motivated by an analysis of habitat suitability (Ayala *et al.*, 2009) and the first author's experience while collecting mosquitoes over two years. The species appears to be largely commensal on humans and strongly concentrated in the villages that I sampled along the transect. Mosquitoes are rarely found in the surrounding habitat, even where nonhuman animals are abundant. The first model is therefore a one-dimensional stepping-stone array with 105 demes of equal size that represent the villages sampled. To account for the variation in distances between demes, the probability of dispersal to the neighboring deme that is at a linear distance d is calculated as the area under the tail of a gaussian distribution with expectation 0 and variance σ_m^2 evaluated at $d/2$.

The second model of migration takes an strongly contrasting view of how mosquitoes are distributed by assuming that the mosquitoes live in a continuous two-dimensional habitat with constant densities. Migration is uniform in all directions with a variance σ_m^2 . Without any loss of accuracy, I modeled this situation using a continuous one-dimensional habitat running north-south (that is, orthogonal to the boundaries between the ecological zones). The simulation model then predicts the genotype

frequencies for a given village based on its latitude, that is, its position along the ecological gradient.

Table 2.1 summarizes the combinations of these assumptions that I simulated. The simulations were written in C++. Runs were initialized with karyotype frequencies at the values observed in each population (except if an observed frequency was zero, where it was adjusted to 10^{-3}). Simulations were stopped when all karyotype frequencies changed less than 10^{-8} per generation. The equilibrium frequencies were evaluated at the point in the life cycle following dispersal, corresponding to when individuals in the field samples were collected. The number of generations needed to reach equilibrium depends on the parameter values and initial conditions. With the maximum likelihood parameter estimates, the system evolves roughly halfway to the equilibrium in 200 generations, corresponding to about 15 years for this species.

Fitting the models: I estimated the parameters for each model by finding the values that give equilibrium frequencies for the inversion that best fit the data. The fit was evaluated using likelihood. Write f_{ij} for the number of individuals sampled from deme (village) i with karyotype j ($= 0, 1, 2$). Denote the vector of eight parameter values as Π , and the equilibrium frequency in deme i of karyotype j from the simulation under those parameter values as $p_{i,j}(\Pi)$. Then the log likelihood for the vector \mathbf{f} of all the data, given the parameter values Π , is found from the multinomial distribution:

$$\ln L(\mathbf{f} | \Pi) = C \sum_{i=1}^{n_p} \sum_{j=1}^3 f_{i,j} \ln(p_{i,j}(\Pi))$$

where C is a constant that is independent of the parameters Π . The maximum likelihood estimate for the parameters is found by maximizing Equation (1) with respect to Π .

I found the parameter values that maximize the likelihood using a simulated annealing method (Press *et al.*, 2007). This heuristic algorithm samples the parameter space sequentially, starting from an arbitrary set of parameters. Random changes are then made to the parameter set. The new values are always accepted if they increase the likelihood of the data. A change that decreases the likelihood is accepted with probability proportional to a control parameter T . These steps are repeated with gradually decreasing values of T , with the goal of reaching the global maximum of the likelihood function.

I obtained confidence regions for the maximum likelihood estimates by calculating likelihood profiles of parameters for the best-fit model. Profiles for migration, mating, and selection parameters were estimated by evaluating each parameter on a lattice of points, then maximizing the likelihood with respect to the remaining parameters. This process is computationally costly, so I fixed the region boundaries at their maximum likelihood estimates to analyze profiles of other parameters. (The boundary estimates correspond closely to those previously defined on ecological grounds (Olivry, 1986), and are the parameters of least interest.) Approximate 95% and 99% confidence regions were identified as the parameter space for which $\ln(L)$ was within 2 and 3 units, respectively, of the maximum. These likelihood profiles were also useful in verifying that the annealing runs did in fact reach optima.

To evaluate the fit of different models I used Akaike's Information Criterion (AIC; Akaike (1974)). This measure takes into account both the likelihood and the number of parameters in the model, penalizing those models that have more parameters. I judged a model to have better fit than another if it has a lower AIC score. Models whose AIC scores differ by less than 6 units were judged to be statistically indistinguishable. Some of the models are nested (i.e. special cases of other models). In these cases, I also

compared models though χ^2 tests of their likelihood ratios (Whitlock and Schluter (2009), chap. 20).

I estimated total reproductive isolation (RI_{total}) in each ecological zone by combining the contributions of prezygotic (RI_{pre}) and postzygotic isolation (RI_{post}): $RI_{\text{total}} = RI_{\text{pre}} + (1 - RI_{\text{pre}}) RI_{\text{post}}$ (Coyne & Orr, 1989; 1997). Prezygotic isolation is the proportional reduction in heterokaryotypic matings caused by assortment in a population consisting of 50% standard and 50% inverted individuals, and is equal to $RI_{\text{pre}} = 2F / (1 + F)$, where F is the fraction of individuals that mate with their own karyotype (with a fraction $1 - F$ mating at random). Postzygotic isolation is the reduction in survival of F_1 hybrids relative to the average of their parents: $RI_{\text{post}} = 1 - 1/[(W_{SS} + W_{II}) / 2]$, where W_{SS} and W_{II} are the viabilities of standard and inverted homozygotes, respectively, relative to the heterozygotes.

RESULTS

The inversion cline: Inversion 3Ra in *Anopheles funestus* shows a dramatic cline in frequency along the transect. As shown in Figure 2.1, the inversion is near fixation in the rainforest, at intermediate frequencies in the highlands, and nearly absent in the savannah. Shifts in frequencies at the boundaries between the zones are abrupt, which is consistent with how rapidly the environment changes at these transition points. Within each of the three zones, inversion frequencies are generally similar. The data also suggest qualitatively that there are heterozygote deficits in the highlands, as can result from selection against inversion heterozygotes and/or assortative mating.

Viability, mating, and reproductive isolation: I fit 10 models that differ in their combinations of assumptions about selection, mating, and migration. Four of these gave very similar fits to the data. However, they also give very similar qualitative (and even

quantitative) results regarding selection and mating, suggesting that those conclusions are robust. I will first describe the model that best fits the data, then turn to the alternatives. They include three models whose fit is statistically no worse than the best-fit model.

Under the best-fit model, viability varies in space and there are three ecological zones corresponding to the savannah, highlands, and rainforest. The estimated locations for the boundaries between the zones lie very close to boundaries that were previously defined on ecological grounds (Olivry, 1986). The inversion is involved with assortative mating, but its strength does not vary in space or with the karyotype. Migration follows a stepping stone model in which the demes correspond to villages along the highway where the mosquitoes were sampled.

The fit of the model to the data is shown in Figure 2.2. Qualitatively, predictions for the average karyotype frequencies within each zone are good. Much of the variation around the prediction (visible particularly in the highlands) results from the small sample sizes in individual villages. Only three of the 105 villages have frequencies that differ significantly from the model, fewer than is expected by chance.

The estimated strength of viability selection is intense. Figure 2.3 and Table 2.1 show estimates for the viabilities of the karyotypes in the three ecological zones. The viabilities of homozygotes range from 25% to 130% relative to those of heterozygotes. Further, the pattern of selection differs strikingly between the three zones.

Heterozygotes have low fitness in the savannah, reduced by about 23% compared to the two homozygotes, which in turn have similar fitness. The evidence for underdominance in the savannah is strong, as the 99% confidence region excludes values for the homozygote fitnesses that are equal to or lower than the heterozygote fitness.

In the highlands and rainforest, by contrast, standard homozygotes are strongly selected against, while heterozygotes have highest fitness. This pattern of overdominance

is statistically significant in the highlands, where the viabilities of standard and inverted homozygotes are respectively 25% and 50% that of heterozygotes (Figure 2.3). Results from the rainforest are consistent with overdominance. The pattern is not significant, however, because the inversion is near fixation there and selection parameters are therefore poorly estimated.

The data suggest that the inversion is involved with strong assortative mating. The maximum likelihood estimate for F is 0.82 (Figure 2.4). The lower bound of the 99% confidence interval is $F = 0.77$, which implies that even the weakest assortment consistent with the data is still strong. To my knowledge, this is the first estimate in any natural population for the overall strength of assortment associated with a chromosomal rearrangement. The analysis provides no suggestion, however, regarding the mechanism of assortment. It might result from active mate choice or from passive mechanisms such as habitat preference and the pattern of diurnal activity (Diabate *et al.*, 2009; Pennetier *et al.*, 2009).

My inferences about viability and assortment are based on the assumption that the inversion (or genes that it carries) is causal for these effects. I am unable to test this assumption formally, for example by a classic genetic analysis using recombinants, because it has been impossible to establish a breeding colony in Cameroon despite several years of effort. In the Discussion I return to the possibility that the patterns in the data are caused by genetic factors outside of 3Ra.

I can combine my inferences about natural selection and assortative mating to arrive at an overall estimate of the strength of reproductive isolation conferred by this inversion. The most interesting context is the savannah, where the premating isolation that results from assortment is reinforced by the postmating isolation that results from underdominance. Using the estimates for viability and assortment with Eq. (2), I estimate

that the total isolation between standard and inverted homozygotes is 92%. Genetic introgression between the two homozygous forms in the savannah continues as the result of both the incomplete isolation and gene flow from neighboring populations in the highlands.

I can also estimate the total reproductive isolation in the rainforest and highlands habitats. Heterozygotes in those two habitats have highest viability, however, which produces negative values for postzygotic isolation. If one is willing to allow for that possibility, the values for total isolation in the highlands and rainforest are 0.74 and 0.88, respectively.

The last parameter estimated in my analysis is the dispersal variance. The maximum likelihood estimate is 7.3 km², corresponding to an average movement of 2.2 km / generation (Figure 2.4). This is the parameter that is estimated with least precision (the 95% confidence region is [6.2, ∞], where the upper bound corresponds to free movement between adjacent villages). This lack of precision presumably results because migration only has strong effects on the predictions for the frequencies in villages near to boundaries between ecological zones. It is, however, the first estimate of the migration rate for this malaria vector, and for any organism using an inversion. In the best-fit model, migration follows a stepping stone model in which the demes correspond to villages. The alternative model for migration, in which space is continuous, fits the data virtually as well and gives very similar parameter estimates (Table 1).

Alternative models: These qualitative conclusions about the patterns of evolutionary forces and the quantitative parameter estimates depend on the model that is assumed. Table 2.1 summarizes results from the 10 models that I analyzed. The two righthand columns give the negative of the log likelihood (where a larger value corresponds to better agreement between model and data) and the AIC (where a smaller

value implies a better fit after accounting for the number of parameters in the model). Both of these scores are measured relative to the model that fit best, as judged by AIC.

Six models can be rejected because they give a significantly worse fit to the data (based on an AIC score and/or a likelihood ratio test). For the remaining four models, the fits to the data are very similar. Importantly, the patterns and strength of selection and assortative mating that they estimate are very similar, suggesting that the central conclusions may be robust.

The four models that fit best share several key features. First, they show strong evidence that the fitness of the inversion varies between three ecological zones. Second, they suggest the inversion is overdominant in some habitats but underdominant in others. All of these models give similar maximum likelihood estimates for the viabilities of the inversion genotypes in the savannah and highlands, where all the estimates fall within the 95% confidence intervals shown for viability parameters in Figure 2.3. Last, all four models implicate the inversion in strong assortative mating.

There are, however, important differences between the four best-fit models regarding mating and dispersal. I cannot distinguish between a model in which the strength of assortment is constant, one in which it varies by genotype, and one in which it varies by ecological zone. The latter two models show slightly improved likelihood over the best-fit model, but the two extra parameters that they each require lead to worse AIC scores. In all cases, however, there is support for relatively strong assortment in the savannah and highlands. (Estimates in the rainforest are imprecise because there is little polymorphism there.)

I also cannot distinguish between two very different sets of assumptions about spatial distribution and dispersal, a linear stepping-stone vs. a continuous two-dimensional habitat. Figure 2.5 shows that the two models give very similar fits to the

data. Inferences about dispersal and population distribution are the weakest point of the analysis. It is encouraging, however, that the qualitative results for viability and assortment are the same under these very different models for population structure.

I can clearly reject several alternative models for how viability selection varies in space. In one alternative there is only a single ecological zone, and fitnesses for the karyotypes do not vary along the cline. This model includes the situation in which the cline is maintained in a “tension zone” where a combination of underdominance and assortative mating causes fixation of standard homozygotes at one end of the cline and inversion homozygotes at the other (Barton & Hewitt, 1985; Barton & Gale, 1993). (The estimated viability of the heterozygote is very slightly less than that of the homozygotes, but the difference is too small to appear in Table 2.1.) In another alternative model of selection, there are two ecological zones. Both the one- and two-zone models are decisively rejected ($p \ll 10^{-6}$, χ^2 test of the likelihood ratio).

I also considered several alternative assumptions about mating. The simplest is random mating. Here departures from Hardy-Weinberg result only from viability selection and migration. This model fits the data poorly ($p \ll 10^{-6}$, χ^2 test of the likelihood ratio). As discussed above, other models make different assumptions about how assortative mating works. In one, the strength of assortment is allowed to vary between the three ecological zones. In the second, the strength of assortment depends on the karyotype but does not vary in space. The fit to the data is marginally improved in both models by the addition of two more parameters, but the increase in likelihood is not significant ($p > 0.05$, χ^2 tests of the likelihood ratios).

Last, I considered the possibility that the results are driven by aberrant samples. Karyotype frequencies depart significantly from the model predictions at the $p < 0.05$ level in three of the 105 villages sampled. While at least that many are expected by

chance alone, I removed those villages from the dataset and reran the analysis for the best-fit model. The parameter estimates changed little (results not shown), and so the analyses reported above are based on the full data set.

DISCUSSION

This study estimates the viability effects and nonrandom mating associated with an inversion. Three key conclusions are suggested by the results. First, viability selection appears to be intense. The maximum likelihood estimates from the best-fit model suggest that survival of homozygotes ranges from 25% to 130% of the heterozygote fitness. Second, the pattern of selection appears to differ strikingly between habitats. In the savannah, inversion 3Ra is strongly underdominant, while in the highlands it is strongly overdominant. Third, assortative mating seems to be strong. The likelihood analysis suggests that between 77% and 91% of matings are assortative based on karyotype. These estimates are based on the assumptions of the models. Independent estimates, based for example on experiments in the field, would be very valuable. The following conclusions can be viewed as hypotheses that might be tested using other approaches.

In the savannah, where selection is underdominant, the inversion segregates as a single genetic element that confers both pre- and postzygotic isolation. Natural selection depresses the frequency of heterozygotes. This enhances the power of assortative mating to inhibit matings between carriers of standard and inverted chromosomes. In short, natural selection reinforces premating isolation because the inversion links viability and mating phenotypes. A trait with these properties is particularly favorable to speciation (Kirkpatrick & Ravigné, 2002), and so is sometimes referred to as a “magic trait” (Gavrilets, 2004). The total reproductive isolation that results in the savannah between chromosome heterozygotes is 92%, which is perhaps the strongest that ever been reported

from sympatric forms of a single species. This degree of isolation is similar to the figure of 94% reported for sympatric pairs of *Drosophila* species when measured in the lab (Coyne & Orr, 1997). It is difficult to compare these numbers directly, however, as the *Drosophila* estimates do not take into account the effect of factors like habitat preferences that are excluded from the lab environment.

It may seem remarkable that an inversion, which segregates like a single locus with two alleles, could at once confer both pre- and postzygotic isolation. Evidence from two other species suggests this situation may not be uncommon, however, at least for large inversions that capture many loci. In *Rhagoletis* flies, inversions influence diapause, which in turn adapts races of the flies to alternative hosts and contributes to assortative mating (Feder *et al.*, 2003; Michel *et al.*, 2010). In the yellow monkeyflower, *Mimulus guttatus*, a recently discovered inversion affects several morphological and phenological traits (Lowry & Willis, 2010). These changes adapt the plant to two very different habitats and incidentally contribute to both pre- and postzygotic isolation between the two chromosomal forms. There is also abundant evidence from other species of *Anopheles* that inversions have diverse phenotypic effects, including aridity tolerance (White *et al.*, 2007; Gray *et al.*, 2009; Cassone *et al.*, 2011; Fouet *et al.*, 2012), temperature tolerance (Rocca *et al.*, 2009), mating preferences (Perevozkin *et al.*, 2012), host preference (Coluzzi *et al.*, 1979; Petrarca & Beier, 1992; Costantini *et al.*, 1999), resting behavior (Coluzzi *et al.*, 1977; Rishikesh *et al.*, 1985; Bryan *et al.*, 1987; Costantini *et al.*, 1999), susceptibility to parasites (Toure *et al.*, 1996), breeding site preference (Manoukis *et al.*, 2008a), insecticide resistance (Brooke *et al.*, 2002), and environmental tolerance (Coluzzi *et al.*, 1979; Toure *et al.*, 1998; Petrarca *et al.*, 2000; Coluzzi *et al.*, 2002; Cohuet *et al.*, 2004; Cohuet *et al.*, 2005; Simard *et al.*, 2009; Ayala *et al.*, 2011b).

What evolutionary events led to a polymorphism for inversion 3Ra that mediates both local adaptation and reproductive isolation? When two or more loci experience local adaptation to environmental conditions that change in space, an inversion that spans the loci will spread under quite general conditions (Kirkpatrick & Barton, 2006). The result predicted by theory is a cline that looks very much like that seen in Figure 2.1. One could imagine that when inversion 3Ra first appeared, it captured alleles that allow its carriers to survive well in more humid environments and also to mate in a different time or place than savannah-adapted mosquitoes do. The data are consistent with this hypothesis for how inversions become established, and it could explain why inversion 3Ra causes both pre- and postzygotic isolation.

The data are, however, also consistent with at least two other hypotheses for how the inversion was established (Kirkpatrick, 2010). One of the inversion breakpoints could be responsible for a mutation at a single locus that has pleiotropic effects on both viability and mating. Alternatively, the inversion's effects could result from genetic divergence at one or more loci that accumulated after the inversion was established by some other mechanism (Navarro & Barton, 2003). Distinguishing between these hypotheses will require other kinds of data, for example molecular markers for the inversion breakpoints and candidate loci for adaptation inside the region it spans. Developing these resources will require substantial effort, however, since the genome of this species has not been sequenced. Regardless of the inversion's history and the genetic basis of its phenotypic effects, the results suggest the inversion experiences very strong local adaptation and nonrandom mating.

Three previous studies have estimated the viability effects of polymorphic inversions. Lewontin and White (1960) studied two inversions segregating on different chromosomes in the grasshopper *Moraba scurra*. They estimated the genotype viabilities

using the deviation between their observed frequencies and those expected under random mating. They found very strong selection, with statistical support for some genotypes differing by 40% in viability. They constructed adaptive topographies (the surface of mean fitness as a function of the inversion frequencies) for two populations and reported that both lie at saddle points. This result is surprising, since under many conditions a saddle point is evolutionarily unstable. Lewontin and White offered several hypotheses, for example that fitnesses are frequency-dependent. They did not, however, consider two hypotheses suggested by my results. First, fitnesses may vary in space. If the localities that they sampled are at midpoints along a cline, genotype frequencies could resemble the pattern of disruptive selection they describe. Second, mating might be assortative. While Lewontin and White assumed random mating, assortment could produce the appearance of disruptive selection. Unfortunately, their data do not seem able to test either of these two possibilities.

The most direct evidence for selection acting on a common inversion comes from a 900 kb inversion on human chromosome 8 (Stefansson *et al.*, 2005). Data on fertilities of the karyotypes in Iceland leads to the estimate that inversion heterozygotes have about 3% greater fitness than standard chromosome homozygotes. An intriguing observation is that the inversion seems to be spreading in Europe but not in Africa, suggesting that its fitness effects vary in space.

Schaeffer (2008) analyzed the famous inversion polymorphisms in *Drosophila pseudoobscura* studied by Dobzansky and colleagues (Dobzhansky, 1944; Anderson *et al.*, 1991). He used an approach similar to ours that fit a population genetic model to karyotype frequencies sampled from populations sampled across the Western US. Two of Schaeffer's major conclusions are qualitatively consistent with ours. First, selection can be strong: the fitnesses of common karyotypes often differ on the order of 20%, and some

karyotypes in some environments are nearly lethal. Second, there is substantial geographical variation in fitnesses. There are, however, differences between the Schaeffer's analysis and ours that make it difficult to compare the results quantitatively. In Schaeffer's model, mating is assumed to be random, migration rates are assumed rather than estimated from the data, and fitnesses are estimated by minimizing the summed differences between observed and predicted frequencies rather than by likelihood. Last, because the *D. pseudoobscura* system involves five different segregating arrangements, it is not possible to describe patterns of selection in the simple terms of over- and underdominance as is done here for inversion 3Ra in *A. funestus*.

A large number of studies from the mid-twentieth century quantified the contribution that inversions make to postzygotic isolation by measuring the fitnesses of hybrids in crosses between species that have karyotypic differences (White, 1973; King, 1993). It is difficult to compare those data to my results for two reasons. It is unknown what evolutionary forces established those inversions. Second, it is not known whether the fitness effects seen now were present when the inversion first appeared, or whether they accumulated after the two species were fixed for alternative arrangements.

Speciation and Inversion 3Ra: Do these results imply that *A. funestus* is speciating, or perhaps already has? Microsatellite loci on chromosome 3 both inside and outside of the inversion show signs of differentiation between standard and inverted chromosomes in some geographical regions but not others (Cohuet *et al.*, 2005; Michel *et al.*, 2006; Ayala *et al.*, 2011b). This suggests that there is some ongoing recombination and genetic introgression between the standard and inverted arrangements, at least in some populations. In Cameroon, the opportunity for recombination is increased in the highlands, where heterozygotes are common. At the genomic level, I expect the amount of genetic exchange between standard and inverted chromosomes to vary along the

inversion. The regions surrounding the breakpoints and genes within the inversion that experience local adaptation should show higher divergence than other regions (Guerrero *et al.*, 2012). Early in the process of speciation, variation across the genome in the degree of differentiation may be quite common (Wu, 2001; Diabate *et al.*, 2009; Nosil *et al.*, 2009; Butlin, 2010; Michel *et al.*, 2010; White *et al.*, 2010). The current situation in *A. funestus* could therefore represent an intermediate step on the path to speciation that will be completed in the near future, perhaps with contributions from other inversions.

It is also possible that we are observing a long-term equilibrium in which genetic isolation will never be completed. The genus *Anopheles* is replete with examples of inversions that are shared across species boundaries (Coluzzi, 1982). These result from both shared ancestral polymorphisms and introgression between species (Besansky *et al.*, 2003; White *et al.*, 2009). One consequence is a mosaic pattern of differentiation within and between species (Turner *et al.*, 2005; Turner & Hahn, 2007; Wang-Sattler *et al.*, 2007; White *et al.*, 2009; Turner & Hahn, 2010; Cheng *et al.*, 2012a).

It may ultimately be possible to test hypotheses regarding the role that this inversion plays in reproductive isolation and local adaptation by analyzing patterns of neutral molecular variation with coalescent models of inversions (Guerrero *et al.*, 2012). In an important recent advance, Cheng *et al.* (2012a) resequenced inversions from *A. gambiae* along a cline in Cameroon that shares many qualitative features with the one studied here. Their results are consistent with the pattern expected for an inversion polymorphism maintained by local adaptation of loci carried by the inversion, but quantitative tests of that hypothesis have not yet been done.

These results imply there is strong assortative mating, but they give no information about the mechanism responsible. In principle, it could involve differentiation between standard and inverted chromosomes in both a signaling system

and a receiving system. A genetically simpler possibility is that the chromosome forms differ in a single trait that leads to assortative mating, for example habitat preference. This is what happens with the *M* and *S* molecular forms of *A. gambiae*, which form mating swarms in different habitats (Diabate *et al.*, 2009). Another mechanism that could contribute to the assortment observed is suggested by the recent discoveries that mated pairs tend to have matching wing beat frequencies (Gibson & Russell, 2006; Pennetier *et al.*, 2009), and that inversion 3Ra affects wing shape in *A. funestus* (Ayala *et al.*, 2011a). Since wing shape could plausibly affect wing beat frequency, this could be one pathway that contributes to the prezygotic isolation between inversion homozygotes (Sanford *et al.*, 2011).

Crossing and adaptive valley: An enduring conundrum in evolutionary genetics is to explain how underdominant chromosome rearrangements are established (White, 1973; King, 1993). Some inversions are unconditionally underdominant, causing decreased heterozygote fitness regardless of the environment. Many pericentric inversions have this property because recombination in inversion heterozygotes generates aneuploid gametes and so decreases fertility (Hoffmann & Rieseberg, 2008a). Paracentric inversions like 3Ra, however, may often be free of these effects. They can then evolve as the result of the alleles that they carry, which may have fitness effects that depend on the environmental context.

These results suggest how a paracentric inversion that is underdominant can become established when its fitness varies in space. Heterozygotes for 3Ra have lowest relative fitness in the savannah, but they have highest fitness in the highlands. Assuming that the savannah-adapted standard chromosome is ancestral (Green & Hunt, 1980), an inverted chromosome that appeared by mutation could have spread in the highlands (and perhaps enabled a range expansion there and into the rainforest). If highland populations

were to evolve reproductive isolation from the savannah populations, the stage would be set for inverted homozygotes to reinvade the savannah if suitable habitat becomes available there. The result would be a pair of sympatric sister species separated by an adaptive valley of low heterozygote fitness. This scenario is hypothetical in the case of *A. funestus*. Nevertheless, the potential for this mechanism is supported by the fact that previous studies (Stefansson *et al.*, 2005; Schaeffer, 2008; Lowry & Willis, 2010) as well as my results show that the fitness of inversions can vary in space.

Caveats and extensions: My strategy here uses statistical analyses of genotype frequencies to make inferences about selection and mating. A fundamental strength of this approach is that the data come from natural populations. Model-based estimation is the basis for the vast majority of inferences about how selection acts on genetic variation in other species, including humans.

This approach, however, also has weaknesses. I noted earlier that there is no formal genetic analysis proving that inversion 3Ra is the causal agent of the patterns seen in Figure 2.1. Controlled crosses to test the effects of the inversion are not possible because no karyotyped breeding colony of *A. funestus* is available. Another technical limitation of this system is that the data are based on classical cytogenetic methods that can be applied only to one sex at one point in the life cycle (half-gravid females). Molecular markers are not available for the inversion breakpoints of *A. funestus*. Those would be valuable to develop in the future since they would allow further tests of the model, for example by sampling inversion frequencies at different life stages and by correlating the genotypes of females and the sperm of the males they mated.

Progress towards identifying the loci involved in local adaptation and isolation is likely to be slow, even in those species where additional genetic resources are available. Since recombination is almost completely blocked in inversion heterozygotes, it is

generally not possible to map phenotypic effects using recombinants. For now, inversion 3Ra is perhaps best regarded as a single (albeit very large) quantitative trait locus segregating for two alleles that have major effects on multiple phenotypes.

It is conceivable that the cline in inversion 3Ra shown in Figure 2.1 is in fact caused by selection on loci outside of the inverted region. If so, then those loci must be in very strong linkage disequilibrium with 3Ra, which would require both strong epistatic selection and tight linkage between the causal locus and the inversion. That would change the quantitative results about selection, but likely would not change the qualitative conclusion that there is strong assortative mating and strong natural selection that varies in space. Previous studies using molecular markers have highlighted the importance of gene flow between natural populations of this mosquito in Cameroon (Cohuet *et al.*, 2005; Ayala *et al.*, 2011b). These results are consistent with the assumptions of very strong linkage disequilibrium between those loci and the inversion.

The inferences depend on the assumptions made in the model. I found that the four models that best-fit the data have significantly better statistical support than several alternatives: models that assume fitnesses do not change in space, that mating is random, and that there are fewer than three ecological zones. There are, however, additional alternatives that might be considered. The model assumes that selection acts on viability but not fertility. That assumption is justified by experimental results showing that karyotypes affect larval survival in *Anopheles* (Rocca *et al.*, 2009) and that many paracentric inversions have little if any effect on fertility in dipterans (Hoffmann & Rieseberg, 2008b). I have run simulations that assume selection acts on fertility rather than viability and found that the results are not much changed (results not shown). I assumed densities are equal everywhere. The position and shape of clines are affected by variation in density (Barton & Gale, 1993), which could bias the parameter estimates. I

have no way to estimate density, however, so it does not seem possible to make progress on this issue with the available data.

The inferences are partly limited by the sampling protocol. Since the data come only from the wet season, I am not able to include seasonal variation in the model. The frequencies of 3Ra in *A. funestus* vary between the wet and dry seasons in the dry savannah of West Africa (Guelbeogo *et al.*, 2009). These changes are relatively small, however, and certainly less dramatic than the frequency differences seen between the three zones in the transect. Consequently, there might be seasonal changes in the parameters of my model, but that the qualitative patterns from the analyses are robust. A second issue involves the geography of the samples. The samples are limited to the one-dimensional transect along a highway. The analyses are not able to distinguish between a one-dimensional stepping-stone and an alternative model in which mosquitoes are distributed uniformly in two-dimensional space. Consequently, I have little confidence about any inference concerning spatial distribution and migration patterns.

Inversion 3Ra is only one of several common inversions in *A. funestus*. Intriguing observations are that their frequencies vary geographically, and that there is strong linkage disequilibrium between some pairs of inversions even when they occur on different chromosomes (Costantini *et al.*, 1999; Dia *et al.*, 2000; Ayala *et al.*, 2011b). Disequilibria could result from epistatic selection between inversions and/or from spatial variation in selection and migration. Extending the modeling framework to multiple inversions could test these possibilities.

These results can be viewed from two perspectives. On the one hand, they provide estimates of the strengths of evolutionary forces and the amount of reproductive isolation that results. An alternative view is that this study generates hypotheses about the phenotypic effects of the inversion. An important next step will be to test these

hypotheses using other approaches. They might include experimental studies of the effects of the inversion on mating behavior and viability, for example, and analyses of patterns of neutral molecular variation inside and outside of the inversion. While the constraints of *A. funestus* discussed above make some approaches infeasible, it should be possible to pursue them in other species (such as *A. gambiae*). Together, these complementary research programs offer the prospect of quantifying the evolutionary forces that act on inversions and ultimately settling some of the oldest debates in evolutionary genetics.

Chapter 3: Cryptic recombination in the ever-young sex chromosomes of Hyliid frogs*

INTRODUCTION

Recombination determines the fate of sex chromosomes. Suppressed recombination between the X and Y (or Z and W) chromosomes facilitates their divergence, ultimately leading to the degeneration of the Y (or W) (Charlesworth & Charlesworth, 2000). Recombining regions of sex chromosomes, on the other hand, remain homomorphic and show little or no signs of degeneration. In general, sex chromosomes are expected to evolve towards suppression of recombination favored by the interaction between the sex-determining locus and sex-specific selection on neighboring genes (reviewed in Otto *et al.* (2011)). This reduction of recombination and subsequent degeneration can happen rapidly after the origin of the sex chromosome pair (Bachtrog *et al.*, 2008). Many homomorphic sex chromosomes are young (e.g. in medaka; (Kondo *et al.*, 2001)) and stickleback (Peichel *et al.*, 2004)), and in these cases there may not yet have been enough time for the X and Y to diverge. Other homomorphic sex chromosomes, however, are old. Some culicid mosquitoes (Rai, 2010), boid snakes (Ohno, 1967), and ratite birds (Janes *et al.*, 2009) show no signs of divergence between the sex chromosomes even though they are tens to hundreds of millions of years old. In these cases, ongoing recombination is thought to prevent the sex chromosomes from diverging (Janes *et al.*, 2009).

Some sex chromosomes, nonetheless, remain homomorphic even in the apparent absence of recombination. This may be the case of the European tree frogs (*Hyla* spp.).

* Significant portions of this work have been previously published as Guerrero, Kirkpatrick & Perrin (2012) *Journal of Evolutionary Biology* 25:1947-1952. Author contributions: Mark Kirkpatrick, supervisor; Nicolas Perrin collected the microsatellite data I analyze in this chapter.

Crossing experiments have never detected recombination in males (Berset-Brändli *et al.*, 2008). Karyotype studies are not able to distinguish the X and Y (Anderson, 1991), however, despite the fact that this X-Y system is at least 5 million years old (Stöck *et al.*, 2011). Recent molecular data present an unexpected pattern: the X and Y chromosomes within a species seem to be more similar than the X chromosomes or the Y chromosomes from different species (Stöck *et al.*, 2011). This result suggests that the sex chromosomes have recombined since the speciation events, an implication at odds with the lack of recombination observed in male frogs.

A possible solution to this apparent paradox is that the X and Y in fact do recombine occasionally, producing Y chromosomes purged of deleterious mutations. These ‘rejuvenated’ Y chromosomes then sweep through the population, maintaining the homomorphism between the sex chromosomes. One mechanism by which recombination between X and Y might happen is the “fountain of youth” hypothesis (Perrin, 2009). Under this hypothesis, recombination is always suppressed in males, but occasionally sex-reversed XY females appear and produce gametes with recombined sex chromosomes. This idea is supported by evidence for sex reversal in *Hyla japonica* (Kawamura & Nishioka, 1977) and by data from *Rana temporaria* showing that recombination depends on an individual’s phenotypic, not genetic, sex (Matsuba *et al.*, 2010). Alternatively, it is possible that recombination happens in males at very low rates that have not been detected in lab crosses. Simulation results suggest that recombination rates as low as 10^{-5} could allow the purge of deleterious mutation and maintenance of homomorphy (Grossen *et al.*, 2012). Therefore, even though hundreds of meiosis events have been observed (Berset-Brändli *et al.*, 2008; Stöck *et al.*, 2011), these may not have been enough to detect such low rates.

Regardless of the mechanism by which it may happen, how much X-Y recombination is consistent with the molecular data in *Hyla*? The available molecular markers can be used to infer X-Y recombination, as they may hold signatures left by this process (Figure 3.1). In the absence of genetic exchange, sex chromosomes within species will diverge just as chromosomes from different species do. If recombination has been suppressed since the origin of the sex-determination system, the X and Y within one species are expected to be as diverged as the X in that species and the Y in another species. A different pattern is expected when X and Y recombine. In this case, divergence within species is inhibited and the sex chromosomes within species can be more similar to each other relative to those of closely related species. In *Hyla*, qualitative patterns consistent with recombination have been found. No rigorous test, however, has been carried out to determine if there is statistically significant evidence for recombination. If there were, a quantitative estimate for its rate would suggest how much recombination might be sufficient to maintain X-Y homology.

Here I use a novel approach to infer X-Y recombination from population-genetic data. The goal is to estimate the rate of X-Y recombination relative to that between X chromosomes in three species of European tree frogs (*Hyla arborea*, *H. intermedia*, and *H. molleri*). For these three species, there are samples from seven sex-linked microsatellite markers at known positions in the genetic map of the X chromosome. I analyze these data using Approximate Bayesian Computation, or ABC (reviewed in Beaumont (2010)). The results provide strong statistical support for the hypothesis that there is recombination between X and Y. I estimate that its rate is about 10^5 times smaller than between X chromosomes.

METHODS

My main interest is to estimate α , the factor by which recombination between X and Y chromosomes is reduced relative to its rate between two X chromosomes. For this purpose, I assume an evolutionary model for the phylogeny, the recombination pattern for the sex chromosomes, and the evolution of microsatellite loci. This model includes seven parameters that specify, for example, the XY recombination rate, the ages of the speciation events, and the microsatellite mutation rates. I then generate a large number of simulations under this model using a large number of parameter values. Next, I determine which of these simulations produce patterns that most closely resemble those seen in the real data. The parameter values used in the best simulations (the 0.1% closest to the data) give us the estimated posterior distributions for all parameters, including (and most importantly) α . I now describe the three components of my approach: the evolutionary model, a brief account of the data, and the estimation using ABC.

Evolutionary model: I built a coalescent-based model to simulate the gene trees of neutral sites linked to sex chromosomes in three species with known phylogeny (Figure 3.2). These sex chromosomes originated from an autosomal pair in the common ancestor species T_Y generations ago (i.e. there is no initial X-Y divergence). Two speciation events follow at times T_1 and T_2 , giving rise to the three species. I assume these species do not hybridize (an assumption supported by molecular data, (Verardi *et al.*, 2009; Stöck *et al.*, 2012)), and they each have a constant population size N . The chromosomes evolve under the standard neutral model for recombining sex chromosomes (Kirkpatrick *et al.*, 2010). The sex-determining region (SDR) is at an unknown position x_{SDR} on the genetic map of females, and it behaves as a single locus with two alleles (X, Y). For a pair of loci that recombine at a rate r in females (XX), their recombination rate in males (XY) is αr . The model makes no assumption about the mechanism by which recombination happens,

however. It is therefore also possible to think of α as the rate of sex-reversal in XY individuals (which would, as females, recombine at rate r). I assume there are no selected sites on the sex chromosomes.

After the gene tree has been simulated, I simulate the occurrence of mutations along its branches. Mutation rates are locus-specific with mean μ and gamma distributed, $\Gamma(a, a/\mu)$, where a is the shape parameter. I assume the generalized step-wise mutation model of microsatellites (reviewed in Estoup *et al.* (2002)), which allows for varying step sizes. In this mutation model, the size of each mutation is a random variable drawn from a geometric distribution. This distribution is described by p , the last parameter in the model. The value of p gives the fraction of mutations that are a single step (e.g. at $p = 0.78$, 78% of mutations are one step, 13% are two steps, etc.), and it ranges from zero to unity. I make the simplifying assumption that mutation rates are equal in males and females.

Data: I used a dataset of seven microsatellite markers previously published by Stöck *et al.* (2011). The samples come from *H. arborea* ($n = 49$ X and 13 Y), *H. intermedia* ($n = 72$ X and 24 Y) and *H. molleri* ($n = 17$ X and 13 Y). The observed distributions of alleles for these microsatellite markers are shown in Figure 3.

Estimation of α using ABC: The Approximate Bayesian Computation consists of three steps. First, I produce a large set of simulations under an assumed evolutionary model with parameter values drawn from prior distributions. Second, I use summary statistics to compare the simulations to the observed frog chromosome data. Third, I take those simulations that most closely resemble the data to produce a posterior distribution for the parameters of interest.

The model has a total of nine parameters: T_1 , T_2 , T_Y , N , x_{SDR} , μ , a , p , and the parameter of interest, α . The prior for T_1 is a normal distribution (mean = 2.75×10^6 , sd =

10^6) truncated between 7.5×10^5 and 6×10^6 . This distribution is based on the estimated species divergence time and credible interval (Stöck *et al.*, 2011), and assuming 2 years per generation. The prior for T_Y is a uniform distribution between 1 and 5×10^6 generations before T_1 (which I denote as $T_Y \sim U[1, 5 \times 10^6] + T_1$). Population size is unknown for these species, so I set a wide uninformative prior: $\log_{10}(N) \sim U[2.5, 5]$, which corresponds to an exponential distribution truncated at $10^{2.5}$ ($= 316$) and 10^5 . The position of the SDR is also unknown, so I allow it to vary uniformly along the consensus genetic map for the X. The prior for the mean mutation rate, μ , is $\log_{10}(\mu) \sim U[-6, -3]$ and the prior for the shape parameter is $a \sim U[8, 15]$. These values produce mutation rate distributions similar to those observed in nature (Seyfert *et al.*, 2008).

I set a prior for the relative recombination rate, α , after evaluating the amount of recombination present in a preliminary set of simulations. For this purpose, I executed 10^6 simulations with $\log_{10}(\alpha) \sim U[-11, -7]$ and other priors as described above, and recorded the number of recombination events that happened in each simulation. In this preliminary set, only simulations with $\alpha > 2.5 \times 10^{-10}$ had one or more recombination events, implying that α is effectively zero at about this value. Consequently, I assume the prior for $\log_{10}(\alpha) \sim U[-10, -2]$ in the main body of the simulations.

The last two parameters in the model, T_2 and p , are treated as fixed values. Experience shows that ABC has difficulty converging as the number of free parameters increases (Beaumont, 2010). Preliminary analyses indicated that T_2 and p have little impact on the estimate of α . For this reason, I fix $T_2 = \frac{1}{2} T_1$, and $p = 0.78$ (based on estimates of the microsatellite step-size distributions from humans (Dib *et al.*, 1996)).

The summary statistic for the ABC analyses measures the amount of X-Y divergence within species relative to X-Y between species:

$$D_{ij} = 2 \frac{\delta_{ii} + \delta_{jj}}{\delta_{ij} + \delta_{ji} + \delta_{ii} + \delta_{jj}},$$

where δ_{ij} is the square of the difference in mean allele length between X of species i and the Y of species j at a given locus. The δ_{ij} are a natural choice here because their expected values increase linearly with the time since divergence under a simple random walk model for microsatellite evolution. High levels of recombination will slow the divergence of X and Y within species (as in the right hand side of Figure 3.1), leading to values of D_{ij} that approach zero. On the other hand, suppressed recombination allows the X and Y to diverge. If recombination is completely suppressed and the sex chromosomes are very much older than the speciation events, D_{ij} is expected to approach unity. In principle, divergence between the X and Y within species may be greater than between species, making it possible to have $D_{ij} > 1$.

I can calculate a value of D_{ij} for each locus between each pair of species. With three species and seven markers, it is possible to calculate 21 values. Some loci in some species are null or uninformative, however, which reduces to twelve the number of D_{ij} values that can be computed. I further reduce the number of statistics by using the mean of *H. intermedia* and *H. molleri* in their comparison to *H. arborea* for three markers (5-22, M2, and M3). After these reductions, there are a total of nine D_{ij} summary statistics for the ABC analyses (Figure 3.4).

By their very nature, summary statistics discard some of the information in the data. For example, ours discard information about variation at the microsatellite loci within species. While additional summary statistics could be added, including a large number of summary statistics tends to increase the variance of estimators and introduce bias (Wegmann *et al.*, 2009). I therefore elected to limit ours to D_{ij} statistics, which quantify the key features of the pattern of interest.

I estimated α by the rejection sampling ABC algorithm (Tavare *et al.*, 1997) implemented in the ABCtoolbox software (Wegmann *et al.*, 2010). I used the ABC-REG algorithm described by Wegmann *et al.* (2009), which executes a post-sampling regression adjustment before estimating the posterior distribution. I simulated a total of 10^7 runs using parameters values drawn from the priors described above. This number of simulations was chosen after preliminary analyses showed that 5 million runs were adequate to obtain convergence on the posterior distribution of α . I use the nine D_{ij} summary statistics to calculate the Euclidean distance between the simulated and observed datasets and retain the best 10^4 simulations. From this retained set, I estimate the posterior distribution of α . To validate the model, I inspected the distributions of the D_{ij} statistics in the retained simulations and verify that the observed D_{ij} values for *Hyla* are within the simulated range.

RESULTS

My main interest is in α , the relative recombination rate of X and Y chromosomes. Its posterior distribution is shown in Figure 3.5. The mode, median and mean are 4.6×10^{-6} , and the 99% highest posterior density interval (HPDI) is $[1.1 \times 10^{-7}, 1.4 \times 10^{-4}]$ (Figure 3.5). This interval does not include the value estimated to be effective zero recombination (2.5×10^{-10}), therefore I conclude that α is significantly different from zero. Using the mode of α , the length of the Y chromosome is approximately 0.0008cM. This implies that recombination between X and Y will happen about once among 10^5 individuals. With population sizes smaller than that number, several generations may pass between recombination events.

The nine D_{ij} statistics observed in *Hyla* suggest a pattern of heterogeneous divergence along the chromosome (Figure 3.4). Towards the center of the genetic map I

observe the highest levels of divergence in X and Y within species (showing values of D_{ij} close to and greater than unity). As we move towards the end of the chromosome we find the lower divergence between X and Y, and the last marker (A-103) has the lowest D_{ij} value. This pattern allows some speculation about the position of the SDR, probably around the 82cM mark in the X chromosome. It is not surprising that my estimate for x_{SDR} is close to that position (although with very low confidence).

To confirm the conclusion that α is significantly greater than zero, I carried out one more set of analyses. I first ran 10^5 simulations fixing $\alpha = 0$ while leaving the prior distributions for the other parameters unchanged. I took each of these simulations as independent observed datasets and carried out 10^5 new ABC estimations (as described for the main analysis and reusing the set of 10^7 simulations for the computation). The purpose is to evaluate how likely is it to obtain a value of $\alpha = 4.6 \times 10^{-6}$ (the estimated mode for *Hyla*) when in fact there is no X-Y recombination. Given that in only 0.01% of the cases I obtained modes of $\alpha > 4.6 \times 10^{-6}$, I can safely reject the possibility of falsely concluding $\alpha > 0$ from the observed data.

To assess potential biases in the estimation, I carried out a test of the coverage properties of the posterior estimates (as recommended by Wegmann *et al.* (2009)). For this purpose, I simulated 1000 datasets drawn from the prior distribution and carried out an ABC estimation for each set (reusing the 10^7 runs from the main analysis). I then obtain the distribution of posterior quantiles (that is, the quantile in the posterior distribution where the real value of the parameter lies). If the parameter estimates are not biased, this quantile distribution is expected to be uniform (Wegmann *et al.*, 2009). In the case of α , I did not find any deviations from a uniform distribution (Kolmogorov-Smirnov test, $p = 0.26$), indicating that there is no evidence of bias in my estimate.

DISCUSSION

I estimate X-Y recombination in *H. arborea*, *H. intermedia*, and *H. molleri* to be significantly greater than zero, confirming the previous suggestion by Stöck *et al.* (2011). My analyses, however, imply that recombination events will be difficult to observe directly. I estimate that recombination occurs only in about 1 out of 10^5 XY individuals. The findings further suggest very low recombination rates might be sufficient to prevent degeneration of the Y and maintain the homomorphism observed in the Hyliid sex chromosomes.

The upper bound for the map length of the Y chromosome is 0.025cM, which is at least an order of magnitude smaller than the recombining regions of sex chromosomes observed in other species (reviewed in Otto *et al.* (2011)). The pseudo-autosomal region of the mosquito *Culex tarsalis*, for example, is about 12cM in heterogametic and 50cM in homogametic individuals (implying $\alpha = 0.24$).

These findings are consistent with the fountain of youth hypothesis, which posits that rare X-Y recombination events in XY sex-reversed individuals rejuvenate the Y chromosome (Perrin, 2009). This analysis is not, however, a test of that hypothesis. There is evidence for recombination of the Y, but can make no inference about whether that results from sex reversal (as envisioned in the hypothesis) or from some other mechanism.

While my model assumes that the X and Y are evolving neutrally, it is certain that some form of selection acts at some loci on these chromosomes. While selection would affect the quantitative estimate, I suggest that my qualitative conclusion that X and Y recombine is robust. In this regard, consider three types of selection that could be at work. The first is background selection, which occurs when deleterious mutations cause a reduction in the effective population size of Y chromosomes and hence lead to a

reduction in diversity at neutrally-evolving loci (reviewed in Charlesworth (2012)). The neutral diversity expected on the sex chromosomes therefore depends not only on the rate of X-Y recombination, but also on the intensity of background selection and the positions of sites under selection. Since these last two quantities are unknown, it is not possible to make clear predictions about how selective forces might change my estimate for the X-Y recombination rate. The changes in neutral diversity, however, should not lead to the observed pattern of similarity between X and Y within each species. Consequently, the conclusion that these chromosomes recombine seems robust to the presence of background selection.

Selective sweeps on the Y driven by positive selection, which are envisioned in the fountain of youth hypothesis, are a second way in which the neutrality assumption in my model might be violated. Sweeps of recombinant Y chromosomes would strongly reduce divergence between X and Y. In this case, a smaller recombination rate than what I estimated would suffice to produce the patterns in the data. Nevertheless, recombination is still needed to explain the basic pattern observed.

Third, these sex chromosomes may carry balanced polymorphisms maintained by sex-specific selection (Kidwell *et al.*, 1977; Rice, 1984; Otto *et al.*, 2011). This type of selection will increase neutral divergence between recombining X and Y chromosomes (Kirkpatrick & Guerrero, unpublished). The result would be to cause the model to underestimate the recombination rate between the X and Y, making the estimate conservative.

It might seem possible to make further quantitative inferences using patterns of neutral diversity observed at the microsatellite loci within each sex chromosome and each species. (Recall that the summary statistics discard this information.) Berset-Brandli *et al.* (2007) found that molecular diversities on Y, X, and autosomes in *H. arborea* fall

roughly in the ratio 1 :: 3 :: 4. Such ratios are expected for neutrally evolving sex chromosomes that do not recombine. As I have just discussed, however, these chromosomes do recombine and are very likely under some form(s) of selection. Therefore, it seems difficult to draw further quantitative conclusions without additional data.

Several caveats apply to the results. I assume a constant recombination rate for the three species through time. These species, however, have evolved independently for a considerable time and differences in present recombination rates are plausible. This possibility is perhaps not such a concern, given that female genetic maps do not differ between species (Stöck *et al.*, 2011). The X-Y recombination rate could also vary along the chromosome, through chromosomal inversions or other local recombination modifiers. The data available, though, do not provide us with enough power to implement more complex models.

The results suggest the possibility that very low rates of recombination suffice to prevent the Y chromosome from degenerating. Grossen *et al.* (2012) used simulations to study the purging of deleterious mutations on the Y by occasional recombination events with the X, and how this process might prevent the X and Y from diverging. They found that recombination rates on the order of 10^{-4} (or smaller, depending on the population size) are enough to keep sex chromosomes homomorphic. Intriguingly, this value is not too far from what I estimate. It is impossible to make direct comparisons between my results and that model, however, since they are based on such different assumptions. The question of how much recombination is adequate to maintain homomorphic sex chromosomes certainly warrants further study.

The current data paint but a coarse picture of the evolution of the whole sex chromosome. It is possible that strong background selection is acting on the Y

chromosome and that genes under sex-specific selection have accumulated in this region. These two forces combined set the stage for the fountain of youth hypothesis, the mechanism that could be responsible for the ongoing X-Y recombination observed. To explore these possibilities and expand on the simple model presented here, however, it is necessary to obtain genomic data along with complementary approaches that inform us about the function of genes that reside on the sex chromosomes.

Chapter 4: The Evolution of Chromosome Fusions by Local Adaptation

INTRODUCTION

Understanding the evolution of chromosomal rearrangements is essential to making sense of genome biology. Variation in chromosomal arrangements, within and between species, is ubiquitous in nature and plays important roles in adaptation and reproductive isolation (White, 1978; Barton & Hewitt, 1985; Kirkpatrick, 2010). Nearly a century ago, chromosomal polymorphisms allowed major breakthroughs in the studies of inheritance and geographic clines (Robertson, 1916; Sturtevant & Beadle, 1936; Dobzhanski, 1947). Classic models of speciation were built around the idea that chromosomal rearrangements fixed between species cause hybrid sterility, due to unbalanced meiotic products or segregation problems in heterozygotes (White, 1973). With the advent of molecular genetics, the study of chromosome rearrangements fell off the spotlight for some years. However, the past decade of genomic studies have highlighted chromosome rearrangements and structural variation as key factors in evolution of the genome (Sankoff & Nadeau, 2003; Heslop-Harrison, 2012).

Chromosome rearrangements may affect large segments of DNA. In fact, the largest involve whole chromosome arms. This is the case of chromosome fusions and whole-arm translocations, which occur in several ways. Centric fusions happen when two telocentric chromosomes join their terminal centromeres to become a metacentric chromosome. Tandem fusions involve the joining of two chromosomes at their telomeres, followed by inactivation of one of the two centromeres in the new chromosome. This occurred in the ancestry of human chromosome 2, which is a fusion that occurred since our most recent common ancestor with chimps (Ijdo *et al.*, 1991). In whole-arm reciprocal translocations, two chromosomes break at the centromeres and switch arms.

When this happens between two acrocentric chromosomes (with near-terminal centromeres), the product is a large metacentric and a small chromosome that is generally lost. These rearrangements are known as Robertsonian (Rb) translocations. Because their outcome is hard to distinguish from a centric fusion, the two terms are commonly used interchangeably.

Fusions are probably the most common rearrangements in animals (King, 1993). Centric fusions are estimated to happen in 0.1% of human meiosis (Hamerton *et al.*, 1975), and occur frequently in ovine and bovine cattle (Bruere & Ellis, 1979; Pagacova *et al.*, 2009). A wide variety of taxonomic groups show labile karyotypes (White, 1973), from plants (Chang *et al.*, 2013), to trematodes (Grossman *et al.*, 1981), ants (Santos *et al.*, 2012), fish (Galetti *et al.*, 2000), and mammals (Medarde *et al.*, 2012). Some groups show extreme karyotype variation. In the house mouse, for example, 106 out of the possible 171 fusions between its 19 autosomes have been observed (Gazave *et al.*, 2003). Fusions are also common between autosomes and sex chromosomes, producing neo- sex chromosomes that allow species differentiation (McAllister, 2003; Kitano *et al.*, 2009). Fusions with autosomes are the main mechanism by which the pseudo-autosomal region of the sex chromosome expands (Otto *et al.*, 2011), allowing for the continuous evolution of sex chromosomes.

How do chromosome fusions evolve? In models of chromosomal speciation it is assumed that rearrangements are underdominant. During meiosis, crossover between chromosomes with different arrangements may result in unbalanced meiotic products. Thus, heterozygotes for an arrangement are expected to have a fitness disadvantage and their evolution would require strong genetic drift and population structure (Bush *et al.*, 1977; White, 1978; Lande, 1979). Meiotic drive may also cause the spread of fusions (Hedrick, 1981; Coyne *et al.*, 1991). An interesting mechanism by which this could

happen involves female meiotic drive at a whole-karyotype scale (Pardo-Manuel de Villena & Sapienza, 2001). Central to this hypothesis is the orientation of the meiotic spindle, which can cause different chromosome forms to be pulled into the egg with higher probability. The polarity of the spindle determines whether acrocentric or fused arrangements are favored in the complement, producing genomes with mostly one type of chromosome. Over evolutionary time, the polarity may switch and cause the whole genome to rearrange. Evidence for this mechanism comes mostly from mammals (Pardo-Manuel de Villena & Sapienza, 2001), but other taxa show patterns consistent with the hypothesis.

Fusions physically link genes previously in different chromosomes, a major feature that could drive their evolution. The fusion not only reduces recombination between formerly unlinked genes, but may also reduce recombination rates within each arm (Dumas & Britton-Davidian, 2002). Therefore, fusions could be favored in situations when tight linkage is beneficial (Charlesworth, 1985). This mechanism has received little consideration as a possible force driving chromosome fusions (see (Dumas & Britton-Davidian, 2002) and (Giménez *et al.*, 2013) for exceptions), but there is some empirical evidence for it. For example, in *Drosophila americana* a fusion between an autosome and the X chromosome appears to have evolved by reducing recombination between sexually antagonistic genes and the sex-determining locus (McAllister, 2003).

Local adaptation favors reduced recombination between alleles adapted to the same habitat (Charlesworth & Charlesworth, 1979). This is a well-studied mechanism for the evolution of chromosome inversions, another type of rearrangement that drastically reduces recombination when in heterozygous form (Kirkpatrick & Barton, 2006). Inversions that capture two or more locally adapted alleles spread in the population where those alleles are beneficial. Individuals that are heterozygote for such inversions avoid

introgression of maladapted alleles from incoming migrants. By reducing recombination, inversions preserve the divergence between populations in the presence of gene flow, setting the stage for potential speciation.

Can fusions evolve by their effect on recombination, as inversions do? Here, I analyze the theoretical conditions under which a chromosome fusion invades a population when it carries locally adapted alleles. For this purpose, I assume a two-locus continent-island model. I find the conditions when a new fusion will spread and find its frequency at equilibrium. I complement the analytical approach with numerical results for a two-deme model. Broadly, I find that under the continent-island model fusions evolve when selection on the locally adapted loci is weaker than recombination between them. The same is not true for the two-deme model, where the fusion is always beneficial. In that case, I find that the recombination rates in homozygotes and heterozygotes for the fusion are key determinants of polymorphism. I focus mostly on centric (or Rb) fusions. The results also apply, however, to tandem fusions, reciprocal translocations, and to holocentric chromosomes.

MODELS AND RESULTS

I investigate the evolution of a chromosome fusion when it reduces recombination between two loci under opposite selective pressures in two different habitats. Prior to the origin of the fusion, two separate chromosomes (unfused form U) carry the two polymorphic loci (**A** and **B**, with alleles A, a and B, b respectively). This polymorphism is maintained by migration and selection, both happening in the haploid phase. Fitness effects (s) are additive and equal at each locus, and the fusion has no effect on viability. The appearance of the fusion (chromosome form F) causes **A** and **B** to be linked, recombining at a rate r_f in fused homozygotes. In heterozygotes for the fusion (carrying U

and F chromosomes), I consider the recombination distance between each loci and the point of fusion. In the case of a centric fusion, for example, the point of fusion would correspond to the centromere. I make the simplifying assumption that both loci are at equal recombination distance (r_h) from the point of fusion. I assume all evolutionary forces are small, and that migration is much weaker than selection and recombination (specifically, $s = O(\epsilon)$, $r_h = O(\epsilon)$, $r_f = O(\epsilon)$, and $m = O(\epsilon^2)$, where $\epsilon \ll 1$).

Continent-Island model

I first analyze a continent-island model in which unfused chromosomes carrying alleles A and B are fixed in the continent. These alleles are deleterious in the island, where the fusion appears capturing the beneficial alleles (a and b). Viabilities in the island are defined as 1 for haplotypes with alleles ab , $1 - s$ for Ab and aB , and $1 - 2s$ for AB . Migration is unidirectional, at rate m from the continent. In the island there are eight haplotypes with frequencies p_{ijk} , for chromosomes of form i ($=U, F$), allele j at locus **A**, and allele k at locus **B**. I denote $p_{(l)}$ as the sum of all haplotypes with form or allele l (e.g., $p_{(F)} = p_{FAB} + p_{FAb} + p_{FaB} + p_{Fab}$).

Conditions for invasion: Now I derive the per-generation change in haplotype frequencies, which allows us to analyze the invasion of the fusion. I am interested in the instability of the system caused by the increase in frequency of the F haplotypes. I therefore reduce the variable system to the frequencies of the four fused haplotypes. I obtain simple equations for the per-generation change in frequency of fused haplotypes, $\Delta p_{Fjk} = \Delta_m p_{Fjk} + \Delta_s p_{Fjk} + \Delta_r p_{Fjk}$, as the sum of the changes at each of the steps in the life cycle (migration, followed by selection and recombination).

Frequency changes due to migration and selection steps are straightforward. All migrants from the continent are in U karyotypes, so after migration the frequency of

every fused haplotype with changes by $\Delta_m p_{Fjk} = -m p_{Fjk}$. The change due to selection is $\Delta_s p_{Fjk} = p_{Fjk} W_{Fjk} / \bar{W}$, where \bar{W} is the mean fitness of the population and W_{Fjk} is the viability of the haplotype. The equations for recombination are slightly more complicated. In the island, eight haplotypes form 64 possible diploid genotypes and recombine at three different rates. The changes in frequency of F haplotypes after recombination are shown in Appendix C.

The selective advantage of the fusion and indicator of its invasion is λ , the leading eigenvalue of the Jacobian matrix for the Δp_{Fjk} system evaluated at the pre-fusion equilibrium. To obtain a closed-form expression for λ , I approximate the Δp_{Fjk} to $O(\epsilon^2)$. I find that the fusion has a relative selective advantage

$$\lambda = m(1 - 2 \frac{r_h}{r_h + s}),$$

which implies that the fusion is favored when $r_h < s$. This result suggests that the spread of the fusion is driven by migration. The selective advantage is maximum at $\lambda = m$ when the rearrangement fully suppresses recombination (i.e., at $r_h = 0$), which is consistent with previous results by Kirkpatrick & Barton (2006). Another implication is that the recombination in fused homozygotes, r_f , does not play any role in determining the spread of the fusion. The rearrangement is beneficial because it reduces introgression from maladapted alleles carried by migrants with the U haplotype. There is no benefit to reducing r_f , however, because there is no source of fused haplotypes with maladapted alleles.

Figure 1 compares the simple analytic approximation for λ (blue dashed lines) and the exact eigenvalue (blue solid lines) obtained numerically for two cases. The approximation is better when evolutionary forces are weak, but still holds at values of selection that are close to violating my assumptions (namely, $s \ll 1$). Given small evolutionary forces and a when a continent-island model is appropriate, chromosome

fusions will evolve as long recombination in heterozygotes is smaller than the strength of selection at locally adapted loci.

In a more general model, I allow recombination and selection to be different between loci. I define $r_{h,A}$ and $r_{h,B}$ as the recombination rates in heterozygotes at loci **A** and **B** respectively, and s_A and s_B as their selective effects. In this case, I find that the fusion is favored when $r_{h,A}r_{h,B} < s_A s_B$. This result implies that the fusion may evolve even if one of the loci is not tightly linked to the point of fusion. In an extreme example, a centric fusion will always evolve if it has one perfectly linked locus (i.e., when $r_{h,A}$ or $r_{h,B}$ equal 0), regardless of the position of the other locus or the strength of selection on them.

Frequency of the fusion at equilibrium: I now find an approximation for the frequency of the fusion, $p_{(F)}$, at equilibrium. The model assumptions are the same as above, with eight haplotypes leading to a system of seven equations. I can obtain equilibrium frequencies when the advantage of the fusion is strong. In this case, I assume maladapted alleles (*A* and *B*) and unfused haplotypes are extremely rare. Specifically, I assume that the frequencies $p_{(U)}$, $p_{(A)}$, and $p_{(B)}$ are $O(\phi) \ll 1$. After approximating the equation system to terms of $O(\epsilon^2, \phi^2)$, I find a solution at

$$p_{(F)} = 1 - \frac{m + 8r_h^2}{2(s - r_h)}$$

This result confirms the observation that the recombination rate in homozygotes, r_f , does not play an important role in this system. Once more, the relationship between r_h and s is highlighted. The fusion approaches fixation as the difference between the two parameters increases, and disappears when $r_h \geq s$.

Two-deme model

Next, I study the evolution of a fusion that captures locally adapted alleles in a system of two demes connected by migration. Forces of selection and migration are

symmetric between the demes. Assumptions on the strength of evolutionary forces are the same as in the continent-island model. Under these assumptions, I am not able to present a closed-form expression for the selective advantage of a new fusion. Numerical exploration of this model shows that the selective advantage of the fusion is always positive, declining towards zero as r_h increases (i.e., the fusion is neutral when it does not reduce recombination).

To explore the equilibrium frequencies under the two-deme model, I carry out numerical recursions of the system. I allow the system to reach an initial equilibrium (when all $\Delta p_{ijk} < 10^{-12}$) before introducing the fusion in deme 1 at $p_{FAB} = 0.001$ and continuing the recursions to equilibrium again. I carry out these recursions at parameter values that satisfy the invasion conditions for the continent-island model, as those are likely to represent situations in which the forces driving the fusion are strong.

The fusion spreads close to fixation in deme 1 always, carrying the alleles locally adapted to that habitat. The rearrangement does not always reach a polymorphic equilibrium, as fused chromosomes may also become fixed in deme 2 because fusions carrying alleles favored in deme 2 are produced (as long as $r_h > 0$) (Figure 2). The potential benefit of a chromosomal rearrangement comes from the reduction of introgression from maladapted migrants. In deme 2, the fusion becomes beneficial if homozygotes for the fusion recombine less than heterozygotes, therefore reducing introgression from migrant fused haplotypes from deme 1. Roughly, this happens when $r_f < 2r_h$ for small recombination rates. (To understand the factor of two in the inequality, recall r_h is the recombination rate between each locus and the fusion point, while r_f is the rate between loci **A** and **B**.) On the other hand, when recombination in heterozygotes is less than in fused homozygotes the rearrangement does not spread in deme 2. Then, the fusion will be in strong association with alleles a and b , and may appear as a nearly-fixed

karyotype difference between demes. The outcome in this case is similar to the one produced by inversions, in which only the recombination in heterozygotes is strongly reduced.

Although conditions for a polymorphic equilibrium in the two-deme model are fairly restrictive, I find that transient polymorphisms can persist for very long times. Figure 3 shows a numerical case for such dynamics. The fusion invades deme 1 in about 10^4 generations, but stays at low frequency in deme 2 for over 10^6 generations before spreading there. The absolute time scale for the evolution of the fusion is determined largely by migration (as the invasion analysis for the continent-island model showed). However, the lag between demes is longer when both recombination rates are small and r_f is just slightly lower than $2r_h$. Under those conditions, the advantage of haplotype F over U is vanishingly small, so the rate of increase is slow.

DISCUSSION

My models show how chromosome fusions can evolve by their effect on recombination of locally adapted alleles. In a continent-island model, rare fusions spread when recombination in heterozygotes is less than the strength of viability selection ($r_h < s$), independent of the recombination in fused homozygotes. When two demes exchange migrants, however, rare fusions are always favored to spread and will ultimately fix in both demes.

The ultimate fate of a fusion varies dramatically depending on the values of recombination rates in heterozygotes and fused heterozygotes (r_h and r_f). Polymorphism is expected at equilibrium only when the heterozygotes for the fusion recombine at lower rates than that of homozygotes. Nevertheless, even if a fusion will ultimately become fixed, transient polymorphism may persist for a very long time. This outcome is

expected when fused chromosomes have greatly reduced recombination, which greatly slows the spread of the fusion from the locally adapted genotype on which it first appeared to other locally adapted genotypes. While effects of fusions on recombination have been known for decades (White, 1973), there is little data on their relative effects when homozygous and heterozygous. Current evidence from house mice indicates that heterozygotes have higher recombination rates than homozygotes (Bidau *et al.*, 2001; Dumas & Britton-Davidian, 2002). If this pattern holds in other species, my results suggest that many fusions evolving by local adaptation in two (or more) demes that exchange migrants will eventually become fixed throughout the species range. More data on the effects of fusions on recombination rates when homozygous and heterozygous would be very useful.

Fusions are polymorphic in many species (King 1993). Under my models, these polymorphisms can be explained by at least four hypotheses. First, they may be at an equilibrium that results from local adaptation in demes that exchange migrants when heterozygotes have lower recombination rates than homozygotes. Second, the polymorphism may be transient. The results show fusions that are ultimately destined to fixation can remain polymorphic for substantial evolutionary periods. This occurs when fused chromosomes have greatly reduced recombination, which greatly slows the spread of the fusion from the locally adapted genotype on which it first appeared to other locally adapted genotypes. Third, a polymorphic equilibrium can result from local adaptation when there is one-way migration between demes (as in the continent-island model). Fourth, some other force, such as meiotic drive or associative overdominance resulting from linkage to recessive deleterious alleles, may oppose the spread of a fusion that is favored by local adaptation.

Some empirical observations are consistent with the hypothesis that fusions evolve under local adaptation. Fusions in blind mole rats (*Spalax*) show two independent geographic clines correlated with aridity (Nevo, 2012). The grasshopper *Dichroplus pratensis* has three polymorphic fusions that show geographic variation and correlations to phenotypic traits (such as body size) thought to be adaptive in extreme environments (Bidau *et al.*, 2012).

My models complement the female meiotic drive hypothesis for the evolution of karyotypes in mammals (Pardo-Manuel de Villena & Sapienza, 2001). Under this hypothesis, the polarity of the meiotic spindle influences the probability of a chromosome being pulled into the egg rather than the polar body. Local adaptation forces, as considered in the models presented here, may reinforce or counter meiotic drive depending on the chromosome form favored by the meiotic spindle. In a species for which acrocentric chromosomes are being favored, locally adapted fusions could be the only ones to overcome meiotic drive and persist in the population. Moreover, fusions carrying alleles under selection weaker than meiotic drive could be maintained at low frequencies. These fusions in the standing genetic variation would allow the genome to change quickly in response to a switch in spindle polarity. The polarity of the spindle may switch, favoring the opposite type of chromosome arrangement.

It is plausible that there is a balance between meiotic drive and other evolutionary forces acting on the karyotype. The meiotic-drive hypothesis predicts that karyotypes should be either mostly metacentric or mostly acrocentric, and that species with intermediate proportions of meta- to acrocentric chromosomes are rare. A pattern consistent with this prediction is found in mammals (Pardo-Manuel de Villena & Sapienza, 2001). While most species have very low or very high proportions of acrocentrics, a number of species do show intermediate proportions. This pattern suggests

that other forces, such as fusions driven by local adaptation, are maintaining fused chromosomes even when they are not favored by spindle polarity. To quantify the importance of these alternative forces, a quantitative analysis of the data in using an adequate phylogenetic framework is necessary.

My results are also relevant to the current search for “islands of speciation” at the centromeres. Recent studies of neutral genetic divergence between hybridizing species have found high differentiation around some centromeres and chromosome inversions (Turner *et al.*, 2005; Cheng *et al.*, 2012b; Ellegren *et al.*, 2012). These regions of low recombination may hold the genes involved in adaptive divergence or hybrid incompatibilities. The neutral genetic divergence, therefore, could represent the signature of incipient speciation in the genome. In this scenario, fusions that link two pericentromeric regions (which hold locally adapted genes) are favored and may sweep in one of the two incipient species. Sweeps in both species will eliminate the neutral divergence in the regions around the new fused centromere, erasing the potential island of speciation. To rigorously test whether locally adapted fusions happened as part of a speciation process, it is necessary to gather karyotype and genomic data for the incipient species and closely related outgroups. Additionally, models to predict the neutral variation around polymorphic fusions are needed to adequately analyze the genetic data from these scenarios.

Here, I confirm the intuition that evolution favors the reduction of recombination between locally adapted alleles. I explore the conditions under which this force drives the evolution of centric (or Rb) fusions, tandem fusions, and reciprocal translocations. Interestingly, I find that in most cases considered the rearrangement will evolve to fixation while promoting the divergence between populations. These results remind us that to further our understanding of genomic divergence and the role rearrangements play

in it, we need studies on karyotype variation in a comparative framework, estimation of fine-scale recombination rates, and proper quantitative analysis of genomic data.

Tables

Table 2.1

Models fit to the data. The preferred model is shown in the first row, and alternative models in the following rows. Left of the vertical line are the assumptions; under M (migration) “SS” is the stepping stone model and “C” is the continuous space model. To the right of the vertical line are the parameter estimates, the negative log likelihood, and the AIC values for each model. The four best-fit models are shown above the heavy horizontal line, with the best-fit model listed first.

Zones	Assort Mating	M	Savannah		Highlands		Rainforest		F	σ^2	$-\ln(L)$	AIC
			W_{SS}	W_{II}	W_{SS}	W_{II}	W_{SS}	W_{II}				
3	Yes	SS	1.29	1.30	0.25	0.50	0.73	0.86	0.82	7.34	-	-
		C	1.44	1.44	0.27	0.51	0.83	0.90	0.82	7.12	-1.4	2.8
	By genotype	SS	1.21	1.21	0.31	0.57	0.95	1.05	SS: 0.95 SI: 0.68 II: 0.002	80.8	1.3	1.4
		SS	1.22	1.23	0.21	0.44	1.65	1.28	Sa: 0.83 Hi: 0.96 Ra: 0.3	5.8	1.1	1.8
1	Yes	SS	1.00	1.00					0.5	3.0	-102.9	195.8
		C	1.00	1.00					0.5	7.12	-117.7	224.8
2	Yes	SS	1.28	1.27	0.33	0.60			0.80	49.4	-80.8	157.6
		C	3.22	3.16	0.81	0.95			0.39	34.0	-96.0	188.0
3	No	SS	1.04	1.04	0.56	0.84	1.08	1.08		37.0	-35.8	69.6
		C	1.33	1.25	0.56	0.84	1.07	1.1		19.0	-36.5	71.0

Figures

Figure 1.1

Expected coalescent times for neutral sites linked to an old inversion with locally adapted breakpoints (left) or alleles (right). Two genes are sampled from the same arrangement (\bar{T}_{II} , \bar{T}_{SS} , filled points), or between arrangements (\bar{T}_{IS} , open points), for the cases where $q = 0.09$ (black lines) or $q = 0.02$ (grey line). The breakpoints are located at 0 and 10 cM, and an arrow indicates the position of one locally adapted allele. Curves are based on analytical results, and the circles show simulation results (10^6 runs each). Squares show evaluations from the generating function method (see text). The dashed line shows the expectation for a neutral subdivided population. In both models, the patterns for the opposite half of the inversion (not shown) are symmetric to the ones presented. Shown on both sides, patterns for a small portion of the flanking region with normal recombination patterns.

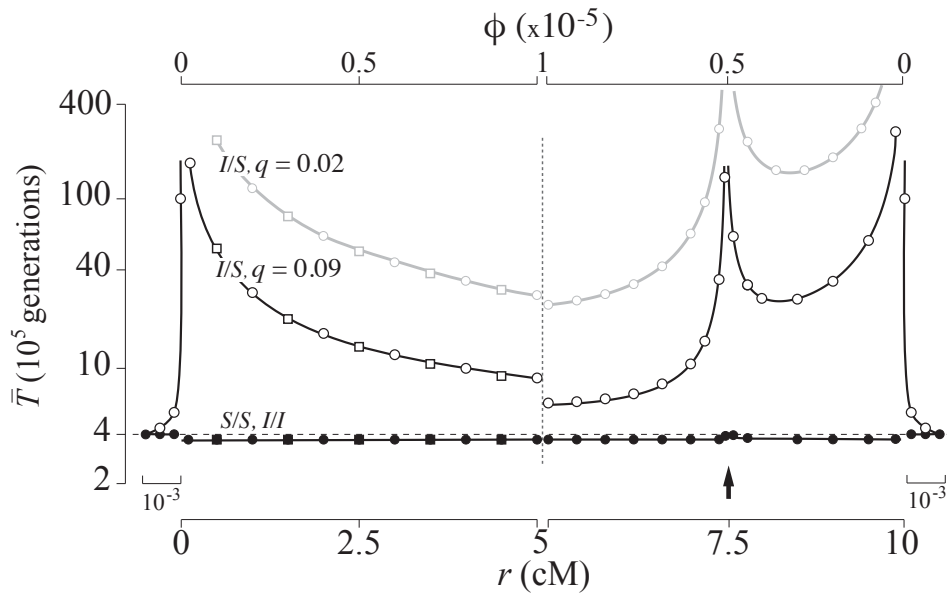


Figure 1.2

The effect of migration (Nm) on F_{ST} and F_{AT} under the model of locally adapted breakpoints. Results are shown for three values of gene flux: $\phi = 10^{-3}$ (dotted lines), $\phi = 10^{-5}$ (dashed lines), $\phi = 10^{-8}$ (solid lines).

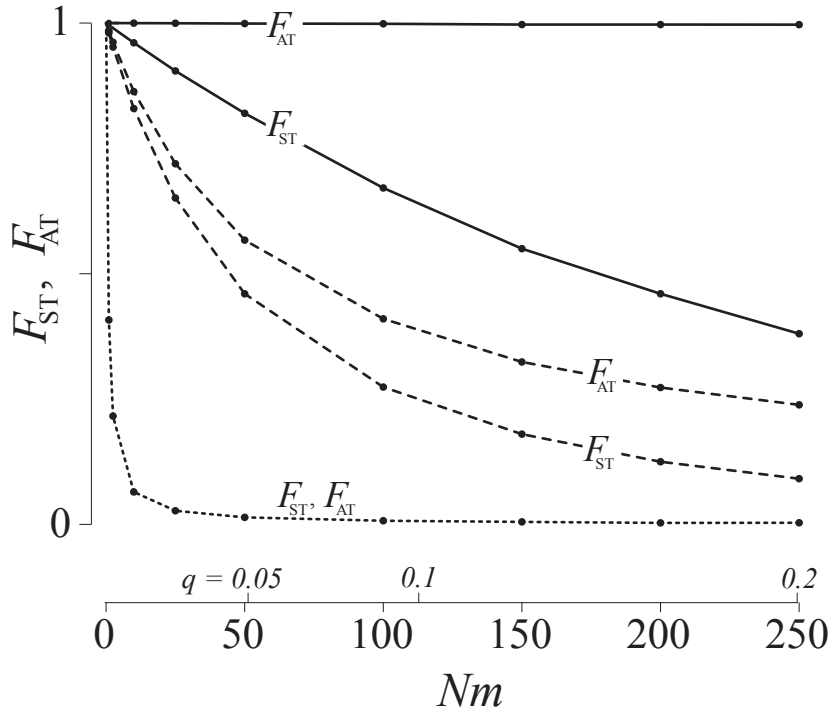


Figure 1.3

Expected coalescent time for pairs of genes linked to a young inversion under three alternative models. Two genes are sampled from the same arrangement (\bar{T}_{II} , \bar{T}_{SS} , filled points), or different arrangements (\bar{T}_{IS} , open points). (a) A neutral inversion in a single population, at frequency $x_0 = 0.5$. (b) An inversion with locally adapted breakpoints. (c) An inversion that carries two locally adapted alleles. In (b) and (c), $T^* = 0.1$ and $q = 0.09$. The breakpoints are at 0 and 10 cM, and an arrow indicates the position of one locally adapted allele. The points show simulation results (10^6 runs each), and lines are added for clarity purposes only. The dashed line shows the SNM for (a), and a neutral subdivided population for (b) and (c). Only the left half of the inversion is shown for each case; patterns for the right half of the inversion are symmetric to the ones presented.

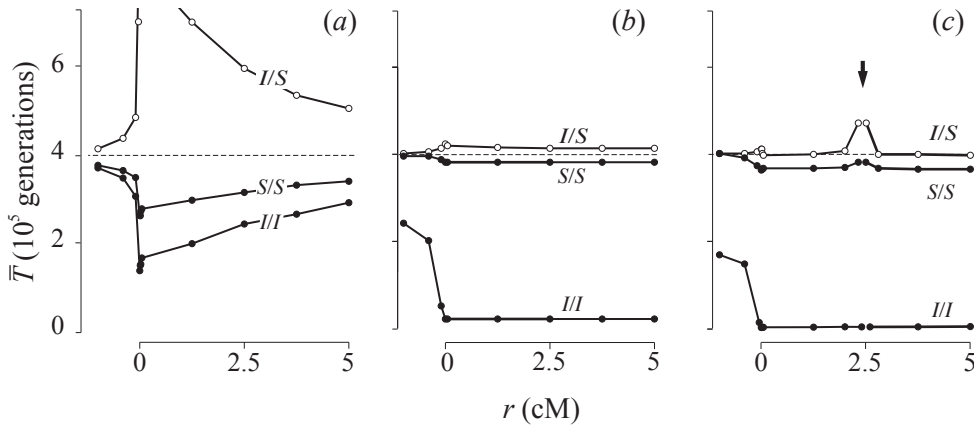


Figure 1.4

Effect of the age inversion, T^* (in N generations), on the expected coalescent time of pairs of genes under the locally adapted breakpoints model (at $\phi = 10^{-5}$, $N = 10^5$). The points show simulation results (10^5 runs each), and lines are added for clarity purposes only. Two cases are presented $q = 0.09$ (black lines), and $q = 0.02$ (grey lines).

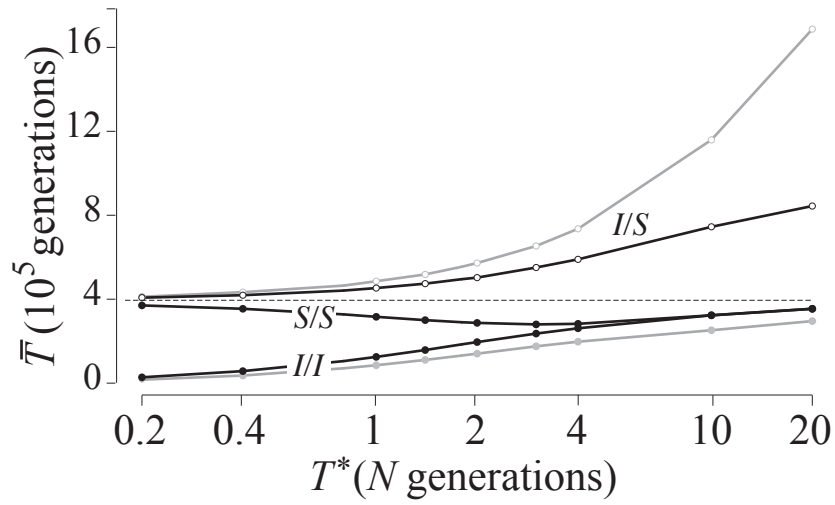


Figure 2.1

Karyotype frequencies for inversion 3Ra in Cameroon. Solid points show villages sampled for *A. funestus*. Pie diagrams show the frequencies of the standard homozygote (SS, white), heterozygote (SI, blue), and inverted homozygote (II, black).

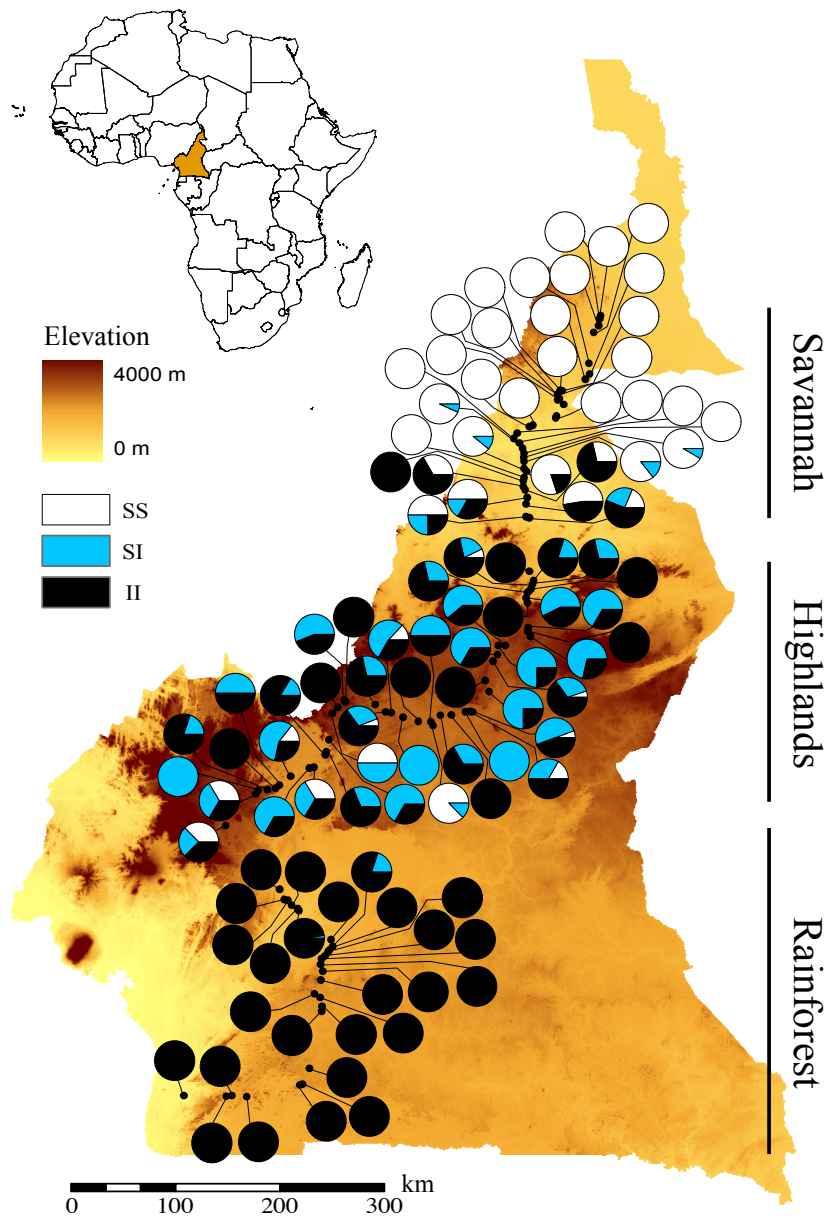


Figure 2.2

Comparison of the observed frequencies (circles) and predicted frequencies from the best-fit model (red curves) for the three karyotypes. SS = standard homozygotes, SI = heterozygotes, II = inverted homozygotes. Populations that differ significantly from the model are shown as filled circles. Further details are given in Table S1.

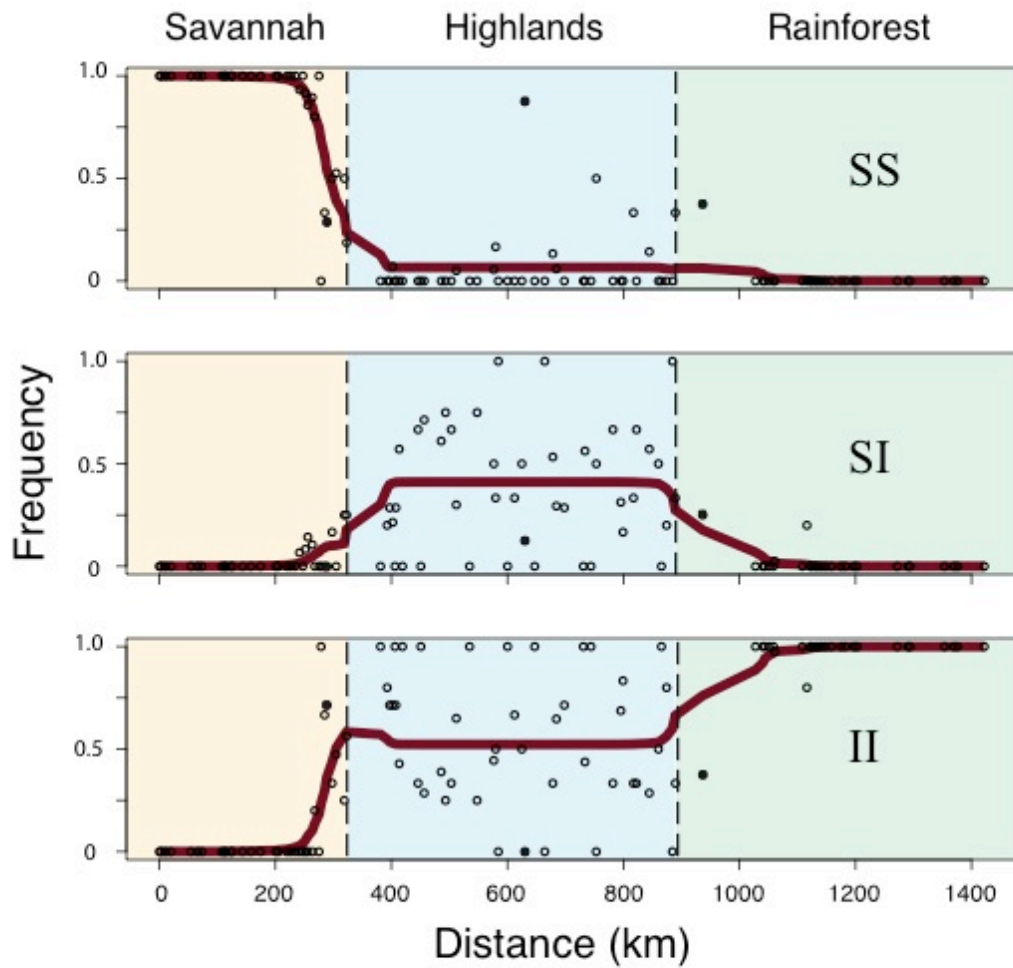


Figure 2.3

Estimated viabilities of standard homozygotes (W_{SS}) and inverted homozygotes (W_{II}) relative to heterozygotes (W_{SI}) in the three ecological regions. Bars show maximum likelihood estimates, and the error bars show 95% and 99% confidence regions (identified respectively by a change of 2 and 3 log likelihood units). In the savannah, the upper 99% limits are 1.94 and 1.95 for standard and inverted homozygotes, respectively.

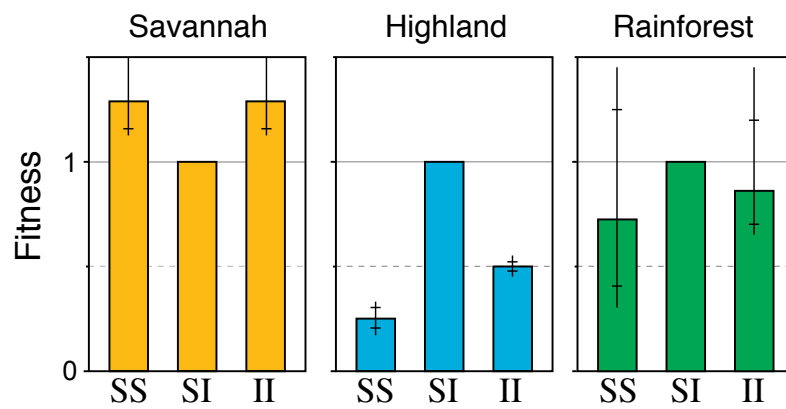


Figure 2.4

Likelihood profiles for the dispersal variance (σ^2 , above) and assortative mating (F , below). The vertical axis shows the decrease (in $\ln(L)$ units) from the maximum likelihood value. The maximum likelihood estimate is shown as the closed circle, and the confidence regions are indicated by crosshairs on the horizontal line (95% for the inner and 99% for the outer crosshairs). The upper limit for the confidence region of F is unbounded.

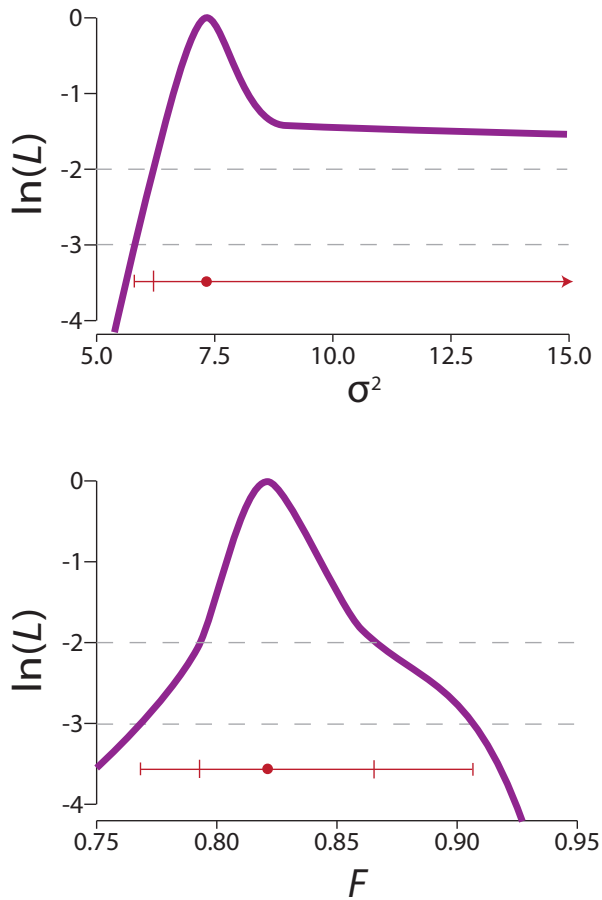


Figure 3.1

The effect of recombination on the gene trees from samples in X and Y chromosomes of two species. In the absence of recombination (left), X and Y start diverging at the origin of the sex-determining region (SDR), while gametologs diverge later (at speciation). X-Y recombination (right) inhibits divergence between opposite sex chromosomes, so they appear more similar to each other than to their gametologs from sister species.

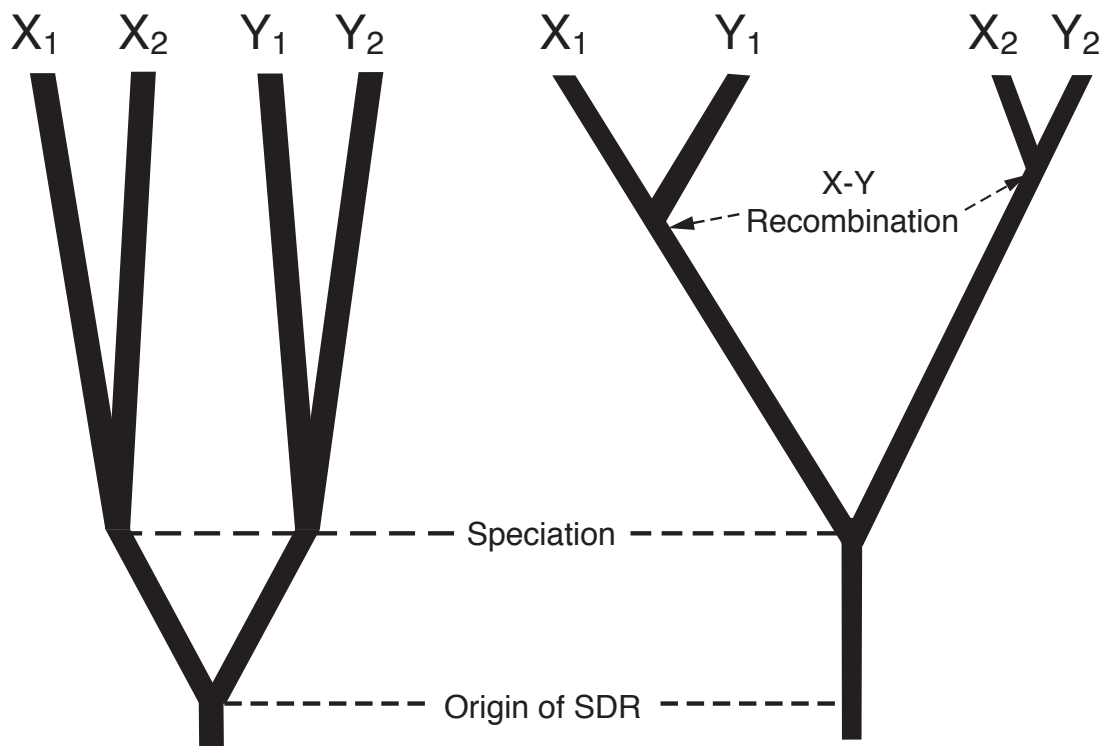
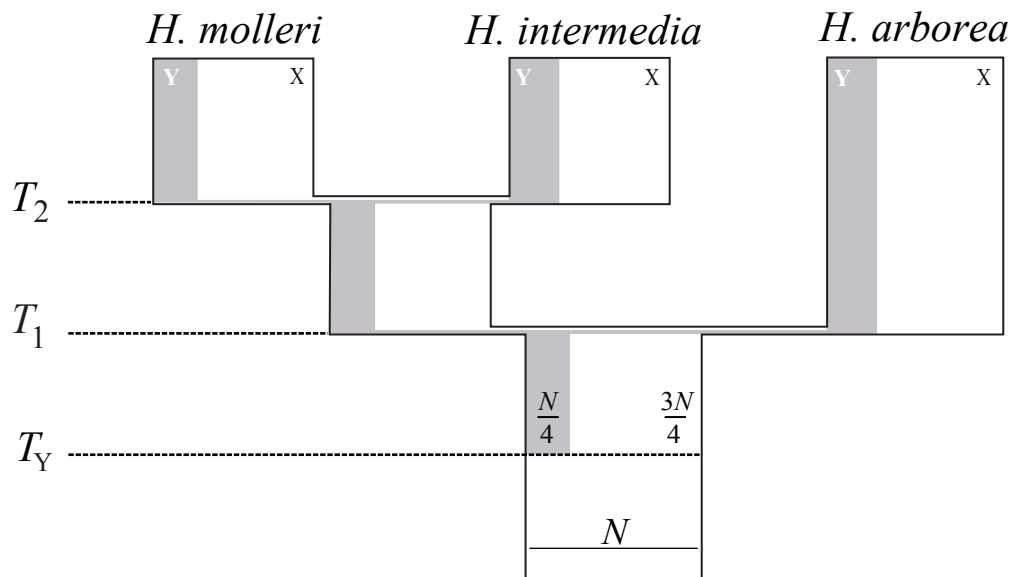


Figure 3.2

Schematic of the assumed coalescent model. Before the origin of the SDR (at T_Y), the ancestral population is composed of N autosomes. At time T_Y , a single mutation gives rise to the new Y chromosome, subdividing the population. Two speciation events give origin to the three studied species. Through time, the X and Y chromosomes recombine at a constant rate.



Allele frequencies at the sex-linked microsatellite markers. The columns show the seven loci. Frequencies on X chromosomes are the filled bars, and those on Y chromosomes are open bars. The stars denote groups for which all chromosomes have a null allele at the given marker.



Figure 3.4

Nine observed summary statistics, D_{ij} , from seven microsatellite markers along the sex chromosomes of three Hylid frogs. Small values of D_{ij} reflect lower X-Y divergence within species, indicative of X-Y recombination. The positions (in cM) of each marker on the consensus genetic map for the X chromosomes are shown below the horizontal axis.

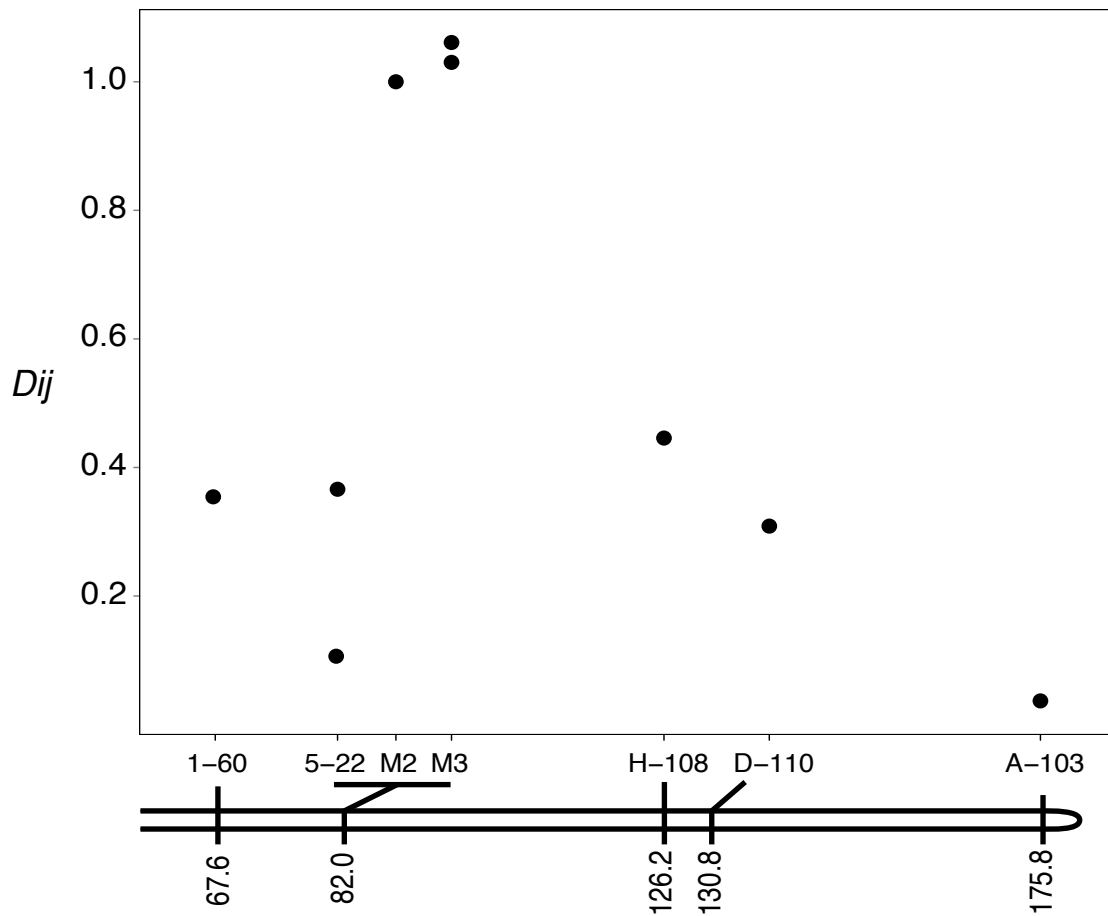


Figure 3.5

Posterior distribution of α , the relative X-Y recombination rate in three species of Hylid frogs (black curve). The distribution of α in simulations retained for ABC estimation (before post-sampling adjustment) is shown in light grey. The horizontal band represents the 99% highest posterior density interval (HPDI), and the horizontal dashed line indicates the prior distribution. The arrow indicates the point at which X-Y recombination is effectively zero in the simulations ($\alpha = 2.5 \times 10^{-10}$).

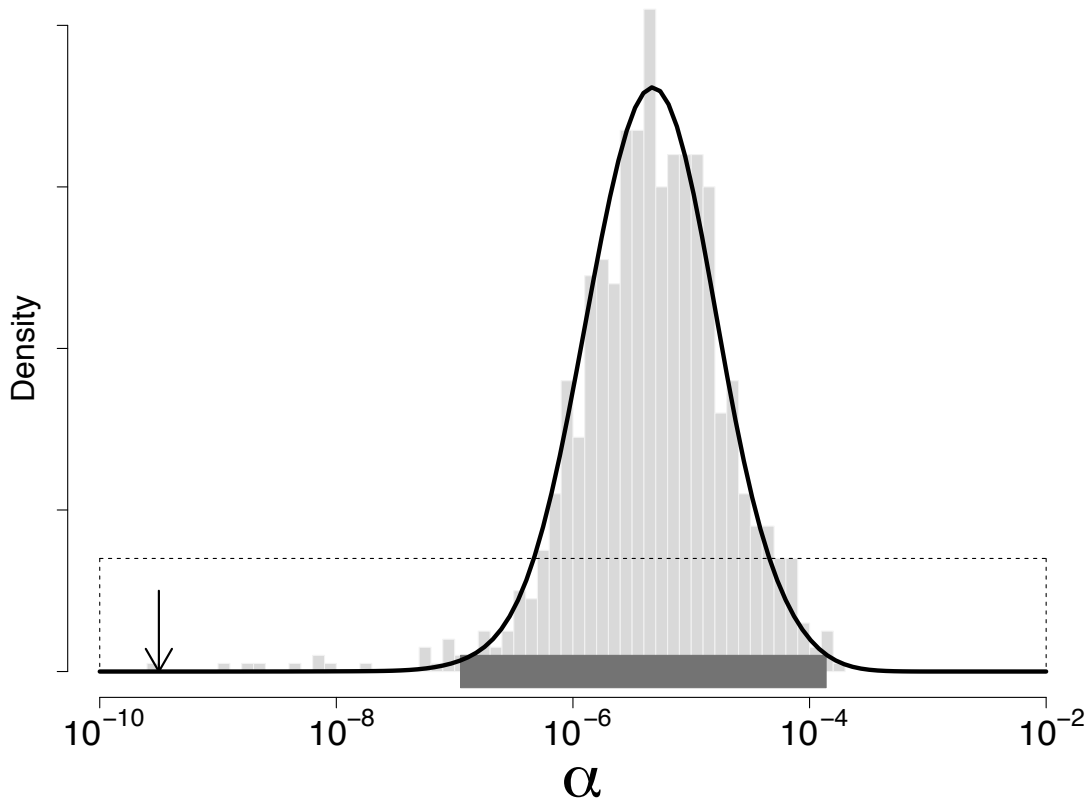


Figure 4.1

Selective advantage of a fusion (λ) across recombination (r_h) for different values of selection and migration (s, m).

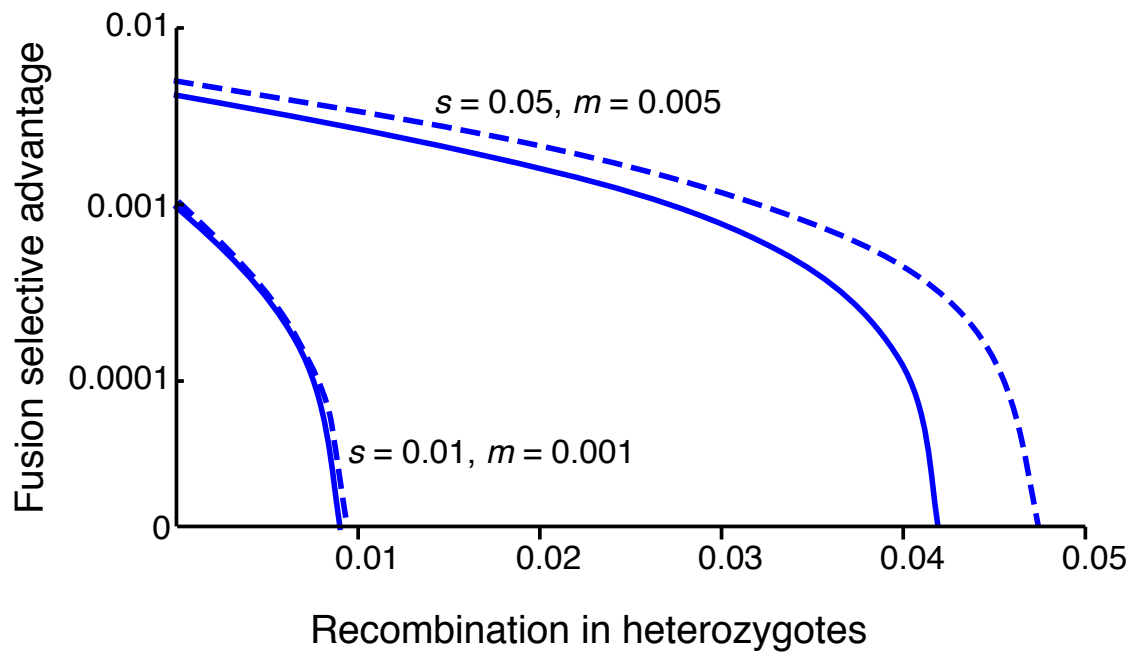


Figure 4.2

Contour plot of the frequency of the fusion, $p_{(F)}$, in deme 2 for different values of recombination in homozygotes and heterozygotes. The dashed line is $r_f = r_h$. Other parameters are $m=0.001$ and $s=0.01$

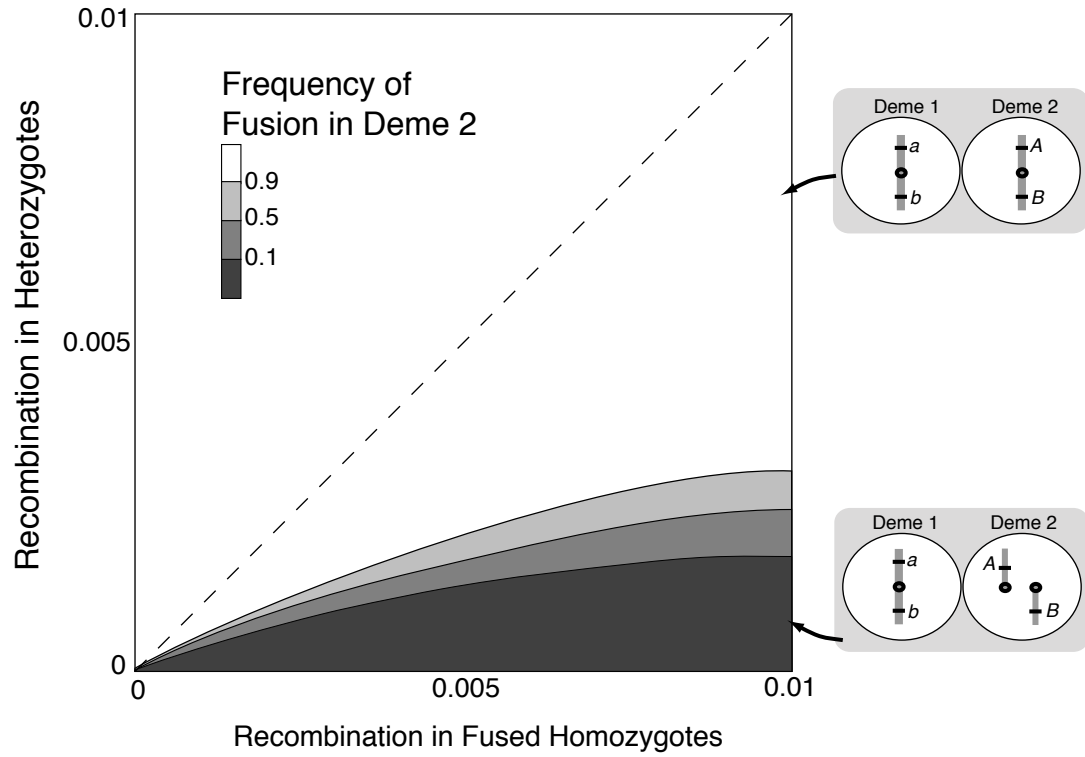
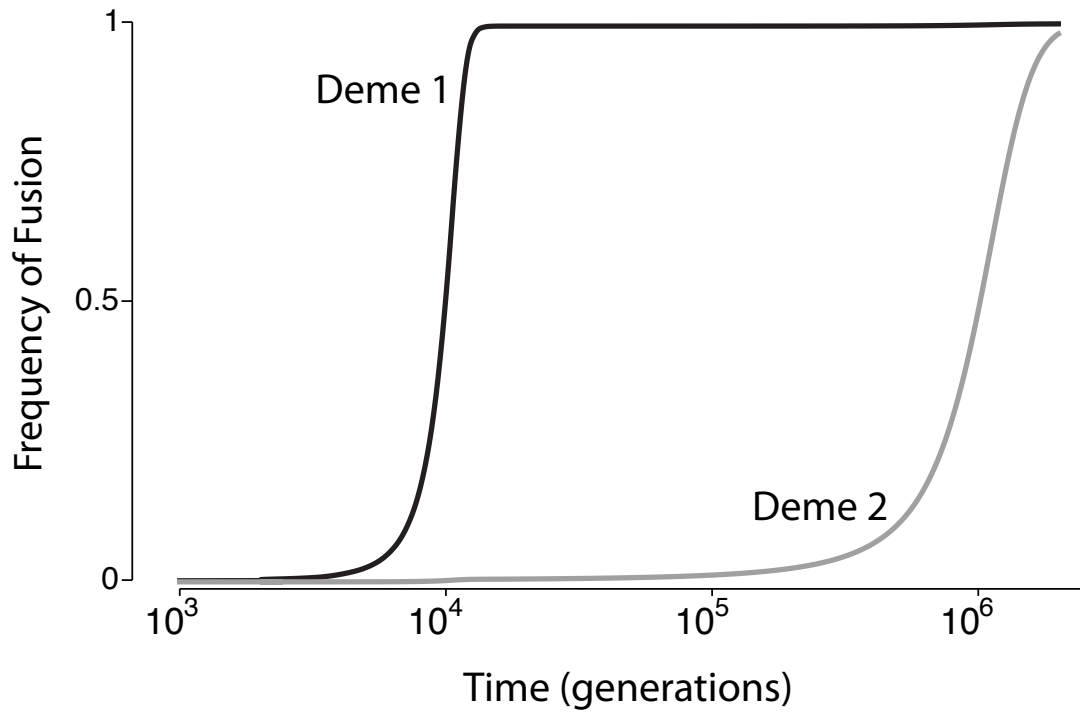


Figure 4.3

The trajectory through time of the frequency of the fusion, $p_{(F)}$, under the two-deme model for parameters $r_f = r_h = 0.001$, are $m = 0.001$ and $s = 0.01$.



Appendix A: Expected coalescent times at sites in old inversions under local adaptation

The Structured Coalescent: In this part of the appendix we summarize the calculations used to find expected coalescent times for pairs of genes. This approach was developed by Kaplan *et al.* (1988) and Hudson and Kaplan (1988), and we present a general description in order to introduce our models.

A sampled gene can appear in different *contexts*, which are determined by the genetic background to which the gene is linked. The context includes the chromosome arrangement (*I* or *S*), and in some of our models also includes the population (1 or 2) and the alleles at the selected loci on that chromosome (for example, $A_1 B_2$). Genes change contexts as a result of gene flux, recombination, or migration. At any point in the past, the state of the system can be described by the genes and their contexts. We denote the contexts as \mathbf{c}_i ($i=1, 2, \dots, n_c$), where n_c is the number of possible contexts.

We use the standard first-step analysis, described in Wakeley (2009; section 5.1.3). For a system that starts in transient state V , we write the expected time back to the most recent common ancestor of two genes as \bar{T}_V . Its value is

$$\bar{T}_V = \bar{H}_V + \sum_U Q_{V \rightarrow U} \bar{T}_U \quad (\text{A.1})$$

Here \bar{H}_V is the expected holding time in transient state V , that is, the waiting time until the system exits that state. $Q_{V \rightarrow U}$ is the jump probability from state V to state U , that is, the probability that when the process leaves V (in the backward sense) it goes to U . The summation is over all states U , both transient and absorbing. The expected holding times and jump probabilities can be written in terms of the transition rates between states of the system:

$$\bar{H}_V = 1 / \sum_U P_{V \rightarrow U}, \quad (\text{A.2})$$

$$Q_{V \rightarrow U} = \frac{P_{V \rightarrow U}}{\sum_U P_{V \rightarrow U}} = \bar{H}_V P_{V \rightarrow U}, \quad (\text{A.3})$$

where $P_{V \rightarrow U}$ is the (backward) transition rate from state V to state U . We define $P_{V \rightarrow V} = 0$.

To solve for the expected coalescence times, we first order the transient states in some arbitrary way. Using this ordering, denote the vector of expected coalescent times as $\bar{\mathbf{t}}$, the matrix of conditional jump probabilities as \mathbf{Q} , and the vector of expected holding times as $\bar{\mathbf{h}}$. Then writing the system of equations (A.2) in matrix notation, simple algebra gives the solution for the expected coalescent times as:

$$\bar{\mathbf{t}} = (\mathbf{I} - \mathbf{Q})^{-1} \bar{\mathbf{h}} \quad (\text{A.4})$$

We evaluated this equation for the different models described in the text by specifying the appropriate transition rates

$P_{V \rightarrow U}$. To obtain these transition rates, we consider two types of events:

- Transition between transient states, from V to U , is only possible if it can be described by a single event that changes a gene from its current context \mathbf{c}_i to a new \mathbf{c}_j . This transition happens at rate $2Nb_{ij}n_V$, where b_{ij} ($i, j = 1, 2, \dots, n_c$) is the probability that a gene is found in \mathbf{c}_j in the previous generation given that it is in \mathbf{c}_i in the current generation, and n_V ($= 1, 2$) is the number of genes in state V that could transition from \mathbf{c}_i to \mathbf{c}_j .

- Transition from a transient to its absorbing state (by coalescence), happens only when the two genes are in the same context \mathbf{c}_i , at a rate $1/p_i$, the inverse of the frequency of \mathbf{c}_i .

The rates described above are linear approximations, in which we assume the transitions happen at very low rate (and quadratic terms are negligible). The critical parameters of these transitions, the probabilities b_{ij} , are specific of each model and we describe them below.

Locally Adapted Breakpoints: In this scenario, two populations of equal size N are polymorphic for a chromosomal inversion under local selection, and have been at migration-selection equilibrium for an infinite time. The system has 4 contexts, defined by the possible combinations of its two populations and two chromosomal arrangements: $\mathbf{c}_1 = \{1, S\}$, $\mathbf{c}_2 = \{1, I\}$, $\mathbf{c}_3 = \{2, S\}$, and $\mathbf{c}_4 = \{2, I\}$. We define κ_i to be the population (1 or 2) and ι_i to be the chromosomal arrangement (S or I) for context \mathbf{c}_i . The frequency of \mathbf{c}_i , defined as p_i (shorthand for $p_{\{\kappa, \iota\}(i)}$), is assumed to be constant. In the main text we have used q to denote the frequency of the disfavoured arrangement (that is, $p_2 = p_3 = q$).

The backward transition probabilities are:

$$b_{ij} = \begin{cases} \frac{mp_j}{mp_j + (1-m)p_i} & \text{if only } \kappa \text{ of } \mathbf{c}_i \text{ and } \mathbf{c}_j \text{ are different} \\ \phi p_j & \text{if only } \iota \text{ of } \mathbf{c}_i \text{ and } \mathbf{c}_j \text{ are different} \\ 0 & \text{otherwise} \end{cases} \quad (\text{A.5})$$

Again, this is an approximation that assumes small m and ϕ .

We used a forward-time model to obtain the context frequencies numerically. In the life cycle for this model, selection is followed by migration. (Recombination is not important, as we are modeling a single locus.) We assume diploid selection, with viabilities:

Population	$W_{I,I}$	$W_{S,I}$	$W_{S,S}$
1	1	$1 - hs$	$1 - s$
2	$1 - s$	$1 - hs$	1

The frequency of \mathbf{c}_i after selection, $p'_{\{\kappa, \iota\}(i)} = (W_{\iota(i), \iota(i)} p_{\{\kappa, \iota\}(i)}^2 + W_{\iota(i), \neg \iota(i)} p_{\{\kappa, \iota\}(i)} p_{\{\kappa, \neg \iota(i)\}}) / \bar{W}_i$, where \bar{W}_i is the mean fitness of population κ_i , and $\neg \iota(i)$ denotes the chromosome arrangement opposite to ι_i (that is, S for I and vice versa).

The populations then exchange m migrants. The frequency of \mathbf{c}_i after migration is $p''_{\{\kappa, \iota\}(i)} = p'_{\{\kappa, \iota\}(i)}(1 - m) + p'_{\{\neg\kappa(i), \iota(i)\}} m$, where $\neg\kappa(i)$ denotes the population opposite to κ_i .

After migration, the life cycle starts over (p''_i becomes p_i in the next generation). We iterate these steps to equilibrium, starting at frequencies $p_1 = 1 - 1/2 N$, $p_2 = 1/2 N$, $p_3 = 1$, $p_4 = 0$.

Locally adapted alleles: Now we consider the situation where there are two loci A and B (with alleles A_1, A_2 and B_1, B_2 respectively) inside of the inversion. Two populations of equal size N are polymorphic for a chromosomal inversion under local selection, and have been at migration-selection equilibrium for an infinite time. The system now has 16 contexts. Each context is defined by four elements: $\mathbf{c}_i = \{\kappa, \iota, \alpha, \beta\}$, where $\kappa (= 1, 2)$ is the population, $\iota (= I, S)$ is the arrangement, $\alpha (= A_1, A_2)$ is the allele at locus A , and $\beta_i (= B_1, B_2)$ is the allele at locus B . Context \mathbf{c}_i has constant frequency $p_{\{\kappa, \iota, \alpha, \beta\}(i)}$. We use \bullet to indicate the summation of the frequencies across a particular element. For example, $p_{\{1, I, A_1, \bullet\}} = p_{\{1, I, A_1, B_1\}} + p_{\{1, I, A_1, B_2\}}$. In the main text we have used q to denote the frequency of the disfavoured arrangement, that is $q = p_{\{1, I, \bullet, \bullet\}} = p_{\{2, S, \bullet, \bullet\}}$.

An explicit genetic architecture is necessary to describe the recombination and gene flux probabilities. For simplicity, we will explain here the case where the position $*$ on the chromosome at which the genes are sampled lies between the two selected loci. That is, the genetic map in a S chromosome is:

$$breakpoint - A - * - B - breakpoint$$

Recombination in homokaryotypes can be described using two rates r_α, r_β , that are the forward recombination rate (genetic map distances) between A and $*$, and between $*$ and B respectively. Gene flux in heterokaryotypes may exchange different segments within the inverted region. There are four forward gene flux rates exchange: $\phi_{A \leftrightarrow B}$, $\phi_{A \leftrightarrow *}$, $\phi_{* \leftrightarrow B}$, $\phi_{* \leftrightarrow -}$, where the subindexes correspond to the portion of chromosome that is exchanged at a particular rate.

The backward transition probabilities are:

$$b_{ij} = \begin{cases} mp_{\{\kappa, \iota, \alpha, \beta\}(j)} / (mp_{\{\kappa, \iota, \alpha, \beta\}(j)} + (1 - m) p_{\{\kappa, \iota, \alpha, \beta\}(i)}) & \text{if only } \kappa \text{ of } \mathbf{c}_i \text{ and } \mathbf{c}_j \text{ are different} \\ \frac{p_{\{\kappa, \iota, \alpha, \beta\}(j)}}{p_{\{\kappa, \iota, \alpha, \beta\}(i)}} (\phi_{A \leftrightarrow B} p_{\{\kappa, \iota, \bullet, \bullet\}(i)} + \phi_{* \leftrightarrow -} p_{\{\kappa, \iota, \alpha, \beta\}(i)} + \phi_{A \leftrightarrow *} p_{\{\kappa, \iota, \bullet, \beta\}(i)} + \phi_{* \leftrightarrow B} p_{\{\kappa, \iota, \alpha, \bullet\}(i)}) & \text{if only } \iota \text{ of } \mathbf{c}_i \text{ and } \mathbf{c}_j \text{ are different} \\ \frac{p_{\{\kappa, \iota, \alpha, \beta\}(j)}}{p_{\{\kappa, \iota, \alpha, \beta\}(i)}} r_\alpha p_{\{\kappa, \iota, \alpha, \bullet\}(i)} & \text{if only } \alpha \text{ of } \mathbf{c}_i \text{ and } \mathbf{c}_j \text{ are different} \\ \frac{p_{\{\kappa, \iota, \alpha, \beta\}(j)}}{p_{\{\kappa, \iota, \alpha, \beta\}(i)}} r_\beta p_{\{\kappa, \iota, \bullet, \beta\}(i)} & \text{if only } \beta \text{ of } \mathbf{c}_i \text{ and } \mathbf{c}_j \text{ are different} \\ \frac{p_{\{\kappa, \iota, \alpha, \beta\}(j)}}{p_{\{\kappa, \iota, \alpha, \beta\}(i)}} (\phi_{* \leftrightarrow B} p_{\{\kappa, \iota, \alpha, \bullet\}(i)} + \phi_{* \leftrightarrow -} p_{\{\kappa, \iota, \alpha, \beta\}(i)}) & \text{if } \iota \text{ and } \alpha \text{ of } \mathbf{c}_i \text{ and } \mathbf{c}_j \text{ are different} \\ \frac{p_{\{\kappa, \iota, \alpha, \beta\}(j)}}{p_{\{\kappa, \iota, \alpha, \beta\}(i)}} (\phi_{A \leftrightarrow *} p_{\{\kappa, \iota, \bullet, \beta\}(i)} + \phi_{* \leftrightarrow -} p_{\{\kappa, \iota, \alpha, \beta\}(i)}) & \text{if } \iota \text{ and } \beta \text{ of } \mathbf{c}_i \text{ and } \mathbf{c}_j \text{ are different} \\ \frac{p_{\{\kappa, \iota, \alpha, \beta\}(j)}}{p_{\{\kappa, \iota, \alpha, \beta\}(i)}} (\phi_{* \leftrightarrow -} p_{\{\kappa, \iota, \alpha, \beta\}(i)}) & \text{if } \iota, \alpha \text{ and } \beta \text{ of } \mathbf{c}_i \text{ and } \mathbf{c}_j \text{ are different} \\ 0 & \text{otherwise} \end{cases} \quad (\text{A.6})$$

Again, we assume weak m, ϕ, r .

As for the case of locally adapted breakpoints, we use a forward-time numerical approach to obtain the context frequencies for our coalescent model. The life cycle is: selection \rightarrow migration \rightarrow recombination. We assume multiplicative selection on diploids. The contribution of locus A to viability is:

Pop	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$
1	1	$1 - hs$	$1 - s$
2	$1 - s$	$1 - hs$	1

The same applies for locus B . The inversion does not contribute directly to fitness in this model.

To find the context frequencies, we first obtain equilibrium for loci A and B in the absence of the inversion. We then introduce the inversion at $p_{\{1,I,A_1,B_1\}} = 1/2 N$ and iterate to the new equilibrium.

Appendix B: A Neutral Inversion

Simulating the coalescent process: Our goal here is to derive an algorithm for simulating realisations of the coalescent process, given that the rates of events change stochastically in time. To apply this machinery to inversions that are evolving by drift, we use a first phase of simulations to obtain stochastic trajectories for the frequencies of the inversions and use those to calculate the rates of events resulting from gene flux and coalescence. In a second phase, we simulate the coalescent process itself.

The basic idea used here comes from a suggestion made by Graham Coop (pers. comm., fall 2009). The approach is to do the first phase of simulation by modeling the cumulative distribution function for the events, going backwards in time.

As we move backwards in time, the system jumps between states in a continuous-time Markov process whose rates change with time. Let $H_i(t_0)$ be the holding time, that is, the random variable representing the time until the system next changes state, given that it is in State V at time t_0 . Denote the CDF for this holding time as $F_V(t)$. If the system leaves State V at time t , the probability that it jumps to State U is called the jump probability, denoted $Q_{V \rightarrow U}(t)$. The process is completely determined by the $F_V()$ and $Q_{V \rightarrow U}()$.

We begin by giving some general results that will be used in our algorithm. An expression that applies to any CDF is

$$\frac{d}{dt} F_V(t) = P_V(t) [1 - F_V(t)] \quad (\text{B.1})$$

where $P_V(t)$ is the sum of all the rates for exiting from State V at time t :

$$P_V(t) = \sum_{V \neq U} P_{V \rightarrow U}(t) \quad (\text{B.2})$$

If the system leaves State V at time t , the probability that it jumps to State U is simply the rate of that transition normalized by the sum of all the rates for exits from State V :

$$Q_{V \rightarrow U}(t) = P_{V \rightarrow U}(t) / P_V(t). \quad (\text{B.3})$$

Equations (B.1) and (B.3) tell us how to calculate the CDF and jump probabilities that describe our process, given that we have expressions for the transition rates $P_{V \rightarrow U}(t)$.

In our coalescent model for an inversion polymorphism in a single population, the transition rates $P_{V \rightarrow U}(t)$ depend on the frequency of the inversion, which changes as the result of drift. If we had deterministic expressions for $P_{V \rightarrow U}(t)$, for example in a model of an inversion spreading deterministically by selection, we could integrate (B.1) and calculate (B.3) analytically. In our case, however, the frequencies of the inversion through time are determined by simulation, and we

used the simulated trajectories to calculate the functions numerically.

Algorithm, Phase 1

The equations from the last section give us the following algorithm for the first phase of the simulation to find the $F_V()$ and $Q_{V \rightarrow U}(t)$. We now use T rather than t to denote time to be consistent with the coalescent time scale used below.

0) Initialize: $F_V(0) = 0$ for all V
 $T = 0$

1) Calculate the $P_{S \rightarrow U}(T)$, and from them calculate the $Q_{S \rightarrow U}(T)$ and store them.

2) Increment all of the $F_S()$ using Eq. (B.1):

$$F_V(T + \Delta T) = F_V(T) + P_V(T) [1 - F_V(T)] \Delta T$$

Store the $F_V(T + \Delta T)$. Discard the $P_{V \rightarrow U}(T)$.

3) Update: $T := T + \Delta T$

4) Stop if $F_V(T) = 1$, otherwise return to (1).

Algorithm, Phase 2

To simulate a coalescent process that begins in State i :

0) Initialize: $T = 0$
 $F_V(0) = 0$

1) Draw a random number X that is uniformly distributed on $[F_V(T), 1]$.

2) Find the time (going backwards) at which the next event happens:

$$T' = F_V^{-1}(X),$$

where $F_V^{-1}(\cdot)$ is the inverse function of $F_V(\cdot)$, which was calculated in Phase 1.

3) Find the **state** j to which the system jumps by choosing randomly between all possible j in proportion to $Q_{V \rightarrow U}(T')$.

4) Update: $T := T'$

$$V := U$$

5) Stop if an absorbing state is reached, otherwise return to (1).

In Step 1 we choose a random number uniformly distributed on $[F_V(T), 1]$. Why is that? From the general relation given by Eq. (B.1) above, the cumulative distribution function for the holding time of any Markov process that begins at time t_0 obeys

$$F(t) = \int_{t_0}^t P(x) [1 - F(x)] dx. \quad (\text{B.4})$$

If we take a time t_1 that lies between t_0 and t (so $t_0 < t_1 < t$), then

$$\begin{aligned} F(t) &= \int_{t_0}^{t_1} P(x) [1 - F(x)] dx + \int_{t_1}^t P(x) [1 - F(x)] dx \\ &= F(t_1) + \int_{t_1}^t P(x) [1 - F(x)] dx. \end{aligned}$$

A trivial bit of algebra gives:

$$\begin{aligned} \frac{F(t) - F(t_1)}{1 - F(t_1)} &= \frac{1}{1 - F(t_1)} \int_{t_1}^t P(x) [1 - F(x)] dx \\ &= \int_{t_1}^t P(x) \left[1 - \frac{F(x) - F(t_1)}{1 - F(t_1)} \right] dx. \end{aligned}$$

$$G(t) = \int_{t_1}^t P(x) [1 - G(x)] dx \quad (\text{B.5})$$

where

$$G(x) = \frac{F(t) - F(t_1)}{1 - F(t_1)}. \quad (\text{B.6})$$

Comparing (B.5) with (B.4) shows that $G()$ is the CDF of a Markov process that starts at time t_1 and that has the same transition rates as the initial process. Further, we can find the CDF of this renewed process using (B.6): it is simply $F()$ that has been linearly rescaled. This last fact leads to Step (1) of Phase 2 of our algorithm.

Simulating drift: Przeworski *et al.* (2005, *Evolution* 59:2312) suggest an algorithm for simulating drift under a diffusion model. Time is again scaled in units of $2N$ generations.

Consider an inversion that is polymorphic with frequency x_0 at time $T = 0$. Looking backwards in time towards its origin, the evolution frequency of the inversion is approximated by a process with discrete jumps in which the new frequency becomes:

$$x_{T+\Delta T} = x_T - x_T \Delta T + I_T \sqrt{x_T(1 - x_T) \Delta T} \quad (\text{B.7})$$

where I_T is a random variable that the values -1 and $+1$ with equal probability. We have used $\Delta T = 1/(4N)$, a conservative value used also by Przeworski *et al.* (2005).

To implement the coalescent model with drift, we follow the algorithm for Phase 1 as described above. In each time step, we calculate the transition rates using the current frequency of the inversion $x(T)$. The rate of coalescence at time T between a pair of sites that share genetic context \mathbf{c}_i (e.g. both linked to the inversion) is

$$A_i(T) = 1 / p_i(T) \quad (\text{B.8})$$

where $p_i(T)$ is $x(T)$ if \mathbf{c}_i is an inverted chromosome and is $1 - x(T)$ if \mathbf{c}_i is a standard chromosome. The rate of gene flux is

$$R_{ij}(T) = 2 N \phi p_j(T) \quad (\text{B.9})$$

These transition rates are used to calculate the $Q_{V \rightarrow U}(T)$ from (B.3) and the $F_V(T + \Delta T)$ (see Step (2) of Phase 1 above). We then update the frequency of the inversion using (B.7).

Appendix C: Dynamic equations for the fusion invasion analysis

These are the equations for the frequency change due to recombination in the four fused haplotypes in the island. The notation is as described in the main text.

$$\begin{aligned}
\Delta_r P_{\{f,A,B\}} &= \\
r_h \big(&P_{\{f,A,b\}} P_{\{u,a,B\}} + P_{\{f,a,B\}} P_{\{u,A,b\}} + P_{\{f,a,B\}} P_{\{u,A,B\}} + P_{\{f,A,b\}} P_{\{u,A,B\}} - P_{\{f,A,B\}} P_{\{u,a,B\}} - P_{\{f,A,B\}} P_{\{u,A,b\}} - \\
&2 P_{\{f,A,B\}} P_{\{u,a,b\}} \big) + r_f \big(P_{\{f,a,B\}} P_{\{f,A,b\}} - P_{\{f,A,B\}} P_{\{f,a,b\}} \big) \\
\Delta_r P_{\{f,a,B\}} &= \\
r_h \big(&P_{\{f,A,B\}} P_{\{u,a,b\}} + P_{\{f,a,b\}} P_{\{u,a,B\}} + P_{\{f,A,B\}} P_{\{u,a,B\}} + P_{\{f,a,b\}} P_{\{u,A,B\}} - P_{\{f,a,B\}} P_{\{u,A,B\}} - P_{\{f,a,b\}} P_{\{u,a,b\}} - \\
&2 P_{\{f,a,B\}} P_{\{u,A,b\}} \big) + r_f \big(P_{\{f,a,b\}} P_{\{f,A,B\}} - P_{\{f,a,B\}} P_{\{f,A,b\}} \big) \\
\Delta_r P_{\{f,A,b\}} &= \\
r_h \big(&P_{\{f,A,B\}} P_{\{u,a,b\}} + P_{\{f,a,b\}} P_{\{u,A,b\}} + P_{\{f,A,B\}} P_{\{u,A,b\}} + P_{\{f,a,b\}} P_{\{u,A,B\}} - P_{\{f,A,b\}} P_{\{u,A,B\}} - P_{\{f,A,b\}} P_{\{u,a,b\}} - \\
&2 P_{\{f,A,b\}} P_{\{u,a,B\}} \big) + r_f \big(P_{\{f,a,b\}} P_{\{f,A,B\}} - P_{\{f,a,B\}} P_{\{f,A,b\}} \big) \\
\Delta_r P_{\{f,a,b\}} &= \\
r_h \big(&P_{\{f,a,B\}} P_{\{u,a,b\}} + P_{\{f,A,b\}} P_{\{u,a,b\}} + P_{\{f,A,b\}} P_{\{u,a,B\}} + P_{\{f,a,b\}} P_{\{u,A,b\}} - P_{\{f,a,b\}} P_{\{u,a,B\}} - P_{\{f,a,b\}} P_{\{u,A,b\}} - \\
&2 P_{\{f,a,b\}} P_{\{u,A,B\}} \big) + r_f \big(P_{\{f,a,B\}} P_{\{f,A,b\}} - P_{\{f,a,b\}} P_{\{f,A,B\}} \big)
\end{aligned}$$

Bibliography

- Akaike, H. 1974. New Look at Statistical-Model Identification. *Ieee T Automat Contr* **Ac19**: 716-723.
- Alphey, L., Beard, C.B., Billingsley, P., Coetzee, M. & Crisanti, A. 2002. Malaria control with genetically manipulated insect vectors. *Science* **298**: 119-121.
- Anderson, K. 1991. Chromosome evolution in Holarctic Hyla treefrogs. In: *Amphibian cytogenetics and evolution* (D. M. Green & S. K. Sessions, eds), pp. 299-312. Academic Press, San Diego.
- Anderson, W.W., Arnold, J., Baldwin, D.G., Beckenbach, A.T., Brown, C.J., Bryant, S.H., et al. 1991. Four decades of inversion polymorphism in *Drosophila pseudoobscura*. *Proceedings of the National Academy of Science USA* **88**: 10367-10371.
- Andolfatto, P., Depaulis, F. & Navarro, A. 2001. Inversion polymorphisms and nucleotide variability in *Drosophila*. *Genet Res* **77**: 1-8.
- Ayala, D., Costantini, C., Ose, K., Kamdem, G.C., Antonio-Nkondjio, C., Agbor, J.P., et al. 2009. Habitat suitability and ecological niche profile of major malaria vectors in Cameroon. *Malaria Journal* **8**.
- Ayala, D., Caro-Riaño, H., Dujardin, J.-P., Rahola, N., Fontenille, D. & Simard, F. 2011a. Chromosomal and environmental determinants of morphometric variation in natural populations of the malaria vector *Anopheles funestus* in Cameroon. *Infection, Genetics and Evolution* **11**: 940-947.
- Ayala, D., Fontaine, M.C., Cohuet, A., Fontenille, D., Vitalis, R. & Simard, F. 2011b. Chromosomal inversions, natural selection and adaptation in the malaria vector *Anopheles funestus*. *Mol Biol Evol* **28**: 745-758.
- Bachtrog, D., Hom, E., Wong, K.M., Maside, X. & de Jong, P. 2008. Genomic degradation of a young Y chromosome in *Drosophila miranda*. *Genome Biol* **9**.
- Barton, N.H. & Charlesworth, B. 1984. Genetic revolutions, founder effects, and speciation. *Annu Rev Ecol Syst* **15**: 133-164.
- Barton, N.H. & Hewitt, G.M. 1985. Analysis of Hybrid Zones. *Annu Rev Ecol Syst* **16**: 113-148.
- Barton, N.H. & Gale, K.S. 1993. Genetic analysis of hybrid zones. In: *Hybrid Zones and the Evolutionary Process* (R. G. Harrison, ed.), pp. 13-45. Oxford University Press, Oxford.
- Barton, N.H. 2000. Genetic hitchhiking. *Philos Trans R Soc Lond B Biol Sci* **355**: 1553-1562.
- Beaumont, M.A. 2010. Approximate bayesian computation in evolution and ecology. *Annu Rev Ecol Evol S* **41**: 379-406.
- Berset-Brandli, L., Jaquiere, J. & Perrin, N. 2007. Recombination is suppressed and variability reduced in a nascent Y chromosome. *J Evolution Biol* **20**: 1182-1188.

- Berset-Brändli, L., Jaquiéry, J., Broquet, T., Ulrich, Y. & Perrin, N. 2008. Extreme heterochiasmy and nascent sex chromosomes in European tree frogs. *Proceedings of the Royal Society B: Biological Sciences* **275**: 1577-1585.
- Besansky, N.J., Krzywinski, J., Lehmann, T., Simard, F., Kern, M., Mukabayire, O., et al. 2003. Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: Evidence from multilocus DNA sequence variation. *P Natl Acad Sci USA* **100**: 10818-10823.
- Bidau, C.J., Gimenez, M.D., Palmer, C.L. & Searle, J.B. 2001. The effects of Robertsonian fusions on chiasma frequency and distribution in the house mouse (*Mus musculus domesticus*) from a hybrid zone in northern Scotland. *Heredity* **87**: 305-313.
- Bidau, C.J., Mino, C.I., Castillo, E.R. & Marti, D.A. 2012. Effects of Abiotic Factors on the Geographic Distribution of Body Size Variation and Chromosomal Polymorphisms in Two Neotropical Grasshopper Species (Dichroplus: Melanoplinae: Acrididae). *Psyche: A Journal of Entomology* **2012**.
- Boccolini, D., Carrara, G.C., Dia, I., Fortes, F., Cani, P.J. & Costantini, C. 2005. Chromosomal differentiation of *Anopheles funestus* from Luanda and Huambo Provinces, western and central Angola. *American Journal of Tropical Medicine and Hygiene* **73**: 1071-1076.
- Boete, C. & Koella, J.C. 2003. Evolutionary ideas about genetically manipulated mosquitoes and malaria control. *Trends in Parasitology* **19**: 32-38.
- Brooke, B.D., Hunt, R.H., Chandre, F., Carnevale, P. & Coetzee, M. 2002. Stable chromosomal inversion Polymorphisms and insecticide resistance in the malaria vector mosquito *Anopheles gambiae* (Diptera : Culicidae). *Journal of Medical Entomology* **39**: 568-573.
- Brown, K.M., Burk, L.M., Henagan, L.M. & Noor, M.A. 2004. A test of the chromosomal rearrangement model of speciation in *Drosophila pseudoobscura*. *Evolution* **58**: 1856-1860.
- Bruere, A.N. & Ellis, P.M. 1979. Cytogenetics and reproduction of sheep with multiple centric fusions (Robertsonian translocations). *Reproduction* **57**: 363-375.
- Bryan, J.H., Petrarca, V., di Deco, M.A. & Coluzzi, M. 1987. Adult behaviour of members of the *Anopheles gambiae* complex in the Gambia with special reference to *An. melas* and its chromosomal variants. *Parassitologia (Rome)* **29**: 221-249.
- Bush, G., Case, S. & Wilson, A. 1977. Rapid Speciation and Chromosomal Evolution in Mammals. *P Natl Acad Sci USA* **74**: 3942-3946.
- Butlin, R.K. 2005. Recombination and speciation. *Molecular Ecology* **14**: 2621-2635.
- Butlin, R.K. 2010. Population genomics and speciation. *Genetica* **138**: 409-418.
- Cassone, B.J., Molloy, M.J., Cheng, C.D., Tan, J.C., Hahn, M.W. & Besansky, N.J. 2011. Divergent transcriptional response to thermal stress by *Anopheles gambiae* larvae carrying alternative arrangements of inversion 2La. *Molecular Ecology* **20**: 2567-2580.

- Chang, Y.-C., Shii, C.-T., Lee, Y.-C. & Chung, M.-C. 2013. Diverse chromosome complements in the functional gametes of interspecific hybrids of MT- and A-karyotype *Lycoris* spp. *Plant Syst Evol*: 1-15.
- Charlesworth, B. 1985. Recombination, genome size and chromosome number. *The evolution of genome size*: 489-513.
- Charlesworth, B., Nordborg, M. & Charlesworth, D. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet Res* **70**: 155-174.
- Charlesworth, B. & Charlesworth, D. 2000. The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **355**: 1563-1572.
- Charlesworth, B. 2012. The Effects of Deleterious Mutations on Evolution at Linked Sites. *Genetics* **190**: 5-22.
- Charlesworth, D. & Charlesworth, B. 1979. Selection on recombination in clines. *Genetics* **91**: 581-589.
- Cheng, C., White, B.J., Kamdem, C., Mockaitis, K., Costantini, C., Hahn, M.W., et al. 2012a. Ecological genomics of *Anopheles gambiae* along a latitudinal cline: a population resequencing approach. *Genetics* **190**: 1417-1432.
- Cheng, C.D., White, B.J., Kamdem, C., Mockaitis, K., Costantini, C., Hahn, M.W., et al. 2012b. Ecological Genomics of *Anopheles gambiae* Along a Latitudinal Cline: A Population-Resequencing Approach. *Genetics* **190**: 1417-+.
- Chovnick, A. 1973. Gene conversion and transfer of genetic information within the inverted region of inversion heterozygotes. *Genetics* **75**: 123-131.
- Clements, A.N. 1999. *The Biology of Mosquitoes: Development, Nutrition and Reproduction v. 1*. CABI Publishing.
- Coetzee, M. & Fontenille, D. 2004. Advances in the study of *Anopheles funestus*, a major vector of malaria in Africa. *Insect Biochem Mol Biol* **34**: 599-605.
- Cohuet, A., Simard, F., Toto, J.C., Kengne, P., Coetzee, M. & Fontenille, D. 2003. Species identification within the *Anopheles funestus* group of malaria vectors in Cameroon and evidence for a new species. *American Journal of Tropical Medicine and Hygiene* **69**: 200-205.
- Cohuet, A., Dia, I., Simard, F., Raymond, M. & Fontenille, D. 2004. Population structure of the malaria vector *Anopheles funestus* in Senegal based on microsatellite and cytogenetic data. *Insect Mol Biol* **13**: 251-258.
- Cohuet, A., Dia, I., Simard, F., Raymond, M., Rousset, F., Antonio-Nkondjio, C., et al. 2005. Gene flow between chromosomal forms of the malaria vector *Anopheles funestus* in Cameroon, Central Africa, and its relevance in malaria fighting. *Genetics* **169**: 301-311.
- Coluzzi, M., Sabatini, A., Petrarca, V. & Dideco, M.A. 1977. Behavioral divergences between mosquitos with different inversion karyotypes in polymorphic populations of *Anopheles gambiae* complex. *Nature* **266**: 832-833.
- Coluzzi, M., Sabatini, A., Petrarca, V. & DiDeco, M.A. 1979. Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae*

- complex. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **73**: 483-497.
- Coluzzi, M. 1982. Spatial distribution of chromosomal inversions and speciation in Anopheline mosquitoes. In: *Mechanisms of Speciation* (C. Barigozzi, ed., pp. 143-153. Alan R. Liss, New York.
- Coluzzi, M., Sabatini, A., della Torre, A., Di Deco, M.A. & Petrarca, V. 2002. A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science* **298**: 1415-1418.
- Costantini, C., Sagnon, N., Ilboudo-Sanogo, E., Coluzzi, M. & Boccolini, D. 1999. Chromosomal and bionomic heterogeneities suggest incipient speciation in *Anopheles funestus* from Burkna Faso. *Parassitologia* **41**: 595-611.
- Coyne, J., Aulard, S. & Berry, A. 1991. Lack of underdominance in a naturally occurring pericentric inversion in *Drosophila melanogaster* and its implications for chromosome evolution *Genetics* **129**: 791-802.
- Coyne, J.A. & Orr, H.A. 1989. Patterns of speciation in *Drosophila*. *Evolution* **43**: 362-381.
- Coyne, J.A., Meyers, W., Crittenden, A.P. & Sniegowski, P. 1993. The fertility effects of pericentric inversions in *Drosophila melanogaster*. *Genetics* **134**: 487-496.
- Coyne, J.A. & Orr, H.A. 1997. "Patterns of speciation in *Drosophila*" revisited. *Evolution* **51**: 295-303.
- Coyne, J.A. & Orr, H.A. 2004. *Speciation*. Sinauer Associates, Inc, Sunderland, MA.
- Csillery, K., Blum, M.G.B., Gaggiotti, O.E. & Francois, O. 2010. Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* **25**: 410-418.
- Dia, I., Lochouarn, L., Boccolini, D., Costantini, C. & Fontenille, D. 2000. Spatial and temporal variations of the chromosomal inversion polymorphism of *Anopheles funestus* in Senegal. *Parasite-Journal de la Societe Francaise de Parasitologie* **7**: 179-184.
- Diabate, A., Dao, A., Yaro, A.S., Adamou, A., Gonzalez, R., Manoukis, N.C., et al. 2009. Spatial swarm segregation and reproductive isolation between the molecular forms of *Anopheles gambiae*. *Proceedings of the Royal Society B-Biological Sciences* **276**: 4215-4222.
- Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., et al. 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**: 152-154.
- Dobzhanski, T. 1947. Genetis of natural populations. XIV. A response of certain gene arrangements in the third chromosome of *Drosophila pseudoobscura* to natural selection. *Genetics* **32**: 142-160.
- Dobzhansky, T. 1944. Chromosomal races in *Drosophila pseudoobscura* and *Drosophila persimilis*. *Carnegie Institute Washington Publication* **554**: 47-144.
- Dobzhansky, T. 1951. *Genetics and the Origin of Species*, 3 edn. Columbia University Press, New York.

- Dumas, D. & Britton-Davidian, J. 2002. Chromosomal Rearrangements and Evolution of Recombination: Comparison of Chiasma Distribution Patterns in Standard and Robertsonian Populations of the House Mouse. *Genetics* **162**: 1355-1366.
- Ellegren, H., Smeds, L., Burri, R., Olason, P.I., Backstrom, N., Kawakami, T., et al. 2012. The genomic landscape of species divergence in Ficedula flycatchers. *Nature* **491**: 756-760.
- Enayati, A. & Hemingway, J. 2010. Malaria management: past, present, and future. *Annu. Rev. Entomol.* **55**: 569-591.
- Estoup, A., Jarne, P. & Cornuet, J.M. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* **11**: 1591-1604.
- Feder, J.L., Roethele, F.B., Filchak, K., Niedbalski, J. & Romero-Severson, J. 2003. Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. *Genetics* **163**: 939-953.
- Feder, J.L. & Nosil, P. 2009. Chromosomal inversions and species differences: when are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? *Evolution* **63**: 3061-3075.
- Feder, J.L. & Nosil, P. 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* **64**: 1729-1747.
- Fouet, C., Gray, E., Besansky, N.J. & Costantini, C. 2012. Adaptation to aridity in the malaria mosquito *Anopheles gambiae*: chromosomal inversion polymorphism and body size influence resistance to desiccation. *PLoS One* **7**: e34841.
- Galetti, P.M., Jr., Aguilar, C.T. & Molina, W.F. 2000. An overview of marine fish cytogenetics. *Hydrobiologia* **420**: 55-62.
- Gavrilets, S. 2004. *Fitness Landscapes and the Origin of Species*. Princeton University Press, Princeton NJ.
- Gazave, E., Catalan, J., Da Gracia Ramalhinho, M., Da Luz Mathias, M., Claudia Nunes, A., Dumas, D., et al. 2003. The non-random occurrence of Robertsonian fusion in the house mouse. *Genet Res* **81**: 33-42.
- Gibson, G. & Russell, I. 2006. Flying in tune: Sexual recognition in mosquitoes. *Current Biology* **16**: 1311-1316.
- Gillespie, J.H. 2000. Genetic drift in an infinite population. The pseudohitchhiking model. *Genetics* **155**: 909-919.
- Giménez, M.D., White, T.A., Hauffe, H.C., Panithanarak, T. & Searle, J.B. 2013. Understanding the basis of diminished gene flow between hybridizing chromosome races of the house mouse. *Evolution*: no-no.
- Gray, E.M., Rocca, K.A.C., Costantini, C. & Besansky, N.J. 2009. Inversion 2La is associated with enhanced desiccation resistance in *Anopheles gambiae*. *Malaria Journal* **8**.
- Green, C.A. & Hunt, R.H. 1980. Interpretation of variation in ovarian polytene chromosomes of *Anopheles funestus* Giles, *A. parvipes* Gillies, and *A. Aruni*? *Genetica* **51**: 187-195.

- Grossen, C., Neuenschwander, S. & Perrin, N. 2012. The evolution of XY-recombination: Sexually antagonistic selection versus deleterious mutation load. *Evolution*: DOI: 10.1111/j.1558-5646.2012.01661.x.
- Grossman, a.I., Short, R.B. & Cain, G.D. 1981. Karyotype evolution and sex chromosome differentiation in Schistosomes (Trematoda, Schistosomatidae). *Chromosoma* **84**: 413-430.
- Guelbeogo, W.M., Grushko, O., Boccolini, D., Ouedraogo, P.A., Besansky, N.J., Sagnon, N.F., et al. 2005. Chromosomal evidence of incipient speciation in the Afrotropical malaria mosquito *Anopheles funestus*. *Medical and Veterinary Entomology* **19**: 458-469.
- Guelbeogo, W.M., Sagnon, N., Grushko, O., Yameogo, M.A., Boccolini, D., Besansky, N.J., et al. 2009. Seasonal distribution of *Anopheles funestus* chromosomal forms from Burkina Faso. *Malaria Journal* **8**.
- Guerrero, R.F., Rousset, F. & Kirkpatrick, M. 2012. Coalescence patterns for chromosomal inversions in divergent populations. *Philosophical Transactions of the Royal Society B-Biological Sciences* **367**: 430-438.
- Hamerton, J., Canning, N., Ray, M. & Smith, S. 1975. A cytogenetic survey of 14,069 newborn infants. *Clinical genetics* **8**: 223-243.
- Hedrick, P. 1981. The Establishment of Chromosomal Variants. *Evolution* **35**: 322-332.
- Heslop-Harrison, J.S.P. 2012. Genome evolution: extinction, continuation or explosion? *Current opinion in plant biology* **15**: 115-121.
- Hoffmann, A. & Rieseberg, L. 2008a. Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of *Annual Reviews of Ecology Evolution and Systematics* **39**: 21-42.
- Hoffmann, A.A. & Rieseberg, L.H. 2008b. Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation? *Annu Rev Ecol Evol S* **39**: 21-42.
- Hudson, R.R. & Kaplan, N.L. 1988. The coalescent process in models with selection and recombination. *Genetics* **120**: 831-840.
- Ijdo, J.W., Baldini, A., Ward, D.C., Reeders, S.T. & Wells, R.A. 1991. Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proceedings of the National Academy of Sciences* **88**: 9051-9055.
- Janes, D.E., Ezaz, T., Graves, J.A.M. & Edwards, S.V. 2009. Recombination and Nucleotide Diversity in the Sex Chromosomal Pseudoautosomal Region of the Emu, *Dromaius novaehollandiae*. *J Hered* **100**: 125-136.
- Kamau, L., Munyekenye, G.O., Koekemoer, L.L., Hunt, R.H. & Coetzee, M. 2003. A survey of the *Anopheles funestus* (Diptera: Culicidae) group of mosquitoes from 10 sites in Kenya with special emphasis on population genetic structure based on chromosomal inversion karyotypes. *Journal of Medical Entomology* **40**: 664-671.
- Kaplan, N., Hudson, R.R. & Iizuka, M. 1991. The coalescent process in models with selection, recombination and geographic subdivision. *Genet Res* **57**: 83-91.
- Kaplan, N.L., Darden, T. & Hudson, R.R. 1988. The coalescent process in models with selection. *Genetics* **120**: 819-820.

- Kaplan, N.L., Hudson, R.R. & Langley, C.H. 1989. The "Hitchhiking Effect" revisited. *Genetics* **123**: 887-899.
- Kawamura, T. & Nishioka, M. 1977. *The reproductive biology of amphibians*. Plenum Publishing Corporation, New York & London.
- Kennington, W.J., Partridge, L. & Hoffmann, A.A. 2006. Patterns of diversity and linkage disequilibrium within the cosmopolitan inversion *In(3R)Payne* in *Drosophila melanogaster* are indicative of coadaptation. *Genetics* **172**: 1655-1663.
- Kidwell, J., Clegg, M., Stewart, F. & Prout, T. 1977. Regions of stable equilibria for models of differential selection in the two sexes under random mating. *Genetics* **85**: 171.
- Kim, Y. & Stephan, W. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* **160**: 765-777.
- Kimura, M. & Ohta, T. 1971. *Theoretical Topics in Population Genetics*. Princeton University Press, Princeton, NJ.
- King, M. 1993. *Species Evolution: The Role of Chromosomal Change*. Cambridge University Press, Cambridge, U.K.
- Kirkpatrick, M. & Ravigné, V. 2002. Speciation by natural and sexual selection: Models and experiments. *Am Nat* **159**: S22-S35.
- Kirkpatrick, M. & Barton, N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* **173**: 419.
- Kirkpatrick, M. 2010. How and why chromosome inversions evolve. *PLoS Biology* **8**: 1000501.
- Kirkpatrick, M., Guerrero, R.F. & Scarpino, S.V. 2010. Patterns of neutral genetic variation on recombining sex chromosomes. *Genetics* **184**: 1141-1152.
- Kitano, J., Ross, J.A., Mori, S., Kume, M., Jones, F.C., Chan, Y.F., et al. 2009. A role for a neo-sex chromosome in stickleback speciation. *Nature* **461**: 1079-1083.
- Kolaczowski, B., Kern, A.D., Holloway, A.K. & Begun, D.J. 2011. Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics* **187**: 245-260.
- Kondo, M., Nagao, E., Mitani, H. & Shima, A. 2001. Differences in recombination frequencies during female and male meioses of the sex chromosomes of the medaka, *Oryzias latipes*. *Genet Res* **78**: 23-30.
- Lande, R. 1979. Effective deme sizes during long-term evolution estimated from rates of chromosomal rearrangement. *Evolution* **33**: 234-251.
- Lawniczak, M.K.N., Emrich, S.J., Holloway, A.K., Regier, A.P., Olson, M., White, B., et al. 2010. Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science* **330**: 512-514.
- Lehmann, T. & Diabate, A. 2008. The molecular forms of *Anopheles gambiae*: a phenotypic perspective. *Infection, Genetics and Evolution* **8**: 737-746.
- Lewontin, R.C. & White, M.J.D. 1960. Interaction between inversion polymorphisms of two chromosome pairs in the grasshopper, *Moraba scurra*. *Evolution* **14**: 116-129.

- Livingstone, K. & Rieseberg, L. 2004. Chromosomal evolution and speciation: a recombination-based approach. *New Phytologist* **161**: 107-112.
- Lochouart, L., Dia, I., Boccolini, D., Coluzzi, M. & Fontenille, D. 1998. Bionomical and cytogenetic heterogeneities of *Anopheles funestus* in Senegal. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **92**: 607-612.
- Lowry, D.B. & Willis, J.H. 2010. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PloS Biology* **8**: e1000500.
- Machado, C.A., Haselkorn, T.S. & Noor, M.A.F. 2007. Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **175**: 1289-1306.
- Manoukis, N.C., Powell, J.R., Toure, M.B., Sacko, A., Edillo, F.E., Coulibaly, M.B., et al. 2008a. A test of the chromosomal theory of ecotypic speciation in *Anopheles gambiae*. *P Natl Acad Sci USA* **105**: 2940-2945.
- Manoukis, N.C., Powell, J.R., Touré, M.B., Sacko, A., Edillo, F.E., Coulibaly, M.B., et al. 2008b. A test of the chromosomal theory of ecotypic speciation in *Anopheles gambiae*. *Proc Natl Acad Sci USA* **105**: 2940-2945.
- Matsuba, C., Alho, J.S. & Merila, J. 2010. Recombination Rate between Sex Chromosomes Depends on Phenotypic Sex in the Common Frog. *Evolution* **64**: 3634-3637.
- Matzkin, L., Merritt, T., Zhu, C. & Eanes, W. 2005. The structure and population genetics of the breakpoints associated with the cosmopolitan chromosomal inversion *In(3R)Payne* in *Drosophila melanogaster*. *Genetics* **170**: 1143-1152.
- McAllister, B.F. 2003. Sequence differentiation associated with an inversion on the neo-X chromosome of *Drosophila americana*. *Genetics* **165**: 1317-1328.
- McGaugh, S.E., Machado, C.A. & Noor, M.A.F. 2012. Genomic impacts of chromosomal inversions in parapatric species of *Drosophila*. *Philos T Roy Soc B* **367**.
- Medarde, N., López-Fuster, M.J., Muñoz-Muñoz, F. & Ventura, J. 2012. Spatio-temporal variation in the structure of a chromosomal polymorphism zone in the house mouse. *Heredity* **109**: 78-89.
- Michel, A.P., Grushko, O., Guelbeogo, W.M., Lobo, N.F., Sagnon, N., Costantini, C., et al. 2006. Divergence with gene flow in *Anopheles funestus* from the Sudan Savanna of Burkina Faso, West Africa. *Genetics* **173**: 1389-1395.
- Michel, A.P., Sim, S., Powell, T.H.Q., Taylor, M.S., Nosil, P. & Feder, J.L. 2010. Widespread genomic divergence during sympatric speciation. *P Natl Acad Sci USA* **107**: 9724-9729.
- Munté, A., Rozas, J., Aguadé, M. & Segarra, C. 2005. Chromosomal inversion polymorphism leads to extensive genetic structure: a multilocus survey in *Drosophila subobscura*. *Genetics* **169**: 1573-1581.

- Murray, C.J.L., Rosenfeld, L.C., Lim, S.S., Andrews, K.G., Foreman, K.J., Haring, D., et al. 2012. Global malaria mortality between 1980 and 2010: a systematic analysis. *Lancet* **379**: 413-431.
- Nagylaki, T. 1980. The strong-migration limit in geographically structured populations. *J Math Biol* **9**: 101-114.
- Navarro, A., Betran, E., Barbadilla, A. & Ruiz, A. 1997. Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. *Genetics* **146**: 695-709.
- Navarro, A., Barbadilla, A. & Ruiz, A. 2000. Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in *Drosophila*. *Genetics* **155**: 685-698.
- Navarro, A. & Barton, N.H. 2003. Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution* **57**: 447-459.
- Neafsey, D.E., Lawniczak, M.K.N., Park, D.J., Redmond, S.N., Coulibaly, M.B., Traore, S.F., et al. 2010. SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. *Science* **330**: 514-517.
- Nevo, E. 2012. Stress, adaptation, and speciation in the evolution of the blind mole rat, *Spalax*, in Israel. *Molecular phylogenetics and evolution*.
- Noor, M.A.F., Grams, K.L., Bertucci, L.A. & Reiland, J. 2001. Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci USA* **98**: 12084-12088.
- Noor, M.A.F., Garfield, D.A., Schaeffer, S.W. & Machado, C.A. 2007. Divergence between the *Drosophila pseudoobscura* and *D. persimilis* genome sequences in relation to chromosomal inversions. *Genetics* **177**: 1417-1428.
- Nordborg, M. 1997. Structured coalescent processes on different time scales. *Genetics* **146**: 1501-1514.
- Nosil, P., Funk, D.J. & Ortiz-Barrientos, D. 2009. Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* **18**: 375-402.
- Ohno, S. 1967. *Sex chromosomes and sex-linked genes*. Springer-Verlag, Berlin, New York etc.
- Olivry, J.C. 1986. *Fleuves et Rivières du Cameroun*. ORSTOM, Paris.
- Onyabe, D.Y. & Conn, J.E. 2001. Genetic differentiation of the malaria vector *Anopheles gambiae* across Nigeria suggests that selection limits gene flow. *Heredity* **87**: 647-658.
- Ortiz-Barrientos, D., Reiland, J., Hey, J. & Noor, M.A. 2002. Recombination and the divergence of hybridizing species. *Genetica* **116**: 167-178.
- Otto, S.P., Pannell, J.R., Peichel, C.L., Ashman, T.L., Charlesworth, D., Chippindale, A.K., et al. 2011. About PAR: The distinct evolutionary dynamics of the pseudoautosomal region. *Trends Genet* **27**: 358-367.
- Pagacova, E., Cernohorska, H., Kubickova, S., Vahala, J. & Rubes, J. 2009. Centric fusion polymorphism in captive animals of family Bovidae. *Conservation Genetics* **12**: 71-77.

- Pardo-Manuel de Villena, F. & Sapienza, C. 2001. Female meiosis drives karyotypic evolution in mammals. *Genetics* **159**: 1179-1189.
- Payne, F. 1924. Crossover modifiers in the third chromosome of *Drosophila melanogaster*. *Genetics* **9**: 327-342.
- Peichel, C.L., Ross, J.A., Matson, C.K., Dickson, M., Grimwood, J., Schmutz, J., et al. 2004. The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Current Biology* **14**: 1416-1424.
- Pennetier, C., Warren, B., Dabire, K.R., Russell, I.J. & Gibson, G. 2009. "Singing on the wing" as a mechanism for species recognition in the malarial mosquito *Anopheles gambiae*. *Current Biology* **20**: 131-136.
- Perevozkin, V.P., Printseva, A.A., Maslennikov, P.V. & Bondarchuk, S.S. 2012. Genetic aspects of sexual behavior in malaria mosquitoes on the basis of specific acoustic signals at mating. *Russian Journal of Genetics* **48**: 587-591.
- Perrin, N. 2009. Sex Reversal: A Fountain of Youth for Sex Chromosomes? *Evolution* **63**: 3043-3049.
- Petrarca, V. & Beier, J.C. 1992. Intraspecific chromosomal polymorphism in the *Anopheles gambiae* complex as a factor affecting malaria transmission in the Kisumu area of Kenya. *American Journal of Tropical Medicine and Hygiene* **46**: 229-237.
- Petrarca, V., Nugud, A.D., Ahmed, M.A.E., Haridi, A.M., Di Deco, M.A. & Coluzzi, M. 2000. Cytogenetics of the *Anopheles gambiae* complex in Sudan, with special reference to *An. arabiensis*: relationships with East and West African populations. *Medical and Veterinary Entomology* **14**: 149-164.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. 2007. *Numerical Recipes: The Art of Scientific Computation*, 3rd edn. Cambridge University Press, Cambridge, UK.
- Przeworski, M., Coop, G. & Wall, J. 2005. The signature of positive selection on standing genetic variation. *Evolution* **59**: 2312-2323.
- Rai, K.S. 2010. Insights from mosquito evolution: Patterns, tempo and speciation. In: *Nature at Work: Ongoing Saga of Evolution in Play* (V. P. Sharma, ed., pp. 197-215. Springer, New Dehli.
- Rice, W.R. 1984. Sex-Chromosomes and the Evolution of Sexual Dimorphism. *Evolution* **38**: 735-742.
- Rice, W.R. 1987. Speciation via habitat specialization: the evolution of reproductive isolation as a correlated character. *Evolutionary Ecology* **1**: 301-314.
- Rieseberg, L.H. 2001. Chromosomal rearrangements and speciation. *Trends Ecol Evol* **16**: 351-358.
- Rishikesh, N., Dideco, M.A., Petrarca, V. & Coluzzi, M. 1985. Seasonal variations in indoor resting *Anopheles gambiae* and *Anopheles arabiensis* in Kaduna, Nigeria. *Acta Tropica* **42**: 165-170.
- Robertson, W.R.B. 1916. Chromosome studies. I. Taxonomic relationships shown in the chromosomes of tettigidae and acrididae: V-shaped chromosomes and their

- significance in acrididae, locustidae, and gryllidae: Chromosomes and variation. *Journal of Morphology* **27**: 179-331.
- Rocca, K.A.C., Gray, E.M., Costantini, C. & Besansky, N.J. 2009. 2La chromosomal inversion enhances thermal tolerance of *Anopheles gambiae* larvae. *Malaria Journal* **8**.
- Sandler, L. & Hiraizumi, Y. 1960. Meiotic drive in natural populations of *Drosophila melanogaster*. IV. Instability at the segregation-distorter locus. *Genetics* **45**: 1269-1287.
- Sanford, M.R., Demirci, B., Marsden, C.D., Lee, Y., Cornel, A.J. & Lanzaro, G.C. 2011. Morphological differentiation may mediate mate choice between incipient species of *Anopheles gambiae* s.s. *PLoS One* **6**.
- Sankoff, D. & Nadeau, J.H. 2003. Chromosome rearrangements in evolution: From gene order to genome sequence and back. *Proc Natl Acad Sci USA* **100**: 11188-11189.
- Santos, I.S., Delabie, J.H.C., Silva, J.G., Costa, M.A., Barros, L.A.C., Pompolo, S.G., et al. 2012. Karyotype Differentiation among Four Dinoponera (Formicidae : Ponerinae) Species. **95**: 737-742.
- Schaeffer, S.W., Goetting-Minesky, M.P., Kovacevic, M., Peoples, J.R., Graybill, J.L., Miller, J.M., et al. 2003. Evolutionary genomics of inversions in *Drosophila pseudoobscura*: evidence for epistasis. *Proc Natl Acad Sci USA* **100**: 8319-8324.
- Schaeffer, S.W. & Anderson, W.W. 2005. Mechanisms of genetic exchange within the chromosomal inversions of *Drosophila pseudoobscura*. *Genetics* **171**: 1729-1739.
- Schaeffer, S.W. 2008. Selection in heterogeneous environments maintains the gene arrangement polymorphism of *Drosophila pseudoobscura*. *Evolution* **62**: 3082-3099.
- Seyfert, A.L., Cristescu, M.E.A., Frisse, L., Schaack, S., Thomas, W.K. & Lynch, M. 2008. The rate and spectrum of microsatellite mutation in *Caenorhabditis elegans* and *Daphnia pulex*. *Genetics* **178**: 2113-2121.
- Sharakhov, I., Braginets, O., Grushko, O., Cohuet, A., Guelbeogo, W.M., Boccolini, D., et al. 2004. A microsatellite map of the African human malaria vector *Anopheles funestus*. *J Hered* **95**: 29-34.
- Sharakhov, I.V., Serazin, A.C., Grushko, O.G., Dana, A., Lobo, N., Hillenmeyer, M.E., et al. 2002. Inversions and gene order shuffling in *Anopheles gambiae* and *A. funestus*. *Science* **298**: 182-185.
- Simard, F., Ayala, D., Kamdem, G.C., Pombi, M., Etouana, J., Ose, K., et al. 2009. Ecological niche partitioning between *Anopheles gambiae* molecular forms in Cameroon: the ecological side of speciation. *BMC Ecology* **9**: Article No.: 17.
- Slatkin, M. 1991. Inbreeding coefficients and coalescence times. *Genet Res* **58**: 167-175.
- Stefansson, H., Helgason, A., Thorleifsson, G., Steinthorsdottir, V., Masson, G., Barnard, J., et al. 2005. A common inversion under selection in Europeans. *Nature Genetics* **37**: 129-137.
- Stöck, M., Horn, A., Grossen, C., Lindtke, D., Sermier, R., Betto-Colliard, C., et al. 2011. Ever-Young Sex Chromosomes in European Tree Frogs. *Plos Biology* **9**.

- Stöck, M., Dufresnes, C., Litvinchuk, S.N., Lymberakis, P., Biollay, S., Berroneau, M., et al. 2012. Cryptic diversity among Western Palearctic tree frogs: Postglacial range expansion, range limits, and secondary contacts of three European tree frog lineages (*Hyla arborea* group). *Molecular Phylogenetics and Evolution* **in press**.
- Strasburg, J.L., Scotti-Saintagne, C., Scotti, I., Lai, Z. & Rieseberg, L.H. 2009. Genomic patterns of adaptive divergence between chromosomally differentiated sunflower species. *Mol Biol Evol* **26**: 1341-1355.
- Stump, A.D., Pombi, M., Goeddel, L., Ribeiro, J.M.C., Wilder, J.A., Torre, A.D., et al. 2007. Genetic exchange in 2La inversion heterokaryotypes of *Anopheles gambiae*. *Insect Mol Biol* **16**: 703-709.
- Sturtevant, A.H. & Beadle, G.W. 1936. The relations of inversions in the X chromosome of *Drosophila melanogaster* to crossing over and disjunction *Genetics* **21**: 554-604.
- Tavare, S., Balding, D.J., Griffiths, R.C. & Donnelly, P. 1997. Inferring coalescence times from DNA sequence data. *Genetics* **145**: 505-518.
- Toure, Y.T., Petrarca, V., Traore, S.F., Coulibaly, A., Maiga, H.M., Sankare, O., et al. 1994. Ecological genetic studies in the chromosomal form Mopti of *Anopheles gambiae* s.str. in Mali, West Africa. *Genetica* **94**: 213-223.
- Toure, Y.T., Traore, S.F., Sankare, O., Sow, M.Y., Coulibaly, A., Esposito, F., et al. 1996. Perennial transmission of malaria by the *Anopheles gambiae* complex in a North Sudan Savanna area of Mali. *Medical and Veterinary Entomology* **10**: 197-199.
- Toure, Y.T., Petrarca, V., Traore, S.F., Coulibaly, A., Maiga, H.M., Sankare, O., et al. 1998. The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. *Parassitologia (Rome)* **40**: 477-511.
- Tripet, F., Wright, J., Cornel, A., Fofana, A., McAbee, R., Meneses, C., et al. 2007. Longitudinal survey of knockdown resistance to pyrethroid (kdr) in Mali, West Africa, and evidence of its emergence in the Bamako form of *Anopheles gambiae* s.s. *American Journal of Tropical Medicine and Hygiene* **76**: 81-87.
- Tsuchimatsu, T., Suwabe, K., Shimizu-Inatsugi, R., Isokawa, S., Pavlidis, P., Städler, T., et al. 2010. Evolution of self-compatibility in *Arabidopsis* by a mutation in the male specificity gene. *Nature* **464**: 1342-1346.
- Turner, T., Hahn, M. & Nuzhdin, S.V. 2005. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*.
- Turner, T.L. & Hahn, M.W. 2007. Locus- and population-specific selection and differentiation between incipient species of *Anopheles gambiae*. *Mol Biol Evol* **24**: 2132-2138.
- Turner, T.L. & Hahn, M.W. 2010. Genomic islands of speciation or genomic islands and speciation? *Molecular Ecology* **19**: 848-850.
- Verardi, A., Canestrelli, D. & Nascetti, G. 2009. Nuclear and mitochondrial patterns of introgression between the parapatric European treefrogs *Hyla arborea* and *H. intermedia*. *Ann Zool Fenn* **46**: 247-258.

- Via, S. 2012. Hiding in plain sight: divergence hitchhiking and the dynamics of genomic isolation between incipient ecological species. *Philos T Roy Soc B* **367**.
- Wakeley, J. 2009. *Coalescent Theory: An Introduction*. Roberts and Company Publishers, Greenwood Village.
- Wang-Sattler, R., Blandin, S., Ning, Y., Blass, C., Dolo, G., Toure, Y.T., et al. 2007. Mosaic genome architecture of the *Anopheles gambiae* species complex. *PLoS One* **2**.
- Wegmann, D., Leuenberger, C. & Excoffier, L. 2009. Efficient Approximate Bayesian Computation Coupled With Markov Chain Monte Carlo Without Likelihood. *Genetics* **182**: 1207-1218.
- Wegmann, D., Leuenberger, C., Neuenschwander, S. & Excoffier, L. 2010. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *Bmc Bioinformatics* **11**: -.
- White, B.J., Hahn, M.W., Pombi, M., Cassone, B.J., Lobo, N.F., Simard, F., et al. 2007. Localization of candidate regions maintaining a common polymorphic inversion (2La) in *Anopheles gambiae*. *PLoS Genetics* **3**: 2404-2414.
- White, B.J., Cheng, C., Sangaré, D., Lobo, N.F., Collins, F.H. & Besansky, N.J. 2009. The population genomics of trans-specific inversion polymorphisms in *Anopheles gambiae*. *Genetics* **183**: 275-288.
- White, B.J., Cheng, C., Simard, F., Costantini, C. & Besansky, N.J. 2010. Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. *Molecular Ecology* **19**: 925-939.
- White, M.J.D. 1969. Chromosomal rearrangements and speciation in animals. *Annu Rev Genet* **3**: 75-98.
- White, M.J.D. 1973. *Animal cytology and evolution*, 3d edn. University Press, Cambridge, Eng.
- White, M.J.D. 1978. *Modes of speciation*. W.H. Freeman, San Francisco, California.
- Whitlock, M.C. & Schluter, D. 2009. *The Analysis of Biological Data*. Roberts and Company Publishers, Greenwood Village.
- Wolfram Research, I. 2008. *Mathematica*, Version 7.0 edn. Wolfram Research, Inc, Campaign, Illinois.
- Wright, S. 1951. The genetical structure of populations. *Ann. Eugen.* **15**: 323-354.
- Wu, C.I. 2001. The genic view of the process of speciation. *J Evolution Biol* **14**: 851-865.
- Yatabe, Y., Kane, N.C., Scotti-Saintagne, C. & Rieseberg, L.H. 2007. Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics* **175**: 1883-1893.

Vita

Rafael Felipe Guerrero Farías was born and raised in Bogotá (Colombia) in 1980, the son of Rafael Guerrero Lozano and Nubia Farías García. After graduating from Colegio San Carlos in 1998, he enrolled in the biology program at the Universidad de Los Andes. There he worked under the supervision of Adolfo Amézquita, both for his undergraduate and M.S. thesis. In 2005, he worked as a field assistant for the Tungara Team lead by Mike Ryan in Panamá. During that time, he heard about the wonders of Texas, and decided to enroll in the Ecology, Evolution and Behavior program at The University of Texas – Austin under the supervision of Mark Kirkpatrick. Rafael's graduate work focused on population genetics, meddling in theoretical questions and data analysis. Provided that immigration authorities allow him, Rafael will be joining Matt Hahn's lab at Indiana University in the autumn of 2013.

Permanent address: r.guerrero@utexas.edu

This dissertation was typed by the author.