THE UNIVERSITY OF CHICAGO

CONSEQUENCES AND SCALINGS OF POPULATION HISTORY

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECOLOGY AND EVOLUTION

BY

EVAN M. KOCH

CHICAGO, ILLINOIS

AUGUST 2018

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

My PhD advisor, John Novembre, has my eternal gratitude for accepting me into his nascent group at the University of Chicago when I was relatively lost after my first year. He has been a tireless and supportive mentor in steering my education and has allowed me the freedom to follow my scientific interests. As my thesis committee, Dick Hudson, Matthew Stephens, and Chung-I Wu, have given invaluable advice on the work presented here. I would also like to acknowledge my undergraduate mentors, Claus Wilke and Mark Kirkpatrick, for first exposing me to the fascinating world of evolutionary and population genetics and trusting me to conduct research.

My officemate, Hussein Al-Asadi, has been an inescapable source of discussion and advice, and my other labmates and fellow fourth floor denizens have contributed in so many ways to my work and intellectual development. I am thankful for all the friends I have made during my time in Chicago. They have provided the necessary support and distraction for work to get done and life to go on. My partner, Jenn Stanley, has given me all the love and encouragement I could ever want or need. Finally, my parents have given guidance since the start and never accepted that there was something I could not do. To the extent that I have accomplished anything I owe it to them.

ABSTRACT

Population genomic data contains a startling amount of information about the demographic history of populations. As advances in statistical methodology for demographic inference and increases in sampling provide greater resolution into the past, empirical studies of various organisms have illuminated histories replete with changes in population size, migrations, population splits, and mergers. Demographic inference generally operates by fitting simplified models of population history to patterns of genetic variation assumed to have no effect on the fitness of the organism. An important use of these models is provide approximate dates for historical events. It is also important to understand what the implications of these models, and the complex histories they attempt to represent, are for the distribution of biologically interesting types of genetic variation. The work presented here first addresses the consequences of population history for deleterious genetic variation and genetic variation affecting quantitative traits. The investigations center around (1) how divergent histories between examples of African and Out-of-Africa populations have led to differences in the distribution of deleterious variation, and (2) how arbitrary models of demographic history affect the distribution of quantitative traits. Finally, (3) the dog/wolf mutation rate is estimated and used to scale demographic models of dog domestication.

CHAPTER 1

INTRODUCTION

Demographic history and population structure have long interested population geneticists. The demographic history of a population is conceptualized as the series of size changes, population splits, and rates of migration with other populations. Population structure refers to deviations from random mating, usually in the context of the spatial organization of populations, and is usually conceptualized as a number of subpopulations with various rates of migration between them. The concepts covered by the notions of population structure and demography are not mutually exclusive, and real populations surely have both spatial distributions and have undergone historical changes. However, the division is useful in allowing scientists to independently pose questions related to history and space.

That species are structured into a large number of smaller units with limited levels of gene flow was an important part of Sewall Wright's Shifting Balance theory of evolution (Wright, 1931). Mathematical models of this type of population structure first analyzed by Wright have been named "island models". The basic idea behind island models is that in each generation, in addition to the usual processes of mating, selection, mutation and Mendelian segregation, some number of individuals migrate out of the subpopulation (deme) and some migrate into it. Island models and their relatives have been studied extensively by population geneticists trying to understand the evolutionary consequences of the geographic structure clearly observable in nearly all species (Slatkin, 1987). Some examples include the study of speciation in the presence of gene flow (Felsenstein, 1981), the spread of adaptive alleles through a population (Fisher, 1937), and the genetic covariance expected at different distances on a landscape (Kimura and Weiss, 1964).

Models of population size change have historically received less attention than those of population structure. The first models to incorporate nonequilibrium demographic histories focused on the effects of relatively simple population bottleneck and growth events. Nei

et al. (1975) numerically solved for the time course of heterozygosity change following a population bottleneck. Subsequent theory by Watterson as well as Maruyama and Fuerst investigated the effects of both bottlenecks and population on neutral genetic variation in greater detail and showed under an infinite-alleles model that bottlenecks would decrease the number of alleles and that growth would increase the number of alleles relative to the overall heterozygosity at a locus (Watterson, 1984; Maruyama and Fuerst, 1984, 1985a,b). From a more biological perspective, bottlenecks and growth events were interesting because of their roles in theories of founder event speciation (Templeton, 1980), their effects on evolutionary rates through influencing the probability of fixation of mutations under selection (Kimura and Ota, 1974; Otto and Whitlock, 1997), and their ability to generate false positives in tests for selection (Watterson, 1986; Tajima, 1989).

Interest in demographic history has taken off more seriously since researchers have increasingly used genetic data to infer the timing and magnitude of past demographic changes. Differences in genetic variation between populations were used before DNA sequencing of populations became feasible. For instance, reduced allozyme variation was observed in the Bogotá population of *Drosophila pseudoobscura* relative to populations in the United States, and this was used to infer the occurrence of a founder event in the history of the Bogotá population (Prakash *et al.*, 1969; Prakash, 1972). However, these data were not informative about the timing and magnitude of bottlenecks associated with founder events. Starting with mtDNA and microsatellite data, parameter estimates for simple models of population growth and decline could be obtained. Slatkin and Hudson (1991) showed how growth rates could be estimated from pairwise differences sequences in a mtDNA sample and in general demonstrated how the distribution of pairwise differences is informative about population growth. Following this, a number of studies estimated the timing and rate of growth in different human populations (Rogers, 1995; Reich and Goldstein, 1998; Kimmel *et al.*, 1998).

Since nuclear DNA sequences have become widely available, a great variety of statistical methods have been devised to fit demographic models to sequence data (Schraiber and Akey, 2015). Different approaches use different aspects of the data. One set uses the proportions of variants at different counts in a sample, called the site frequency spectrum (Marth *et al.*, 2004; Gutenkunst *et al.*, 2009; Excoffier *et al.*, 2013). Another set attempts to reconstruct genealogies in short segments of the genome under the assumption of no recombination and then find the population history that best fits these individual genealogies (Rannala and Yang, 2003; Gronau *et al.*, 2011). Information about past population sizes is also contained in the sequence of correlated genealogies along a sample of chromosomes, and demographic inference methods have been developed to take advantage of this as well (Li and Durbin, 2011). Methods that use both the frequencies of variants as well as the correlation of genealogies at linked sites have also been developed, but due to the difficulty of specifying a full likelihood these methods rely on simulations (Voight *et al.*, 2005; Wegmann *et al.*, 2009).

The application of demographic inference methods has uncovered highly nonequilibrium histories in humans and many other species to which they have been applied. For example, Gravel *et al.* (2011) inferred a demographic history for humans from individuals of European descent from Utah, Han Chinese individuals from Beijing, Japanese individuals from Tokyo, and Yoruba individuals from Ibadan. They found evidence for population growth in the ancestral human population, a bottleneck associated with the Out-of-Africa split, ongoing migration after population splits, and recent population growth. That populations sizes have experienced great changes over time is not surprising given that the range of times to find common ancestors in many species encompasses major climactic chances. In a study of population size histories in multiple avian species Nadachowska-Brzyska *et al.* (2015) found evidence for a decrease in population size during the last glacial period. Given the extensive evidence for complex, nonequilibrium demographic histories, it is important to ask

what effects, if any, these have for patterns of biologically interesting genetic variation. In the first two chapters of this work we ask what the consequence of complex demographies are for patterns of deleterious genetic variation and for the distribution of quantitative trait values in the absence of selection.

The frequencies of slightly deleterious mutations are impacted both by natural selection and genetic drift (Kimura *et al.*, 1963). The effective population size therefore has a strong impact on their distribution in the population. As an explanation for an excess of low-frequency allozyme alleles and a high variance in the substitution rate, Ohta (1972, 1973) suggested that slightly deleterious (nearly neutral) mutations might be common. While it is still unclear what impact the fixation of deleterious mutations has on the substitution rate, a large number of studies have confirmed that slightly deleterious alleles are a common component of genetic variation (Fay *et al.*, 2001; Eyre-Walker and Keightley, 2007; Boyko *et al.*, 2008; Chen *et al.*, 2017). Since this variation is sensitive to the effective population size, it makes sense to ask what effect histories of large-scale population size changes have had on its distribution.

Empirical studies have found differences in the distribution of deleterious genetic variation between human populations (Lohmueller *et al.*, 2008; Casals *et al.*, 2013; Peischl *et al.*, 2013; Henn *et al.*, 2016; Peischl *et al.*, 2017), where deleteriousness is usually crudely measured by whether an exonic mutation is synonymous or nonsynonymous. Large differences were described in the proportion of sample variants that were classified as deleterious, with non-African samples showing a greater proportion of deleterious variants than African samples (Lohmueller, 2014a). Differences in this proportion reflect differences in the frequency spectrum of deleterious variation (Boyko *et al.*, 2008). However, it has been shown that, although differences in the proportion of sample variants that are deleterious is a consequences of differences in demographic history, this does not mean that individual genomes from certain human populations have accumulated more deleterious mutations

(Simons *et al.*, 2014; Do *et al.*, 2015).

The first chapter of this work takes a temporal look at how summaries of slightly deleterious variation have changed in example demographic histories of African and Out-of-Africa human populations (Tennessen *et al.*, 2012). The distribution of neutral genetic variation has obviously also diverged between human populations, and we ask the question of whether changes in different summaries of deleterious variation can be described as "due to selection".

When a large number of genetic variants affect a quantitative trait, the trait follows a normal distribution as expected from infinitesimal models (Fisher, 1918), and the average genetic distance between individuals is sufficient to describe their similarity in a neutral quantitative trait. However, when a trait is controlled by smaller numbers of variants the details of demographic history may matter and could impact the deviation of that trait from normal distribution. Khaitovich *et al.* (2005) and Schraiber and Landis (2015) developed neutral models for quantitative traits where the number of loci controlling the trait is a parameter to be varied and where any distribution of mutational effects is allowed. In the second chapter we expand the modeling approach of these studies to arbitrary demographic histories and population structure. We then show the extent to which the distribution of a neutral quantitative trait is affected by demography. Deviations from normality, and therefore demographic history, can impact the results of tests for selection on quantitative traits.

Demographic history and the reconstruction of past events is of course interesting in its own right. The demographic history of dogs and wolves has received special attention because dogs were the first animals domesticated by humans. Central questions in dog domestication research surround the timing and location of domestication (Freedman and Wayne, 2017). Researchers have attempted to answer these questions by using demographic inference methods to find which wolf population dogs split from most recently and at what time this occurred (Wang *et al.*, 2013; Freedman *et al.*, 2014). Scaling demographic models

to units of generations requires knowing the mutation rate. In the third chapter of this work we directly measure the mutation rate in wolves by sequencing a pedigree of wolves and identifying de novo mutation between parents and offspring.

In addition to their importance for scaling demographic models, mutation rate estimates are useful for understanding the evolutionary forces behind differences in mutation rates between species. In particular, the drift-barrier hypothesis for the evolution of mutation rates predicts that the primary determinant of mutations rates is the effective population size (Lynch *et al.*, 2016). This occurs because the strength of selection against mutator alleles is proportional to the rate at which deleterious mutations are produced (Lynch, 2011). When the strength of selection becomes small relative to the inverse of the effective population size selection can no long effectively remove mutator alleles from the population. In line with the drift-barrier hypothesis, a negative relationship between effective population size and mutation rate is observed across species throughout the tree of life (Sung *et al.*, 2012). This pattern holds among the increasing number of species with mutation rate estimates from whole-genome sequencing (Smeds *et al.*, 2016; Lynch *et al.*, 2016; Pfeifer, 2017).

Other hypotheses exist to explain the evolution of mutation rates. It may be that there is a fitness cost at the cellular level to replication fidelity that prevents the indefinite optimization of mutation rates (Kondrashov, 1995). Metabolic rate may influence the mutation rate by increasing concentrations of mutagenic molecules or the rate of DNA synthesis (Martin and Palumbi, 1993). In animals, increased sperm competition may increase the number of cell divisions during sperm production and therefore the possibilities for error during replication (Wong, 2014). The proposed factors affecting mutation rates are not mutually exclusive and their relative importance is an empirical question. Our estimate of the mutation rate in wolves adds another data point to help understand this important area of molecular evolution.

CHAPTER 2

A TEMPORAL PERSPECTIVE ON THE INTERPLAY OF DEMOGRAPHY AND SELECTION ON DELETERIOUS VARIATION IN HUMANS

2.1 Abstract

When mutations have small effects on fitness, population size plays an important role in determining the amount and nature of deleterious genetic variation. The extent to which recent population size changes have impacted deleterious variation in humans has been a question of considerable interest and debate. An emerging consensus is that the Out-of-Africa bottleneck and subsequent growth events have been too short to cause meaningful differences in genetic load between populations; though changes in the number and average frequencies of deleterious variants have taken place. To provide more support for this view and to offer additional insight into the divergent evolution of deleterious variation across populations, we numerically solve time-inhomogenous diffusion equations and study the temporal dynamics of the frequency spectra in models of population size change for modern humans. We observe how the response to demographic change differs by the strength of selection, and we then assess whether similar patterns are observed in exome sequence data from 33,370 and 5,203 individuals of non-Finnish European ancestry and West African respectively. Our theoretical results highlight how even simple summaries of the frequency spectrum can have complex responses to demographic change. These results support that some apparent discrepancies between previous results have been driven by the behaviors of the precise summaries of deleterious variation. Further, our empirical results make clear the difficulty of inferring slight differences in frequency spectra using recent next generation sequence data.

2.2 Introduction

Inferences comparing the relative strengths of selection in different populations are particularly difficult in populations with nonequilibrium demographic histories (Brandvain and Wright, 2016). Yet understanding differences in selection on deleterious variation between populations is vital to explaining observed patterns of genetic variation.

One major context for research on deleterious variation has been in studies of human populations (Lohmueller, 2014a). Nonsynonymous variants in humans show many patterns that are similar to those seen for neutral variants (Yu *et al.*, 2002). Studies have observed decreased nonsynonymous heterozygosity and increased derived allele homozygosity in European populations relative to African ones (Lohmueller *et al.*, 2008). These effects increase as distance from Africa increases (Henn *et al.*, 2012, 2016). However, considerable debate and interest has focused on, beginning from Lohmueller *et al.* (2008), the finding that European populations have proportionally more deleterious variation than neutral variation when compared to African populations (Peischl *et al.*, 2013; Torkamani *et al.*, 2012; Tennessen *et al.*, 2012; Subramanian, 2012). Simons *et al.* (2014) and Do *et al.* (2015) show that this does not imply a biologically important difference in the deleterious allele burden or genetic load. These two studies fail to detect a difference between populations in the mean number of putatively deleterious variants contained in a single genome. Lohmueller (2014a) explains that the apparent differences between the Simons *et al.* (2014) and Do *et al.* (2015) studies versus previous ones is due to different ways of summarizing patterns of genetic variation because previous studies did not use statistics for detecting a difference in genetic load. However, subsequent work has generated conflicting observations of the derived allele burden. Fu *et al.* (2014) and Henn *et al.* (2016) observe greater numbers of deleterious variants in genomes from Out-of-Africa (OOA) populations, but the first did not perform any formal statistical test, and the test used by the second does not account for variation in the evolutionary process (as noted by Simons and Sella (2016) and Gravel

(2016)). Simons and Sella (2016) analyze different data and conclude that there is little or no difference in nonsynonymous allele burden or genetic load among contemporary human populations.

Recently, Brandvain and Wright (2016) summarize the results of many of these studies (and related ones in non-humans) and emphasize that, while many insights have been gained by describing differences in the distribution of deleterious variation between populations, further work is necessary in order to interpret these differences in terms of natural selection. Studies of deleterious variation often focus on the genetic load or the evolution of fitness differences between populations and species. However, the distribution of deleterious variation has other important consequences including effects on the trait architecture (Simons *et al.*, 2014; Gazave *et al.*, 2013; Eyre-Walker, 2010) and the future evolution of the population.

Here, we use numerical solutions of diffusion equations for the frequency spectrum and analyze the empirical site frequency spectrum of deleterious alleles in a large-scale human sequencing study. Analyzing numerical solutions can help illustrate the response of frequency spectrum summaries to changes in population size; our goal is to show how equilibrium logic can mislead because deleterious genetic variation has complex responses to different demographies.

We do this by stratifying expected changes by the strength of purifying selection and by time. We first investigate two simple demographic events: 1) a size reduction and 2) a size reduction followed by growth. We then examine previously fitted demographic models for West African and European population history. Like many studies, we take these two models as examples of African and OOA population histories and hereafter refer to them as such. We elaborate further on differences between African and OOA populations by addressing the relative impacts of drift versus selection and by analyzing the properties of a commonly used statistic: the proportion of sample variants that are deleterious. We

then compare theoretical predictions to patterns of heterozygosity, derived allele burden, and homozygosity in the Exome Aggregation Consortium data set (Lek *et al.*, 2016) when stratified by a measure predictive of evolutionary constraint (Cooper *et al.*, 2005; Davydov *et al.*, 2010). Relative to previous work examining the response of deleterious variation to demographic events, we emphasize the temporal pattern and sensitivity to the degree of purifying selection.

2.3 Methods

2.3.1 Basic model assumptions and numerical solutions

To model the evolution of the site frequency spectrum through time, we use the diffusion approximation to a Wright-Fisher model with selection in an infinitely many sites model with no linkage (Poisson random field model) (Ewens, 2004; Sawyer and Hartl, 1992; Hartl *et al.*, 1994). Since the demographies of many human populations are far from equilibrium, we numerically solve for the time evolution of the derived allele frequency density using a forward Kolmogorov diffusion equation similar to that first widely applied in population genetics by Motoo Kimura (Kimura and Weiss, 1964). Specifically, we use a numerical solution described by Evans *et al.* (2007) to obtain the function $f(x,t)$ such that $f(x,t)dx$ gives approximately the expected number of derived alleles in the small range $[x, x+dx]$, where x is the frequency of a derived mutation. $f(x,t)$ is the *frequency spectrum* of the population. This approach is similar to that used in frequency-spectrum based methods of estimating demographic histories but differs trivially in the boundary conditions (Williamson *et al.*, 2005; Gutenkunst *et al.*, 2009). We use an additive model of selection where derived allele heterozygotes and homozygotes have relative fitnesses $1-s$ and $1-2s$ respectively. The parameter s is the strength of selection against the derived variant. The system we solve

numerically is

$$\frac{\partial}{\partial t}f(x,t) = \frac{\partial}{\partial x}\left[Sx(1-x)f(x,t)\right] + \frac{\partial^2}{2\partial x^2}\left[\frac{x(1-x)}{\rho(t)}f(x,t)\right], \tag{2.1}$$

$$\lim_{x\to 0} x(1-x)f(x,t) = \theta\rho(t),$$
$$\lim_{x\to 1} x(1-x)f(x,t) = 0.$$

Here, $S = 2N_0 s$ and $\theta = 4N_0 l\mu$ where N_0 is the initial population size, l is the number of sites under consideration, and μ is the per base pair mutation rate. Time (t) is measured in $2N_0$ generations. $\rho(t)$ is the population size at time t relative to N_0. The numerical solutions use a nonuniform grid on x with finer spacing at low values to account for the highly peaked nature of the frequency spectrum that arises during strong selection and population growth (see Appendix A.2 for details). The grid on t uses a step size of 10^{-3} coalescent units, and increasing the resolution does not affect results.

To validate our implementation of this methodology we performed numerical solutions for a constant size population and compared them to equilibrium expressions to determine the deviation due to inherent error in the numerical scheme. Doing so shows that low-order moments of the frequency spectrum are stable enough that subsequent results are nearly unaffected (Figure A2). We also compare diffusion results to the Wright-Fisher markov model it approximates and find very close agreement (see Appendix A.3 for details).

For many results we consider the expected *site frequency spectrum* (SFS) in a sample of n haploid genomes. Assuming sampling with replacement from the population this is given by

$$E[q_{n,k}] = \int_0^1 \binom{n}{k} x^k (1-x)^{n-k} f(x,t)\mathrm{d}x \tag{2.2}$$

(Hartl *et al.*, 1994), where $q_{n,k}$ is the number of alleles at count k in a sample of size n. It can be seen that $E[q_{n,k}]$ depends on moments n through k of the frequency spectrum. For instance, $E[q_{7,5}] = \int_0^1 \binom{n}{k}(x^5 - 2x^6 + x^7)f(x,t)dx$, where $E[q_{k,k}]$ is the k^{th} moment

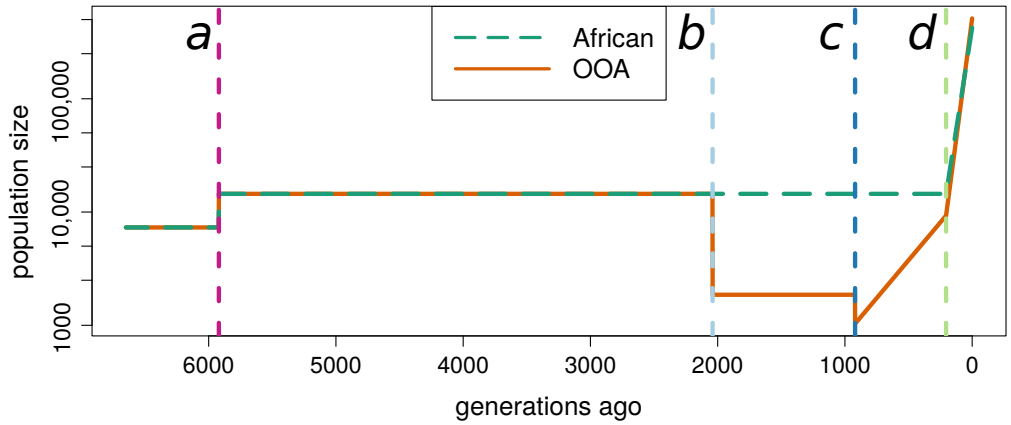


Figure 2.1: **Representative population histories used for African and Out-of-Africa demographic models** The estimated effective population sizes as a function of time estimated by Tennessen *et al.* (2012). Demographic events are shown by dashed vertical lines. Lines of the same color denote the same event in subsequent plots. Event a is an approximate doubling of the population size before the OOA split. Event b is the OOA bottleneck (13%). Event c is a bottleneck (55%) followed by exponential growth (0.31%). Event d corresponds to recent and rapid population growth experienced by both populations (African: 1.66%, OOA: 1.95%). This is a simplification of the Tennessen *et al.* (2012) model because it ignores a low rate of migration inferred to have occurred post bottleneck between the African and OOA populations.

of $f(x,t)$. Equation 2.2 is computed by numerical integration over the grid on x for which $f(x,t)$ (equation 2.1) was solved.

2.3.2 Demographic scenarios

As examples of African and OOA population histories we use the demographies inferred by Tennessen *et al.* (2012) from a large sample of 1,088 African-American and 1,351 European-American individuals. These demographies are characterized by an OOA bottleneck (13%, 2,000 generations ago), a second European bottleneck (55%, 920 generations ago) with immediate recovery at a rate of 0.31%, and recent exponential growth in both populations (1.95% in European, 1.66% in African, 205 generations ago, Figure 2.1).

2.3.3 Analysis of exome sequence data

We analyze exome sequence data from 33,370 individuals taken from a non-Finnish European (NFE) ancestry cluster and 5,203 individuals from a West African (AFR) ancestry cluster from the Exome Aggregation Consortium (ExAC)(Lek *et al.*, 2016). An advantage of the large size of the ExAC sample is that it provides more precise estimates of heterozygosity and derived allele frequency at low-frequency sites such as those under strong purifying selection. Ancestry clusters were determined by the ExAC authors using principal components analysis. We use the NFE and AFR clusters to roughly correspond to the OOA and African population models respectively. Variants were called in the data by the ExAC authors, and we filtered variants based on information they provide. This involves filtering variants by their variant quality score log-odds (VQSLOD) to obtain a set with a tranche sensitivity level of 99.6% in the ExAC training set and then removing sites with missing data in more than 90% samples in both the African and non-Finnish European groups (see Appendix A.4 for more details).

To obtain a set of high-confidence derived alleles we first subset the data by only considering sites where the human-chimpanzee ancestral state is inferred with high confidence in the six primate EPO (Enredo, Pecan, Ortheus) alignments (Durbin *et al.*, 2010). In a sample as large as ExAC some sites will have both experienced a substitution along the human lineage and be polymorphic in the sample. At such sites the identity of the last substitution (the ancestral state) is not obvious. In such cases, if the human-chimpanzee ancestral allele is present at a site, then we call all other alleles derived. Otherwise, we assign the highest frequency allele as ancestral and call all other alleles are derived. Using a more sophisticated procedure, similar to that of Hernandez *et al.* (2007), to correct the site frequency spectrum for misidentification of derived alleles does not substantially affect results (results not shown), and so for efficiency we use the simpler procedure.

As a measure of selection against derived alleles, we used rejected substitution scores obtained through Genomic Evolutionary Rate Profiling (GERP) (Cooper *et al.*, 2005; Davy-

DOV *et al.*, 2010), which we hereafter refer to as GERP scores. High GERP scores indicate greater levels of phylogenetic constraint. When analyzing data by GERP score, we divide the observed range into 20 equally spaced bins along the GERP axis and following HENN *et al.* (2016) put all sites with a score < -1.8 into a separate bin because these are a mix of highly constrained and poorly aligned sites. Standard deviations were calculated by bootstrapping across sites within a GERP bin.

2.4 Results

We analyze the dynamics of deleterious variation by first exploring evolution within populations and then move on to differences between populations and compare results from the OOA and African models to data. To begin our analysis we use two basic example demographies, taking parameters from a model of OOA demographic history TENNESSEN *et al.* (2012).

2.4.1 The response of heterozygosity to simple demographies

Bottleneck

At equilibrium under the infinite sites model a smaller population will always have lower expected heterozygosity regardless of the strength of selection. When the population size drops from N_0 to $N_0\rho$ and equilibrium is reached, heterozygosity at neutral sites is reduced by the fraction ρ. At sites under purifying selection heterozygosity is also reduced, but the fractional reduction is less than ρ. Sites under very strong selection ($2Ns \gg 1$) experience almost no heterozygosity decrease because these are maintained by a mutation-selection balance that is insensitive to population size (SIMONS *et al.*, 2014).

After a bottleneck, the approach to the new equilibrium heterozygosity is not always monotonic through time, as shown here by by the numerical solutions for a population

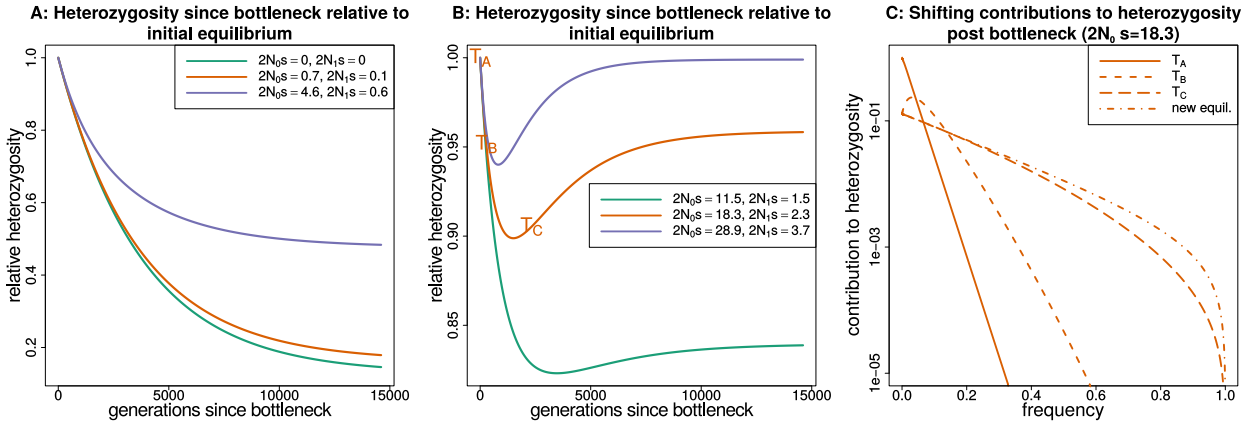


Figure 2.2: **The response of heterozygosity at sites under purifying selection to a prolonged bottleneck.** In each panel, N_0 corresponds to the the population size at the start of the trajectory (before the bottleneck) and N_1 to the population size after the bottleneck. Panel A shows the response of heterozygosity to a population bottleneck for selection coefficients that are neutral or nearly neutral throughout. Panel B shows heterozygosity trajectories for selection coefficients that are strongly deleterious before the bottleneck and nearly neutral afterwards. Panel C shows how the total heterozygosity is distributed across different frequency alleles by plotting the contribution to heterozygosity ($x(1-x)f(x)$) at different time points in (T_A, T_B, T_C) for the orange trajectory from panel B. Integrating over the contributions to heterozygosity from zero to one gives the total expected heterozygosity. The frequency spectrum shifts from being strongly deleterious before the bottleneck to nearly neutral afterwards. The initial loss of variation at low frequency (times T_B, T_C) is not immediately compensated for by increased variation at higher frequency as would occur at equilibrium. The difference between the time it takes to lose rare variants and the time it takes to accumulate variants at higher frequencies explains the heterozygosity dip and recovery in panel B.

starting at equilibrium and going through a prolonged bottleneck (Figure 2.2). If sites are neutral or have a $2Ns$ value with magnitude < 1 following the bottleneck, then the approach to the new equilibrium value is monotonic. When the magnitude of $2Ns$ is > 1, then the bottleneck causes heterozygosity to undershoot its new equilibrium value (Figure 2.2B). Breaking down the expected heterozygosity into its contributions from alleles at different frequencies shows that the undershooting is due to a faster heterozygosity loss from fixation of low-frequency variants than heterozygosity increase from variants drifting to intermediate frequencies (Figure 2.2C). Heterozygosity later increases as many variants drift to higher frequencies and compensate for the initial loss.

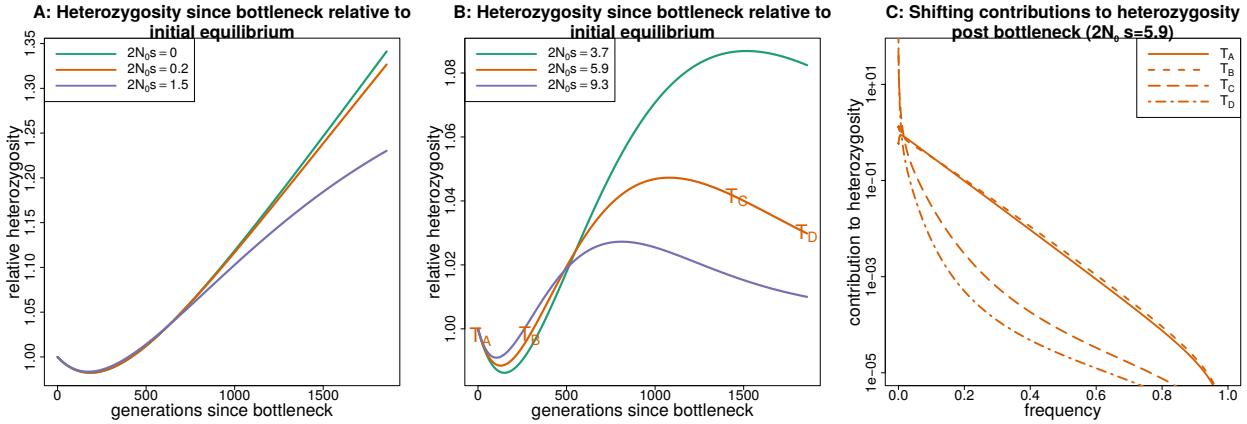


Figure 2.3: **The response of heterozygosity at sites under purifying selection to a bottleneck followed by exponential growth.** Panels A and B show the response of heterozygosity to a bottleneck of about 50% followed by exponential growth. N_0 corresponds to the population size at the start of the trajectory (before the bottleneck). Heterozygosity initially drops due to the bottleneck, regardless of selection coefficient, but begins to increase as the population size grows. When the population size becomes large relative to the selection coefficient, heterozygosity overshoots the equilibrium value that it would approach for mutation-selection balance. Panel C shows how the total heterozygosity is distributed across different frequency alleles by plotting the contribution to heterozygosity ($x(1-x)f(x)$) at different points in time (T_A, T_B, T_C, T_D) for the orange trajectory in panel B. This demonstrates how the contribution to heterozygosity shifts towards lower frequency alleles as the population size grows.

Bottleneck+growth

Figure 2.3 shows how heterozygosity approaches mutation-selection balance in a population starting from equilibrium, going through a bottleneck, and then growing exponentially. The initial heterozygosity drop is similar to that in the lone bottleneck model, but the recovery is rapid as the population grows exponentially. Similar to how heterozygosity in the bottleneck model can undershoot its equilibrium value, in the bottleneck+growth model it overshoots the asymptotic value at mutation-selection balance. Both cases demonstrate how easily equilibrium intuition can fail. For a period following a bottleneck heterozygosity may be increasing, and conversely heterozygosity can be decreasing during a period of population growth.

2.4.2 The response of $\mathbf{P_N/P_T}$ *to complex demography*

The complex behavior of heterozygosity in simple demographic scenarios suggests the difficulty of comparing deleterious genetic variation between populations. When we consider the OOA trajectory fitted by Tennessen *et al.* (2012) (Figure 2.1), the response of heterozygosity to the bottleneck and bottleneck+growth events is similar to when these events are considered in isolation (Figure A1). The response becomes complex when we consider the evolution of a more elaborate summary: the proportion of sample variants that are predicted to be deleterious.

The proportion of sample variants that are predicted to be deleterious has been used as a statistic to look for differences in the distribution of deleterious genetic variation between human populations (reviewed by Lohmueller (2014a)). We write this proportion as P_N/P_T, where P_N is the number of deleterious polymorphic sites, and $P_T = P_N + P_S$ is the total number of polymorphic sites (see Appendix A.6 for details of calculating this quantity). P_N/P_T depends on sample size because it counts all variants equally regardless of frequency, and larger sample sizes will contain a greater proportion of rare variants. In empirical human studies, a higher P_N/P_T is found in OOA populations compared to African ones when the sample size is relatively small (Peischl *et al.* (2013), $n = 17$ African and $n = 25$ non-African ancestry individuals). However, a study with a much larger sample size of found no difference in P_N/P_T between African and European ancestry samples ((Tennessen *et al.*, 2012), $n = 1088$ African and $n = 1351$ non-African ancestry individuals).

Figure 2.4 shows how demography, selection, and sample size interact to determine the evolution of P_N/P_T. The OOA bottleneck 2000 generations ago initially causes a drop in P_N/P_T (Figure 2.4) because deleterious alleles are more likely than neutral ones to be at low frequencies and therefore lost during the bottleneck. Interestingly, if s is large and the sample size is small enough (e.g. $s = 0.01$, $n = 200$), then OOA P_N/P_T can increase above African levels within the duration of the bottleneck.

For OOA populations, P_N/P_T increases during the growth period following the second bottleneck, but whether this increase is sufficient to surpass P_N/P_T for an African sample depends on s and the sample size. In both phases, P_N/P_T often decreases when the population size has decreased, and increases when the population size has increased. These are both transient, nonequilibrium behaviors. We observe that the OOA P_N/P_T becomes greater than the African P_N/P_T during the growth period after the OOA bottleneck. This behavior was originally noted by Lohmueller *et al.* (2008) and was advanced by Do *et al.* (2015) as the main cause of the higher P_N/P_T in OOA populations. This observation argues against the interpretation that a greater P_N/P_T reflects deleterious variants drifting to higher frequencies during the OOA bottleneck.

The magnitude of P_N/P_T and the expected difference between an African and OOA sample vary dramatically with sample size (Figure 2.5). P_N/P_T increases with sample size as more low frequency deleterious variants are discovered. Recent exponential growth in both population trajectories produces a large number of rare variants, and as the sample size becomes large these eventually overwhelm the signal from common ones because the majority of variants at the population level are rare. Since rare variants are only slightly affected by selection, P_N/P_T in a very large sample will eventually resemble the ratio of the input mutation rate between synonymous and nonsynonymous changes. In concordance with this prediction, both a large number of rare variants and a smaller difference in P_N/P_T between African and OOA samples have been observed in sequencing studies with large sample sizes (Tennessen *et al.*, 2012; Nelson *et al.*, 2012). As noted above, we do not predict a greater P_N/P_T in the OOA versus African model for all combinations of sample size and s, and Figure 2.5 shows combinations of s and sample size where we expect the opposite (where dots exceed crosses, e.g. sample size 40, $s \approx 1 \times 10^{-5}$).

It has been previously appreciated that P_N/P_T differences do not correspond to changes in the mean deleterious allele frequency or differences in genetic load between populations

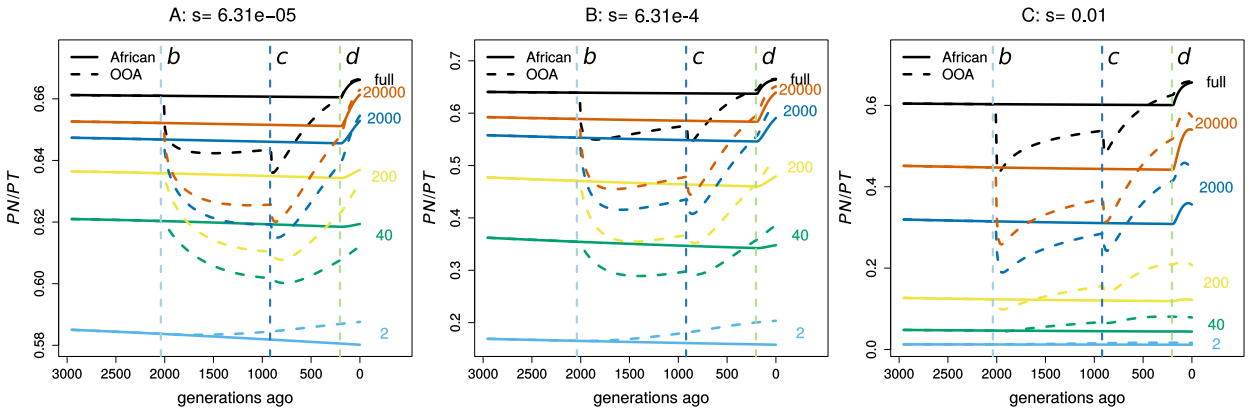


Figure 2.4: **Temporal trajectories of P_N/P_T.** We plot P_N/P_T trajectories over time in order to show how patterns in Figure 2.5 are created for weak ($s = 6.31 \times 10^{-5}$) (A), intermediate ($s = 6.31 \times 10^{-4}$) (B), and strong ($s = 0.01$) selection (C). Samples of size two (heterozygosity proportions) result in increased P_N/P_T following the OOA bottleneck, while larger samples show a decrease first and may or may not begin to rise before the bottleneck ends. The OOA model P_N/P_T outpaces the African one for most sample sizes during the population growth following event *c*.

(Lohmueller, 2014b; Do *et al.*, 2015). P_N/P_T instead reflects the site frequency spectra of putatively deleterious alleles in a complex manner. Specifically, what P_N/P_T differences reflects about the evolution of deleterious variation depends on the sample size and strength of selection (a sensitivity also recently noted by Simons and Sella (2016)). In our example, a greater P_N/P_T may primarily reflect deleterious variants drifting to higher frequency during a period of small population size, or it may reflect a proportionally faster recovery of deleterious variation during the growth phase.

2.4.3 The contribution of selection to population differences

The extent to which differences in site frequency spectra between populations can be attributed to selection is an open problem (Brandvain and Wright, 2016). We investigate two examples of this problem as instances where we are interested in the distribution of deleterious variation but are not focused on the genetic load or mean fitness.

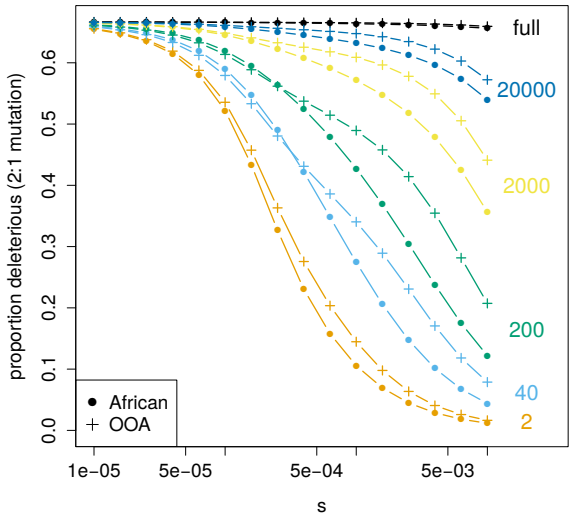


Figure 2.5: **Expected current differences in P_N/P_T at different sample sizes.** The expected P_N/P_T decreases as s gets larger, but how this happens is dependent on the number of sampled chromosomes (shown on the right). Notice the range of s and sample size for which the expected P_N/P_T is actually greater in the African population model than in the OOA one (dots are slightly higher than crosses). The ratio of deleterious to neutral mutations is assumed to be two to one.

Simple summaries

Do *et al.* (2015) argue that differences between synonymous and nonsynonymous frequency spectra in African and OOA populations can be largely explained without needing to invoke selection following their split. We investigate whether this is true for simple summaries of the site frequency spectrum. To do so we revisit how heterozygosity, derived allele homozygosity, and the derived allele burden evolve following the OOA split (Figure 2.6).

Separating the effects of mutation, selection, and drift is not straightforward. The diffusion process specified by equation (2.1) describes the instantaneous change in the frequency spectrum due to selection and drift. However, over any appreciable length of time the effects of these evolutionary forces are not separable because each distorts the shape of the frequency spectrum and affects the operation of the other. Despite this, there are two ways one can investigate the importance of drift versus selection. One is to calculate the selection and drift terms of equation (2.1) each generation and compare these between populations. Another is to turn selection off in one or both populations and see if this affects outcomes.

This is closer to the separation of drift and selection which Do *et al.* (2015) refer to. If similar patterns of heterozygosity, derived allele homozygosity, and derived allele burden are seen in the absence of selection, then we might conclude that any observed differences are primarily a product of drift and mutation.

We calculate numerical solutions turning off selection or both selection and mutation following the OOA split. This means that the initial frequency spectrum is set to the equilibrium distribution under selection and evolves under selection up until event b in Figure 2.1. Without selection there is still a greater expected heterozygosity in the African model relative to the OOA model. When new mutations are included, this provides a good approximation to the differences in heterozygosity and homozygosity below about $s = 5 \times 10^{-4}$ (Figure 2.6). Above this, the heterozygosity difference is about five percent greater than that in the model with selection, and the magnitude of this deviation increases rapidly with s. The derived allele burden difference in models with no selection is zero because selection is necessary for differences to accumulate. For the heterozygosity difference at nonsynonymous sites, results suggest we can ignore selection for alleles with $2Ns < 1$, where N is the size of the bottlenecked population. Overall, these results show it is difficult to conclusively say whether simple differences between the OOA and African models are due to selection because it depends on what level of selection one is interested in an what magnitude of difference one considers important.

$\mathbf{P_N/P_T}$

The question of whether differences between the OOA and African model are due to selection can also be asked of P_N/P_T. Do *et al.* (2015) claimed using simulations that the higher P_N/P_T in European versus in West African samples reflects neutral processes. To make this claim they use the first approach mentioned above: they calculate the changes due to selection and neutral forces separately each generation and compare them between

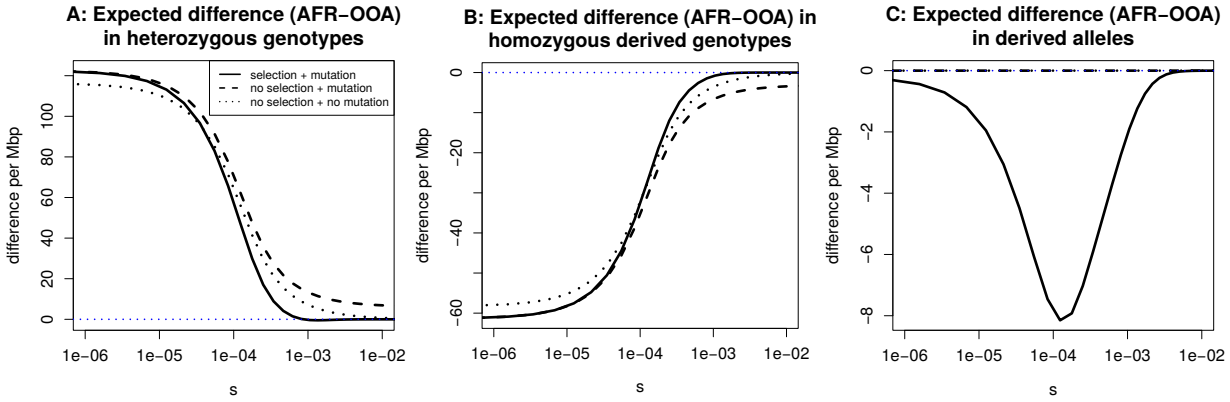


Figure 2.6: **Stratification of expected differences by selection coefficient.** We show, for a range of selection coefficients, the expected nonsynonymous difference per Mbp between the OOA and African model in (A) heterozygous genotypes, (B) homozygous genotypes, and (C) derived alleles. We obtain a number in terms of nonsynonymous differences by setting the mutation rate to the approximate amount of human coding sequence times a mutation rate of 1.2×10^{-8} then multiply by two thirds to approximate the number of new mutation that induce nonsynonymous changes. The vertical axis gives the expected difference per Mbp per diploid genome. For derived allele count and derived allele homozygosity this includes fixations since the start of the population histories shown in Figure 2.1. *No selection + mutation* refers to numerical solutions setting $s = 0$ following the OOA bottleneck in the European trajectory. *No selection + no mutation* refers to the same, but turning off new mutations as well. The difference in derived allele count is small, meaning the heterozygosity and homozygosity differences must be nearly the same, though with opposite signs, as can be seen.

populations. We investigate these rates in an equilibrium population to see if they agree with the intuition that selection is more effective in larger populations.

In an equilibrium population, P_N/P_T decreases both with increasing s and increasing population size, and this result does not depend on sample size (Figure 2.7). This is because a greater value of $S = 2Ns$ corresponds to a greater ability to remove deleterious alleles relative to their accumulation through mutation and drift. Lower P_N/P_T at greater S can therefore be taken to reflect a greater efficacy of selection in removing deleterious alleles.

At equilibrium, mutation, selection, and drift cancel out and cause the expected value of P_N/P_T to remain constant. We analyze this process by calculating the instantaneous rate due to selection. The equilibrium rate of change in P_N/P_T per $2N$ generations due to selection is

$$\frac{d}{dt}\left(\frac{P_N}{P_T}\right)^k_\gamma = -\theta_N \int_0^1 \left(1 - x^k - (1-x)^k\right) \frac{2S^2 e^{-2Sx}}{1 - e^{-2S}} dx \left(\frac{P_S^k}{\left(P_N^k + P_S^k\right)^2}\right), \tag{2.3}$$

or

$$\frac{d}{dt}\left(\frac{P_N}{P_T}\right)_\gamma = -S\theta_N \left(\frac{P_S}{(P_N + P_S)^2}\right). \tag{2.4}$$

Here, θ_N is population scaled mutation rate to deleterious alleles, k is the sample size, and the subscript γ denotes that this is the portion of the rate of change that is due to selection (see Appendix A.7 for details). We calculate these rates for different sample sizes in a population resembling that before the OOA bottleneck ($2N = 29,240$) and in one resembling the bottlenecked size ($2N = 3,801$) (Figure 2.1). Surprisingly, even though P_N/P_T is greater in smaller populations, the *per generation* rate at which selection decreases this value can actually be greater in smaller populations (Figure 2.7B). This is true for a large range of s at small samples sizes and persists at strongly constrained sites even when we consider the full population (Figure 2.7C).

In their simulations Do *et al.* (2015) observe that the rate at which selection decreases

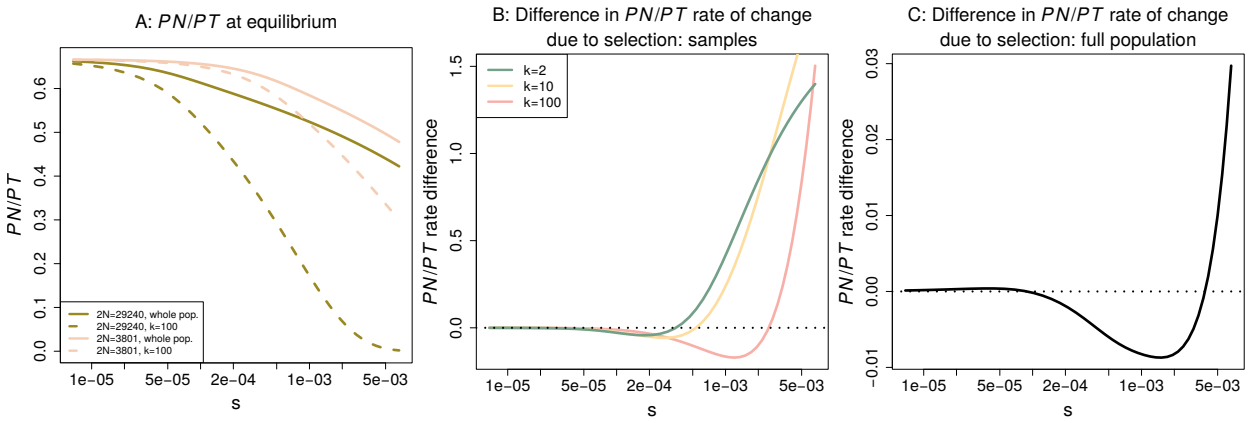


Figure 2.7: **The equilibrium behavior of P_N/P_T.** The equilibrium behavior of P_N/P_T is contrasted between two population sizes ($2N = 29240$, $2N = 3801$) roughly corresponding to the effective population size pre and post OOA bottleneck. The sample size is denoted by k. Panel A shows how P_N/P_T decreases with s and it greater at a lower equilibrium effective size. Panels B and C analyze the difference in the per generation rate at which selection acts to decrease P_N/P_T. Panel B does so for samples from the population, while panel C shows the relationship for the population as a whole. The vertical axis gives the per generation rate of P_N/P_T change in the small population minus that in the larger one. Positive values indicated that the per generation rate at which selection acts to reduce P_N/P_T is greater in the smaller population.

P_N/P_T on a per generation basis is stronger in an OOA population trajectory than an African one. They use this to conclude that primarily non-selective forces have driven the dynamics of this statistic. Given our equilibrium results, we note that a greater per generation change in P_N/P_T due to selection does not necessarily imply evidence for a greater efficacy of selection or the primacy of drift versus selection. The sign of this rate difference between two populations depends on the strength of selection and the sample size.

2.4.4 Empirical relationship between strength of selection and the distribution of deleterious variation

The strength of selection, represented by s, greatly impacts how both heterozygosity and P_N/P_T respond to the demographic events that lead to differences between populations (Figures 2.2, 2.3, and 2.4). We investigate how well GERP scores, a putative measure of the strength of selection based on phylogenetic conservation (Cooper *et al.*, 2005; Davydov

et al., 2010), predict heterozygosity. In order to do this we compare the expected differences, stratified by s, in heterozygosity, derived allele homozygosity, and derived allele burden between the African and OOA trajectories with empirical results from the ExAC data for the AFR and NFE ancestries. We switch to using more simple summaries than P_N/P_T because these are not sensitive to sample size.

We first examine how the expected differences between Africa and OOA in heterozygosity, derived allele homozygosity, and the derived allele burden evolve over time and depend on s. Doing so shows that the present increased homozygosity and decreased heterozygosity in OOA versus Africa originates during the OOA bottleneck. This effect persists at present, but the recovery of the OOA population size beginning around 1000 generations ago decreases the magnitude of this difference and relatively more so for more strongly selected alleles (Figure 2.8 B,C). For derived alleles we predict a slight excess for all s in OOA, but this difference decreases during the recovery when s is large (at least ≥ 0.001). This emphasizes the need to consider demography following the OOA bottleneck when studying selected variation in human populations (Gravel, 2016). The variation between different s values is not as great here as for P_N/P_T, but it is clear that the differences evolve on a faster time scale for sites under greater selection.

The present expected heterozygosity difference between the OOA and African models decreases with s, while the expected derived allele homozygosity difference mirrors this and increases with s (Figure 2.6). The expected difference in derived allele burden is small and peaks at an intermediate value around $s = 10^{-4}$. This is consistent with results from Simons *et al.* (2014) showing only a very small expected increase in genetic load in an OOA model.

In the ExAC, heterozygosity and homozygosity show similar trends (Figure 2.9) as the theoretical prediction (Figure 2.6) with heterozygosity higher in AFR and derived allele homozygosity higher in NFE. Differences between AFR and NFE also decrease with increasing GERP score, similar to how the expected differences decrease with increasing s. However,

there is no clear relationship between GERP score and the mean difference in derived allele burden between AFR and NFE individuals. This is perhaps not surprising because the expected burden difference is so small for all s.

Another approach to look for a relationship between derived alleles and GERP scores is to calculate a GERP score burden, which weighs the frequency of each derived allele by its GERP score ($\sum GERP_i f_i$) (Marsden *et al.*, 2016). We calculate a cumulative GERP score burden and use bootstrapping across sites to assess confidence. While the NFE sample does appear to have an excess GERP burden for mildly deleterious alleles in the GERP score range of 2 to 4, this trend is not apparent in most bootstrap replicates, and we do not see an significant overall difference between the AFR and NFE samples (Figure A9).

The lack of difference between AFR and NFE samples in derived allele frequency or GERP score burden could be attributable to a sensitivity to quality filters (Figure A6). This might then also explain the lack of a relationship between the derived allele frequency and GERP score. Additionally, there is only a weak trend in relative heterozygosity with GERP score (Figure A8). This suggests that GERP scores better reflect the probability a variant is strongly selected rather the selection coefficients of weakly deleterious variants and that the majority of heterozygous sites within each GERP bin are neutral. This make GERP scores less useful when we are interested in stratifying genetic variants by whether they are strongly or weakly selected.

2.5 Discussion

In this article we have demonstrated a number of ways in which equilibrium population genetic logic can mislead when applied to populations with nonequilibrium histories. These four points are: (1) Heterozygosity can be increasing for some period following a population decline and can be decreasing while the population is growing. (2) In the OOA and African population models, differences in P_N/P_T depend strongly on the strength or selection and

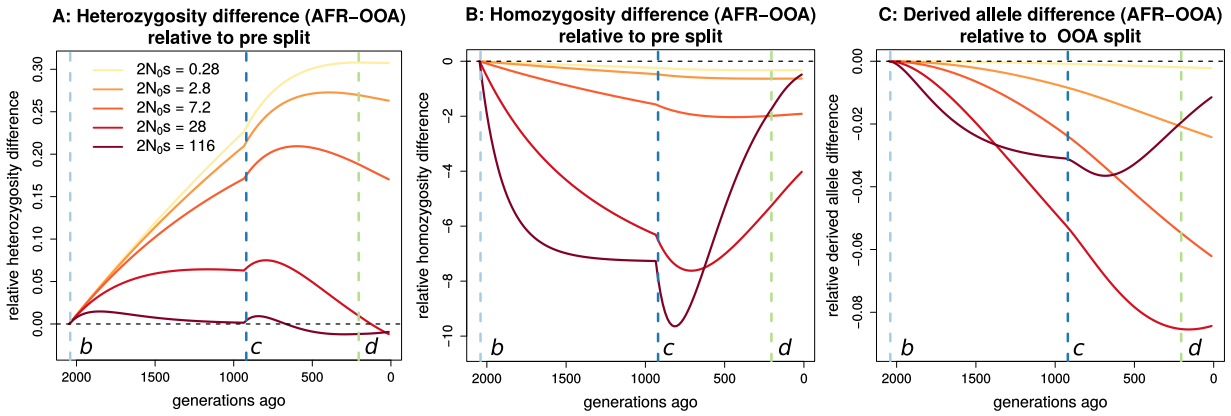


Figure 2.8: **Stratification of expected differences by time.** We show for a range of $2N_0s$, where N_0 is the population size preceding event b, how the expected difference in (A) heterozygous genotypes, (B) homozygous derived genotypes, and (C) derived alleles changes over time, relative to their levels in the ancestral population that existed before event b. Vertical lines indicate demographic events shown in Figure 2.1. Substantial changes in heterozygosity and homozygosity differences occur following event c, emphasizing the importance of the recovery from the OOA bottleneck.

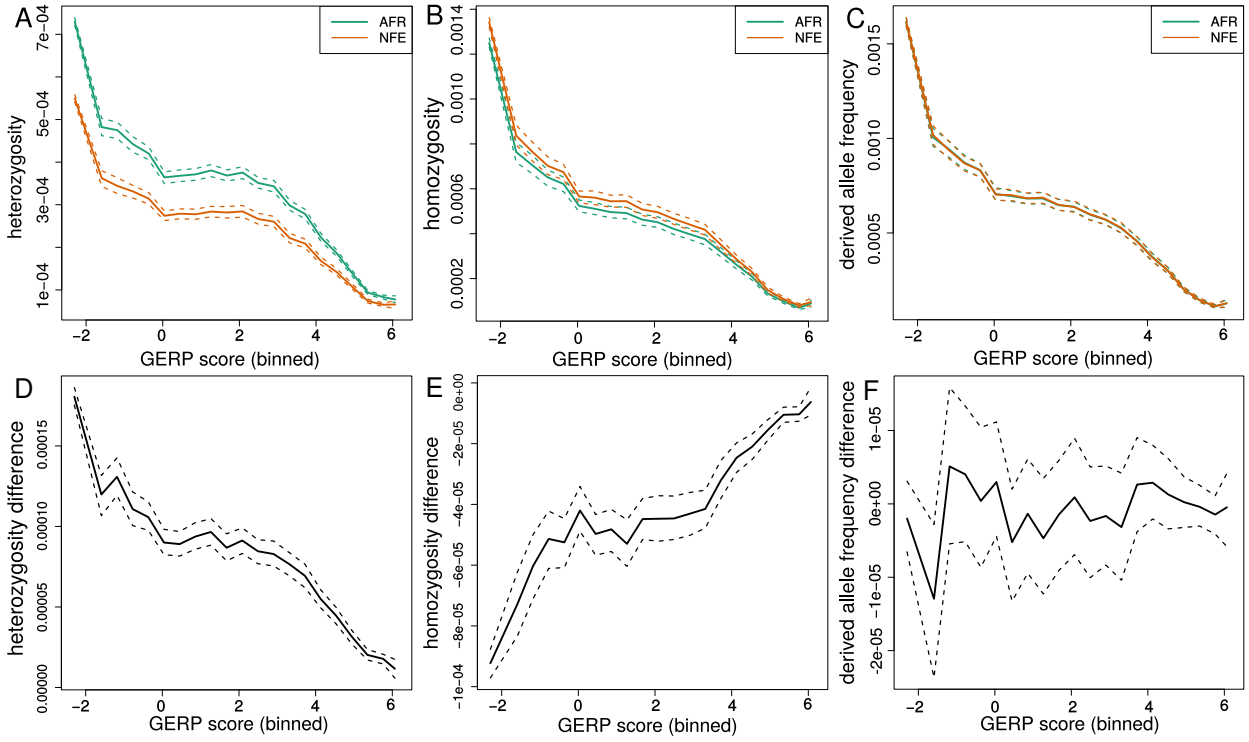


Figure 2.9: **Observed differences in ExAC by GERP score.** The top row (A-C) shows heterozygosity, homozygosity, and derived allele frequency for the African and non-Finnish European population groups in ExAC plotted against binned GERP scores. The bottom row (D-F) shows the differences between them (AFR-NFE). Dotted lines provide 95% confidence intervals obtained by bootstrapping across sites within each bin.

sample size. In particular, which demographic event has the greatest effect on the P_N/P_T difference is greatly influenced by these parameters. (3) The interpretation of differences in P_N/P_T, heterozygosity, and derived allele homozygosity depends on the strength of selection, but we find that GERP scores are imprecise predictions of selection coefficients at particular sites. Additionally, we find that (4) it is difficult to decide whether differences in deleterious variation between nonequilibrium populations are due to drift versus selection.

A number of recent theoretical investigations have supplied useful intuition into the effects of bottlenecks and population growth on deleterious variation (Simons *et al.*, 2014; Balick *et al.*, 2015; Peischl *et al.*, 2013; Gazave *et al.*, 2013; Lohmueller, 2014b; Gravel, 2016). These have been spurred by particular interest in the effects of human demographies (Lohmueller, 2014a) and have used forwards-in-time models as these allow selection to be easily incorporated.

Similar in spirit to our point (1), Balick *et al.* (2015) develop analytical approximations to the change in the mean derived allele frequency following a short population bottleneck with the purpose of contrasting additive and recessive modes of selection. They find non-monotonic behavior when selection is recessive, wherein deleterious variants are purged after recovery from the bottleneck before mutation builds them up again. The non-monotonic behavior we see in heterozygosity (Figure 2.2, 2.3) is less severe than this because it is most pronounced at strongly constrained sites that have low expected heterozygosity to begin with. However, it is interesting because it results from a simpler, additive model and makes the counter-intuitive prediction that heterozygosity should sometimes be decreasing in a growing population.

Regarding point (2), P_N/P_T differences in nonequilibrium populations do not have the same interpretation as in equilibrium ones. This was noted previously in simulation studies that showed the bottleneck effect of initially decreasing P_N/P_T before causing it to increase due to deleterious variants drifting to higher frequencies, and the fact that recovery from

a bottleneck can also increase P_N/P_T (Lohmueller *et al.*, 2008; Lohmueller, 2014b; Do *et al.*, 2015; Simons and Sella, 2016). Our analyses have looked in greater detail at how demography and selection have interacted to produce these patterns and emphasize the strong dependence of P_N/P_T evolution on the sample size and strength of selection. We emphasize that whether P_N/P_T is increasing or decreasing at any point in a demographic history is highly dependent on both the sample size and strength of selection (Figure 2.7).

While human studies often use P_N/P_T (Lohmueller, 2014a), investigations in other species use the ratio of nonsynonymous to synonymous heterozygosity instead. Many have found a negative relationship between this ratio and synonymous heterozygosity (Elyashiv *et al.*, 2010; Cao *et al.*, 2011; Marsden *et al.*, 2016; Li *et al.*, 2014). The heterozygosity ratio has the obvious advantage of not being dependent on the sample size, and a sample of size two in P_N/P_T appears to chiefly respond to deleterious variants drifting to higher frequencies during the OOA bottleneck as opposed to subsequent growth events (Figures 2.4, 2.5). In the recovery from a bottleneck, deleterious alleles reach their equilibrium heterozygosity before neutral ones. This may cause P_N/P_T to be misleading because its value can increase transiently while the effectiveness of selection to remove deleterious alleles is actually increasing.

Regarding point (3), we did not find a strong indication that GERP scores could reflect intermediate selection coefficients, but other studies have noticed interesting patterns between measures of selection and the distribution of deleterious variation. Racimo and Schraiber (2014) fit selection coefficients to sites binned by a different measure of selection, CADD scores (Kircher *et al.*, 2014). They found that the fitted selection coefficient did decrease monotonically with severity of CADD scores. In another study, Henn *et al.* (2016) observe a greater number of putatively deleterious derived alleles (classified using GERP scores) in East Asian and American genomes and explain this in terms of a serial founder model that can produce a higher genetic load with more founder events (Peischl *et al.*, 2013). The

difference in putatively deleterious alleles came from sites with intermediate GERP scores similar to the theoretical expectation in Figure 2.6. This is in contrast to findings of no significant increase in derived allele burden in any contemporary human population compared to an African one (Do *et al.*, 2015; Simons and Sella, 2016). Future research should seek to resolve these conflicting observations because they are based on samples from different populations, used different predictors of deleteriousness, and different forms of statistical tests. More extensive sampling of American, Asian, and other populations geographically far from Africa will also add clarity.

In point (4) we note that the per generation rate at which selection changes P_N/P_T is not an indicator of the efficacy of selection, as it can be greater in a smaller than in a larger population for the same s. Because it has such complex behavior in an equilibrium model, we argue that the per generation rate at which selection changes P_N/P_T should not be used to say whether changes in P_N/P_T have been caused by neutral of selective forces (Do *et al.*, 2015).

The model we have analyzed here is simple in many regards. For instance, the Tennessen *et al.* (2012) demographic model includes some migration between the African and European populations after the OOA split, and we did not include this. Other studies analyzing the response of deleterious variation to this particular demographic scenario come to similar conclusions when including this migration (Fu *et al.*, 2014; Simons *et al.*, 2014; Gravel, 2016). A larger second issue is that we have only considered alleles acting additively within genotypes. It is well known that a large amount of strongly deleterious variation acts recessively, and that there is likely to be a negative relationship between the degree of dominance and how deleterious a mutation is (Agrawal and Whitlock, 2011). However, the dominance effects of nearly neutral mutations are still mostly unknown because these are much less amenable to analysis. Studies that have inferred a distribution of fitness (DFE) effects of new mutations from the site frequency spectrum have almost exclusively used ad-

ditive models (Eyre-Walker *et al.*, 2006; Boyko *et al.*, 2008). Balick *et al.* (2015) find evidence of recessive selection in humans at some sets of genes known to act recessively in causing disease. Specifically, they found a higher derived allele burden in recessive disease genes in a European sample compared to an African-American one. That we do not observe a difference in the derived allele burden overall may thus be partially explained by this effect canceling out the effects of additive mutations. Henn *et al.* (2016) fit a model of serial founder effects under the assumption of complete recessivity and found that such a model is consistent with the observed heterozygosity cline away from Africa in different GERP scores bins.

An additional simplification in the model we used is that sites are modeled to be independent. Linked variants under selection will interfere with each other, reducing the effectiveness of selection and levels of polymorphism (Comeron and Kreitman, 2002). Although there is substantial evidence for background selection influencing patterns of variation in humans (McVicker *et al.*, 2009; Hernandez *et al.*, 2011; Lohmueller *et al.*, 2011), if interference between deleterious alleles is rare it should not substantially affect our results. If there is substantial interference between deleterious alleles, it is not clear how such interference would affect the response of deleterious variation to demographic events like bottlenecks and growth periods, as even equilibrium models of interference selection can be quite complex (Good *et al.*, 2014). In species with larger population sizes than humans it is likely that linked selection and interference cannot be ignored so easily.

Studying fitness differences and the predicted accumulation of deleterious mutations in smaller populations remains a challenge. The derived allele burden, at least for semidominant alleles, is likely to be a robust statistic for identifying populations accumulating deleterious variation, and it has been used to identify a number of cases where this seems to have occurred (Henn *et al.*, 2016; Xue *et al.*, 2015; Marsden *et al.*, 2016; Do *et al.*, 2015; Schubert *et al.*, 2014). The precise interpretation of these results is much more difficult because converting

them to genetic load or fitness differences requires knowing something about the underlying fitness effects of mutations, and differences in dominance can yield opposite results (Henn *et al.*, 2016; Balick *et al.*, 2015). Differences in the adaptive substitution rate might also confound inference from the derived allele burden because new adaptive alleles will also be counted as derived (Brandvain and Wright, 2016). A more rigorous approach would be to jointly infer the DFE, demography, dominance, and mutational load. Future work should explore the possibility of doing so, and in the meantime any inference based on summary statistics is best supported by extensive simulations (e.g. Marsden *et al.*, 2016).

One final factor not considered here was the effects of deleterious alleles introgressing into human populations. Sequencing of ancient DNA has strongly suggested that two archaic humans, Neanderthals and Denisovans, accumulated significantly more deleterious mutation than contemporary humans prior to their extinction (Castellano *et al.*, 2014; Do *et al.*, 2015). Harris and Nielsen (2016) estimate that the average Neanderthal would have had at least 40% lower fitness than the average human at the time of admixture. This admixture would then have introduced a large number of deleterious alleles into human populations, resulting in a contemporary load of deleterious alleles that arose in Neanderthals (Harris and Nielsen, 2016; Juric *et al.*, 2016). It will be interesting to see whether this contributes to the excess derived allele burden in East Asian and American populations (Henn *et al.*, 2016), given that East Asian populations contain a greater fraction of Neanderthal DNA (Wall *et al.*, 2013; Sankararaman *et al.*, 2014) likely due to a greater gene flow from Neanderthal populations (Kim and Lohmueller, 2015).

CHAPTER 3

THE EFFECTS OF DEMOGRAPHY AND GENETIC ARCHITECTURE ON THE NEUTRAL DISTRIBUTION OF QUANTITATIVE TRAITS

3.1 Abstract

Neutral models for quantitative trait evolution are useful for identifying phenotypes under selection in natural populations. Models of quantitative traits often assume phenotypes are normally distributed. This assumption may be violated when traits are affected by relatively few genetic variants or when those variants have skewed or heavy-tailed distributions of effects on the trait. Traits such as gene expression levels and other molecular phenotypes may fall into this category. To accommodate deviations from normality, models making minimal assumptions about genetic architecture and patterns of genetic variation are needed. Here, we develop a general neutral model for quantitative trait variation using a coalescent approach by extending the framework developed by Schraiber and Akey (2015). This model allows interpretation of trait distributions in terms of familiar population genetic parameters because it is based on the coalescent. We show how the normal distribution resulting from the infinitesimal limit, where the number of loci grows large as the effect size per mutation becomes small, depends only on expected pairwise coalescent times. We then demonstrate how deviations from normality depend on demography through the distribution of coalescence times as well as through the genetic architecture. In particular, population growth events exacerbate deviations while bottlenecks reduce them. This model also has practical applications which we demonstrate by designing an approach to simulate from the null distribution of Q_{ST}, the ratio of the trait variance between subpopulations to that in the overall population. We further show that it is likely impossible to distinguish sparsity from skewed or heavy-tailed distributions of mutational effects using only trait values sampled

from a population. The model analyzed here greatly expands the parameter space for which neutral trait models can be designed.

3.2 Introduction

Neutral models of quantitative traits provide a null distribution against which various goodness-of-fit tests can be used to test for the action of natural selection (Lande, 1976). Neutral models can also clarify the effects of purely neutral forces such as population size and mutation rate on trait distributions (Lynch and Hill, 1986). A common approach is to first model phenotypes as normally distributed, either among offspring within a family, among members of a population, or between species (Turelli, 2017). Indeed, it has been suggested that the normality assumption is the defining characteristic of quantitative genetics (Rice, 2004). This might be justified if phenotypes are influenced by a large number of sufficiently independent Mendelian factors (Fisher, 1918), or normality may simply appear approximately true in practice.

Neutral models for quantitative traits have been developed in a variety of contexts. The goal of these models is ask to whether phenotypic differentiation between groups can be reasonably explained by processes other than natural selection. On macroevolutionary time scales, models stemming from Lande (1976) have used Brownian motion to model the change in the mean value of quantitative traits. These models are used in statistical methods to test for extreme trait divergence between species (Turelli *et al.*, 1988), test for phylogenetic signal in trait distributions (Freckleton *et al.*, 2002), and correct for phylogenetic dependence when calculating correlations between trait values (Felsenstein, 1985). On shorter time scales, neutral distributions assuming multivariate normality of trait values also underlie tests for spatially varying selection in structured populations such as the method developed by Ovaskainen *et al.* (2011). Other neutral models for quantitative traits have not assumed normality (Chakraborty and Nei, 1982; Lynch and Hill, 1986; Lande, 1992), and the

dynamics of phenotypic evolution are examined forwards in time as a balance between mutation creating variance, migration spreading variance among subpopulations, and fixation removing it. However, the mean and variance resulting from these models can still be used to parameterize normal distributions.

Under the normality assumption, quantitative trait dynamics can be modeled without concern for the number of causal loci influencing a trait, the genealogies at these sites, or the distribution of mutational effects. However, heritable phenotypic variation is ultimately due to discrete mutations at discrete locations in the genome, and how the phenotypic variance arising from these mutations is distributed depends on the genealogies at these loci. For instance, the distribution of genealogies in the genome might be strongly influenced by recent population growth while the distribution of mutational effects could be skewed for biological reasons. When the number of mutations affecting a trait is large the central limit theorem ensures that the distributions of genealogies and mutational effects can be ignored, but a full model of phenotypic variation would have to include them. Importantly, deviations from normality may affect the outcomes of goodness-of-fit tests that necessarily aim to identify outliers from a normal model.

A more complete model of neutral phenotypic variation can begin by modeling the genealogies at causal loci. The principle modeling framework for genealogical variation is the coalescent process (Wakeley, 2008), but few studies have connected the coalescent to quantitative genetics. Whitlock (1999) argued that under the coalescent, measures of trait (Q_{ST}) and genetic (F_{ST}) differentiation have the same expected value given general models of population subdivision. By simulating from the coalescent with recombination, Griswold *et al.* (2007) investigated the effects of shared ancestry and linkage disequilibrium on the matrix of genetic variances and covariances between traits ($\mathbf{G}$). They found that linkage disequilibrium and small numbers of causal loci can cause phenotypic covariances not predicted by the mutational covariance matrix. Although not explicitly connected to

the coalescent, Ovaskainen *et al.* (2011) developed their test for spatially varying selection by assuming that the covariance in trait values, conditional on the G-matrix in the ancestral population, depends only on the pairwise coancestry coefficients, which have a clear interpretation in terms of the coalescent process (Slatkin and Hudson, 1991).

Two studies have asked how the shape of the distribution of mutational effect sizes, beyond just the mutational variance, impacts trait distributions. Khaitovich *et al.* (2005) modeled the evolution of gene expression values on phylogenetic trees assuming a single non-recombining causal locus but allowed for an arbitrary distribution of mutational effects. Using this model they were able to detect, using deviations from normality, asymmetries in the distribution of mutational effects on gene expression in great apes. More recently, Schraiber and Akey (2015) developed a similar general model of quantitative trait evolution at the population level based on the coalescent and allowing for any number of causal loci. They derived the characteristic function for the distribution of phenotypic values in a sample and showed how such values can deviate strongly from normality when the number of loci is small or the mutational distribution has heavy tails. Schraiber and Akey (2015) note that the possibility for multimodal trait distributions could lead to incorrect inferences of divergent selection.

Schraiber and Akey (2015) derived their results for a panmictic, constant-size population. Natural populations rarely have stable population sizes and show considerable spatial structure, and it is unclear how these violations of the constant-size, panmictic model might influence deviations from normality. We take advantage of the ability of coalescent theory to handle nonequilibrium demographies and population structure to relax these modeling assumptions.

Extending coalescent models of quantitative traits to structured populations is important because the analysis of structured populations provides an opportunity to infer the incidence of local adaptation or stabilizing selection. The Q_{ST}/F_{ST} paradigm was developed to this

end (Spitze, 1993; Whitlock, 2008; Leinonen *et al.*, 2013). Q_{ST}, defined as the ratio of the trait variance between subpopulations to the total trait variance, is compared to F_{ST}, which measures the same property for genetic variation and is calculated using neutral markers to provide a null distribution. If the observed Q_{ST} is sufficiently far from the null expectation, it is concluded that natural selection has acted. Ovaskainen *et al.* (2011) developed a modern extension of the Q_{ST}/F_{ST} paradigm for genetic values measured in breeding experiments and Berg and Coop (2014) also extended the paradigm to make use of genetic values computed from GWAS summary statistics. An advantage of the Berg and Coop (2014) approach is that by using computing genetic values from GWAS loci it makes no assumptions about normality at the population level. However, since suitably sized GWAS's have only been performed in humans, the approach has not yet been extended to other species. Understanding the neutral distribution of trait values at the sample and population level therefore is necessary for the development of goodness-of-fit tests suited to populations with complicated histories and traits with sparse genetic architectures.

We generalize the work of Schraiber and Akey (2015) by deriving the form of the moment generating function (mgf) for arbitrary distributions of coalescent times (e.g. those arising under exponential growth or an island model of migration). The key result of Schraiber and Akey (2015), the characteristic function of the sampling distribution of phenotypic values, is a special case of this general generating function. We then show how a normal model arises by taking the infinitesimal limit where the effect size per mutation becomes small as the number of loci potentially affecting the trait becomes large. We then calculate the third and fourth central moments of the trait distribution in panmictic populations to illustrate how departures from normality depend both on genetic parameters and genealogical distributions. For instance, in exchangeable populations the expected third central moment is proportional to the third moment of the mutational distribution times the expected time to the first coalescent event in a sample of size three.

Finally, we discuss the consequences of these results for Q_{ST} tests and the inference of genetic architecture. We find an improved null distribution for Q_{ST} tests that can be derived simply by using the normal distribution that arises in the infinitesimal limit of our coalescent model. Additionally, we show that it is likely not possible to infer most features of the mutational distribution using only trait values sampled from a population. Future work will be necessary to develop tests for selection that take into account both demography and genetic architecture, but the model developed here provides the groundwork for such an undertaking.

3.3 Model

In the model we investigate here, there are L unlinked causal loci at which mutations influence the trait value. An infinite number of mutations are possible within each locus and the mutation rate per unit of coalescent time to mutations affecting the trait is $\frac{\theta}{2}$. That is, $\frac{\theta}{2}$ is the mutation rate for the entire locus and not necessarily per nucleotide. Mutations are randomly assigned effects from a distribution of effect sizes, and effects are additive both within and between loci. The moment generating function (mgf) of this distribution is written as ψ and the i^{th} (non-central) moment is m_i. The sum of all mutations occurring in a haploid individual's history determines the trait value of the individual. Correlations between individuals arise when mutations fall on shared portions of genealogies at specific loci. Because the loci are unlinked we assume their genealogies are independent. This model is shown schematically in Figure 3.1.

The genealogy at a locus is represented by the random vector of branch lengths, $\mathbf{T}$. An element T_ω of $\mathbf{T}$ is the branch length subtending only individuals in the set ω and no others in the sample. For example, $T_{\{a,b\}}$ is the length of the branch subtending only individuals a and b. If a branch subtending only a and b does not exist for a given genealogy, $T_{\{a,b\}}$ is set to zero. In this way $\mathbf{T}$ encodes both the branch lengths and the topology of a genealogy. Ω

is the set of all possible branches. If there are three sampled individuals, a, b, and c, then $\Omega = \{\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}\}$ and $\mathbf{T} = (T_{\{a\}}, T_{\{b\}}, T_{\{c\}}, T_{\{a,b\}}, T_{\{a,c\}}, T_{\{b,c\}})$. The mgf for the distribution of branch lengths is denoted $\varphi_{\mathbf{T}}$.

Phenotypic trait values are the random quantities we are interested in and result from mutations occuring along the branches of genealogies. We will hereafter refer to the phenotypic trait simply as trait values and ignore any environmental component. Starting with a trait controlled by a single locus, the random vector of trait values in the sampled individuals is $\mathbf{Y}$, such that for sampled individuals a, b, and c, $\mathbf{Y} = (Y_a, Y_b, Y_c)$. The trait values are modeled as the change relative to the value in the most recent common ancestor (MRCA) of the sample at that locus. Since we do not know the ancestral value, we cannot be directly observe the change in trait values. Thus, for a trait controlled by multiple loci, $\mathbf{Y}$ is the sum over contributions from these loci, each measured with respect to an arbitrary value. However, $\mathbf{Y}$ is sufficient to determine measurable quantities such as differences in trait values between individuals and the sample variance. The moment generating functions for the distribution of trait values is denoted $\varphi_{\mathbf{Y}}$.

Here we refer to the genetic architecture of a trait as the combination of genetic parameters affecting the trait distribution (L, $\frac{\theta}{2}$, ψ). Therefore, the realized distribution for a given genetic architecture is a random quantity. Another useful way to describe a trait is by its sparsity which aims to describe the number of segregating mutations influencing the trait. A more 'sparse' trait is one affected by fewer segregating mutations. Formally, we measure sparsity as the average number of pairwise differences between two randomly chosen haplotypes at loci affecting the trait. A trait with fewer causal pairwise differences is more sparse. Sparsity thus depends both on the genetic architecture through the mutation rate, the number of causal loci, and the distribution of coalescence times.

For populations of exchangeable individuals a concise way to summarize the distribution of genealogies is $\mathbb{T}_{i,j}$ which gives the amount of time that i lineages remain in the genealogy

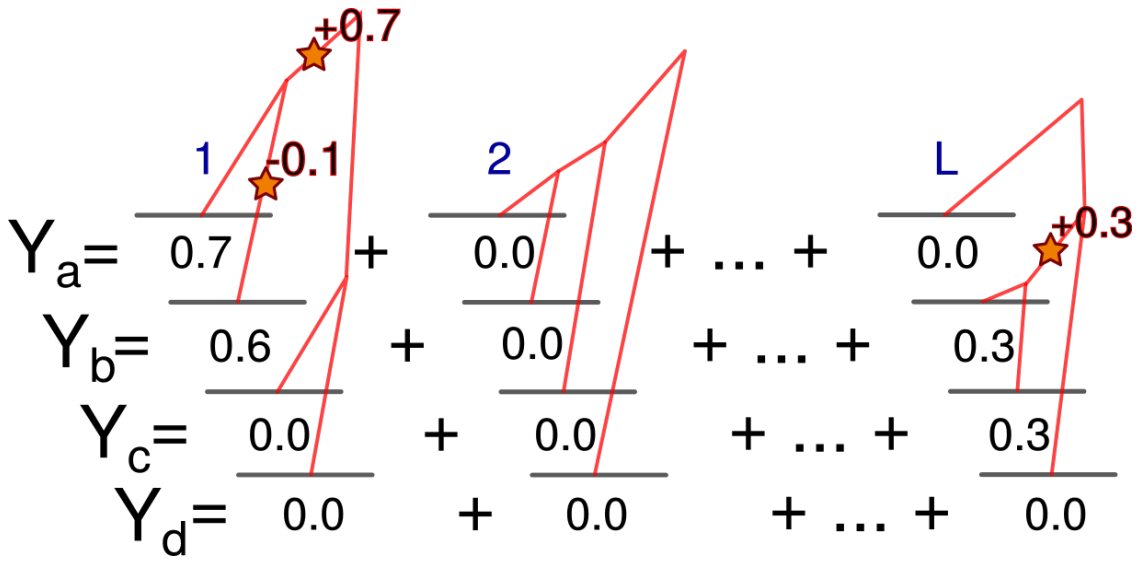


Figure 3.1: **A schematic representation of the model of how we model a trait distributions arising from genealogical and mutational distributions.** L loci affect the trait in a set of individuals and have independent genealogies. Mutations occur within loci as a Poisson process and act additively to give individual trait values. Many loci with the potential to affect the trait may receive no mutations.

of a sample of size j. The pairwise coalescent time between a lineage in individual i and in individual j is written as $\mathcal{T}_{i,j}$. When considering structured populations $\mathcal{T}_{a,b}$ is also used to denote the coalescence time between a randomly chosen lineage from subpopulation a and a randomly chosen lineage from subpopulation b. A final set of quantities are defined for sums of branch lengths. Let τ_{a+b} be the sum of all branches ancestral to both a and b, and $\tau_{a/b}$ be the sum of all branches ancestral to a but not b. Extensions of this for more than two individuals are also used. The same notation is used when referring to sets of branch indices. So Ω_{a+b} and $\Omega_{a/b}$ would be the sets of branches summed to give τ_{a+b} and $\tau_{a/b}$ respectively.

3.4 The moment generating function for the distribution of trait values

We first derive the mgf of the distribution of trait values following closely the approach of Schraiber and Akey (2015) and Khaitovich *et al.* (2005), but generalizing to arbitrary

demographies and population structure. We consider the distribution of trait values over evolutionary realizations of the combined random processes of drift and mutation. This trait distribution is complex in its general form, having a point mass at zero corresponding the possibility that no mutations affecting the trait occur, and mutational effects that could be drawn from discrete or continuous distributions. Correlations between individuals arise because of shared history in the genealogies at individual loci with discrete topologies as well as because of where on these genealogies mutations occur. An analytical expression for the probability distribution of trait values does not exist except under certain limits such as when the number of mutations affecting the trait becomes large.

However, even in the absence of a probability distribution function we can use the mgf approach to learn something about the distribution of trait values. In some cases, the mgf can be fully specified. Following the definition of the mgf for a vector-valued random variable, the mgf for a trait controlled by a single nonrecombining locus is

$$\varphi_{\mathbf{Y}}(\mathbf{k}) = \mathrm{E}\left[e^{\mathbf{k}\cdot\mathbf{y}}\right] = \int e^{\mathbf{k}\cdot\mathbf{Y}}\mathrm{P}(\mathbf{Y}=\mathbf{y})\mathrm{d}\mathbf{y}. \tag{3.1}$$

The vector $\mathbf{k}$ contains dummy variables for each individual, and the whole operation is an intergral transform of the probability distribution of trait values. Equation (3.1) can be rewritten by conditioning on the genealogy to give

$$\begin{aligned}\varphi_{\mathbf{Y}}(\mathbf{k}) &= \int_{\mathbf{Y}} e^{\mathbf{k}\cdot\mathbf{y}} \int_{\mathbf{T}} \mathrm{P}(\mathbf{Y}=\mathbf{y}|\mathbf{T}=\mathbf{t})\mathrm{P}(\mathbf{T}=\mathbf{t})\mathrm{d}\mathbf{t}\mathrm{d}\mathbf{y} \\ &= \int\int e^{\mathbf{k}\cdot\mathbf{y}}\mathrm{P}(\mathbf{Y}=\mathbf{y}|\mathbf{T}=\mathbf{t})\mathrm{d}\mathbf{y}\mathrm{P}(\mathbf{T}=\mathbf{t})\mathrm{d}\mathbf{t}. \end{aligned} \tag{3.2}$$

To proceed it is necessary to make assumptions about the mutational process. The first is that mutations occur as a Poisson process along branches and the second is that mutations at a locus are additive. Under these assumptions, the changes in the trait value along each branch are conditionally independent given the branch lengths. Khaitovich

et al. (2005) noted that this describes a compound Poisson process. The mgf of a compound Poisson process with rate λ over time t is $\exp(\lambda t(\psi(k) - 1))$, where ψ is the mgf of the distribution of the jump sizes caused by events in the Poisson process. In this case the jump sizes are the effects on the trait value caused by new mutations. Using this expression of the mgf of a compound Poisson process, along with the fact that the mgf of two perfectly correlated random variables with the same marginal distribution is $\varphi_{X_1}(k_1 + k_2)$, we can rewrite equation (3.2) as

$$\varphi_{\mathbf{Y}}(\mathbf{k}) = \int \prod_{\omega \in \Omega} \exp\left(\frac{\theta}{2} t_\omega \left(\psi\left(\sum_{a \in \omega} k_a\right) - 1\right)\right) \mathrm{P}(\mathbf{T} = \mathbf{t}) \mathrm{d}\mathbf{t}. \tag{3.3}$$

We recognize equation (3.3) as the mgf for $\mathbf{T}$ with the dummy variable T_ω for branch s_ω equal to $\frac{\theta}{2}\left(\psi(\sum_{a \in \omega} k_a) - 1\right)$. Or,

$$\varphi_{\mathbf{T}}(\mathbf{s})\Big|_{s_\omega = \frac{\theta}{2}\left(\psi(\sum_{a \in \omega} k_a) - 1\right)}. \tag{3.4}$$

Thus, if the mgf of the distribution of branch lengths is known, equation (3.3) allows us to obtain the mgf of the trait values through a simple substitution. When the trait is controlled by L independent loci, the mgf is obtained by raising equation (3.4) to the power L. Lohse *et al.* (2011) derived the mgf of the genealogy in various population models including migration and splitting of subpopulations. Using their result for a single population we can obtain equation (1) of Schraiber and Akey (2015) using equation (3.4). Similarly, we could define the mgf for models with migration between subpopulations although the number of terms in the recursion for the genealogy mgf blows up as the sample size increases.

3.5 The infinitesimal limit

This general model converges to a normal model when we take the infinitesimal limit. We accomplish this by first substituting a Taylor series for the genealogical and mutational distributions in equation (3.3) (see Appendix B.1). The infinitesimal limit corresponds to the situation where the effect sizes of mutations becomes small as the number of loci becomes large. The resulting distribution is multivariate normal where the expected trait value is $E[T_{MRCA}]\frac{\theta}{2}\mu$, the variance is $E[T_{MRCA}]\frac{\theta}{2}\sigma^2$, and the covariance between trait values in two individuals a and b is $E[\tau_{a+b}]\frac{\theta}{2}\sigma^2$. This limit requires that the products of L and moments three and greater of the mutational distribution go to zero as the number of loci becomes large and the effect size per mutation becomes small. We can think of this as requiring the mutational distribution to not have too heavy of tails. Details of the derivation are given in Appendix B.1.

In this limiting normal distribution, $\frac{\theta}{2}\mu$ can be interpreted as the rate of change in the mean trait value per time unit per genome due to mutational pressure. $\frac{\theta}{2}\sigma^2$ can be interpreted as the rate of accumulation of variance in trait values per unit time per genome. Interestingly, the rate of variance accumulation is proportional to the second moment of the mutational distribution but not to the variance. We can see the intuition for this by considering a degenerate distribution where each mutation has the same effect. Here, we would still expect variation among individuals due to differences in the number of mutations each individual receives, even though the variance of the mutational distribution is zero. The variance among individual trait values thus has one component due to differences in the number of mutations and an additional component due to differences in the effects of these mutations. The first component is proportional to the square of the mean mutational effect, while the second is proportional to the mutational variance. Therefore, the sum of the two components is proportional to m_2, the mean squared effect.

Since the trait values are normally distributed, any linear combination of sampled trait

values will be as well. This includes the distributions of observable quantities like the differences in trait values from a reference individual or from a sample mean. The distribution of trait differences between individuals is multivariate normal with mean zero and covariance between any pair of trait differences given by

$$\mathrm{Cov}[Y_i - Y_j, Y_k - Y_l] = \sigma^2 \frac{\theta}{2} \left(\mathrm{E}[\mathcal{T}_{i,l}] + \mathrm{E}[\mathcal{T}_{j,k}] - \mathrm{E}[\mathcal{T}_{i,k}] - \mathrm{E}[\mathcal{T}_{j,l}] \right), \tag{3.5}$$

where $\mathcal{T}_{i,i} = 0$. Classical theory in quantitative genetics uses a univariate normal distribution of phenotypes in a panmictic population. We can recover this by considering a population of exchangeable individuals. In this case $\mathrm{E}[T_{i,j}]$ is the same for all pairs $i \neq j$. Individual trait values are then conditionally independent given the mean value in the population and are normally distributed with variance $\mathrm{E}[T_{2,2}]\frac{\theta}{2}\sigma^2$.

The normal model in the infinitesimal limit provides additional theoretical justification for studies using normal models to look for differences in selection on quantitative traits between populations (Ovaskainen *et al.*, 2011; Praebel *et al.*, 2013; Robinson *et al.*, 2015). Additionally, equation (3.5) implies that a covariance matrix based on mean pairwise coalescent times rather than population split times should be used when modeling traits as normally distributed in phylogenetics.

3.6 Low-mutation-rate approximation

Thus far, the model assumes that loci are unlinked and can experience an infinite number of mutations. However, a useful simplification is to ignore the possibility of more than one mutation per locus. This approximation is reasonable as long as the nucleotide positions affecting the trait are loosely linked throughout the genome. The low-mutation-rate approximation greatly simplifies the mgf of the trait distribution such that it is no longer necessary

to know the full form of the mgf of the genealogy:

$$\varphi_{\mathbf{Y}}(\mathbf{k}) \approx \left[1 + \sum_{\omega \in \Omega} \mathrm{E}[T_\omega] \frac{\theta}{2} \left(\psi \left(\sum_{a \in \omega} k_a \right) - 1 \right) + O\left(\theta^2\right)) \right]^L . \tag{3.6}$$

Equation (3.6) ignores terms that are order two and above in the mutation rate. Conveniently this equation depends only on the expected length of each branch, whereas equation (3.3) requires moments of branch lengths order two and greater. We can use equation (3.6) to express moments of the trait distribution in terms of expected branch lengths calculated from coalescent models.

3.7 Moments of the trait distribution

For most population genetic models and reasonable sample sizes, the recursive nature of the trait distribution mgf makes it computationally unfeasible to solve under general parameter values. However, it is not necessary to have an expression for the full mgf in order to derive moments of the trait distribution in terms of moments of branch lengths and mutational effects. Under the low mutation rate approximation moments can be calculated by differentiating equation (3.6). Even without making this approximation, moments can be calculated by taking Taylor expansions in equation (3.2) and only considering terms contributing to the desired moment's order. We implemented a symbolic math program to calculate trait moments using this procedure, and the details are given in Appendix B.2. As the normal distribution is completely defined by its first two moments, the extent to which a trait distribution deviates from normality can be measured by the extent to which its moments deviate from those of a normal distribution with the same mean and variance.

We have so far considered the distribution of a trait value Y_a over evolutionary realizations. The expectation of Y_a is $L\frac{\theta}{2}m_1\mathrm{E}[T_{MRCA}]$, and the variance is $L\frac{\theta}{2}m_1\mathrm{E}[T_{MRCA}] + L(m_1^2\frac{\theta}{2})\mathrm{Var}[T_{MRCA}]$. Although simple to derive using computer algebra, expressions for the

higher central moments of Y_a are complicated even under the low mutation rate approximation, and there is not much to be gained by showing them here.

However, in a given evolutionary realization there will be a distribution of trait values in the population. The population-level trait distribution can also be described by its moments, but since this distribution is random the moments at the population level are also random quantities. Since Y_a is relative to a value that is not directly observed, the expected population-level moments offer more insight. In particular, we are interested in how the trait distribution at the population level might deviate from normality. Schraiber and Akey (2015) computed the expected first four central moments of a constant-size population. We derived the same expectations under an arbitrary demographic history,

$$\mathrm{E}[M_2] = L\frac{\theta}{2}\mathrm{E}[\mathbb{T}_{2,2}]m_2 \tag{3.7a}$$

$$\mathrm{E}[M_3] = L\frac{\theta}{2}\mathrm{E}[\mathbb{T}_{3,3}]m_3 \tag{3.7b}$$

$$\begin{aligned}\mathrm{E}[M_4] &= 3\left(L\frac{\theta}{2}\mathrm{E}[\mathbb{T}_{2,2}]m_2\right)^2 \\ &+ 3L\left(\frac{\theta}{2}m_2\right)^2\mathrm{Var}[\mathbb{T}_{2,2}] + \frac{1}{3}L\left(\frac{\theta}{2}m_2\right)^2\left(\frac{11}{9}\mathrm{E}[\mathbb{T}_{2,4}^2] - \frac{1}{3}\mathrm{E}[\mathbb{T}_{2,4}\mathbb{T}_{3,4}] - \frac{1}{4}\mathrm{E}[\mathbb{T}_{3,4}^2]\right) \\ &+ Lm_4\frac{\theta}{2}(\mathrm{E}[\mathbb{T}_{4,4}] + \frac{1}{3}\mathrm{E}[\mathbb{T}_{3,4}] + \frac{2}{9}\mathrm{E}[\mathbb{T}_{2,4}]).\end{aligned} \tag{3.7c}$$

The normal limit corresponds to equation (3.7b) and the second two lines of equation (3.7c) going to zero. To give insight into the expressions in equation (3.7), moment calculations done by hand under the low mutation rate approximation are presented in Appendix B.4.

Equation (3.7a) gives the expected trait variance in the population. However, the amount of variance will vary over realizations of the evolutionary process. The variation in the population variance depends on the sparsity of the trait and the number of causal loci. A certain expected variance can arise either by multiple mutations at one locus or by multiple mutations each at a different, independent locus. The variation in the variance can be

quantified using its coefficient of variation (CVV), the standard deviation of the variance divided by its expectation. For a constant-size, panmictic population

$$\text{CVV} = \sqrt{\frac{4}{3}\frac{1}{L} + \frac{1}{6}\frac{m_4/m_2^2}{L\theta \text{E}[\mathbb{T}_{2,2}]}}. \tag{3.8}$$

Equation (3.8) shows a contribution due to linkage that decreases like $1/\sqrt{L}$ and a contribution due to sparsity that decreases like $1/\sqrt{L\theta \text{E}[\mathbb{T}_{2,2}]}$. Even when the sparsity is low, i.e., when a large number of variants affect the trait, if the trait is only controlled by a single locus there will be considerable variation in the population variance (CVV= $\sqrt{\frac{4}{3}}$). On the other hand, the CVV of a sparse trait controlled by many loci will depend on the ratio of the fourth and squared second non-central moments of the mutational distribution (m_4/m_2^2). For a mutational distribution with mean zero this is equivalent to the kurtosis.

3.8 Comparison to normal distribution

Deviations of the population distribution from normality depend on the distribution of coalescent times and the genetic architecture, and they can be investigated by comparison to the expected moments under normality. Even though recombination is not included in the model, we can form an idea about how linkage might impact deviations from normality. Line two of equation (3.7c) corresponds to the contribution from two mutations occurring at a single locus. The first quantity indicates that the expectation of the fourth moment increases with the variance of the pairwise coalescence time. The second part does not have a clear interpretation. If the sum of these two terms is positive this agrees with the intuition that linkage disequilibrium increases deviations from normality by reducing the effective number of independent loci.

When the mutation rate is low, the ratio of the expected fourth moment to that under

normality is

$$\frac{\mathrm{E}[M_4]}{3\left(L\frac{\theta}{2}\mathrm{E}[\mathbb{T}_{2,2}]m_2\right)^2} \approx 1 + \frac{m_4/m_2^2}{6L\frac{\theta}{2}\mathrm{E}[\mathbb{T}_{2,2}]}\left(\frac{2\mathrm{E}[\mathbb{T}_{4,4}] + \frac{2}{3}\mathrm{E}[\mathbb{T}_{3,4}] + \frac{4}{9}\mathrm{E}[\mathbb{T}_{2,4}]}{\mathrm{E}[\mathbb{T}_{2,2}]}\right), \tag{3.9}$$

The expected excess in M_4 is inversely related to the expected sparsity of the trait which is proportional to $L\frac{\theta}{2}\mathrm{E}[\mathbb{T}_{2,2}]$. This excess depends on demography through a factor $Q = \frac{2\mathrm{E}[\mathbb{T}_{4,4}]+\frac{2}{3}\mathrm{E}[\mathbb{T}_{3,4}]+\frac{4}{9}\mathrm{E}[\mathbb{T}_{2,4}]}{\mathrm{E}[\mathbb{T}_{2,2}]}$. The extent to which demography increases or decreases deviations from normality can be investigated by calculating Q in different models. In a constant-size and panmictic population Q is equal to one. In a population where lineages are exchangeable, $\mathrm{E}[\mathbb{T}_{i,j}]$ can be calculated numerically using expressions from Griffiths and Tavaré (1998) or Polanski and Kimmel (2003). Values of Q in an exponentially growing population are shown in Figure 3.2A. Holding sparsity and the mutational distribution constant, a population which underwent exponential growth will have a greater expected deviation from normality in its trait distribution. However, the growth rate must exceed the present coalescent rate to increase Q by more than ten percent. Another useful example demography is a population that goes through a step change at some point in the past. Figure 3.2B shows that when the population size decreases looking into the past, Q is increased similarly to the exponential growth scenario. When the population size increases looking into the past (a population bottleneck), Q is decreased below one. However, Q appears more sensitive to population growth than to bottlenecks.

As a concrete example, we can consider the differences in the expected fourth moment produced by different demographic histories in different human populations. In the demographic model fit by Tennessen *et al.* (2012), the generic European population experiences a bottleneck associated with out-of-Africa and recent growth while the generic African population experiences a more stable history also with recent growth. Differences in demographic history between the two populations has resulted in a lower heterozygosity in European

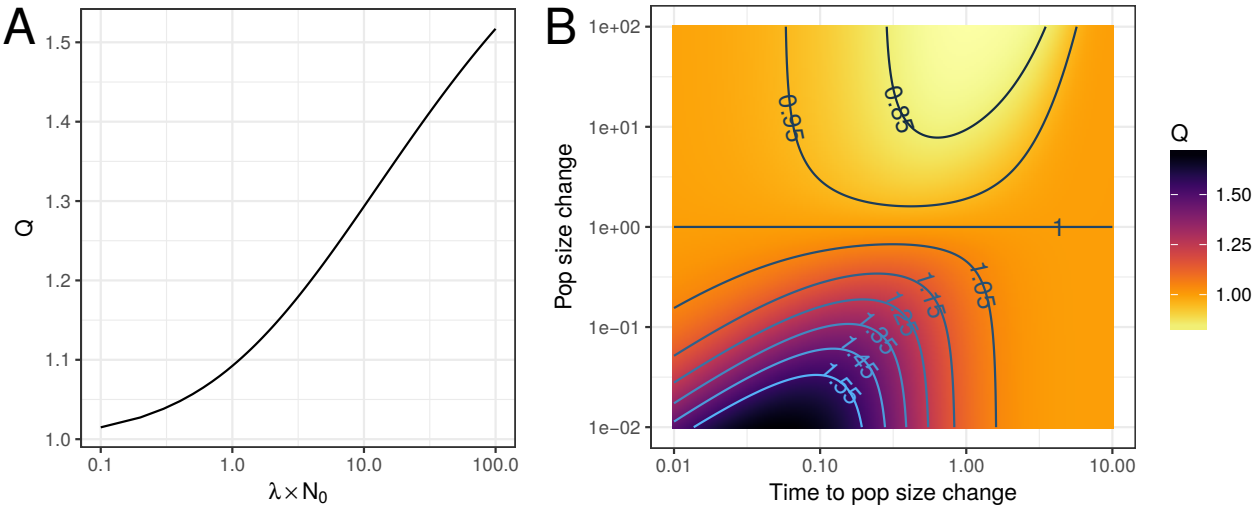


Figure 3.2: **The effects of demography on deviations of the expected fourth central moment of the population trait distribution for normality.** Q measures the effect due to demography on the expected fourth central moment (equation (3.9)). (**A**): Q increases as the exponential growth rate increases relative to the current population size. λ is the growth rate and N_0 is the initial effective population size. (**B**): Q values when the population undergoes an instantaneous step change at some point in the past. The time and magnitude of this change are given in units of the current effective population size. Q increases when the population grows and decreases when it declines.

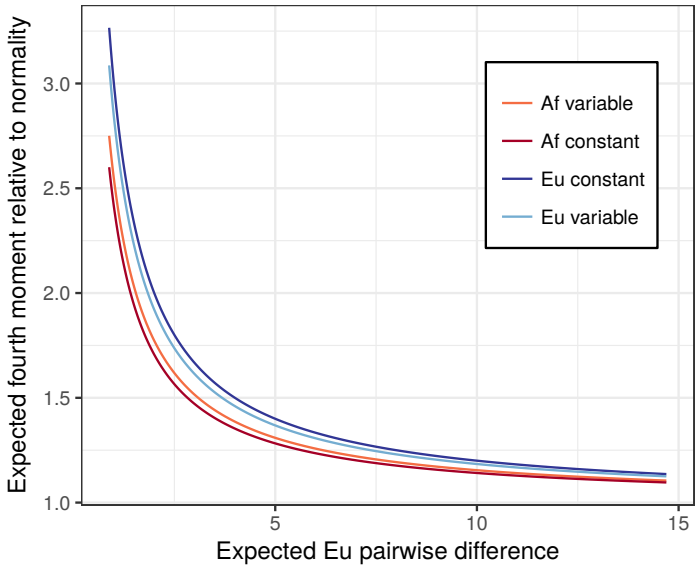


Figure 3.3: **A comparison between the expected fourth moment at different levels of sparsity in the African and European demographic models fit by** Tennessen ***et al.*** **(2012).** Trait sparsity is varied by changing the expected number pairwise differences at sites affecting the trait in the European model. The mutational kurtosis is set to six. The darker lines show the predicted relationship for populations with the same heterozygosity as the European and African models but with constant size.

populations due to the out-of-Africa bottleneck (Yu *et al.*, 2002).

For a given sparsity, the African population model predicts a smaller deviation from normality than the European model (Figure 3.3). The expected fourth moment in constant-size populations with the same heterozygosity as the African and European models is lower for the African model and higher for the European model. This is because the African model is dominated by population growth that leads to a Q greater than one, while the European model is dominated by a bottleneck event that leads to a Q less than one (Figure 3.2). However, differences due to demography are small and the overall deviations from normality are mostly driven by differences in heterozygosity at causal loci.

Another natural way to quantify the deviation of a distribution from normality is its kurtosis. The kurtosis, the ratio of the fourth central moment to the square of the variance, measures the tendency of a distribution to produce outliers (Westfall, 2014). Since the kurtosis of a trait distribution, κ_Y, is ratio of two random quantities, its expectation is not straightforward to calculate. A first order approximation to the kurtosis suggests that κ_Y will be greater than under normality when external branches are longer and and less than under normality when they are longer (Appendix B.3).

However, simulations show that the first order approximation is very poor (Figure B1). The mean kurtosis increases when a trait becomes sparse regardless of whether the population size is constant or growing. Additionally, there is substantial variance in the kurtosis with about a quarter of simulated populations having a kurtosis less than three even as the mean kurtosis increases to almost nine. This high variance in the kurtosis is likely due to a high variance in both the trait variance and fourth moment. This, along with the fact that deviations from the infinitesimal model inflate the fourth moment (equation (3.9)), leads to a situation where the kurtosis increases with trait sparsity but the variance is high across evolutionary realizations.

3.9 Trait divergence in structured populations

Due to the build up of linkage disequilibrium between alleles affecting the trait, the divergence in trait values among different subpopulations in a structured population is about as variable as the variance in allele frequencies (Rogers and Harpending, 1983). We can apply the neutral theory of quantitative traits developed above to obtain a null distribution for the divergence between groups in structured populations. In obtaining this null distribution we use the normal model found in the infinitesimal limit. A common way to quantify the divergence in trait value between groups is Q_{ST}, defined as the variance between the group means divided by the total variance in the population. The normal model does not provide an analytic form for the neutral distribution of Q_{ST}, but it does provide an efficient way to sample from this distribution. The null distribution we obtain through this informed approach performs better than the null distribution suggested by Whitlock and Guillaume (2009).

Since we use a haploid model, we define $Q_{ST} = \frac{V_{between}}{V_{between}+V_{within}}$, where

$$V_{between} = \frac{1}{K}\sum_i \left(\bar{Y}_i - \bar{Y}\right)^2$$

and

$$V_{within} = \frac{1}{\sum_k N_k}\sum_i \sum_j \left(Y_{i,j} - \bar{Y}_i\right)^2.$$

Here, $Y_{i,j}$ is the trait value of individual j in population a.

In the normal model, all $Y_{i,j}$ are normally distributed. Therefore, $\bar{Y}_i - \bar{Y}$ and $Y_{i,j} - \bar{Y}_i$ are also normally distributed. When population sizes are large, individual deviations from population means are nearly uncorrelated as are $V_{between}$ and V_{within}. V_{within} is nearly constant across evolutionary realizations and is equal to $\sum_k N_k E[\mathcal{T}_{k,k}]/\sum_k N_k$ because the within population variances are approximately uncorrelated and their variances are order $1/N_k$. While we do not have an explicit form for the between group variance, we can

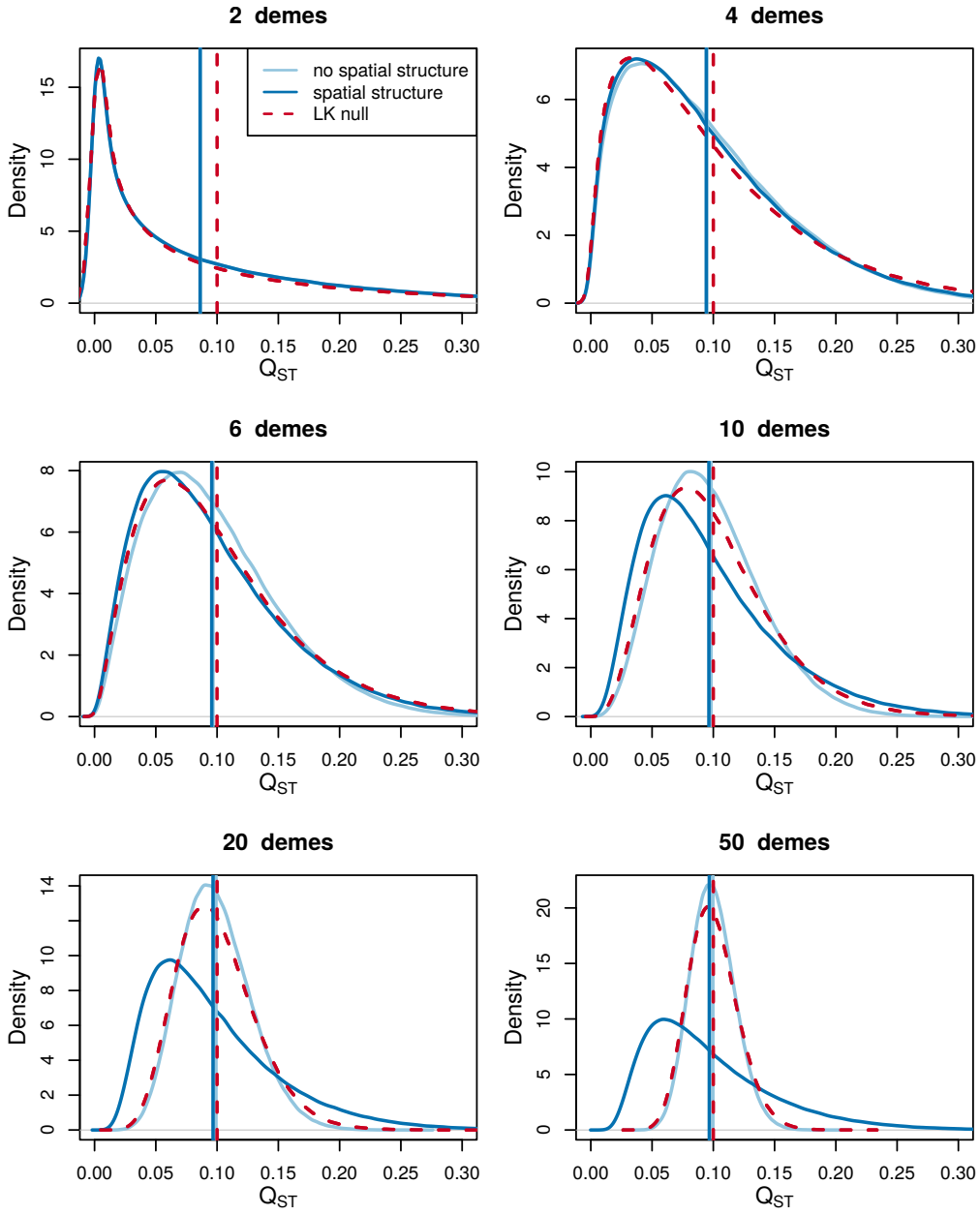


Figure 3.4: **A comparison of neutral sampling distributions for Q_{ST} at the population level.** The Lewontin-Krakauer (LK) distribution for Q_{ST} is compared to the null distribution in the infinitesimal limit in a case with and a case without spatial structure. The case with no spatial structure assumes the migration rate is equal between all demes, and the case with spatial structure arranges demes in a ring with migration only between neighboring demes. Migration rates and subpopulation sizes are all equal and are set such that $F_{ST} = 0.1$ even as the number of demes is increased (Slatkin and Hudson, 1991). Q_{ST} values for these models are simulated by drawing vectors from a multivariate normal distribution as described in the main text. Under the LK distribution Q_{ST} is distributed as $F_{ST}/(n_d-1)$ times a chi-square distribution with n_d-1 ($Q_{ST} \sim F_{ST}\chi^2_{n_d-1}/(n_d-1)$), where n_d is the number of demes. Vertical lines show the mean Q_{ST} under the different null distributions. The mean Q_{ST} values for the two neutral models are nearly identical. Discordance between these lines and that for the LK distribution illustrates that $Q_{ST} \neq F_{ST}$.

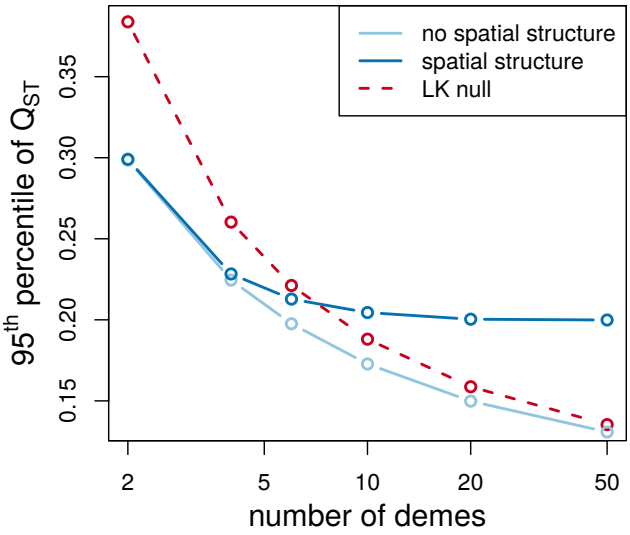


Figure 3.5: **Differences in the 95$^{\text{th}}$ percentile of different neutral sampling distributions for Q_{ST} at the population level.** The Lewontin-Krakauer (LK) distribution is compared to neutral null distributions for structured populations with and without a spatial component using the multivariate normal model that arises in the infinitesimal limit.

simulate from its distribution by drawing a vector of $\bar{Y}_i - \bar{Y}$ values from a multivariate normal distribution with mean zero and with a covariance matrix whose element between populations a and b is

$$\text{Cov}[\bar{Y}_a - \bar{Y}, \bar{Y}_b - \bar{Y}] = \sigma^2(\text{E}[\mathcal{T}_{a,\cdot}] + \text{E}[\mathcal{T}_{b,\cdot}] - \text{E}[\mathcal{T}_{\cdot,\cdot}] - \text{E}[\mathcal{T}_{a,b}]). \tag{3.10}$$

To simulate Q_{ST} values we do not need to know σ^2 because the scale of the trait variance cancels in the Q_{ST} ratio. Therefore, all that is necessary to simulate from the null distribution of Q_{ST} is to have estimates of the expected coalescent time within and between populations. Since only the relative coalescent times matter it is not necessary to scale these estimates to units of years or generations by using a mutation rates.

A classic result in evolutionary quantitative genetics is that $Q_{ST} = F_{ST}$ (Whitlock, 1999). F_{ST}, in this context, refers to a parameter of the population. In particular, $F_{ST} = \frac{\bar{t}-\bar{t}_0}{\bar{t}}$, where $\bar{t}$ is the expected coalescent time for two loci sampled at random from the entire population and $\bar{t}_0$ is the expected coalescent time for two loci sampled within a subpopulation

(Slatkin and Hudson, 1991). This value is constant over realizations of the evolutionary process. Q_{ST} for a particular trait could refer to a state of the population or to an estimate of this state. As shown above, Q_{ST}, as a state of the population, varies across evolutionary realizations. Thus, there is no sense in which Q_{ST} can be defined as a constant parameter in the way that F_{ST} can be. The expectation of $V_{between}$ is $\bar{t} - \bar{t}_0$, and the expectation of V_{within} is $\bar{t}_0$. $\frac{\mathrm{E}[V_{between}]}{\mathrm{E}[V_{between}]+\mathrm{E}[V_{within}]}$ is equal to F_{ST}, but due to Jensen's inequality, the expectation of this ratio ($\mathrm{E}[Q_{ST}]$) is always less than F_{ST}.

The null distribution of Q_{ST} derived here could be useful in testing whether an observed value is unlikley under neutrality. Current goodness-of-fit tests either compare Q_{ST} to an empirical distribution of F_{ST} values or to a χ^2 distribution. In the second case, an identical distribution to that developed by Lewontin and Krakauer (1973) is used as the null distribution for Q_{ST}. The χ^2 testing procedure was suggested by Whitlock and Guillaume (2009) and is implemented in the program *QstFstComp* (Gilbert and Whitlock, 2015). The Lewontin-Krakauer (LK) distribution assumes independence between demes and provides a good approximation in populations without spatial structure (Figure 3.4). When demes are strongly correlated, such as in spatial structured populations, the LK distribution is a very poor approximation. Even when the distributions appear qualitatively similar, there are substantial differences in tail probabilities (Figure 3.5).

The infinitesimal null distribution described here is similar to the extension of the LK F_{ST} test developed by Bonhomme *et al.* (2010) to account for the correlation structure between subpopulations. The Bonhomme *et al.* (2010) method treats allele frequencies as multivariate normal with covariance matrix parameteried by coancestry coefficients. Ovaskainen *et al.* (2011) use a normal model similar to that found in the infinitesimal limit here, but the covariance matrix is also based on coancestry coefficients. When phenotypic and genetic divergence is mostly driven by changes in allele frequency, the coalescent and coancestry based models should be very similar. However, the coalescent model is ultimately preferable

since it is the correct null model at any scale of population divergence in the infinitesimal limit. When only allele frequency data are available, a coancestry model is the only option, but it is still better to model shared ancestry between populations than to use a single value of F_{ST}.

3.10 The response to selection

A situation in which higher order moments of the trait distribution can be relevant is in the response of the population to selection. Evolutionary quantitative genetics often assumes the distribution of additive genetic values in the population remains normally distributed as selection alters the mean and variance. Turelli and Barton (1990) used a multilocus population genetic model to show how departures from normality affect the response to selection. In their analysis, departures from normality are due to the build up of linkage disequilibrium. However, their results are valid regardless of how departures from normality arise. The Turelli and Barton (1990) theory can be used to analyze the response to selection in the toy situation of a trait that has evolved neutrally up to the current time and is then subjected to one generation of selection under a particular fitness function.

According to Turelli and Barton (1990), in the absence of environmental effects the response of the mean phenotype in the population is

$$\Delta\bar{Y} = M_2 L_1 + M_3 L_2 + \gamma_4 M_2^2 L_3 + (M_5 - 4M_3 M_2)\, L_4 + \ldots. \tag{3.11}$$

γ_4 is the excess kurtosis of the trait values in the population above a normal distribution and the M_i terms are again the i^{th} central moments of the trait value distribution in the population. The L_i are selection gradients in terms of the moments of the genetic component of the trait value distribution and describe the shape of the fitness function on the trait values. In the absence of environmental effects, the L_i are selection gradients in terms of the

moments of the trait value distribution that we have so far considered in this study.

Equation (3.11) shows that whether higher order moments of the trait distribution contribute to the selection response depends on the shape of the fitness function through the L_i. The excess kurtosis affects the response to selection linearly with L_3, which depends on the third moment of the trait value distribution.

When selection is cubic ($W(Y) = b_0 + b_3(Y - \bar{Y})^3$), the response to selection is

$$\Delta\bar{Z} = \frac{M_4\beta}{1 + M_3\beta}, \tag{3.12}$$

where $\beta = b_3/b_0$. Cubic selection represents an idealized fitness function to investigate the effects of selection acting on the tails of the population trait distribution. Equation (3.12) is a ratio of random quantities, so calculating the expected response to selection is not feasible. However, we conjecture that demographic and mutational process increasing the expected fourth central moment relative to the third would increase the response to selection. For instance, a high mutational kurtosis and a low skew would likely increase the response to cubic selection.

3.11 Inferring genetic architecture

Schraiber and Akey (2015) suggested it might be possible to infer the shape of the distribution of mutational effects through its moments for sparse traits whose distributions deviate from normality. Using the expressions for the expected moments of the trait value distribution in equation (3.7), we could try and design a method of moments estimator for the moments of the mutational distribution. If $\hat{D}_{i,j}$ is an estimator of $\frac{\theta}{2}\mathrm{E}[\mathbb{T}_{i,j}]$, the system of equations for the first three central moments under the low mutation rate approximation

is

$$\begin{aligned}
\widehat{M_2} &= \widehat{D}_{2,2} Lm_2 \\
\widehat{M_3} &= \frac{1}{3}\widehat{D}_{3,3} Lm_3 \\
\widehat{M_4} &= 3\widehat{M_2^2} + (\widehat{D}_{4,4} + \frac{1}{3}\widehat{D}_{3,4} + \frac{2}{9}\widehat{D}_{2,4}) Lm_4.
\end{aligned} \tag{3.13}$$

From equation (3.13) we can see that the moments of the trait distribution only enter through products with the number of loci potentially affecting the trait (Lm_i). If these products were estimated as $\widehat{Lm_2}$, $\widehat{Lm_3}$, and $\widehat{Lm_4}$, it would be possible to estimate the ratios m_2/m_4 and m_3/m_4 of the moments of the mutational distribution. These ratios are meaningless on their own because any value could be obtained by changing the scale on which the trait is measured. The quantity that is identifiable in this system of equation is the compound parameter $\frac{m_3^2}{m_2 m_4}$. This quantity reflects something about the mutational bias relative to the spread of the distribution. However, it is likely not possible to distinguish sparsity from skewed or heavy-tailed distributions of mutational effects.

3.12 Discussion

Neutral models of quantitative trait evolution are important for establishing a baseline against which to test for selection. Schraiber and Akey (2015) recently analyzed a neutral model of trait evolution that made few assumptions about the number of loci potentially affecting the trait and the distribution of mutational effects at these loci. However, they only derived results for constant-size, panmictic populations. We extend their results to populations with arbitrary distribution of coalescent times and therefore varying demographies and population structures. As their key result, Schraiber and Akey (2015) derived the characteristic function of the distribution of trait values in a sample. In this paper, we instead work with the moment generating function, but the two approaches are interchangeable as long as

the mgf for the distribution of mutational effects exists (which it will provided the moments are all finite). Our main result, given in equation (3.4) is to show that the generating function obtained by Schraiber and Akey (2015) is a special case of a general procedure whereby the moment generation function for a trait distribution can be obtained by making a simple substitution into the moment generating function for a distribution over genealogies. The moment generating functions for many demographic histories of interest and sample sizes above two are sufficiently complex that solving for them is impractical (Lohse *et al.*, 2011). However, progress can still be made by using Taylor expansions to write moments of the trait distribution in terms of moments of the genealogical and mutational distributions.

This result extends previous work using coalescent theory to investigate neutral models of quantitative traits (Whitlock, 1999; Schraiber and Akey, 2015). Ours is the most general model yet analyzed. As a natural first step, we show that the infinitesimal limit suggested by Fisher (1918) leads to a model where phenotypes are normally distributed as the number of loci becomes large and the variance of effect sizes becomes small. In the limiting distribution, the variance of the difference in trait values between two individual is proportional to the expected pairwise coalescent time between them, and the covariance between a pair of differences more generally is $\mathrm{Cov}[Y_a - Y_b, Y_c - Y_d] \propto \mathrm{E}[\mathcal{T}_{a,d}] + \mathrm{E}[\mathcal{T}_{b,c}] - \mathrm{E}[\mathcal{T}_{a,c}] - \mathrm{E}[\mathcal{T}_{b,d}]$. The resulting covariance matrix completely specifies the neutral distribution under the infinitesimal model. This is similar to classic models in evolutionary quantitative genetics considering the neutral divergence of trait values after population splits (Lande, 1976; Lynch, 1989) but holds regardless of the precise details of population structure and history. Schraiber and Akey (2015) derive essentially the same distribution using a central limit theorem argument. It is worth noting that the covariance matrix is scaled by the second moment of the mutational distribution and not the variance.

Since this normal model is the limiting distribution under the coalescent, it is not subject to the problems surrounding the use of species trees to generate covariance matrices in

phylogenetic comparative methods (Mendes *et al.*, 2018). A matrix based on pairwise coalescence times rather than a species tree based on population split times takes into account the effects of all lineages at causal loci, even those that do not follow the species topology. The covariances specified by equation (3.5) could also be used to generate within-species in a similar manner to the method suggested by Felsenstein (2002).

We compared sparse traits to the normal model by calculating how the first four expected central moments different from those expected under normality. This showed how demography and the genetic architecture separately influence the expected deviation from normality. For a fixed expected trait sparsity, population growth produces greater deviations in the fourth central moment while population bottlenecks produce lower deviations (Figures 3.2). However, for realistic demographic scenarios, we find that the effects attributable to demography are small (Figure 3.3). We only analyzed cases where the individuals in the population were exchangeable, but adding population structure would increase deviations from normality as drifting trait means between subpopulations will yield a multimodal distribution.

We next apply the above theory to three simple problems where a coalescent perspective on the neutral distribution of a quantitative trait provides useful intuition. The first of these is the question of the appropriate null distribution for Q_{ST} at the population level. We show how the null distribution under the normal model can be easily simulated from by taking advantage of the theory presented here, providing a much better approximation when populations are correlated than previous approaches (Whitlock and Guillaume, 2009) (Figure 3.4). The second case shows how the single-generation response to cubic selection depends on the third and fourth central moments of the trait distribution (Turelli and Barton, 1990), moments whose expectation depends on genetic architecture and demography. Lastly we show that the shape of the mutational distribution is confounded by the number of loci affecting the trait, making it unlikely that mutational parameters could be inferred from

trait values sampled from a population.

Even though we have broadened the model space for neutral traits, many features of real populations have not yet been incorporated. Linkage between loci is a particular concern as there is substantial linkage disequilibrium between QTLs (Bulik-Sullivan *et al.*, 2015). Lohse *et al.* (2011) derived the form of the moment generating function for linked loci and future work will attempt to incorporate this using equation (3.4). In particular, it will be important for future work to show how this affects the distribution in the infinitesimal limit. Diploidy, dominance, and epistasis have also been ignored thus far. The qualitative effects described here should hold under diploidy, but having trait values within individuals summed over loci from two copies of the genome will decrease deviations from normality. Dominance will also tend to produce a normal distribution as the effects are independent between loci, but future work is needed to examine how this interacts with the distribution of genealogies to affect the trait distribution.

Barton *et al.* (2017) recently performed a deep mathematical investigation of a more formal "infinitesimal model" different from the infinitesimal limit considered here. They proved conditions under which the trait values of offspring within a family are normally distributed with variance independent of the parental trait values conditional on the pedigree and segregation variance in the base population. Interestingly, they found the normality for offspring trait values still holds under some forms of pairwise epistasis that are not too extreme. This implies it may be possible to include epistasis in the infinitesimal limit considered here. It would be important to know how epistasis affects the neutral divergence of trait between populations and species.

Although GWAS of many traits have shown them to be controlled by large numbers of loci (Boyle *et al.*, 2017), this will not necessarily be the case for every trait of interest to biologists. It has been suggested, for instance, that gene expression levels have a sparse genetic architecture (Wheeler *et al.*, 2016). Since there is much interest in testing whether

natural selection has acted on gene expression levels (Whitehead and Crawford, 2006; Gilad *et al.*, 2006; Yang *et al.*, 2017), well-calibrated goodness-of-fit tests will need to take into account the complications that arise when trait distributions deviate from normality (Khaitovich *et al.*, 2005). Direct measurements of mutational distributions (Gruber *et al.*, 2012; Metzger *et al.*, 2016) could aid in such as calibration. Finally, equation (3.7) suggests a means to determine whether sparsity is impacting trait distributions. Populations with a greater $\mathbb{T}_{3,3}$ to $\mathbb{T}_{2,2}$ ratio are also expected to show a greater M_3 to M_2 ratio, so this comparison could be made in studies of multiple populations. Because gene expression studies generally measure a large number of traits, observing such a trend on average could be a sign of deviations from normality.

CHAPTER 4

ESTIMATING THE WOLF MUTATION RATE USING PEDIGREE SEQUENCING

4.1 Abstract

Mutation rates are difficult to measure but have enormous implications for timing events in evolutionary history and for understanding the evolutionary forces surrounding the fidelity of DNA replication. In wolves, uncertainty in the mutation rate has contributed substantially to uncertainty in the timing and location of dog domestication. Here, we directly estimate the per-generation mutation rate in wolves by sequencing a family of seven wolves from one mother and two fathers. Although some putative mutations are still awaiting independent testing, we offer a preliminary estimate of 4.0×10^{-9} per site per generation. This value is consistent with previous estimates based on ancient DNA and suggests a Pleistocene domestication of dogs predating the advent of agriculture.

4.2 Introduction

Dogs were the first animals domesticated by humans and are nearly as widespread across the globe as our own species. It is not surprising then that the history and biology of dog domestication is of great interest to anthropologists, archaeologists, and geneticists, as well as the general public. Central questions in the history of dog domestication have been which species were dogs domesticated from? Did domestication occur only once, or were dogs domesticated multiple times independently? And where and when did the domestication(s) occur? Genetic studies have long since established the fact that dogs are most closely related to the gray wolf (*Canis lupus*) out of all other canids (Wayne and O'Brien, 1987; Vilà *et al.*, 1997). However, despite substantial progress the other questions surrounding dog

domestication still lack definitive answers (Freedman and Wayne, 2017).

The most basic information providing an upper bound on the age of dog domestication is the point at which anatomically modern humans left Africa because the gray wolf is found only in Eurasia and North America. Current population genetic analyses place the divergence of African and Out-of-Africa human populations around 100,000 years ago (Jouganous *et al.*, 2017), so dog domestication must have happened some time after this period. On the other hand, fossil evidence should be able to provide a lower bound on the timing of domestication through the radiocarbon dating of the oldest known dog remains. However, there is disagreement over the assignment of bones to dogs or wolves that makes the position of this lower bound unclear. The oldest remains that archaeologists agree belong to dogs are around 15,000 years old (Pionnier-Capitan *et al.*, 2011; Frantz *et al.*, 2016). Older remains with some dog-like features dated between 30,000 and 40,000 years ago have been found in Belgium, the Altai Mountains, and the Czech Republic (Germonpré *et al.*, 2009; Ovodov *et al.*, 2011; Germonpré *et al.*, 2015), but whether these actually represent dogs has been disputed on morphological grounds (Napierala and Uerpmann, 2012). Ancient mtDNA has placed the Belgian specimens as an outgroup to other dogs and wolves, while the Altai sample was nested within the range of dog and wolf mtDNA diversity (Thalmann *et al.*, 2013). Although this would seem to rule out the Belgian specimens as ancestral to dogs and keep the Altai specimen as a candidate, the low resolution of any mtDNA analysis makes it impossible to drawn any definite conclusions from these fossils. Archaeological findings are also equivocal as to the location of domestication, with remains older than 10,500 years old found in Western Europe, the Middle East, and East Asia. Moreover, the fossil record can only say where and when dogs were living and not anything about the ancestry of these animals. It is therefore uninformative about whether dogs were domesticated once or arose multiple times independently.

Population genetic studies hold great promise for unraveling the mysteries of dog domes-

tication. Genetic studies of demographic history in humans have illuminated broad-scale patterns of Out-of-Africa migrations and subsequent population growth as well as much more recent and subtle occurrences of population movement and admixture (Nielsen *et al.*, 2017). A conceptually straightforward approach to the study of dog domestication would be to fit a demographic model to extant dog and wolf populations. The wolf population from which dogs split most recently would identify the geographic location of domestication, and the time of this split would give a good upper bound on the timing. Skoglund *et al.* (2011) fit a very simple population model to estimate the population split time between dogs and wolves from China, Spain and India. Although the estimated divergence time with the Indian wolf was the earliest, all three times were very similar and given the simplicity of the model no strong conclusion could be drawn regarding the location of domestication. Freedman *et al.* (2014) fit a more detailed model to high quality whole genome sequences from three dogs (Boxer, Basenji, and Dingo) and wolves from three locations in Eurasia representing hypothesized locations of dog domestication (Europe: Croatia, Middle East: Israel, and East Asia: China). Similar to Skoglund *et al.* (2011) this study found that dog populations have not split more recently from any extant wolf population and that any affinities dogs have to extant wolf populations are most likely due to interbreeding following domestication. A subsequent study by Fan *et al.* (2016) expanded the analysis from Freedman *et al.* (2014) to include a greater number of modern wolves and came to the same conclusion.

Model-free analyses have also been used to try and infer which wolf population dogs were domesticated from. Shannon *et al.* (2015) observed a gradient in linkage disequilibrium (LD) that decays away from Central Asia to argue for a Central Asian origin of domestication. A gradient of increases in LD and decreasing heterozygosity would be expected under a model where domestication occurred once and dogs spread across the globe through a series of founder events. However, when Wang *et al.* (2016a) reanalyzed the data and included data from southern China they found equally low levels of LD as in Central Asia. Wang *et al.*

(2013) and Wang *et al.* (2016b) observed a greater heterozygosity in South Asian indigenous dogs than in breed dogs from around the world. Principal components analyses in these studies also suggested that the South Asian indigenous dogs were more closely related to wolves. Results from these model-free analyses are suggestive of a South East Asian location of dog domestication, but it is difficult to interpret them given the likely complex history of domestication and certainty of subsequent gene flow from wolves. Population genetic models of demography, although currently restricted to fewer samples, have the advantage of explicitly weighing the effects of population splits and migrations while also providing measures of model fit.

Population genetic demographic models also output split times. Freedman *et al.* (2014) estimated a split between 14,000-16,000 years ago and the follow-up study by Fan *et al.* (2016) found a similar time. Wang *et al.* (2013) also fit a demographic model of isolation-with-migration to the joint site frequency spectrum of the four extant wolves and three Chinese indigenous dogs. They estimate a split time of about 32,000 years ago. A split between dogs and wolves of 11,000 - 16,000 years ago is cutting it close given the clear appearance of dogs in the fossil record around 15,000 years ago. The Wang *et al.* (2013) estimate might therefore seem more reasonable, but the small sample size used to construct the site frequency spectrum and the fact that wolves from a wide geographic range were treated as a single population might have influenced this estimate. However, as noted by Freedman *et al.* (2014), the biggest factor adding uncertainty to split times in estimates of demographic history is the mutation rate.

In demographic inference in population genetics, the effects of a particular model of history is specified through its effects on genealogies. Genealogies themselves are not observable and the only information available about them, and therefore the demographic history one wants to learn, is through the mutations that occur along branches of individual genealogies. Having some information about the mutation rate is therefore the only way, at least

in a probabilistic sense, to know the lengths of branches in a genealogy in unites of time. The mutation rate therefore allows the conversion of time from mutational units to years or generations, and any uncertainty in the mutation rate is directly propagated to estimates of the divergence time between populations. A commonly used a mutation rate for dogs and wolves has been 1×10^{-8} per generation (Lindblad-Toh *et al.*, 2005; Skoglund *et al.*, 2011; Freedman *et al.*, 2014). Lindblad-Toh *et al.* (2005) and Skoglund *et al.* (2011) used this value without any justification. However, Freedman *et al.* (2014) noted that 1×10^{-8} per generation is close to the value of 6.6×10^{-9} obtained if one multiples the average mammalian mutation rate of 2.2×10^{-9} per year measured by Kumar and Subramanian (2002) using substitution on a fossil-calibrated phylogeny by a generation time of three years. Considering the range of mutation rates from 6.6×10^{-9} to a values of 1.8×10^{-8} per generation measured from humans by Sun *et al.* (2012), Freedman *et al.* (2014) expanded their range for the split time between dogs and wolves to 11,000 - 34,000 years ago. In their estimate Wang *et al.* (2013) used the mutation rate of 6.6×10^{-9} per generation. If they had used a rate of 1×10^{-8} instead their estimated split time would have been about 21,000 years ago. Much of the discordance in estimated split times was therefore due to different assumptions about the mutation rate.

Two recent studies have estimated the mutation rate specifically for dogs and wolves using ancient DNA. Skoglund *et al.* (2015) and Frantz *et al.* (2016) used an approach developed in Green *et al.* (2010) to estimate the divergence time between the ancestral populations of humans and Neanderthals. Green *et al.* (2010) used a fossil-calibrated divergence time between humans and Orangutans as a reference to obtain a human-Neanderthal divergence time in units of years. Alternatively, if the divergence time between populations is known in years, it is possible to obtain an estimate of the per-year mutation rate. The procedure works by calculating the proportion of sites that have the derived allele in the ancient individual conditional on that site being heterozygous in a modern individual. Skoglund *et al.* (2015)

denote this proportion $F(A|B)$. This is equivalent to the average derived allele frequency at the time of population divergence conditional on that allele having count one in a sample of two chromosomes from the extant population. If the demographic history and mutation rate of the extant population are known then allele frequencies can be simulated and a divergence time can be chosen so that derived allele frequencies at this time match the observed $F(A|B)$. The same process can estimate the mutation rate if the divergence time is known.

Skoglund *et al.* (2015) and Frantz *et al.* (2016) estimate the demographic history for modern dog and wolf populations using the pairwise sequentially Markovian coalescent approach (Li and Durbin, 2011). The divergence time between the ancient and extant populations is set to the age of the bone from which the ancient DNA was extracted. Skoglund *et al.* (2015) estimate a mutation rate of 4.0×10^{-9} per generation using a DNA from a 35,000-year-old wolf from the Taimyr peninsula in Siberia. Frantz *et al.* (2016) estimate a mutation rate of $3.0 - 4.5 \times 10^{-9}$ per generation using DNA from a 4,800 year old dog from the Newgrange site in Ireland. Both of these rates are lower than those used previously, pushing the divergence time between dogs and wolves further into the past. In particular, when calibrating their model using the mutation rate from Skoglund *et al.* (2015), Fan *et al.* (2016) estimate a divergence time around 29,000 years ago.

However, there are problems with the above approach to mutation rate estimation. It is not clear how accurately the demographic models predict the true ancestral allele frequencies, and using the age of the ancient specimen will underestimate the divergence time if the individual died after the populations split. While the effect of the first problem likely depends on the method used to infer demography, the second problem will lead to overestimation of the mutation rate. Additionally the Green *et al.* (2010) approach assumes a lack of post-divergence gene flow. The presence of gene flow would increase the $F(A|B)$ statistic and lead to underestimation of the mutation rate. To remedy these issues, we estimate the mutation rate using whole-genome sequencing of parents and offspring. This approach is insensitive to

the issues of fossil calibration and demographic assumptions surrounding previous estimates.

Estimating the mutation rate by sequencing parents and offspring is conceptually straightforward: count the number of sites where the offspring is heterozygous while the parents are homozygous and divide by the number of sites in the genome. In practice it can be difficult to distinguish true de novo mutations (DNMs) from sequencing errors, missed heterozygous genotypes in parents, and alignment issues in repetitive regions of the genome. After identifying DNMs, calculating a rate also requires knowledge of the proportion of the genome in which DNMs would have been detected in had they occurred. In spite of these difficulties, pedigree-based estimation of mutation rates has been performed in a growing list of species. Estimates based on pedigree sequencing are available for *Homo sapiens* (Kong *et al.*, 2012), *Pan troglodytes* (Venn *et al.*, 2014), *Drosophila melanogaster* (Keightley *et al.*, 2014), *Heliconius melpomene* (Keightley *et al.*, 2015), *Apis mellifera* (Yang *et al.*, 2015), *Arabidopsis thaliana* (Yang *et al.*, 2015), *Ficedula albicollis* (Smeds *et al.*, 2016), and *Chlorocebus pygerythrus* (Pfeifer, 2017). In humans, pedigree studies revised the mutation rate from a value of 2.3×10^{-8} per generation that had been estimated using a fossil-calibrated divergence time with Chimpanzees to 1.2×10^{-8} per generation (Ségurel *et al.*, 2014). This has potentially profound effects on the timing of various events during human evolution because the estimated times of all events would have to be scaled back by a factor of two. However, an increase in the generation time could compensate for this by decreasing the average number of mutations occurring per year (Scally and Durbin, 2012; Amster and Sella, 2016).

The evolution of mutation rates themselves is an important topic in molecular evolution. Zuckerkandl and Pauling (1965) and originally observed that the rate of substitutions is roughly constant per year across a diverse set of organisms. In his theoretical treatment of this observation, Kimura (1969) noted that the constancy per year was surprising given that a neutral model of evolution predicts a constant rate per generation. One explanation

then for the observation of constancy per year is that per-generation mutation rates are themselves evolving. Mutation rates do indeed differ drastically among organisms (Drake *et al.*, 1998), but the forces driving this evolution remain obscure. The drift-barrier hypothesis proposes a general explanation for mutation rate evolution whereby species with lower effective population sizes accumulate more mutator alleles because they are less able to purge them due to the deleterious mutations they cause (Sung *et al.*, 2012). Part of assessing the drift-barrier hypothesis means investigating the extent to which mutation rates are negatively correlated with effective population size across the tree of life (Lynch *et al.*, 2016). The utility of high-throughput sequencing for directly estimating mutation rates in pedigree and mutation-accumulation studies means that the relationship between mutation rate and effective population size can now be directly investigated. The available estimates so far have followed the prediction of the drift-barrier hypothesis that species with higher effective population sizes have lower mutation rates (Smeds *et al.*, 2016; Lynch *et al.*, 2016; Pfeifer, 2017). Mutation rates have so far been directly estimated in 37 species and 13 multicellular eukaryotes, and more estimates are needed to further test the drift-barrier hypothesis.

In this study we estimate the mutation rate in wolves by sequencing a family from Yellowstone National Park consisting of four children, a single mother, and two fathers. We identified DNMs by applying strict filters based on genomic context and independently verifying a large number of candidate sites using Sanger sequencing. The general pipeline is similar to that developed by Keightley *et al.* (2014). The posterior probability of different mutation rates was then calculated based on the number of sites in the genome passing all filters and estimated false negative rates. Although not all candidate DNMs have yet been tested, we can bound the mutation rate between 2×10^{-9} and 8×10^{-9} per generation, with a preliminary point estimate at 4×10^{-9} per generation. This is consistent with estimates based on ancient DNA. There is also some evidence for a paternal age effect or mutation rate heterogeneity among individuals.

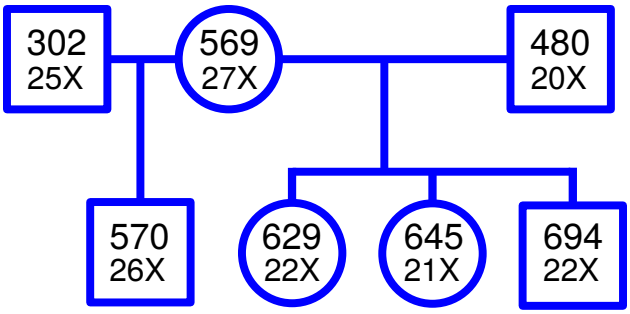


Figure 4.1: **The wolf pedigree used in this study.** Whole genome sequences from these four wolves were analyzed to detect DNMs. Top numbers give Yellowstone National Park ID (YNPID) and bottom numbers give the average sequencing depth.

4.3 Methods

4.3.1 Samples and sequencing

A pedigree of seven wolves containing one mother and her four offspring from two different fathers was sequenced (Figure 4.1). Samples from 569F and 570M were sequenced as part of Fan *et al.* (2016) using the HiSeq 2000 platform and are described in that paper as the "Yellowstone trio". The initial sequencing of 302M in that study has over 50% PCR duplicates, and was redone on four lanes of the Illumina HiSeq 2500 platform. The other four individuals were sequenced on one lane each using the Illumina HiSeq 4000 platform.

4.3.2 Calculating genotype likelihoods

To identify DNMs we employed a strategy of liberally calling potential DNMs and validating them with independent sequencing in order to limit the false negative rate. To do this, sequencing data from all seven individuals were passed through a series of processing and filtering steps to generate a set of putative DNMs that were then tested through by Sanger sequencing (Figure 4.2). Reads were aligned to the dog reference genome version CanFam3.1 using BWA 0.7.12 (Li and Durbin, 2009). Although a wolf reference genome exists, it has been shown that using this does not substantially impact analyses when compared to Can-

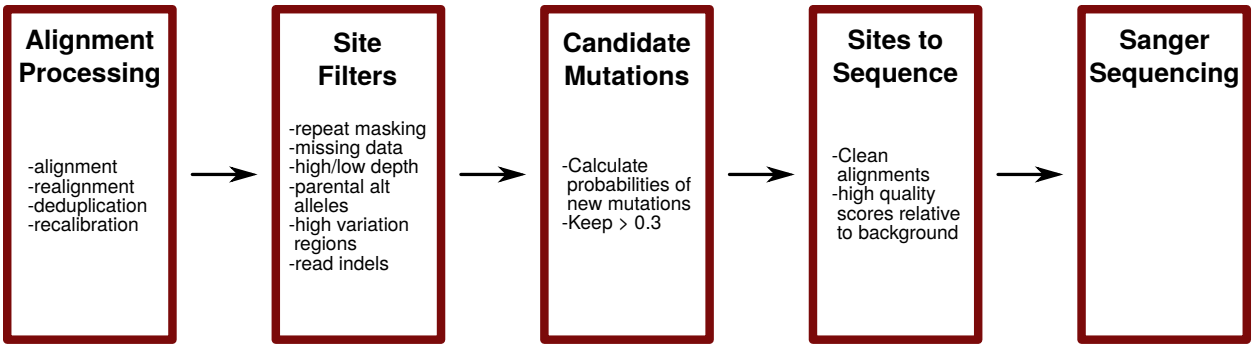


Figure 4.2: **Schematic representation of the bioinformatics pipeline.** We developed a bioinformatics pipeline to go from raw reads to a verified set of candidate DNMs. Genotype likelihoods were calculated after alignment processing and before site filters were applied.

Fam3.1 (Gopalakrishnan *et al.*, 2017). Alignment files were realigned around indels and deduplicated using GATK 3.5.0 (DePristo *et al.*, 2011). We first called a set of variants from the data for the sole purpose of recalibrating base quality scores. This was done using GATK's UnifiedGenotyper and SNP sites were kept if they passed the recommended hard filtering thresholds ($\mathrm{QD} > 2, \mathrm{FS} < 60, \mathrm{MQ} > 40, \mathrm{MQRankSum} > -12.5, \mathrm{ReadPosRankSum} > -9$). Sites in repetitive regions were then filtered from this set using Repeatmasker 4.0.6 and a dog-specific repeat library (Smit *et al.*, 2013-2015). We then treated the remaining set of variants as known variants to generate a recalibration table which was then used to recalibrate base quality scores using the algorithm provided by GATK (DePristo *et al.*, 2011). After recalibrating base quality scores, genotype likelihoods were calculated at all sites using the GATK UnifiedGenotyper algorithm, a minimum base quality score of 15, and the "emit all sites" option. Genotype likelihoods calculated in this manner are independent for each individual. All sites were retained regardless of their variant quality scores in order to avoid bias against variable sites and therefore potential DNMs.

4.3.3 Site filters

Filters were then applied per trio in an attempt to select a set of observed sites at which we could confidently detect DNMs if they had occurred. Our filters were therefore chosen

to remove genomic regions likely to be enriched for false positives and to retain sites with sufficient coverage and sequencing quality. Sites were filtered on a per-trio basis and without reference to genotype likelihoods. As long as filters are not correlated with the probability a site has a DNM they should not bias the mutation rate estimate. As with the variant set used in recalibration, we first filtered sites in repetitive regions using Repeatmasker (Smit *et al.*, 2013-2015). We then removed sites marked by GATK as missing. This included sites with a sequencing depth of zero as well as sites where the fraction of reads spanning the locus containing deletions exceeded 0.05. Sites with less than ten-fold or greater than 100-fold coverage in any individual were removed because we cannot be confident in genotypes at low coverage sites and because high coverage sites tend to have many mismapped reads. Sites with one or more alternative alleles observed in the parents were removed so as to only examine sites where the parents were confidently homozygous. Two additional filters were applied in order to account for base quality and mismapped reads. We removed sites with four or more variant sites within a 200 base window on either side, and we then also removed sites where three or more of the reads mapping to that site contained gaps in their alignments. The first of these has the potential to bias the mutation rate estimate either because areas with high diversity may be regions of the genome with elevated mutation rates or if heterozygosity itself has an impact on the mutation rate (Yang *et al.*, 2015). However, because the fraction of the genome removed by this filter was small (4%) the elevation of the mutation rate in these regions would have to be large to have a meaningful bias. For instance, the mutation rate in these regions would need to be about 26 times greater to double our estimated mutation rate.

4.3.4 Identification of DNMs

The above procedure yields a set of sites for each trio in the family where the the parental individuals appear homozygous. The vast majority of these sites are homozygous in the

offspring as well, but a small number contain DNM. To find these mutations, we narrowed our search to those sites where one or more alternative alleles were observed in the child. Even among sites with at least one alternative read in the child and none in the parents, other studies have found that the vast majority are sequencing errors, missed heterozygous genotypes in the parents, or most often due to reads that have mismapped from elsewhere in the genome (Keightley *et al.*, 2015; Smeds *et al.*, 2016; Pfeifer, 2017). To distinguish sites with true DNMs from those with sequencing errors or missed parental heterozygotes we calculated a de novo score (DN_p) that can be roughly interpreted as the probability that each site contains DNM. This calculation was made by considering the probability of each genotype combination between parents and offspring (Ramu *et al.*, 2013).

$$P(G_C, G_M, G_F|D) \propto \overbrace{P(D_M|G_M)P(D_F|G_F)P(D_C|G_C)}^{\text{genotype likelihoods of observed individuals}} \times \underbrace{P(G_C|G_M, G_F)}_{\text{transmission probability}} \underbrace{P(G_M, G_F|\theta)}_{\text{parental heterozygosity}} \tag{4.1}$$

The term for transmitting different genotypes to the offspring contains an assumed mutation rate and the term for the parental heterozygosity contains a parameter for the heterozygosity in the population (Ramu *et al.*, 2013). The purpose of this calculation is to weigh evidence from genotype likelihoods, which take into account both sequencing depth and quality, with our prior belief about how often mutations occur and how likely the parents are to be homozygous at a given site. Given the fact that base qualities may be a poor reflection of the actual probability of sequencing errors, we do not interpret the de novo scores calculated using this formula as true probabilities but rather as scores with which to rank potential DNMs. We first calculated DN_p using a mutation rate of 4.0×10^{-9} per generation as estimated by Skoglund *et al.* (2015) and a heterozygosity of 0.00015 since values close to this have been observed in many wolf populations (Freedman *et al.*, 2014; Fan *et al.*, 2016).

These realistic parameters yielded very few candidate DNMs, potentially because base quality score recalibration was overly conservative. We then calculated DN_p with the mutation rate prior set to 1×10^{-6} per generation and the heterozygosity set to 0.008. This yielded a reasonable number of potential DNMs. Only the genotype combination where the offspring is heterozygous while both parents are homozygous for the reference allele was considered compatible with a DNM. Sites with a DN_p greater than 0.3 were examined further.

Even among sites passing all filters and having a DN_p greater than 0.3, many contained, upon visual inspection, obvious sequencing errors or had a high proportion of mismapped reads. This is an issue faced by many studies estimating the mutation rate using pedigree sequencing (Keightley *et al.*, 2014, 2015; Smeds *et al.*, 2016; Pfeifer, 2017). Following the example set by Keightley *et al.* (2014) we removed these sites from further analysis by examining read alignments manually using IGV 2.3.79 (Robinson *et al.*, 2011; Thorvaldsdóttir *et al.*, 2013) and the `igv_plotter` library (Weisburd, 2017). IGV plots often clearly showed, in the form of high numbers of mismatches and gaps in the alignment, whether a site was in a region where reads tended to mismap to. Some examples of the kind of plots that allowed us to diagnose sites with mismapped reads and sequencing errors are show in Appendix C.2. In addition to looking at alignments, the QualByDepth (QD) and MappingQualityRankSumTest (MQRankSum) quality metrics output by GATK of potential mutations were examined in the context of the other variants in the sample. QD reflects the sequencing quality at a site normalized by the depth and MQRankSum reflects how well alternative versus reference reads map. Potential mutations falling within the typical QD and MQRankSum ranges of the sample variants and that had clean alignments were selected for Sanger sequencing of the parents and offspring to verify their status. As stated above, Sites were chosen liberally for validation in order to minimize the number of false negatives. To confirm out interpretation of alignment plots, some sites were also sequenced that from inspection of alignment plots appeared be clear examples of mismapping or sequencing er-

rors. A preliminary version of this pipeline that was not applied on a per-trio basis and allowed GATK to call variants was used to identify initial candidates. These initial candidates included both sites with clean alignments and some appearing to be errors The pipeline described above was implemented using Snakemake (Köster and Rahmann, 2012).

4.3.5 Mutation rate calculation

In order to calculate an estimate of the mutation rate given a set of verified DNMs it is necessary to know at how many sites mutations could have potentially been observed at in each trio. We calculated this number by taking the number of sites in each trio that passed all filters and multiplied it by one minus an estimated false negative rate for that trio. We took a random 0.01% of all sites in the remaining portion of the genome and estimated the fraction of sites that pass the filter for gaps in read alignment based on this subset of the data.

The false negative rate for each trio was defined as the probability that a site containing a true DNM and passing all filters would have a DN_p less than 0.3. This calculation was based on the assumption that all true DNMs DNMs with a DN_p greater than 0.3 would be chosen for validation by independent sequencing. That is, it is assumed that no true DNM was discarded when its alignment was examined. We calculated the false negative rate by generating a set of simulated DNMs from the genotype likelihoods in each trio and calculating DN_p for each. The simulated set was created by first taking genotype likelihoods in each child at sites where one or both of their parents were heterozygous as well. These transmitted heterozygous sites were, of course, the majority of heterozygous sites in each child. These sites were then assigned at random genotype likelihoods from sites in the parents passing all filters and at which no alternative alleles were observed. DN_p values were calculated using equation (4.1) just as for the real data. To calculated the false negative rate for a trio, we further assumed that the joint coverage distribution at real DNMs would be the same as

that for the randomized, simulated set.

Using the number of sites passing all filters and an estimate of the false negative rate, we calculated a posterior distribution on the per generation mutation rate using

$$\sum_i X_i \sim \text{Poisson}\left(\sum_i 2L_i(1-\beta_i)\mu\right)$$

$$p(\mu) \propto \sqrt{\frac{1}{\mu}},$$

where X_i is the total number of DNM we observed in trio i, L_i is the number of sites passing all filters in trio i, and β_i is the estimated false negative rate for trio i. μ denotes the mutation rate per generation. We used a prior on μ proportional to the inverse square root. This is the Jeffrey's prior for a Poisson rate.

Given evidence that mutation rates increase with paternal age and the variation in paternal age among our trios (Kong *et al.*, 2012; Venn *et al.*, 2014; Rahbari *et al.*, 2016), we also fit a model allowing the mutation rate to depend on paternal age. This model can be written as

$$X_i \sim \text{Poisson}\left(\sum_i 2L_i(1-\beta_i)(b+aY_i)\right)$$

$$b \sim \text{Normal}(0,1)$$

$$a \sim \text{Cauchy}(0,1),$$

where Y_i is the age of the father at birth in trio i. The posterior distribution on the per-generation mutation rate was analyzed in this model by looking at the distribution of $b+a\bar{Y}$ where (a,b) is a sample from the posterior distribution and $\bar{Y}$ is the average generation time in wolves. This model was implemented using Stan (Stan Development Team, 2016).

4.4 Results

4.4.1 Sequencing filtering and identification of de novo mutations

Filtering within each trio yielded 1.04 Gb, representing a reduction of over one half from the 2.39 Gb of sequence in the dog reference genome (Figure 4.3). The majority of sites were removed when filtering repetitive regions using the dog-specific repeat library and when removing sites where three or more reads mapping there contained gaps. The overall number of sites remaining after filtering did not differ substantially among trios.

Table 4.1 shows how many sites in each trio had at least one alternative read and thus were considered potential DNMs. Although 570M had fewer such mutations than the other offspring, an approximately equal number passed the DN_p cutoff and around twice as many passed the manual inspection. In total 86 sites were chosen for sequencing, 22 of which were chosen using a preliminary version of the pipeline described above. All true DNMs in the preliminary set also showed up in the final set. Sequencing has so far been completed at 71 of the 86 sites. Figure 4.4 shows the current status of this sequencing as whether potential DNMs confirmed, were false positives, or have yet to be sequenced. The quality scores at potential sites are shown against a background distribution from sites passing all filters and where at least one alternative read was observed in each trio. Sites with low quality scores ($QD < 4$) tended to be false positives, as well as those with low mapping qualities for reads with alternative alleles ($MQRankSum < -2$). Among potential DNMs with quality scores within the typical range there were still many false positives.

At present, 13, 5, 5, and 3 DNMs have been confirmed in the offspring 570M, 629M, 645F, and 694F respectively. If all sites yet to be sequenced turn out to be DNMs these numbers will be raised to 17, 7, 9, and 8 (Table 4.1). To give a sense for the range of possible mutation rate estimates, analyses are presented using both these minimum and maximum possible counts. What is clear so far is that 570M has about twice as many DNMs as the

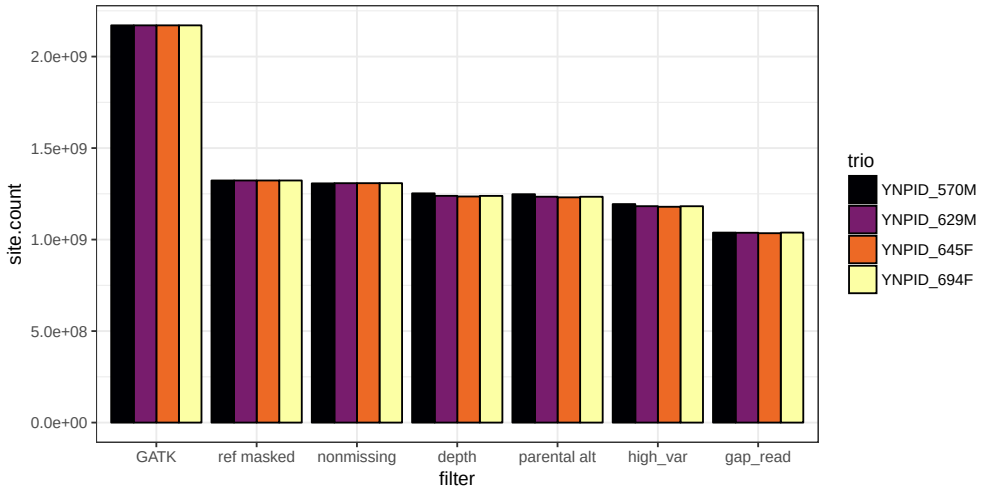


Figure 4.3: **Number sites considered remaining after the sequential application of filters.** The final bars represent the number of sites ultimately considered for DNMs. The "gap_read" filter was only applied to a random sample of 0.01% sites so this number is an estimate. It is worth noting that the order the filters were applied could obscure the effects that each might have had if applied individually to the raw set of sites.

other offspring.

4.4.2 De novo mutation rate

In order to estimate the mutation rate it was necessary to calculate the false negative rate for each trio. Individual 570M had a lower false negative rate than its half-siblings due to the fact that it and its father were sequenced to higher coverage (Figure C1). In general, the fraction of simulated DNMs with $\mathrm{DN}_p < 0.3$ does not decrease to zero as the sequencing

	YNP 570M	YNP 629M	YNP 645F	YNP 694F
≥ 1 alt. read	2,676	3,529	3,935	3,225
$\mathrm{DN}_p > 0.3$	112	109	106	108
Sanger sequenced	22 (12)	14 (1)	15 (3)	12 (7)
Confirmed de novo	13-17	5-7	5-9	3-8

Table 4.1: **Examination of potential DNMs.** The number of sites in each trio after all filtering steps, having a DN_p score greater than 0.3, and chosen for Sanger sequencing. Number in parentheses give the number of sites that were sequenced based on a preliminary version of the pipeline described here. The final row gives the number of confirmed DNMs found in each trio.

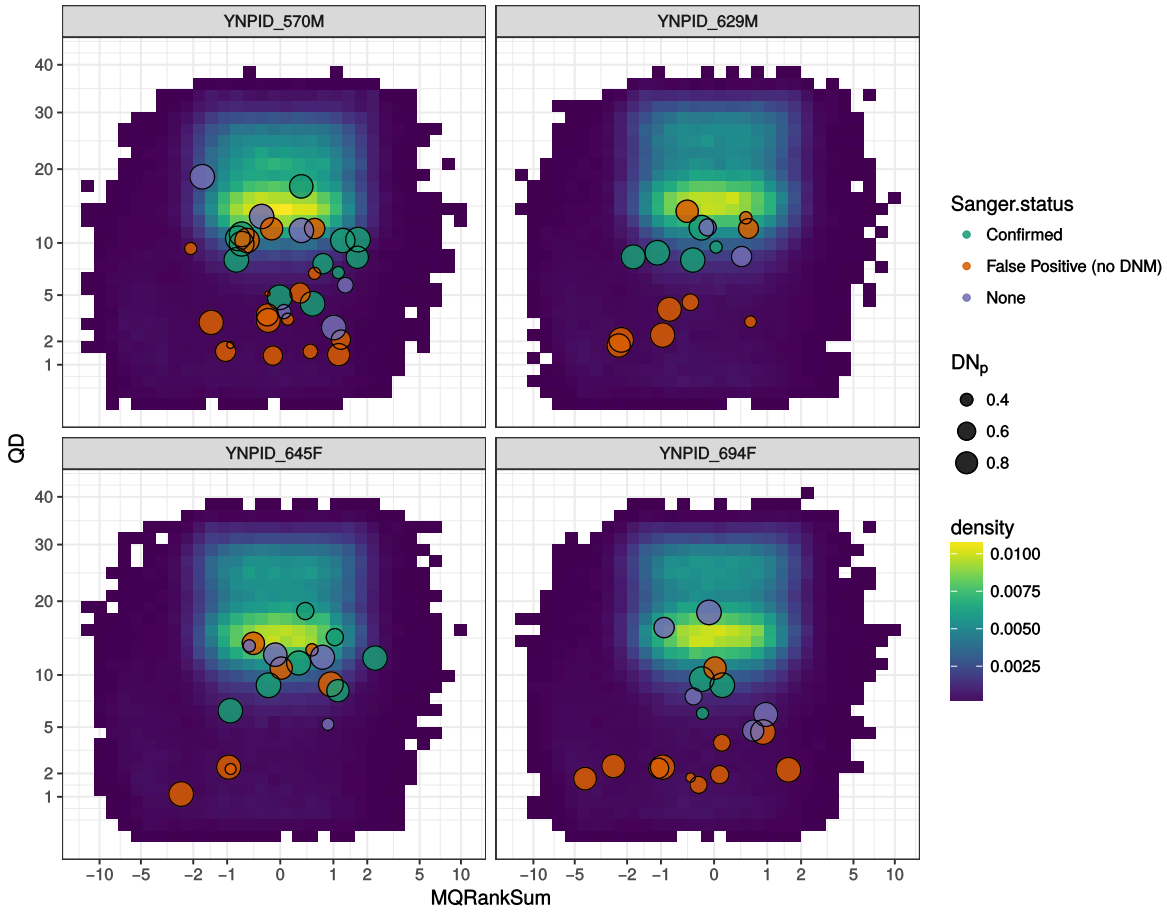


Figure 4.4: **Current results from Sanger sequencing of potential DNMs.** Quality scores at sites chosen for independent sequencing are plotted against the background distribution of quality scores from sites passing all filters and having at least one alternative read in one individual in each trio. MQRankSum is a measure of quality and negative values indicate reads with alternative alleles mapped less well than reads with reference alleles. This can reflect mismapped reads. QD measures base quality score normalized by sequencing depth and can be thought of as a metric for sequencing quality. Sanger sequencing has not yet been completed at all sites.

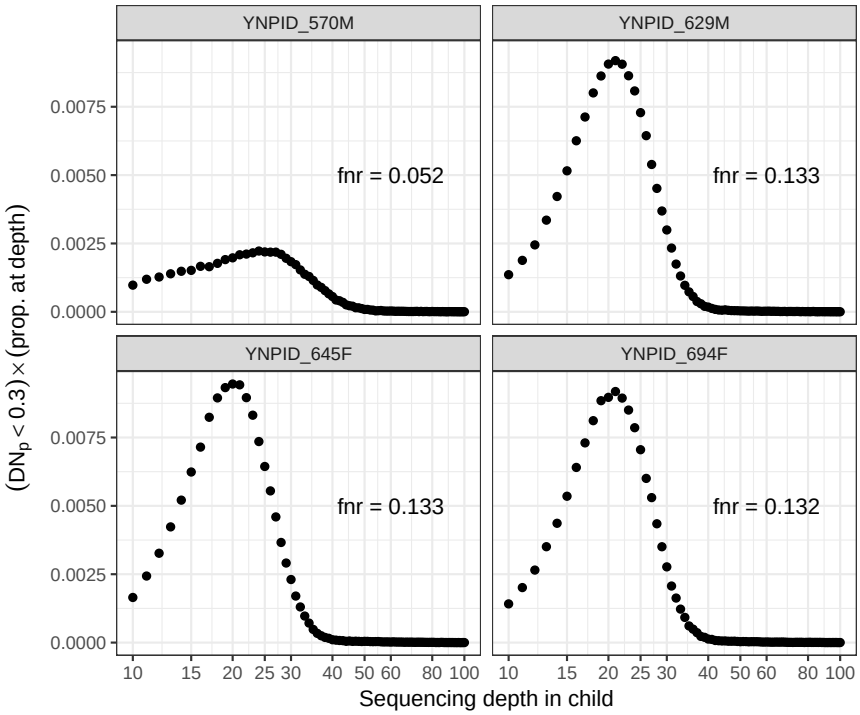


Figure 4.5: **False negative rates in each trio broken up into contributions from different sequencing depths in the child.** False negative rates were estimated for each possible sequencing depth in each child and these were multiplied by the fraction of sites in the child with that depth of coverage. This provides the contribution from each sequencing depth to the overall false negative rate at sites passing all filters. The overall false negative rate is the sum of these points.

depth increases because the possibility of lower coverage in the parents means that the sites must be considered as missed parental heterozygotes. The false negative rate increases for higher sequencing depths likely because higher depths are enriched for mismapped reads that appear as low-frequency alternative alleles within the reads in the offspring. Low-frequency alleles are more likely to have lower DN_p scores because they are more likely to be sequencing errors. However, because the fraction of sites with high enough read depths to elevate the false negative rate are so low, they do not contribute much to the overall false negative rate (Figure 4.5).

The false negative rates shown in Figure 4.5 for each trio were used in two different statistical models for calculating the posterior probability distribution of the per-generation

mutation rate. In the first model all parents were assumed to have the same mutation rate, and in the second model the mutation rate was allowed to depend on the father's age. Posterior distributions under maximum and minimum possible DNM counts are shown in Figure 4.6A. Under the minimum DNM count the posterior mean mutation rate is about 3×10^{-9} per generation and under the maximum count it is about 5×10^{-9} per generation. In both cases the mutation rate we would estimate if only 570M were used would fall far into the tail of the posterior distribution calculated using all offspring at once.

A possible explanation for the excess DNMs observed in 570M is that there is a paternal age effect on the mutation rate and this individual may have had an older father than the other offspring. Approximate birth dates of the wolves in our pedigree are shown in Figure 4.7. Although 570M was born as part of the same litter as 629M and 694M, it was sired by 302M, an older wolf than 480M who sired 629M and 694M. 645F was also sired by 480M but in a litter the subsequent year. Unfortunately, the birth year of the 480M is only known to be either 2001 or 2002. If 480M was born in 2002 it would have been younger at the time its offspring were born than 302M. We fit a simple model for a potential paternal age effect by letting the mutation rate have a linear relationship to the age of the father at birth. Similarly to our uncertainty in the number of mutations, we fit models using both 2001 and 2002 as the birth year of 480M. Fitting this model provides some evidence for a positive relationship of mutation rate with paternal age (Figure 4.8). Under this model each paternal year adds about 5×10^{-10} to the per-generation mutation rate. Samples from the posterior distribution on the slope and intercept of this relationship can be used to generate a posterior distribution on the per generation mutation rate provided the average paternal age is known. The generation time in wolves is usually assumed to be three years, but no justification is given for this value. The average age of mothers in wolves has measured around 4.3 years (Stahler *et al.*, 2013). Plotting samples from the posterior linear model yields the distribution shown in Figure 4.6B.

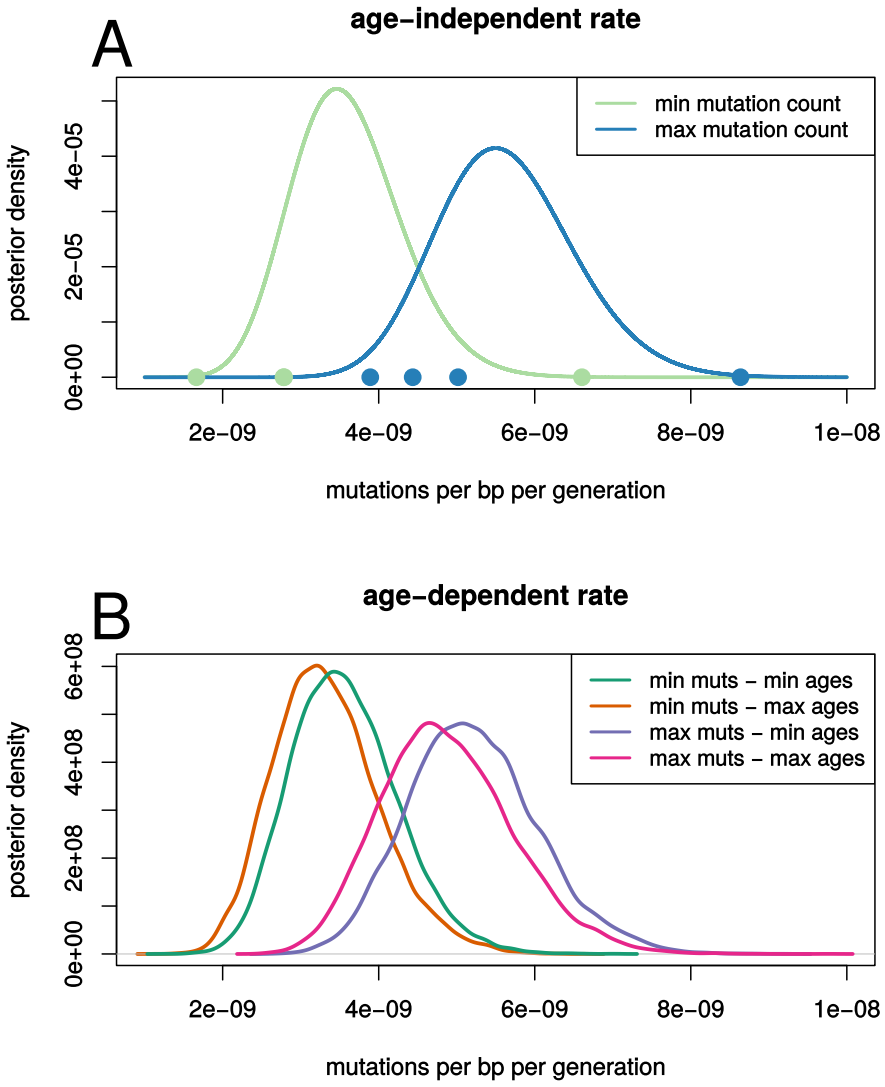


Figure 4.6: **Posterior distributions on the mutation rate under models with and without a paternal age effect.** (**A**): Posterior distribution on the per generation mutation rate for a model without paternal age dependence. The curve corresponding to the minimum mutation count is the posterior distribution we would calculate if no further DNMs are confirmed by independent sequencing. The curve corresponding to the maximum mutation count is the posterior distribution we would calculate if all potential DNMs are confirmed by independent sequencing. Dots beneath the curves show the point estimates of the mutation rates that would be obtained from the different trios. (**B**): Posterior distribution on the per generation mutation rate for a model with paternal age dependence. Since the age of the father with YNPID 480M is not known we calculate posterior distributions for the maximum and minimum age of this individual. To convert to a per-generation mutation rate we assume an average paternal age of 4.3 years.

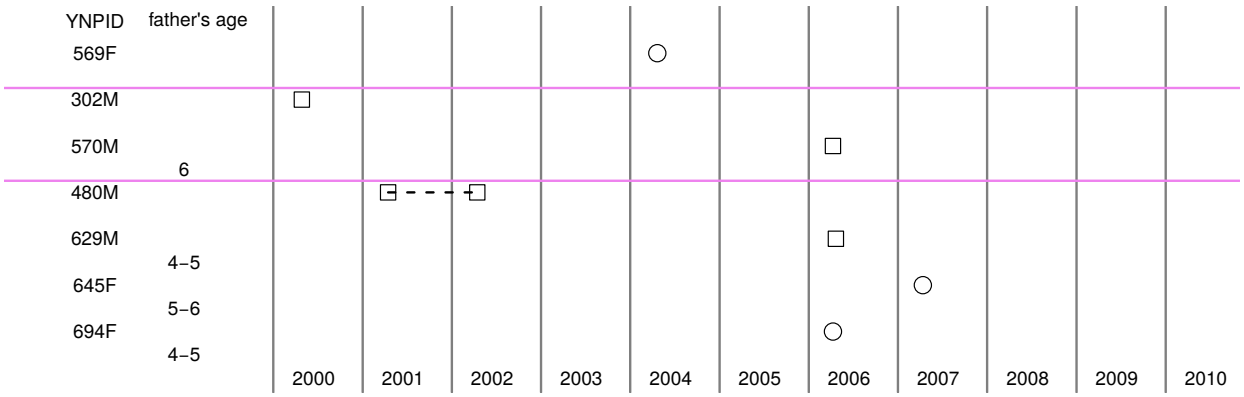


Figure 4.7: **Birth dates and paternal ages of wolves in the pedigree.** Exact dates of birth are not known, but litters are born in April of each year. 480M was born in either 2001 or 2002.

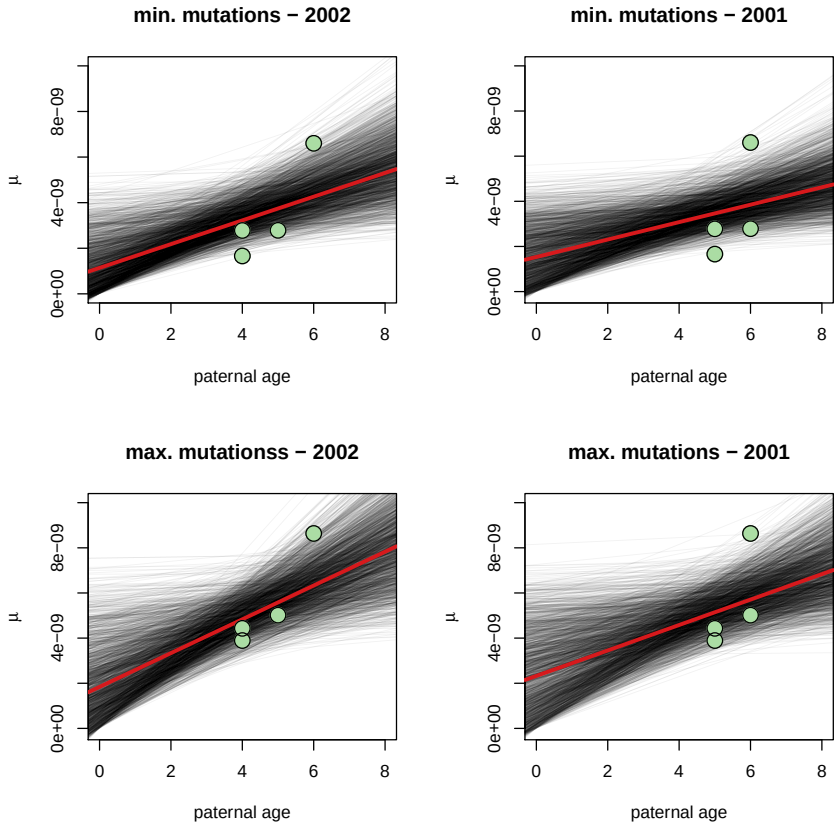


Figure 4.8: **Samples from the posterior distribution of the relationship of mutation rate with paternal age.** The four combinations capture the uncertainty in the number of potential DNMs that will ultimately validate as well as in the age of 480M.

4.5 Discussion

Mutation rates are necessary to scale times in population genetic models from units of mutations to units of generations. When combined with a generation time things can then be scaled in units of years. The divergence time between dog and wolf populations provides a strong upper bound on the timing of dog domestication. It is therefore necessary to scale population genetic models of canine history by the correct mutation rate. Although previous estimates of the mutation rate are available based on fossil calibrated mammal phylogenies (Kumar and Subramanian, 2002) and based on ancient DNA from wolves and dogs (Skoglund *et al.*, 2015; Frantz *et al.*, 2016), we provide the first direct estimate based on sequencing parents and offspring. Although not all putative mutations have yet been tested with independent sequencing, we can currently bound the mutation rate in wolves to $2.0 - 7.0 \times 10^{-9}$ per generation with the posterior mean likely to fall around 4.0×10^{-9} once all putative DNMs have been tested. Crucially, this excludes 1×10^{-8} which had been used by studies prior to the estimates from ancient DNA (Lindblad-Toh *et al.*, 2005; Skoglund *et al.*, 2011; Freedman *et al.*, 2014). Our estimate aligns closely with rates estimated using ancient DNA. It may be that the upward bias in the mutation rate from underestimating the population divergence time is canceled out somewhat by the downward bias in the mutation rate from the assumption of no post-divergence gene flow.

Our estimate of the mutation rate, like those from ancient DNA studies, firmly puts the estimated divergence between dog and wolf populations before the first appearance of dogs in the fossil record around 15,000 years ago (Pionnier-Capitan *et al.*, 2011). On phylogenetic time scales changes in the generation time and other life-history traits affect substitution rates (Wu and Li, 1985; Sayres *et al.*, 2011; Moorjani *et al.*, 2016). It is possible that changes in the generation time associated with dog domestication would lead to biased estimates of divergence times when scaling genealogies by our estimated mutation rate. However, given the relatively recent history of dog domestication this is not likely to

have a large effect.

Based on an analysis using the multiple sequentially Markovian coalescent (MSMC) (Schiffels and Durbin, 2014) and a mutation rate of 4.0×10^{-9}, Frantz *et al.* (2016) estimated a divergence time between East Asian and Western Eurasian dogs around 14,000-6,400 years ago. They use the fact that this predates the earliest findings of dogs in Europe to argue that domestication occurred independently in Europe and Asia and that Western Eurasian dogs were largely replaced by those with East Asian origin. Although our analysis also indicates a mutation rate around 4.0×10^{-9}, we cannot yet rule out a mutation rate a factor of two lower and hence an older divergence between East Asian and Western Eurasian dogs. However, because the MSMC-based divergence time is estimated by eyeballing a graph of the ratio of the within- and between-population coalescence rates the divergence time estimated by Frantz *et al.* (2016) could equally be questioned.

The biggest caveat to the work presented here is the approximately twofold greater number of mutations observed in 570M relative to the offspring of 480M. 570M did not have a greater amount of sequence passing filters than the other offspring (Figure 4.3). Additionally, although the estimated false negative rate was lower for 570M, we would only expect this to increase the number of mutations found by about ten percent. It is possible that the overall false negative rate in the other offspring was underestimated, but it would have to be around 50% to explain the observed number of mutations. One possibility we considered was that more mutations might accumulate with paternal age. However, the birth of 480M is only known to have occurred in either April of 2001 or 2002. If the earlier year is correct there is at most one year's age differences between the offspring of 570M and 480M. Although there is good evidence for a paternal age effect on the mutation rate in humans, it takes 40 years for the mutation rate to double (Kong *et al.*, 2012; Rahbari *et al.*, 2016). Another possible explanation is that there is simply variation in the mutation rate among individuals. With this possibility in mind, the mutation rate estimated here may be quite crude given

that only three parents were sampled. Finally, it is possible that some number of the excess DNMs identified in 570M are actually somatic mutations. Without seeing if any of these were transmitted to a future generation we cannot be sure that the mutations identified here occurred in the parents' germ line.

To see where our estimate of the mutation rate falls with respect to the prediction of the drift-barrier hypothesis we also have to estimate the long-term effective population size in wolves. Heterozygosity in most wolf populations ranges between 0.001 and 0.002 differences per base pair. Using a mutation rate of 4.0×10^{-9} per generation this gives an effective population size between $62{,}500 - 125{,}000$. Crudely, this implies the mutation rate in wolves falls below the regression line from other studies of direct mutation rates (Smeds *et al.*, 2016; Pfeifer, 2017) as a population with an effective size of 1×10^5 would have a predicted mutation rate around 0.8×10^{-9} per generation. However, there is evidence that the ancestral wolf population was actually much larger than that which exists today (Freedman *et al.*, 2014; Fan *et al.*, 2016).

Deciphering the causes of mutation rate evolution will require more mutation rate estimates. In species that not amenable to being kept in laboratories or gardens, or where the generation time is too long, pedigree sequencing will remain the only way to directly estimate mutation rates. An unfortunate step in this process, as it currently exits, is the manual examination of alignments at putative DNMs. Filtering is necessary to limit the number of alignment plots that must be examined. The development of computational methods to eliminate the need for manual inspection would make it much easier to estimate mutation rates, especially in non-model organisms lacking highly curated repeat libraries.

In a study of variation in substitution rates on a mammalian phylogeny Sayres *et al.* (2011) found a significant negative correlation between the substitution rate and lifespan. If lifespan is positively related to the generation time then it is possible that changes in generation time explain variation in the substitution rate. However, the per generation

mutation rate itself differs between mice, wolves, and humans (5.5×10^{-9}, 4.0×10^{-9}, 1.3×10^{-8} per generation). A possible explanation for differences in the per generation mutation rates is that the mutation rate increases with paternal age, which is itself correlated with generation time, owing to the number of cell divisions involved in sperm production (Wu and Li, 1985; Amster and Sella, 2016). In theory, increases in generation time can then have either positive and negatives effects on the substitution rate depending on life-history traits and the details of replication fidelity. Direct estimates of mutation rates and paternal age effects are necessary to understand how substitution rates arise from these various biological factors. Outside of primates the only other mutation rate that has been directly estimated in a mammal is from a mutation accumulation study in mice (Uchimura *et al.*, 2015). Our estimate therefore represents another step towards understanding the evolution of mutation rates, though additional mammalian studies are need.

APPENDIX A

APPENDIX: A TEMPORAL PERSPECTIVE ON THE INTERPLAY OF DEMOGRAPHY AND SELECTION ON DELETERIOUS VARIATION IN HUMANS

A.1 Out-of-Africa Demography

In Figure 2.2 and 2.3, changes in the frequency spectrum were examined starting from equilibrium, the parameters for bottleneck sizes and growth rates in these examples were chosen to match those in the OOA demography from Tennessen *et al.* (2012) which contains a bottleneck period (between events b and c) and a bottleneck+growth period (between events c and d). We next ask how well these two periods, which we examined in isolation in Figures 2.2 and 2.3, describe phases of heterozygosity change in the full OOA demography. The full demography also contains other differences; the population size doubles before the split, and the OOA bottleneck lasts only about 1,000 generations before a second bottleneck and growth event occurs (Figure 2.1).

Figure A1 shows changes in expected heterozygosity during this period for a range of s. Qualitatively, the heterozygosity dynamics seen in the isolated periods of OOA demography (Figures 2.2 and 2.3) are also seen in numerical solutions over the full trajectory. Heterozygosity decreases following the first bottleneck and temporarily undershoots its equilibrium value when selection is strong. Heterozygosity again drops after the second bottleneck but rapidly begins to recover during the following exponential growth period. It is only for very strongly deleterious variation that we see the over- and undershooting behavior that appear in the isolated bottleneck and bottleneck plus growth models. The timescale of the OOA demography is not long enough for these behaviors to occur when selection is weaker. As is clear from the lower heterozygosity of non-African populations (Yu *et al.*, 2002), the growth phase does not persist long enough for neutral variation to recover. However, heterozygosity

at strongly selected sites is predicted to recover more quickly.

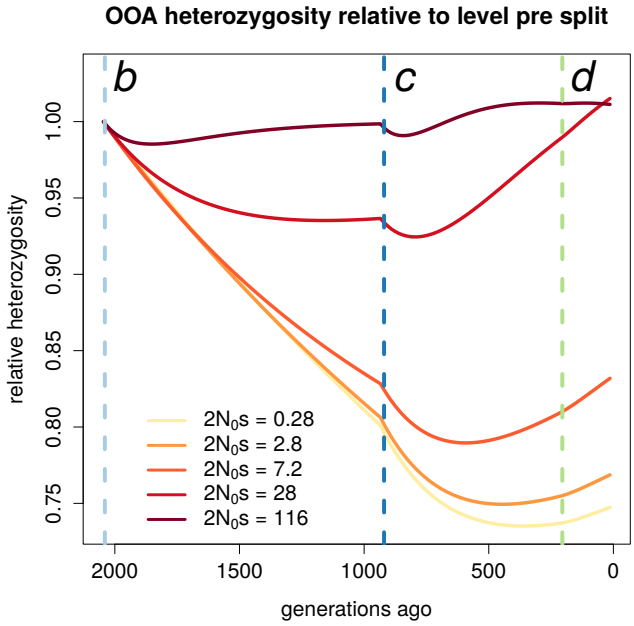


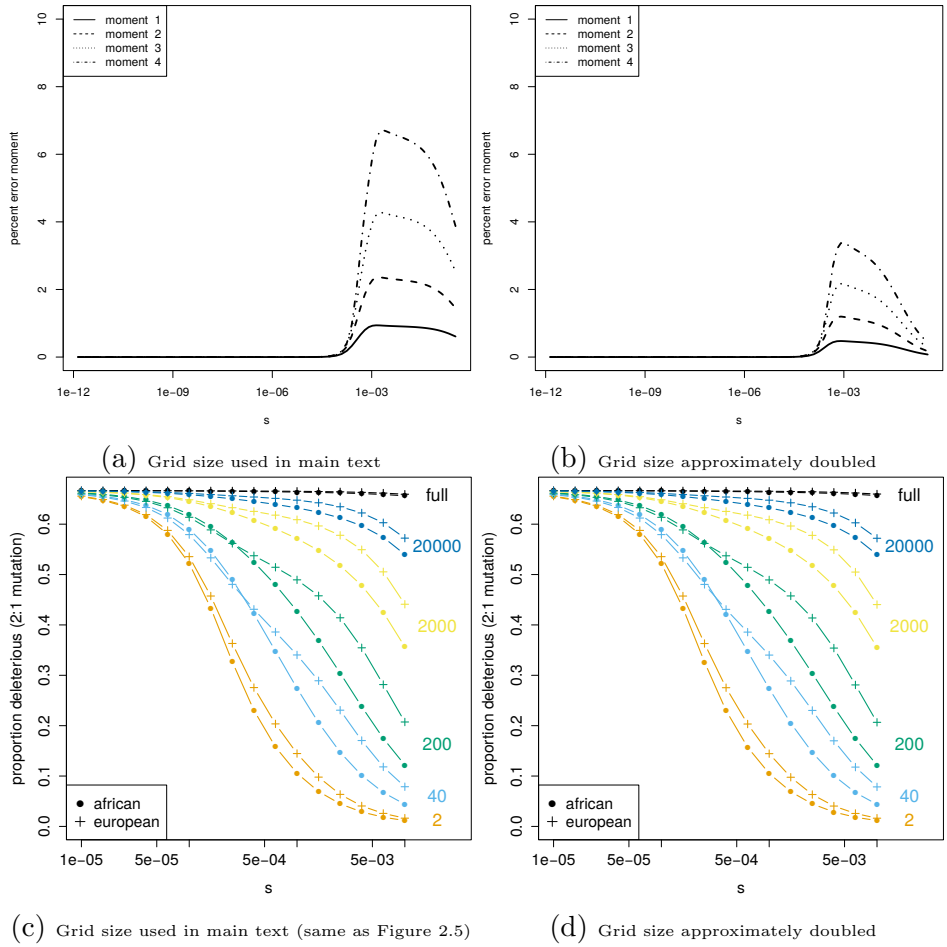
Figure A1: **The response of heterozygosity at sites under purifying selection to events following the OOA bottleneck.** The three vertical lines here correspond to events b, c, and d in Figure 2.1. N_0 corresponds to the population size preceding event b. For the strongest selection coefficients heterozygosity can be seen to undershoot and begin to increase, but for most the decrease is monotonic following b. Following c, heterozygosity only overshoots its value at mutation-selection balance and begins to decrease when selection is strongest ($2N_0s = 116$).

A.2 Evaluation of numerical precision

For the numerical analyses of equation 2.1 it was necessary to choose a grid of points on the derived allele frequency x and a time step for t. Due to the highly peaked nature of the frequency spectrum as one approaches zero it was more important to have a dense grid of values at small x than at large x (Evans *et al.*, 2007). Specifically, we required an algorithm that generates a nonuniform grid on x such that the grid density doubles at any change-point in density (Evans *et al.*, 2007). The algorithm takes a maximum step size and number of grid points after which the grid interval should double. We then search for an initial interval size such that the final grid point is $x = 1$. The grid for all figures of the main text uses an initial step size of $x_0 = 1.564 \times 10^{-10}$, a maximum step size of 10^{-3}, and doubles after 80

iterations. This resulted in a grid with 2,525 points. The t interval used was 5×10^{-4} in units of the effective population size. Lowering this time interval did not affect results.

We investigated the sensitivity of numerical solutions to the grid on x by starting with the equilibrium solution to equation 2.1 and solving this forward in time to evaluate the accumulation of numerical error. Figure A2 shows the percent error in the first four moments of the frequency spectrum for different selection coefficients after the same amount of time as in Figure 2.1. Error is greatest as one considers higher order moments of the frequency spectrum, and is peaked at an intermediate value of s. Even though error is smaller for the finer grid, the qualitative results in Figure 2.5 are unaffected (Figure A2). As P_N/P_T is influenced by higher moments of the frequency spectrum it may be more sensitive to numerical error. Results in the main text that are dependent on only the first two moments are also nearly identical.



(a) Grid size used in main text

(b) Grid size approximately doubled

(c) Grid size used in main text (same as Figure 2.5)

(d) Grid size approximately doubled

Figure A2: **Little effect of numerical errors.** Panels (a) and (b) show the accumulation of errors in the first four moments of the frequency spectrum after a time period equivalent to that in Figure 2.1, with the same initial population size, with (b) having about twice as many points as (a). Panels (c) and (d) show Figure 2.5 using the same grids as (a) and (b).

A.3 Comparison to the Wright-Fisher model

We compare a few cases of diffusion results to a Wright-Fisher (WF) model in order to check our numerical solutions. For the WF model we solve for the expected site frequency spectrum using the Markov chain approach described by Evans *et al.* (2007) with the standard Wright-Fisher transition matrix (Ewens, 2004). Figure A3 compares the evolution of heterozygosity shown in the middle line of Figure 2.2B (orange, $2N_0s = 18.3$) to the expected heterozygosity

in the WF model. The results show the same qualitative behavior and only small-scale error (¡ 0.1% difference in relative heterozygosity). Figure A4 compares the evolution of heterozygosity shown in the middle line (orange, $2N_0 = 5.9$) of Figure 2.3B to the expected heterozygosity in the WF model. It was necessary in this case to scale the population size down because the large size of the population after exponential growth makes the transition matrix very large. The models should have approximately the same behavior as long as the product of N and s is the same each generation and that time is rescaled. We again find very close agreement. We finally compare WF and diffusion results for P_N/P_T over the OOA trajectory for $s = 6.31e-4$. Figure A5 shows that the agreement between the models is very good except when the sample size is very large ($k = 20,000$). The period when agreement is poor occurs during the OOA bottleneck. At this time the effective size of the population $2N = 3,722$ is much less than the sample size, and this will create discordance between the diffusion and WF models.

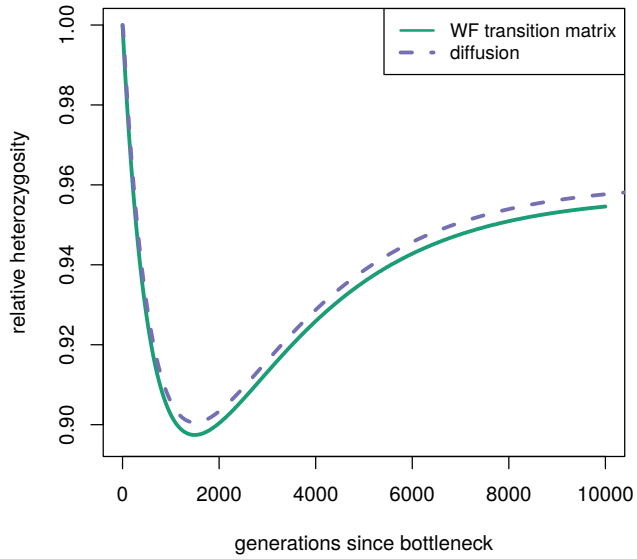


Figure A3: **Comparison of heterozygosity under the WF model and numerical solutions to the forward diffusion equation in a bottleneck scenario.** Bottleneck has the same magnitude as that in Figure 2.2.

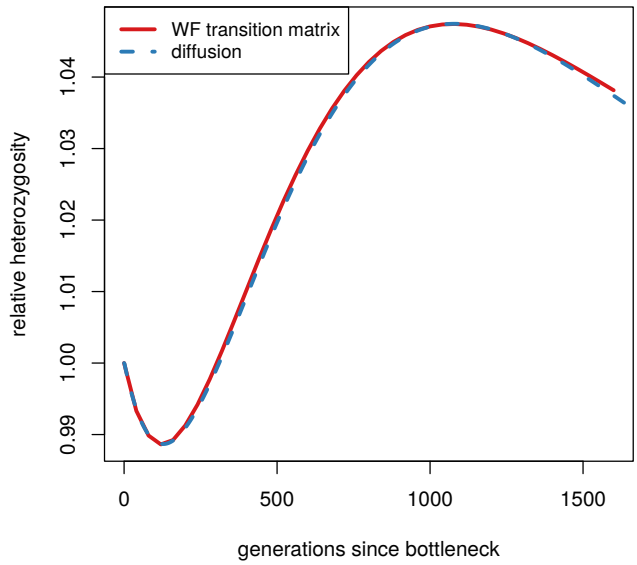


Figure A4: **Comparison of heterozygosity under the WF model and numerical solutions to the forward diffusion equation in a bottleneck scenario.** Bottleneck and growth rate parameters are the same as in Figure 2.3.

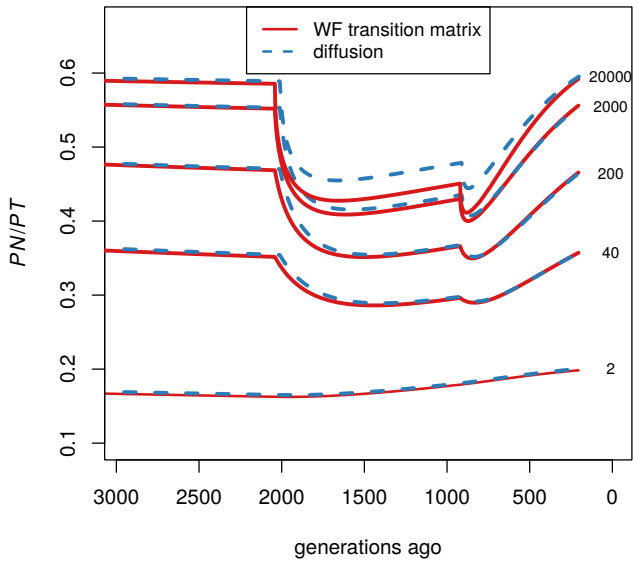


Figure A5: **Comparison of P_N/P_T in the WF model and numerical solutions to the forward diffusion equation in over the OOA trajectory.** The trajectory of effective population sizes is taken from the OOA model shown in Figure 2.1.

A.4 Sensitivity of derived allele count to quality filters

Substantial care was taken by the ExAC curators to provide high quality genotype calls (Lek *et al.*, 2016). However, we find that the difference in the derived allele count between AFR and NFE clusters in the ExAC data is sensitive to two quality measures. The first of these is the tranche level which is calculated when recalibrating variant quality scores against a training set of known variants. A tranche level of 99.6% means that variants are chosen with a log-odds of being a true variant threshold such that there is 99.6% sensitivity of true variants in the training set (DePristo *et al.*, 2011). Thus, choosing a higher tranche level means a greater number of both false positives and true variants. The second filter was applied after the tranche level had been chosen. For this we removed sites that did not successfully genotype in a certain fraction of individuals in both the African and European clusters. For both filters increasing stringency tended to decrease the excess number of derived alleles in the African cluster, and whether there is an excess of derived alleles in the African versus European cluster depends on the combination used (Figure A6). For the analysis in the main text we use a tranche level of 99.6% and cutoff of 80%.

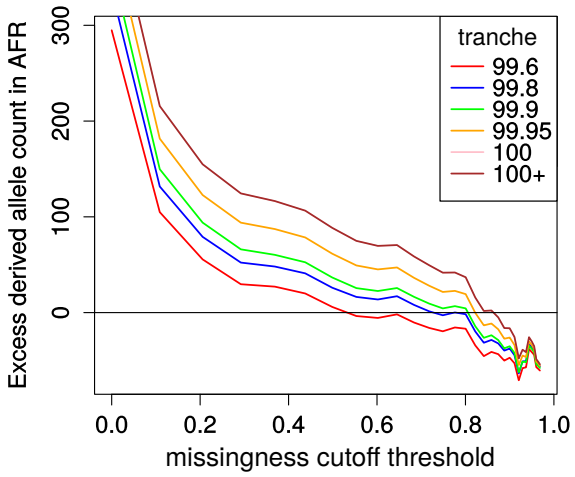


Figure A6: **The dependence of the derived allele count on sequence quality filters.** The effects of removing sites according to two quality filters on the difference in derived allele count between African and European samples. The overall difference shrinks as expected as we remove sites from consideration, and for very loose criteria on missingness (i.e. removing sites where the fraction of samples with no genotype is less than 0.8) the sign of the difference changes.

A.5 Relative differences between between African and Out-of-Africa SFS summaries

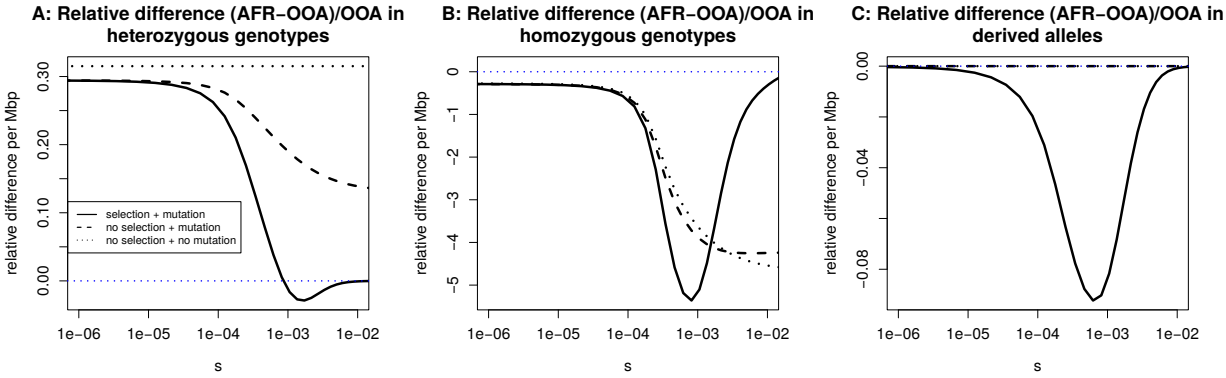


Figure A7: **Stratification of expected differences by selection coefficient, relative to value in the OOA trajectory.** The same situation as in Figure 2.6 but differences are given relative to the OOA value. We show, for a range of selection coefficients, the expected difference per Mbp between the OOA and African model, relative to the OOA value, in (A) heterozygous genotypes, (B) homozygous genotypes, and (C) derived alleles. The vertical axis gives the expected difference per Mbp per diploid genome. For derived allele count and derived allele homozygosity this includes fixations since the start of the population histories shown in Figure 2.1. *No selection + mutation* refers to numerical solutions setting $s = 0$ following the OOA bottleneck in the European trajectory. *No selection + no mutation* refers to the same, but turning off new mutations as well.

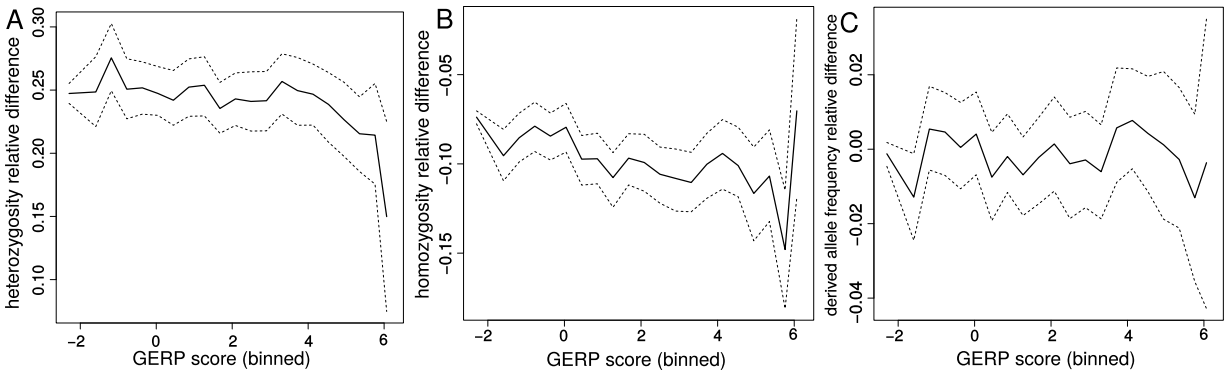


Figure A8: **Relative differences ((AFR − NFE)/NFE) in heterozygosity, homozygosity, and derived allele frequency stratified by GERP score.** The same situation as in the bottom row of Figure 2.9 but differences are given relative to the NFE value. Relative Heterozygosity (A), homozygosity (B), and derived allele frequency (C) differences for the African and non-Finnish European population groups in ExAC plotted against binned GERP scores. Dotted lines provide 95% confidence intervals obtained by bootstrapping across sites within each bin.

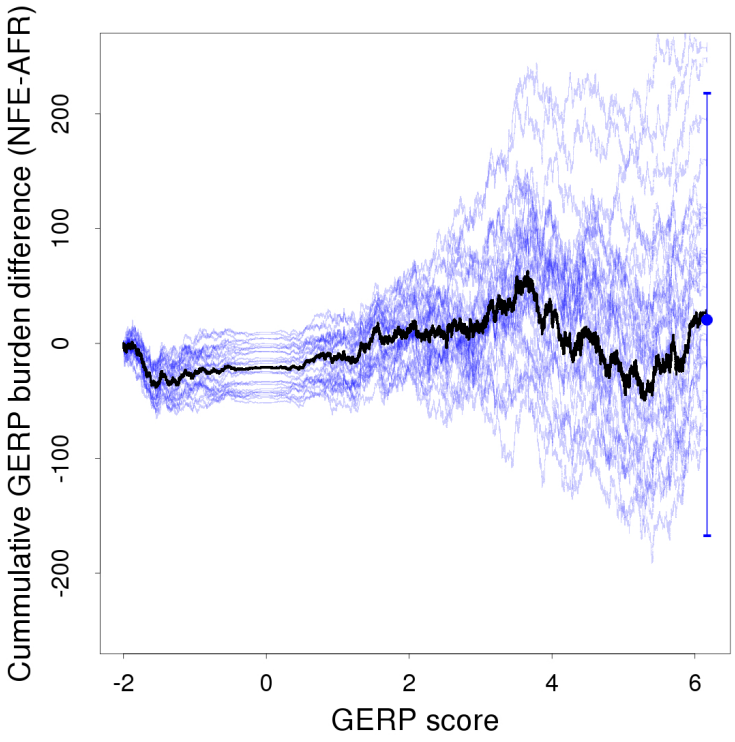


Figure A9: **Cumulative difference in GERP score burden.** The cumulative difference in the GERP score burden starting with -2. Blue lines show thirty samples bootstrapped across sites. The final blue point and bars show the mean difference in GERP burden and 95% confidence interval from 200 bootstrap replicates.

A.6 Approximating the expectation of P_N/P_T

Since the simplest prediction of deleteriousness is whether a mutation is synonymous or nonsynonymous, we write the proportion of variants that are deleterious as

$$E[P_N/P_T] \approx \frac{E[P_N^k]}{E[P_N^k] + E[P_S^k]}. \tag{A.1}$$

P_N^k and P_S^k are the expected total numbers of variants in a sample of size k that are nonsynonymous and synonymous respectively, and P_T is their sum. These correspond to polymorphism counts such as those used in a McDonald-Kreitman test McDonald and Kreitman (1991). Superscripts are dropped when considering all variants in the population. These

quantities can be computed for a given site frequency spectrum as

$$P^k(t) = \int_0^1 \left(1 - x^k - (1-x)^k\right) f(x,t)dx \tag{A.2}$$

or

$$P(t) = \int_{\frac{1}{2N}}^1 f(x,t)\mathrm{d}x \tag{A.3}$$

(Ewens, 2004) depending on whether we consider a sample of size k or the entire population. We want to be able to calculate the expectation of P_N/P_T, where

$$E[P_N/P_T] = E\left[\frac{P_N}{P_N + P_S}\right]. \tag{A.4}$$

One difficulty in calculating this value is that the random variables in the numerator and the denominator can both be zero. We first make the approximation that

$$E\left[\frac{P_N}{P_N + P_S}\right] \approx E\left[\frac{P_N}{P_N + P_S + 1}\right]. \tag{A.5}$$

Under the Poisson random field model P_N and P_S are both Poisson distributed. Writing their means as λ_N and λ_S, we can calculate

$$\begin{aligned} E\left[\frac{P_N}{P_N + P_S + 1}\right] &= \frac{\lambda_N\left[e^{-\lambda_N-\lambda_S} + \lambda_N + \lambda_S + 1\right]}{(\lambda_N + \lambda_S)^2} \\ &\approx \frac{E[P_N]}{E[P_S] + E[P_N]}. \end{aligned} \tag{A.6}$$

The final approximation works as long as P_T is large because $e^{-\lambda_N-\lambda_S}$ will be large. Since this includes neutral alleles as well as deleterious ones, the approximation should work even when selection is strong.

A.7 Equilibrium properties of P_N/P_T

Knowing that $E[P_N/P_T] \approx \frac{E[P_N]}{E[P_N]+E[P_S]}$ is a good approximation we can now ask how the forces of mutation, selection, and drift affect this value. These forces will cancel out at equilibrium, but they can still be separated out within the diffusion equation. Dropping the expectation notation and applying the chain rule we can write

$$\frac{d}{dt}\left(\frac{P_N}{P_T}\right) = \frac{P_N}{P_N+P_S}\left(\frac{P_N'}{P_N} - \frac{P_N'+P_S'}{P_N+P_S}\right). \tag{A.7}$$

Since we are assuming that only nonsynonymous mutations are selected against only the P_N' terms are affected by selection. If $f_N(x,t)$ is the frequency spectrum at nonsynonymous sites, then we can write

$$\begin{aligned} P_N' &= \frac{d}{dt}\int_{\frac{1}{2N}}^{1} f_N(x,t)dx \\ &= \int_{\frac{1}{2N}}^{1}\left(\frac{d}{dx}\left[Sx(1-x)f_N(x,t)\right] + \frac{1}{2}\frac{d^2}{dx^2}\left[x(1-x)f_N(x,t)\right]\right)dx \end{aligned}$$

The left term of this gives the instantaneous change due to selection which we write as $\left(P_N'\right)_\gamma$. The notation $()_\gamma$ is used to indicate the portion of a rate that is due to selection. This rate is negative and is balanced out by drift and selection at equilibrium.

$$\left(P_N'\right)_\gamma = \int_{\frac{1}{2N}}^{1}\frac{\partial}{\partial x}\left[Sx(1-x)f_N(x,t)\right]dx. \tag{A.8}$$

Substituting the equilibrium equation for the frequency spectrum,

$$f_N(x) = \theta_N\frac{e^{-2S}(1-e^{2S(1-x)})}{(e^{-2S}-1)x(1-x)}, \tag{A.9}$$

this integral evaluates to

$$\left(P'_N\right)_\gamma = -S\theta_N \frac{e^{-2S} - e^{-S/N}}{e^{-2S} - 1} \approx -S\theta_N \tag{A.10}$$

if selection is not too strong, and where θ_N is the population-scaled mutation rate to non-synonymous alleles. We can then calculate the equilibrium change in P_N/P_T that is due to selection by only taking the P'_N terms in equation A.7 and only considering the change in P_N that is due to selection $\left(\left(P'_N\right)_\gamma\right)$. The equilibrium decrease in P_N/P_T that is due to selection can then be written as

$$\begin{aligned}\frac{d}{dt}\left(\frac{P_N}{P_T}\right)_\gamma &= -S\theta_N \left(\frac{1}{P_S + P_N} - \frac{P_N}{(P_S + P_N)^2}\right) \\ &= -S\theta_N \left(\frac{P_S}{(P_S + P_N)^2}\right)\end{aligned} \tag{A.11}$$

This rate does not depend on θ, and we can show this by writing

$$\theta = \theta_N + \theta_S = \pi\theta + (1 - \pi)\theta, \tag{A.12}$$

where π is the proportion of mutations that are nonsynoymous, and θ_S is the population-scaled mutation rate to synonymous alleles. If $P_S := \theta_S F_S$ and $P_N := \theta_N F_N$, we can see that the rate does not depend on the population mutation rate θ by making substitutions into equation A.11.

$$\frac{d}{dt}\left(\frac{P_N}{P_T}\right)_\gamma = -S\left(\frac{\pi(1-\pi)F_S}{(\pi F_N + (1-\pi)F_S)^2}\right). \tag{A.13}$$

Although the F are the same as the P but with $\theta = 1$. When comparing this value between different population sizes, it is important to note that this is a rate per $2N$ generations, so we need to scale to generations when comparing rates.

The rate for a sample of size k

When considering the rate of change due to selection of P_N/P_T in a sample of size k, the same basic equation applies, except that we have

$$\begin{aligned}
\frac{d}{dt}\left(\frac{P_N}{P_T}\right)_\gamma^k &= \left(P_N'\right)_\gamma^k \left(\frac{1}{P_S^k + P_N^k} - \frac{P_N^k}{\left(P_S^k + P_N^k\right)^2}\right) \\
&= -\theta_N \int_0^1 \left(1 - x^k - (1-x)^k\right) \frac{2S^2 e^{-2Sx}}{1 - e^{-2S}} dx \left(\frac{P_S^k}{\left(P_N^k + P_S^k\right)^2}\right) \\
&= -\int_0^1 \left(1 - x^k - (1-x)^k\right) \frac{2S^2 e^{-2Sx}}{1 - e^{-2S}} dx \left(\frac{\pi(1-\pi)F_S^k}{\left(\pi F_N^k + (1-\pi)F_S^k\right)^2}\right). \quad \text{(A.14)}
\end{aligned}$$

This is solved by numerical integration.

APPENDIX B

APPENDIX: THE EFFECTS OF DEMOGRAPHY AND GENETIC ARCHITECTURE ON THE NEUTRAL DISTRIBUTION OF QUANTITATIVE TRAITS

B.1 A central limit theorem in the infinitesimal limit

Recall that the moment generating function for the distribution of trait values from a single locus is

$$\varphi_{\mathbf{Y}}(\mathbf{k}) = \int \exp\left(\sum_{\omega\in\Omega} s_\omega t_\omega\right) \mathrm{P}(\mathbf{T}=\mathbf{t})\mathrm{d}\mathbf{t}\Big|_{s_\omega=\frac{\theta}{2}\left(\psi\left(\sum_{a\in\omega}k_a\right)-1\right)}.$$

If we substitute in the Taylor series expansions for the moment generating function of the trait value distribution we get

$$\int \prod_{\Omega} \exp\left[t_\omega \frac{\theta}{2}\left(\sum_{n=1}^{\infty}\frac{m_n}{n!}\left(\sum_{a\in\omega}k_a\right)^n\right)\right] \mathrm{P}(\mathbf{T}=\mathbf{t})\mathrm{d}\mathbf{t}.$$

If we then write the Taylor series of each exponential function we get

$$\int \prod_{\Omega} \left[\sum_{j=0}^{\infty}\frac{t_\omega^j}{j!}\left(\frac{\theta}{2}\right)^j\left(\sum_{n=1}^{\infty}\frac{m_n}{n!}\left(\sum_{a\in\omega}k_a\right)^n\right)^j\right] \mathrm{P}(\mathbf{T}=\mathbf{t})\mathrm{d}\mathbf{t},$$

which is equivalent to

$$1+\sum_{\Omega}\mathrm{E}[T_\omega]\frac{\theta}{2}\sum_{n=1}^{\infty}\frac{m_n}{n!}\left(\sum_{a\in\omega}k_a\right)^n + \sum_{\Omega\times\Omega}\frac{1}{2}E[T_{\omega_1}T_{\omega_2}]\sum_{n=1}^{\infty}\frac{\theta}{2}\frac{m_n}{n!}\left(\sum_{a\in\omega_1}k_a\right)^n\sum_{n=1}^{\infty}\frac{\theta}{2}\frac{m_n}{n!}\left(\sum_{a\in\omega_2}k_a\right)^n+\ldots$$

This is raised to the power L for a trait controlled by L loci. We want the limit as the number of loci increases while the size of mutational decreases. This can be expressed by

the limits $Lm_1 \to \mu$, $Lm_2 \to \sigma^2$ and $Lm_i \to 0$ for $i > 2$ as $L \to \infty$. Knowing we will not be retaining m_3 and above we can rewrite the mgf as

$$\left(1 + \sum_{\Omega} \mathrm{E}[T_\omega]\frac{\theta}{2}\left(m_1 \sum_{a\in\omega} k_a + \frac{m_2}{2}\left(\sum_{a\in\omega} k_a\right)^2\right)\right)^L.$$

The result of taking these limits is

$$\exp\left(\sum_{\omega\in\Omega} \mathrm{E}[T_\omega]\frac{\theta}{2}\left(\mu \sum_{a\in\omega} k_a + \frac{\sigma^2}{2}\left(\sum_{a\in\omega} k_a\right)^2\right)\right). \tag{B.1}$$

This is multivariate normal distribution with mean equal to $\mathrm{E}[T_{MRCA}]\frac{\theta}{2}\mu$, variance equal to $\mathrm{E}[T_{MRCA}]\frac{\theta}{2}\sigma^2$, and covariance between Y_a and Y_b equal to $\mathrm{E}[\tau_{a+b}]\frac{\theta}{2}\sigma^2$. This can be seen from equation (B.1) by noting that in the mgf for a multivariate normal distribution the coefficient in the exponential of k_a is the mean of Y_a and the coefficient of $k_a k_b$ is $2\mathrm{Cov}[Y_a, Y_b]$ if $a \neq b$ and $\mathrm{Var}[Y_a]$ if $a = b$.

These Y values are not directly observed because in the theory presented here they are measured as the sum of differences since the T_{MRCA}s at the causal loci affecting the trait. Rather, differences between individual trait values are what analyses would be based on. Since the Y are normally distributed differences between trait values such as $Y_a - Y_b$ will be as well.

$$\begin{aligned}\mathrm{Var}[Y_a - Y_b] &= Var[Y_a] + \mathrm{Var}[Y_b] - 2\mathrm{Cov}[Y_a, Y_b] \\ &= 2(\mathrm{E}[T_{MRCA}] - \mathrm{E}[\tau_{a+b}])\sigma^2.\end{aligned}$$

This also tells us that the distribution of trait differences will be the same in the infinitesimal limit with and without a low mutation rate approximation.

A much simpler heuristic derivation of the limiting normal distribution can be done by

calculating the variance and covariance at a single locus. This derivation is very similar to that done by Schraiber and Akey (2015). Using the law of total variance we can write

$$\mathrm{Var}[Y] = \mathrm{E}\left[\mathrm{Var}[Y|T]\right] + \mathrm{Var}\left[\mathrm{E}[Y|T]\right]$$

The variance conditional on T can be calculated again using the law of total variance and conditioning on the number of mutation at the locus.

$$\begin{aligned}
\mathrm{Var}[Y|T] &= \mathrm{E}[\mathrm{Var}[Y|M]|T] + \mathrm{Var}[E[Y|M]|T] \\
&= \mathrm{E}[M(m_2 - m_1^2)|T] + \mathrm{Var}[Mm_1|T] \\
&= \frac{\theta}{2}T(m_2 - m_1^2) + \frac{\theta}{2}Tm_1^2 \\
&= \frac{\theta}{2}Tm_2
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[Y|T] &= \frac{\theta}{2}Tm_1 \\
\mathrm{Var}[\frac{\theta}{2}Tm_1] &= \left(\frac{\theta}{2}m_1\right)^2 \mathrm{Var}[T]
\end{aligned}$$

Therefore we have

$$\mathrm{Var}[T] = \frac{\theta}{2}m_2\mathrm{E}[T_{MRCA}] + (\frac{\theta}{2}m_1)^2\mathrm{Var}[T_{MRCA}]. \tag{B.2}$$

The same procedure can be done for the covariance.

$$\mathrm{Cov}[Y_a, Y_b] = \mathrm{E}[\mathrm{Cov}[Y_a, Y_b|T]] + \mathrm{Cov}[E[Y_a|T], E[Y_b|T]]$$

We can break Y_a and Y_b into a shared part, Y_S and unshared parts for each, $Y_{\delta a}$ and $Y_{\delta b}$.

$$\mathrm{Cov}[Y_S + Y_{\delta a}, Y_S + Y_{\delta b}|T] = \mathrm{Var}[Y_s|T] = \frac{\theta}{2}\tau_{a+b}m_2.$$
$$E[\frac{\theta}{2}\tau_{a+b}m_2] = \frac{\theta}{2}m_2\mathrm{E}[\tau_{a+b}]$$

$$\mathrm{Cov}[E[Y_a|T], E[Y_b|T]] = \mathrm{Cov}[\frac{\theta}{2}Tm_1, \frac{\theta}{2}Tm_1] = \left(\frac{\theta}{2}m_1\right)^2 \mathrm{Var}[T_{MRCA}]$$

Therefore we have

$$\mathrm{Cov}[Y_a, Y_b] = \frac{\theta}{2}m_2\mathrm{E}[\tau_{a+b}] + \left(\frac{\theta}{2}m_1\right)^2 \mathrm{Var}[T_{MRCA}]. \tag{B.3}$$

The terms proportional to the variance of the T_{MRCA} disappear because $Lm_1^2 \to 0$ as $L \to \infty$.

B.2 Automatic moment derivations

Moments of the trait distribution can be calculated by differentiating equation (3.3). To automate the process, symbolic math programs were written using sympy (Meurer *et al.*, 2017). To derive moments for an arbitrary distribution of coalescent times we take Taylor series of the mgf of the mutational distribution to get

$$\left(\int \prod_{\Omega} \exp\left[t_\omega \frac{\theta}{2}\left(\sum_{n=1}^{\infty}\frac{m_n}{n!}\left(\sum_{a\in\omega}k_a\right)^n\right)\right]\mathrm{P}(\mathbf{T}=\mathbf{t})\mathrm{d}\mathbf{t}\right)^L. \tag{B.4}$$

We then also Taylor expand the exponential function appearing in this integral to get

$$\left(1 + \sum_{\Omega}\mathrm{E}[T_\omega]\frac{\theta}{2}\sum_{n=1}^{\infty}\frac{m_n}{n!}\left(\sum_{a\in\omega}k_a\right)^n + \frac{1}{2}\mathrm{E}\left[\left(\sum_{\Omega}T_\omega\frac{\theta}{2}\sum_{n=1}^{\infty}\frac{m_n}{n!}\left(\sum_{a\in\omega}k_a\right)^n\right)^2\right] + \frac{1}{6}\mathrm{E}\left[\left(\sum_{\Omega}T_\omega\frac{\theta}{2}\sum_{n=1}^{\infty}\frac{m_n}{n!}\left(\sum_{a\in\omega}k_a\right)^n\right)^3\right] + \ldots\right)^L. \tag{B.5}$$

Of course, the infinite sums appearing in this expression pose a problem. To calculate a particular moment, we only consider terms that will contribute terms of that moment's order when equation (B.5) is differentiated and zero is substituted for all dummy variables. For example, to calculate $E[Y_a Y_b Y_c]$ we only want terms that contain an order three product of k when the polynomial in equation (B.5) is expanded. For a moment of order m, we only need consider terms of size m and less in the series expansion of the exponential, and within the expansions of the mutational mgf we only need consider terms up to order $m - m_t + 1$ where m_t is the order of the exponential term being considered.

B.3 Kurtosis simulations

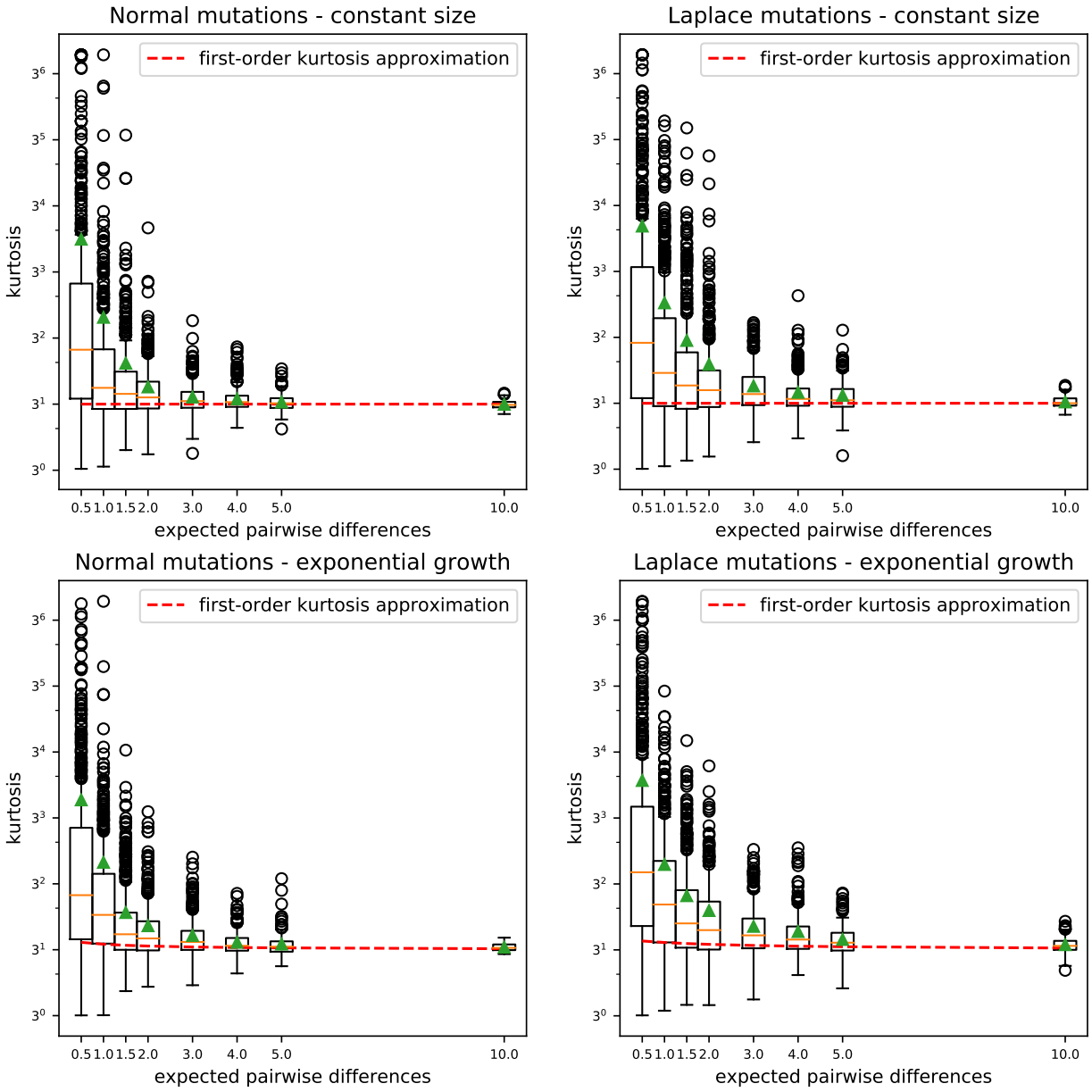


Figure B1: **The distribution of the population kurtosis under different genetic architectures, mutational kernels, and demographies.** The genetic architecture is varied by changing the expected number of pairwise differences at sites affecting the trait. Normal and Laplace distributions of mutational effects are compared. A constant size population is compared to an exponential growth scenario with growth rate equal to the reciprocal of the final effective population size. Green triangles denote the mean kurtosis. The dashed red lines give the first order approximation to the expected kurtosis given in equation B.6.

Entire populations were simulated using msprime (Kelleher *et al.*, 2015) and mutations

were assigned effects from a zero-centered normal or Laplace distribution. The effective population size and mutation rate were kept constant and the expected number of pairwise difference was increased by increasing the number of loci affecting the trait.

A first order approximation to the kurtosis is

$$\kappa_Y^{(1)} = 3 + \frac{\kappa_M \frac{\mathrm{E}[\mathbb{T}_{4,4}] - \frac{1}{6}\mathrm{E}[\mathbb{T}_{3,4}] - \frac{1}{9}\mathrm{E}[\mathbb{T}_{2,4}]}{\mathrm{E}[\mathbb{T}_{2.2}]}}{L\frac{\theta}{2}\mathrm{E}[\mathbb{T}_{2.2}] + \kappa_M \frac{\frac{1}{9}\mathrm{E}[\mathbb{T}_{2,4}] + \frac{1}{6}\mathrm{E}[\mathbb{T}_{3,4}]}{\mathrm{E}[\mathbb{T}_{2.2}]}}. \tag{B.6}$$

Although this expression suggests that the expected κ_Y will be greater than under normality when external branches are longer ($\mathrm{E}[\mathbb{T}_{4,4}] > \frac{1}{6}\mathrm{E}[\mathbb{T}_{3,4}] + \frac{1}{9}\mathrm{E}[\mathbb{T}_{2,4}]$) and less than under normality when they are longer ($\mathrm{E}[\mathbb{T}_{4,4}] < \frac{1}{6}\mathrm{E}[\mathbb{T}_{3,4}] + \frac{1}{9}\mathrm{E}[\mathbb{T}_{2,4}]$), simulations show that the approximation is actually quite poor (Figure B1).

B.4 Central moment derivations

We can use the low mutation rate approximation to the moment generating function to calculate moments of the distribution of trait vales. We will start by calculating the first and second moments. We start, as we did in deriving the normal distribution, by substituting the Taylor series of the mutational mgf.

$$\varphi_{\mathbf{Y}}(\mathbf{k}) \approx \left[1 + \sum_{\omega \in \Omega} \mathrm{E}[T_\omega] \frac{\theta}{2} \left(m_1 \sum_{a \in \omega} k_a + \frac{m_2}{2!} \left(\sum_{a \in \omega} k_a \right)^2 + \frac{m_3}{3!} \left(\sum_{a \in \omega} k_a \right)^3 + \frac{m_4}{4!} \left(\sum_{a \in \omega} k_a \right)^4 \ldots \right) \right]^L \tag{B.7}$$

We can expand this out using multinomial coefficients to get

$$\begin{aligned}\varphi_{\mathbf{Y}}(\mathbf{k}) \approx 1 + L\frac{\theta}{2}\sum_{\omega\in\mathcal{O}}\mathrm{E}[T_\omega]\left(m_1\sum_{a\in\omega}k_a + \frac{m_2}{2}\left(\sum_{a\in\omega}k_a\right)^2 + \dots\right)\\ + \frac{L(L-1)}{2}\left(\frac{\theta}{2}\right)^2\sum_{\omega\in\Omega}E[T_\omega]^2\left(m_1\sum_{a\in\omega}k_a + \frac{m_2}{2}\left(\sum_{a\in\omega}k_a\right)^2 + \dots\right)^2\\ + L(L-1)\left(\frac{\theta}{2}\right)^2\sum_{\omega_1,\omega_2\in\Omega}\mathrm{E}[T_{\omega_1}]\mathrm{E}[T_{\omega_2}]\left(m_1\sum_{a\in\omega_1}k_a + \dots\right)\left(m_1\sum_{a\in\omega_2}k_a + \dots\right) + \dots.\end{aligned} \tag{B.8}$$

The first coefficient is $\binom{L}{L-1,1,\mathbf{0}}$, the second is $\binom{L}{L-2,2,\mathbf{0}}$, and the third is $\binom{L}{L-2,1,1,\mathbf{0}}$. To calculate the moments of this distribution one takes the partial derivatives of the mgf and sets the dummy variables to zero.

$$\mathrm{E}[Y_1^{r_1}\dots Y_n^{r_n}] = \frac{\partial^{r_1+\dots+r_n}}{\partial k_1^{r_1}\dots\partial k_n^{r_n}}\varphi_{\mathbf{Y}}(\mathbf{k})\Big|_{\mathbf{k}=0} \tag{B.9}$$

Using this to calculate the first moment of the trait distribution we get

$$\mathrm{E}[Y_a] \approx L\frac{\theta}{2}m_1\sum_{\omega\in\Omega_a}\mathrm{E}[T_\omega]. \tag{B.10}$$

The second moment is more complicated because there are k_a^2 terms in all three lines of equation B.8.

$$\begin{aligned}\mathrm{E}[Y_a^2] \approx L\frac{\theta}{2}m_2\sum_{\omega\in\Omega_a}E[t_\omega]\\ + \frac{L(L-1)}{2}\left(\frac{\theta}{2}\right)^2 m_1^2\sum_{\omega\in\Omega_a}2\mathrm{E}[T_\omega]^2\\ + L(L-1)\left(\frac{\theta}{2}\right)^2 m_1^2\sum_{\omega_1,\omega_2\in\Omega_{a+b}}2\mathrm{E}[T_{\omega_1}]E[T_{\omega_2}]\end{aligned} \tag{B.11}$$

Terms with $(\frac{\theta}{2})^2$ are kept because they also include a second order term of L in front of them. We can now calculated the variance using $\mathrm{Var}[Y] = \mathrm{E}[Y^2] - \mathrm{E}[Y]^2$. The squared first moment can be written as

$$\left(L\frac{\theta}{2} m_1 \sum_{\omega \in \Omega_a} \mathrm{E}[t_\omega] \right)^2 = L^2 \left(\frac{\theta}{2}\right)^2 m_1^2 \sum_{\omega \in \Omega_a} \mathrm{E}[T_\omega]^2 + L^2 \left(\frac{\theta}{2}\right)^2 m_1^2 \sum_{\omega_1, \omega_2 \in \Omega_{a+b}} \mathrm{E}[T_{\omega_1}]\mathrm{E}[T_{\omega_2}]. \tag{B.12}$$

Subtracting this from the second moment gives

$$\begin{aligned}
\mathrm{Var}[Y_a] &\approx L\frac{\theta}{2} m_2 \sum_{\omega \in \Omega_a} \mathrm{E}[T_\omega] \\
&\quad - L \left(\frac{\theta}{2}\right)^2 m_1^2 \sum_{\omega \in \Omega_a} \mathrm{E}[T_\omega]^2 \\
&\quad - 2L \left(\frac{\theta}{2}\right)^2 m_1^2 \sum_{\omega_1, \omega_2 \in \Omega_{a+b}} \mathrm{E}[T_{\omega_1}]\mathrm{E}[T_{\omega_2}] \\
&= L\frac{\theta}{2} m_2 \sum_{\omega \in \Omega_a} \mathrm{E}[T_\omega] - L \left(\frac{\theta}{2} m_1 \sum_{\omega \in \Omega_a} \mathrm{E}[T_\omega] \right)^2 \\
&= L\frac{\theta}{2} m_2 \mathrm{E}[T_{MRCA}] - L \left(\frac{\theta}{2} m_1 \mathrm{E}[T_{MRCA}] \right)^2 \\
&\approx L\frac{\theta}{2} m_2 \mathrm{E}[T_{MRCA}].
\end{aligned} \tag{B.13}$$

Due to the large number of terms we only derive the fourth moment of the trait value distribution for the case when the mean mutational effect is zero. The terms of (B.7) appearing in the fourth moment after we apply (B.9) are

$$L \left(\frac{\theta}{2}\right) \frac{m_4}{24} \sum_{\omega \in \Omega_a} \mathrm{E}[T_\omega] \left(\sum_{a \in \omega} k_a \right)^4$$

for the fourth moment along one branch,

$$\binom{L}{L-2,2,\mathbf{0}}\left(\frac{\theta}{2}\right)^2\left(\frac{m_2}{2}\right)^2 24\sum_{\omega\in\Omega_a}\mathrm{E}[T_\omega]^2\left(\sum_{a\in\omega}k_a\right)^4$$

for the second moment of the same branch chosen twice, and

$$\binom{L}{L-2,1,1,\mathbf{0}}\left(\frac{\theta}{2}\right)^2\left(\frac{m_2}{2}\right)^2 24\sum_{\omega_1,\omega_2\in\Omega_{a+b}}\mathrm{E}[T_{\omega_1}]\mathrm{E}[T_{\omega_2}]\left(\sum_{a\in\omega_1}k_a\right)^2\left(\sum_{a\in\omega_2}k_a\right)^2$$

for the second moments on two different branches. Taking the fourth derivatives of these in terms of the desired branch we get

$$\begin{aligned}
\mathrm{E}[Y_a^4] &= L\frac{\theta}{2}m_4\mathrm{E}[T_{MRCA}] \\
&\quad + \frac{L(L-1)}{2}\left(\frac{\theta}{2}\right)^2\left(\frac{m_2}{2}\right)^2 24\sum_{\omega\in\Omega_a}\mathrm{E}[t_\omega]^2 \\
&\quad + L(L-1)\left(\frac{\theta}{2}\right)^2\left(\frac{m_2}{2}\right)^2 24\sum_{\omega_1,\omega_2\in\Omega_{a+b}}\mathrm{E}[T_{\omega_1}]\mathrm{E}[T_{\omega_2}] \\
&= L\frac{\theta}{2}m_4\mathrm{E}[T_{MRCA}] + 3L(L-1)\left(\frac{\theta}{2}m_2\sum_{\omega\in\Omega_a}\mathrm{E}[T_\omega]\right)^2 x \\
&= L\frac{\theta}{2}m_4\mathrm{E}[T_{MRCA}] + 3L(L-1)\left(\frac{\theta}{2}m_2\mathrm{E}[T_{MRCA}]\right)^2 \\
&\approx L\frac{\theta}{2}m_4\mathrm{E}[T_{MRCA}] + 3\left(L\frac{\theta}{2}m_2\mathrm{E}[T_{MRCA}]\right)^2.
\end{aligned} \tag{B.14}$$

The variance and the fourth moment derived in equation (B.13) and (B.14) can be used to derive the kurtosis of Y_a. The kurtosis of a random variable X is defined as

$$\mathrm{Kurt}[X] = \frac{\mathrm{E}[(X-\mathrm{E}[X])^4]}{\mathrm{E}[(X-\mathrm{E}[X])]^2]^2}.$$

This is the fourth central moment divided by the variance. It is possible to calculate the

kurtosis of a single trait value over evolutionary realization. For ease of calculation, we'll examine this in the case where the mean mutation effect (and therefore trait value) is zero. If we plug (B.13) and (B.14) into the expression for the kurtosis we get

$$\begin{aligned}\text{Kurt}[Y_a] &= \frac{L\frac{\theta}{2}m_4\text{E}[T_{MRCA}]}{\left(L\frac{\theta}{2}m_2\text{E}[T_{MRCA}]\right)^2} + \frac{3L(L-1)\left(\frac{\theta}{2}m_2\text{E}[T_{MRCA}]\right)^2}{\left(L\frac{\theta}{2}m_2\text{E}[T_{MRCA}]\right)^2} \\ &= \frac{m_4}{L\frac{\theta}{2}m_2^2\text{E}[T_{MRCA}]} + \frac{3(L^2-L)}{L^2} \\ &= \frac{\kappa}{L\frac{\theta}{2}\text{E}[T_{MRCA}]} + 3\left(1-\frac{1}{L}\right).\end{aligned}$$

We also calculate some additional moments having less clear interpretations but are useful later on when calculating the expected fourth central moment in the population. The first of these is $\text{E}[Y_a^3Y_b]$. The terms of (B.7) appearing in this are

$$L\left(\frac{\theta}{2}\right)\frac{m_4}{24}4k_a^3k_b\sum_{\omega\in\Omega_{a+b}}\text{E}[T_\omega]$$

and

$$L(L-1)\left(\frac{\theta}{2}\frac{m_2}{2}\right)^2k_a^2\times 2k_ak_b\left(\sum_{\omega\in\Omega_{a+b}}\text{E}[T_\omega]\right)\left(\sum_{\omega\in\Omega_a}\text{E}[T_\omega]\right).$$

This ultimately gives

$$E[Y_a^3Y_b] = L\frac{\theta}{2}m_4\text{E}[\tau_{a+b}] + 3L(L-1)\left(\frac{\theta}{2}m_2\right)^2\text{E}[T_{MRCA}]\text{E}[\tau_{a+b}]. \tag{B.15}$$

The next fourth moment of interest is $\text{E}[Y_a^2Y_bY_c]$. The terms of (B.7) are

$$L\frac{\theta}{2}\frac{m_4}{24}12k_a^2k_bk_c\sum_{\omega\in\Omega_{a+b+c}}\text{E}[T_\omega],$$

$$L(L-1)\left(\frac{\theta}{2}\frac{m_2}{2}\right)^2 k_a^2 \times 2k_b k_c \left(\sum_{\omega\in\Omega_a} \mathrm{E}[T_\omega]\right)\left(\sum_{\omega\in\Omega_{a+b}} \mathrm{E}[T_\omega]\right),$$

and

$$L(L-1)\left(\frac{\theta}{2}\frac{m_2}{2}\right)^2 2k_a k_b \times 2k_a k_c \left(\sum_{\omega\in\Omega_{a+b}} \mathrm{E}[T_\omega]\right)\left(\sum_{\omega\in\omega_{b+c}} \mathrm{E}[T_\omega]\right).$$

Taking the appropriate derivatives of these gives

$$L\frac{\theta}{2}m_4\mathrm{E}[\tau_{a+b+c}]+L(L-1)\left(\frac{\theta}{2}m_2\right)^2\mathrm{E}[T_{MRCA}]\mathrm{E}[\tau_{b+c}]+2L(L-1)\left(\frac{\theta}{2}m_2\right)^2\mathrm{E}[\tau_{a+b}]\mathrm{E}[\tau_{a+c}]. \tag{B.16}$$

Individuals in the population are exchangeable as long as it is not structured. The pairwise expected shared branch lengths are in that case all equal and we can write (B.16) as

$$L\frac{\theta}{2}m_4\mathrm{E}[\tau_{a+b+c}] + L(L-1)\left(\frac{\theta}{2}m_2\right)^2\mathrm{E}[T_{MRCA}]\mathrm{E}[\tau_{a+b}] + 2L(L-1)\left(\frac{\theta}{2}m_2\right)^2\mathrm{E}[\tau_{a+b}]^2. \tag{B.17}$$

The final moment we'll look at is $\mathrm{E}[Y_aY_bY_cY_d]$ which has relevant terms

$$L\frac{\theta}{2}\frac{m_4}{24}24k_ak_bk_ck_d\sum_{\omega\in\Omega_{a+b+c+d}}\mathrm{E}[T_\omega],$$

$$L(L-1)\left(\frac{\theta}{2}\frac{m_2}{2}\right)^2 2k_ak_b\times 2k_ck_d\left(\sum_{\omega\in\Omega_{a+b}}\mathrm{E}[T_\omega]\right)\left(\sum_{\omega\in\Omega_{c+d}}\mathrm{E}[T_\omega]\right),$$

$$L(L-1)\left(\frac{\theta}{2}\frac{m_2}{2}\right)^2 2k_ak_c\times 2k_bk_d\left(\sum_{\omega\in\Omega_{a+c}}\mathrm{E}[T_\omega]\right)\left(\sum_{\omega\in\Omega_{b+d}}\mathrm{E}[T_\omega]\right),$$

and

$$L(L-1)\left(\frac{\theta}{2}\frac{m_2}{2}\right)^2 2k_ak_d\times 2k_bk_c\left(\sum_{\omega\in\Omega_{a+d}}\mathrm{E}[T_\omega]\right)\left(\sum_{\omega\in\omega_{b+c}}\mathrm{E}[T_\omega]\right).$$

When the appropriate fourth order partial derivatives are taken we get

$$\begin{aligned} & L\frac{\theta}{2}m_4\mathrm{E}[\tau_{a+b+c+d}] \\ +L(L-1)&\left(\frac{\theta}{2}m_2\right)^2\mathrm{E}[\tau_{a+b}]\mathrm{E}[\tau_{c+d}] \\ +L(L-1)&\left(\frac{\theta}{2}m_2\right)^2\mathrm{E}[\tau_{a+c}]\mathrm{E}[\tau_{b+d}] \\ +L(L-1)&\left(\frac{\theta}{2}m_2\right)^2\mathrm{E}[\tau_{a+d}]\mathrm{E}[\tau_{b+c}]. \end{aligned}$$

We can again simplify this expression for populations if we assume that individual are exchangeable. This gives

$$L\frac{\theta}{2}m_4\mathrm{E}[\tau_{a+b+c+d}] + 3L(L-1)\left(\frac{\theta}{2}m_2\right)^2\mathrm{E}[\tau_{a+b}]^2. \tag{B.18}$$

The expected kurtosis in the population is a quotient of random variables and calculating a second order approximation requires calculating eight order moments of the trait distribution. Instead we will calculate the expected fourth central moment.

$$\mathrm{E}[M_{4,Y}] = \mathrm{E}\left[\frac{1}{N}\sum\left(Y_i - \frac{\sum Y_j}{N}\right)^4\right]. \tag{B.19}$$

Examining the sum inside the expectation we see that

$$\begin{aligned} \mathrm{E}\left[\left(Y_i - \frac{\sum Y_j}{N}\right)^4\right] &= \mathrm{E}[Y_i^4] - 4\mathrm{E}\left[Y_i^3\frac{\sum Y_j}{N}\right] + 6\mathrm{E}\left[Y_i^2\left(\frac{\sum Y_j}{N}\right)^2\right] - \\ &\quad 4\mathrm{E}\left[Y_i\left(\frac{\sum Y_j}{N}\right)^3\right] + \left(\frac{\sum Y_j}{N}\right)^4 \\ &= \mathrm{E}[Y_i^4] - \frac{4}{N}\sum_j \mathrm{E}[Y_i^3Y_j] + \frac{6}{N^2}\sum_{j,k}\mathrm{E}[Y_i^2Y_jY_k] \\ &\quad - \frac{4}{N^3}\sum_{j,k,l}\mathrm{E}[Y_iY_jY_kY_l] + \frac{1}{n^4}\sum_{j,k,l,d}\mathrm{E}[Y_jY_kY_lY_d]. \end{aligned} \tag{B.20}$$

In calculating these expectations we have to remember that the value depends only on the number of times each variable appears in the expectation. That is, $\mathrm{E}[Y_1^2 Y_2 Y_3]$ is equivalent to $\mathrm{E}[Y_1 Y_2^3 Y_3]$ as long as all individuals in the population are exchangeable. The resulting expansion of (B.20) can be simplified by only considering terms of order one. Other terms can be ignored since we are assuming there are a large number of individuals in the population. This yields

$$\begin{aligned}
\mathrm{E}\left[\left(Y_i - \frac{\sum Y_j}{N}\right)^4\right] &= \mathrm{E}[Y_i^4] - \frac{4(n-1)}{n}\mathrm{E}[Y_i^3 Y_j] + \frac{6(n-1)(n-2)}{n^2}\mathrm{E}[Y_i^3 Y_j Y_k] \\
&\quad - \frac{4(n-1)(n-2)(n-3)}{n^3}\mathrm{E}[Y_i Y_j Y_k Y_l] \\
&\quad + \frac{(n-1)(n-2)(n-3)(n-4)}{n^4}\mathrm{E}[Y_j Y_k Y_l Y_d] + O(n^{-1}) \\
&\approx \mathrm{E}[Y_i^4] - 4\mathrm{E}[Y_i^3 Y_j] + 6\mathrm{E}[Y_i^2 Y_j Y_k] - 3\mathrm{E}[Y_i Y_j Y_k Y_l]. \qquad \text{(B.21)}
\end{aligned}$$

The first term, $\mathrm{E}[Y_i^4]$ was derived in equation (B.14) as

$$L\frac{\theta}{2}m_4\mathrm{E}[T_{MRCA}] + 3L(L-1)\left(\frac{\theta}{2}m_2\mathrm{E}[T_{MRCA}]\right)^2.$$

The second term, $\mathrm{E}[Y_i^3 Y_j]$ was derived in equation (B.15) as

$$L\frac{\theta}{2}m_4\mathrm{E}[\tau_{a+b}] + 3L(L-1)\left(\frac{\theta}{2}m_2\right)^2\mathrm{E}[T_{MRCA}]E[\tau_{a+b}].$$

The third term, $\mathrm{E}[Y_i^2 Y_j Y_k]$ was derived in equation (B.16) as

$$L\frac{\theta}{2}m_4\mathrm{E}[\tau_{a+b+c}] + L(L-1)\left(\frac{\theta}{2}m_2\right)^2\mathrm{E}[T_{MRCA}]\mathrm{E}[\tau_{a+b}] + 2L(L-1)\left(\frac{\theta}{2}m_2\right)^2\mathrm{E}[\tau_{a+b}]^2.$$

The fourth term, $\mathrm{E}[Y_i Y_j Y_k Y_l]$ was derived in equation (B.18) as

$$L\frac{\theta}{2}m_4\mathrm{E}[\tau_{a+b+c+d}] + 3L(L-1)\left(\frac{\theta}{2}m_2\right)^2 \mathrm{E}[\tau_{a+b}]^2.$$

Plugging these into (B.21) we get

$$\begin{aligned}\mathrm{E}[M_4] \approx\ & L\frac{\theta}{2}m_4\left(\mathrm{E}[T_{MRCA}] - 4\mathrm{E}[\tau_{a+b}] + 6\mathrm{E}[\tau_{a+b+c}] - 3\mathrm{E}[\tau_{a+b+c+d}]\right) \\ & + 3\left(L\frac{\theta}{2}m_2\right)^2 \left(\mathrm{E}[T_{MRCA}] - \mathrm{E}[\tau_{a+b}]\right)^2. \end{aligned} \tag{B.22}$$

With some simple manipulation this can be rewritten in terms of $\mathbb{T}_{i,j}$ to give

$$\begin{aligned}\mathrm{E}[M_4] = & \ 3\left(L\frac{\theta}{2}\mathrm{E}[\mathbb{T}_{2,2}]m_2\right)^2 \\ & + Lm_4\frac{\theta}{2}(\mathrm{E}[\mathbb{T}_{4,4}] + \frac{1}{3}\mathrm{E}[\mathbb{T}_{3,4}] + \frac{2}{9}\mathrm{E}[\mathbb{T}_{2,4}]). \end{aligned} \tag{B.23}$$

APPENDIX C

APPENDIX: ESTIMATING THE WOLF MUTATION RATE USING PEDIGREE SEQUENCING

C.1 Sequencing coverage in the wolf pedigree

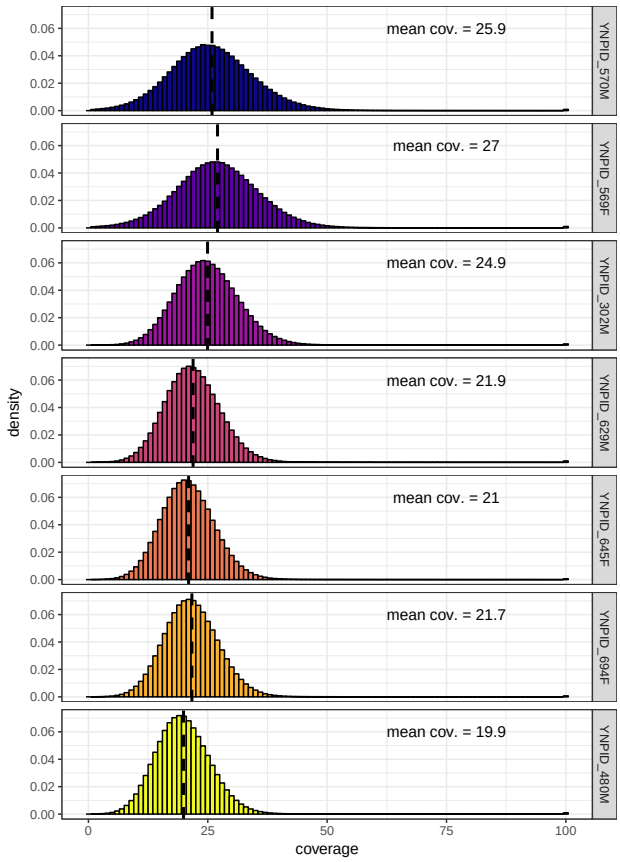


Figure C1: **Coverage distributions in each sequenced individual before the application of any site filters.** Sites with coverage 100 or greater are all counted in the last bin.

C.2 Alignment plots for potential de novo mutations

Plots of local alignments around sites of with posterior probabilities greater than 0.3 were used to decide whether they would be validated by Sanger sequencing. The majority of sites

in this candidate set showed signs of reads from elsewhere in the genome mismapping there. A typical alignment for a case like this is shown in Figure C2. These sites tended to have a couple of variants perfectly linked and not observed in the parent alignments. A smaller number of sites appeared to a have large number of sequencing errors in the vicinity. An example alignment for this sort of site is shown in Figure C3. These sites tended also to be located around repetitive places in the genome.

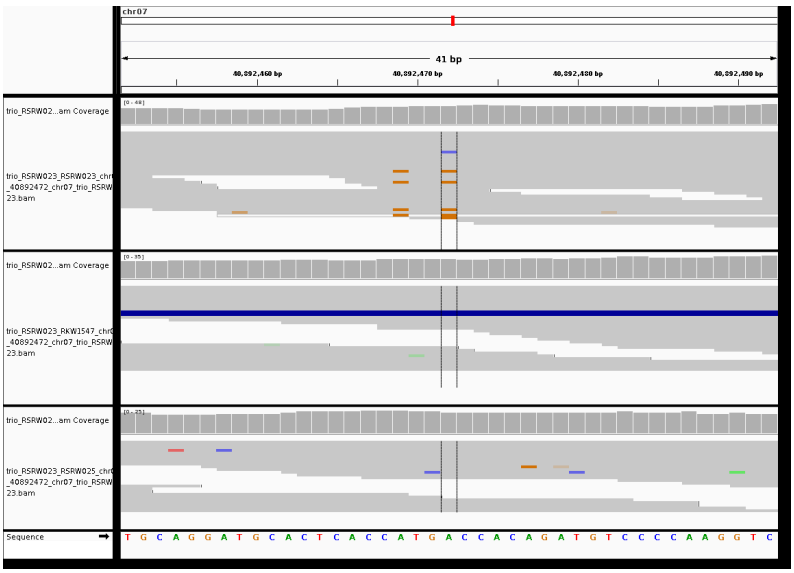


Figure C2: **Example alignment plot for a site with mismapped reads.** Site 40892472 on chromosome seven in the individual with YNPID 645. The fact that the A to T change always appears linked to an A to T change two nucleotides away indicates that the reads containing these bases originate from elsewhere in the genome where the sequence is identical except for these two substitutions.

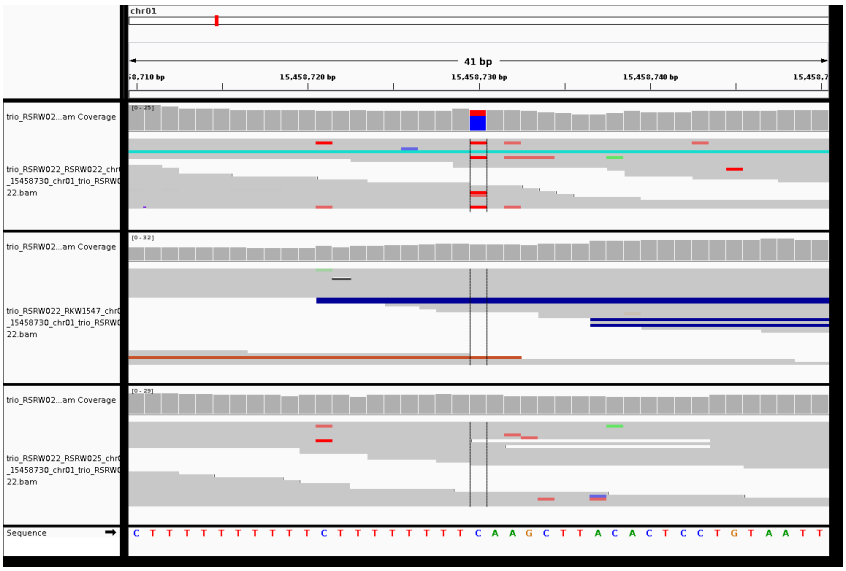


Figure C3: **Example alignment plot for a site with a large number of sequencing errors.** Site 15458730 on chromosome one in the individual with YNPID 629. Variation at this site appears to be due to sequencing errors because the site exists at the end of a string of Ts and other C to T differences exist in the region.

C.3 False negative rate estimation

The procedure described in the main test to calculate the false negative rate creates a set of simulated DNMs by pairing genotype likelihoods at sites where alternative alleles were observed in the offspring as well as one of the parents with genotype likelihoods from sites where no alternative alleles were observed in either parent. The false negative rate is then estimated by the proportion of simulated DNMs with a DN_p score less than 0.3. However, some of the alternative alleles we observe at the sites used to create the simulated DNMs are likely due to sequencing errors or mismapped reads. Because these erroneous alleles are likely to be at low frequency within the reads in an individual those sites will get a lower DN_p scores and therefore cause us to overestimate the false negative rate. To examine the magnitude of this effect we tried to select genotype likelihoods in the child from sites we are more confident are truly heterozygous. We did this first by requiring that alternative reads were observed in both parents and then made the more stringent requirement that four or

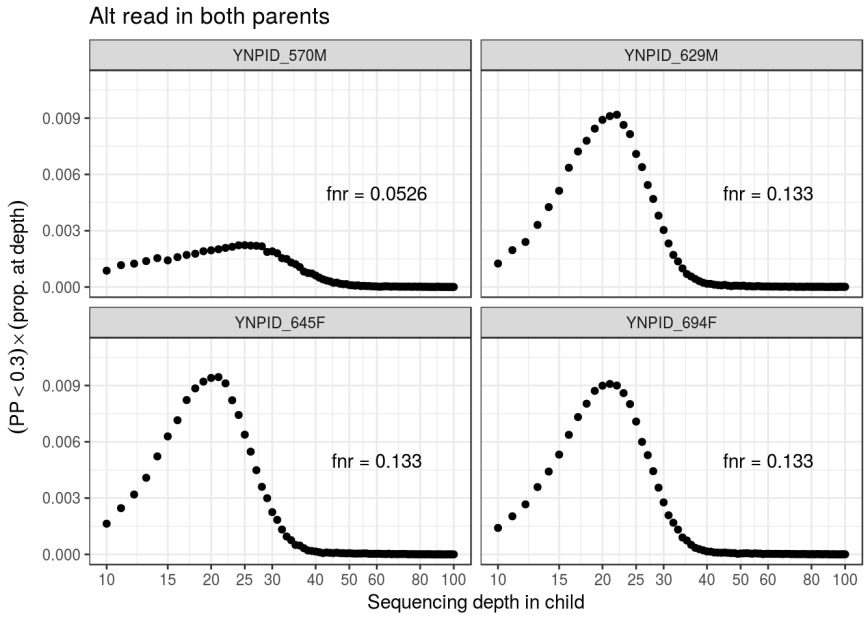


Figure C4: **False negative rates in each trio broken up into contributions from different sequencing depths in the child depth when both parents are required to have at least one alternative allele.**

more alternative alleles were observed in both parents.

In the original set of simulated DNMs used to generate Figure 4.5 we only required that an alternative allele was observed in one parent in addition to the offspring. By requiring that the alternative allele is observed in both parents we will hopefully increase the proportion of true heterozygous sites. Doing so and applying the same method for estimating false negative rates as described in the main text yields the false negative rate estimates shown in Figure C4. Being even more stringent and requiring that both parents have at least four reads with the alternative allele yields the false negative rate estimates shown in Figure C5. The false negative rate estimates do decrease when we are most strict and require four alternative allele reads in each parent for sites used as heterozygous genotypes. However, the rates differ at most by 5% from those in Figure 4.5.

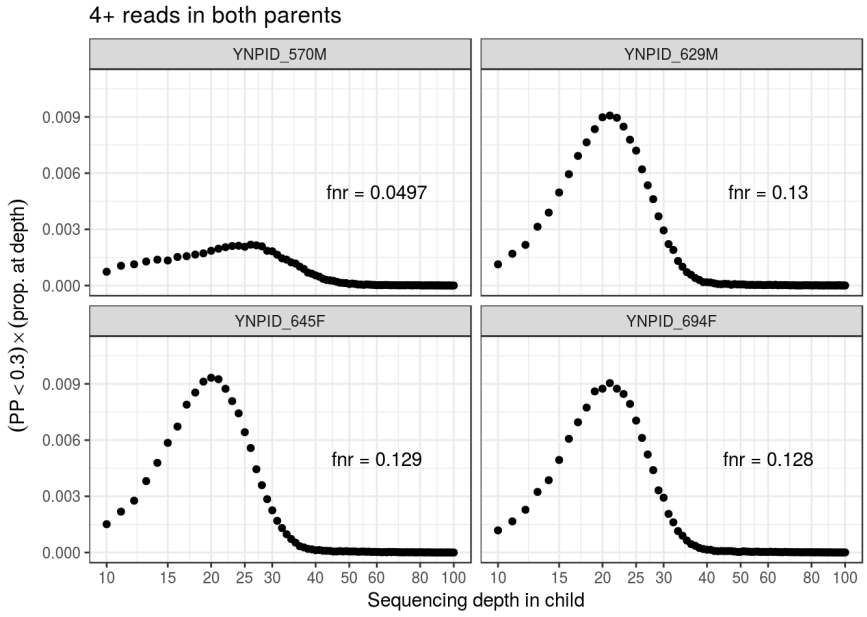


Figure C5: **False negative rates in each trio broken up into contributions from different sequencing depths in the child depth when both parents are required to have at least four alternative alleles.**

REFERENCES

Agrawal, A. F., and M. C. Whitlock, 2011 Inferences about the distribution of dominance drawn from yeast gene knockout data. Genetics **187**: 553–566.

Amster, G., and G. Sella, 2016 Life history effects on the molecular clock of autosomes and sex chromosomes. Proceedings of the National Academy of Sciences **113**: 1588–1593.

Balick, D. J., R. Do, C. A. Cassa, D. Reich, and S. R. Sunyaev, 2015 Dominance of Deleterious Alleles Controls the Response to a Population Bottleneck. PLoS Genetics **11**: e1005436.

Barton, N., A. Etheridge, and A. Véber, 2017 The infinitesimal model: Definition, derivation, and implications. Theoretical Population Biology **118**: 50–73.

Berg, J. J., and G. Coop, 2014 A Population Genetic Signal of Polygenic Adaptation. PLoS Genetics **10**: e1004412.

Bonhomme, M., C. Chevalet, B. Servin, S. Boitard, J. Abdallah, *et al.*, 2010 Detecting selection in population trees: The Lewontin and Krakauer test extended. Genetics **186**: 241–262.

Boyko, A. R., S. H. Williamson, A. R. Indap, J. D. Degenhardt, R. D. Hernandez, *et al.*, 2008 Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genetics **4**: e1000083.

Boyle, E. A., Y. I. Li, and J. K. Pritchard, 2017 An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell **169**: 1177–1186.

Brandvain, Y., and S. I. Wright, 2016 The limits of natural selection in a nonequilibrium world. Trends in Genetics **32**: 201 – 210.

Bulik-Sullivan, B. K., P.-r. Loh, H. K. Finucane, S. Ripke, J. Yang, *et al.*, 2015 LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature Genetics **47**: 291–295.

Cao, J., K. Schneeberger, S. Ossowski, T. Günther, S. Bender, *et al.*, 2011 Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nature Genetics **43**: 956–963.

Casals, F., A. Hodgkinson, and J. Hussin, 2013 Whole-exome sequencing reveals a rapid change in the frequency of rare functional variants in a founding population of humans. PLoS Genetics **9**: 1003815.

Castellano, S., G. Parra, F. A. Sanchez-Quinto, F. Racimo, M. Kuhlwilm, *et al.*, 2014 Patterns of coding variation in the complete exomes of three Neandertals. Proceedings of the National Academy of Sciences **111**: 6666–6671.

Chakraborty, R., and M. Nei, 1982 Genetic Differentiation of Quantitative Characters Between Populations of Species. Genetics Research **39**: 303–314.

Chen, J., S. Glémin, and M. Lascoux, 2017 Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species. Molecular Biology and Evolution **30**: 1229–1235.

Comeron, J. M., and M. Kreitman, 2002 Population, evolutionary and genomic consequences of interference selection. Genetics **161**: 389–410.

Cooper, G. M., E. A. Stone, G. Asimenos, E. D. Green, S. Batzoglou, *et al.*, 2005 Distribution and intensity of constraint in mammalian genomic sequence. Genome Research **15**: 901–913.

Davydov, E. V., D. L. Goode, M. Sirota, G. M. Cooper, A. Sidow, *et al.*, 2010 Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Computational Biology **6**: 1001025.

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, *et al.*, 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics **43**: 491–498.

Do, R., D. Balick, H. Li, I. Adzhubei, S. Sunyaev, *et al.*, 2015 No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. Nature Genetics **47**: 126–131.

Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow, 1998 Rates of spontaneous mutation. Genetics **148**: 1667–1686.

Durbin, R. M., D. L. Altshuler, G. R. Abecasis, D. R. Bentley, A. Chakravarti, *et al.*, 2010 A map of human genome variation from population-scale sequencing. Nature **467**: 1061–1073.

Elyashiv, E., K. Bullaughey, S. Sattath, Y. Rinott, M. Przeworski, *et al.*, 2010 Shifts in the intensity of purifying selection: An analysis of genome-wide polymorphism data from two closely related yeast species. Genome Research **20**: 1558–1573.

Evans, S. N., Y. Shvets, and M. Slatkin, 2007 Non-equilibrium theory of the allele frequency spectrum. Theoretical Population Biology **71**: 109–119.

Ewens, W. J., 2004 *Mathematical Population Genetics 1: I. Theoretical Introduction.*. Springer Science & Business Media.

Excoffier, L., I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, and M. Foll, 2013 Robust Demographic Inference from Genomic and SNP Data. PLoS Genetics **9**: 1003905.

Eyre-Walker, A., 2010 Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. Proceedings of the National Academy of Sciences **107**: 1752–1756.

Eyre-Walker, A., and P. D. Keightley, 2007 The distribution of fitness effects of new mutations. Nature Reviews Genetics **8**: 610–618.

Eyre-Walker, A., M. Woolfit, and T. Phelps, 2006 The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics **173**: 891–900.

Fan, Z., P. Silva, I. Gronau, S. Wang, A. S. Armero, *et al.*, 2016 Worldwide patterns of genomic variation and admixture in gray wolves. Genome Research **26**: 163–173.

Fay, J. C., G. J. Wyckoff, and C. I. Wu, 2001 Positive and negative selection on the human genome. Genetics **158**: 1227–34.

Felsenstein, J., 1981 Skepticism towards santa rosalia, or why are there so few kinds of animals? Evolution **35**: 124–138.

Felsenstein, J., 1985 Phylogenies and the comparative method. American Naturalist **125**: 1–15.

Felsenstein, J., 2002 Contrasts for a Within-Species Comparative Method. In M. Slatkin and M. Veuille, editors, *Modern Developments in Theoretical Population Genetics*. Oxford University Press, 118–129.

Fisher, R. A., 1918 XV.The Correlation between Relatives on the Supposition of Mendelian Inheritance. Transactions of the Royal Society of Edinburgh **52**: 399–433.

Fisher, R. A., 1937 The wave of advance of advantageous genes. Annals of Eugenics **7**: 355–369.

Frantz, L. A. F., V. E. Mullin, M. Pionnier-Capitan, O. Lebrasseur, M. Ollivier, *et al.*, 2016 Genomic and archaeological evidence suggest a dual origin of domestic dogs. Science **352**: 1228–1231.

Freckleton, R. P., P. H. Harvey, and M. D. Pagel, 2002 Phylogenetic analysis and comparative data: a test and review of evidence. American Naturalist **160**: 712–726.

Freedman, A., I. Gronau, R. M. Schweizer, D. O.-D. Vecchyo, E. Han, *et al.*, 2014 Genome sequencing highlights the dynamic early history of dogs. PLoS genetics **10**: 1004016.

Freedman, A. H., and R. K. Wayne, 2017 Deciphering the Origin of Dogs: From Fossils to Genomes. Annual Review of Animal Biosciences **5**: 281–307.

Fu, W., R. Gittelman, M. Bamshad, and J. Akey, 2014 Characteristics of Neutral and Deleterious Protein-Coding Variation among Individuals and Populations. The American Journal of Human Genetics **95**: 421–436.

Gazave, E., D. Chang, A. G. Clark, and A. Keinan, 2013 Population growth inflates the per-individual number of deleterious mutations and reduces their mean effect. Genetics **195**: 969–78.

Germonpré, M., M. Lázničková-Galetová, R. J. Losey, J. Räikkönen, and M. V. Sablin, 2015 Large canids at the Gravettian Pedmostí site, the Czech Republic: The mandible. Quaternary International **359-360**: 261–279.

Germonpré, M., M. V. Sablin, R. E. Stevens, R. E. Hedges, M. Hofreiter, *et al.*, 2009 Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes. Journal of Archaeological Science **36**: 473–490.

Gilad, Y., A. Oshlack, and S. A. Rifkin, 2006 Natural selection on gene expression. Trends in Genetics **22**: 456–461.

Gilbert, K. J., and M. C. Whitlock, 2015 QST-FST comparisons with unbalanced half-sib designs. Molecular Ecology Resources **15**: 262–267.

Good, B. H., A. M. Walczak, R. a. Neher, and M. M. Desai, 2014 Genetic diversity in the interference selection limit. PLoS Genetics **10**: e1004222.

Gopalakrishnan, S., J. A. Samaniego Castruita, M. H. S. Sinding, L. F. Kuderna, J. Räikkönen, *et al.*, 2017 The wolf reference genome sequence (Canis lupus lupus) and its implications for Canis spp. population genomics. BMC Genomics **18**: 1–11.

Gravel, S., 2016 When Is Selection Effective? Genetics **203**: 451–462.

Gravel, S., B. M. Henn, R. N. Gutenkunst, A. R. Indap, G. T. Marth, *et al.*, 2011 Demographic history and rare allele sharing among human populations. Proceedings of the National Academy of Sciences **108**: 11983–11988.

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, *et al.*, 2010 A draft sequence of the neandertal genome. Science **328**: 710–722.

Griffiths, R., and S. Tavaré, 1998 The age of a mutation in a general coalescent tree. Communications in Statistics. Stochastic Models **14**: 273–295.

Griswold, C. K., B. Logsdon, and R. Gomulkiewicz, 2007 Neutral evolution of multiple quantitative characters: a genealogical approach. Genetics **176**: 455–66.

Gronau, I., M. J. Hubisz, B. Gulko, C. G. Danko, and A. Siepel, 2011 Bayesian inference of ancient human demography from individual genome sequences. Nature Genetics **43**: 1031–4.

Gruber, J. D., K. Vogel, G. Kalay, and P. J. Wittkopp, 2012 Contrasting properties of gene-specific regulatory, coding, and copy number mutations in saccharomyces cerevisiae: Frequency, effects, and dominance. PLoS Genetics **8**: e1002497.

Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante, 2009 Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genetics **5**: e1000695.

Harris, K., and R. Nielsen, 2016 The genetic cost of neanderthal introgression. Genetics **203**: 881–891.

Hartl, D. L., E. N. Moriyama, and S. A. Sawyer, 1994 Selection intensity for codon bias. Genetics **138**: 227–234.

Henn, B. M., L. R. Botigué, S. Peischl, I. Dupanloup, M. Lipatov, *et al.*, 2016 Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. Proceedings of the National Academy of Sciences **113**: E440–E449.

Henn, B. M., L. L. Cavalli-Sforza, and M. W. Feldman, 2012 The great human expansion. Proceedings of the National Academy of Sciences of the United States of America **109**: 17758–17764.

Hernandez, R. D., J. L. Kelley, E. Elyashiv, S. C. Melton, A. Auton, *et al.*, 2011 Classic selective sweeps were rare in recent human evolution. Science **331**: 920–924.

Hernandez, R. D., S. H. Williamson, and C. D. Bustamante, 2007 Context dependence, ancestral misidentification, and spurious signatures of natural selection. Molecular Biology and Evolution **24**: 1792–1800.

Jouganous, J., W. Long, A. P. Ragsdale, and S. Gravel, 2017 Inferring the joint demographic history of multiple populations: Beyond the diffusion approximation. Genetics **206**: 1549–1567.

Juric, I., S. Aeschbacher, and G. Coop, 2016 The Strength of Selection against Neanderthal Introgression. PLoS Genetics **12**: 1–25.

Keightley, P. D., R. W. Ness, D. L. Halligan, and P. R. Haddrill, 2014 Estimation of the spontaneous mutation rate per nucleotide site in a Drosophila melanogaster full-sib family. Genetics **196**: 313–320.

Keightley, P. D., A. Pinharanda, R. W. Ness, F. Simpson, K. K. Dasmahapatra, *et al.*, 2015 Estimation of the Spontaneous Mutation Rate in Heliconius melpomene. Molecular Biology and Evolution **32**: 239–243.

Kelleher, J., A. M. Etheridge, and G. Mcvean, 2015 Efficient coalescent simulation and genealogical analysis for large sample sizes. PLoS Comput Biol **12**: e1004842.

Khaitovich, P., S. Pääbo, and G. Weiss, 2005 Toward a Neutral Evolutionary Model of Gene Expression. Genetics **170**: 929–939.

Kim, B. Y., and K. E. Lohmueller, 2015 Selection and reduced population size cannot explain higher amounts of Neandertal ancestry in East Asian than in European human populations. American Journal of Human Genetics **96**: 454–461.

Kimmel, M., R. Chakraborty, J. P. King, M. Bamshad, W. S. Watkins, *et al.*, 1998 Signatures of population expansion in microsatellite repeat data. Genetics **148**: 1921–1930.

Kimura, M., 1969 The rate of molecular evolution considered from the standpoint of population genetics. Proceedings of the National Academy of Sciences of the United States of America **63**: 1181–1188.

Kimura, M., T. Maruyama, and J. F. Crow, 1963 The mutation load in small populations. Genetics **48**: 1303–1312.

Kimura, M., and T. Ota, 1974 Probability of gene fixation in an expanding finite population. Proceedings of the National Academy of Sciences of the United States of America **71**: 3377–3379.

Kimura, M., and G. H. Weiss, 1964 The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance. Genetics **49**: 561–576.

Kircher, M., D. M. Witten, P. Jain, B. J. O'Roak, G. M. Cooper, *et al.*, 2014 A general framework for estimating the relative pathogenicity of human genetic variants. Nature Genetics **46**: 310–315.

Kondrashov, A. S., 1995 Modifiers of mutation-selection balance: General approach and the evolution of mutation rates. Genetical Research **66**: 53–69.

Kong, A., M. L. Frigge, G. Masson, S. Besenbacher, P. Sulem, *et al.*, 2012 Rate of de novo mutations and the importance of father's age to disease risk. Nature **488**: 471–475.

Köster, J., and S. Rahmann, 2012 Snakemake-a scalable bioinformatics workflow engine. Bioinformatics **28**: 2520–2522.

Kumar, S., and S. Subramanian, 2002 Mutation rates in mammalian genomes. Proceedings of the National Academy of Sciences of the United States of America **99**: 803–808.

Lande, R., 1976 Natural Selection and Random Genetic Drift in Phenotypic Evolution. Evolution **30**: 314–334.

Lande, R., 1992 Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. Evolution **46**: 381–389.

Leinonen, T., R. J. McCairns, R. B. O'Hara, and J. Merilä, 2013 QST -FST comparisons: Evolutionary and ecological insights from genomic heterogeneity.

Lek, M., K. J. Karczewski, K. E. Samocha, E. Banks, T. Fennell, *et al.*, 2016 Analysis of protein-coding genetic variation in 60,706 humans. Nature **536**: 285–291.

Lewontin, R., and J. Krakauer, 1973 Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics **73**: 175–195.

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**: 1754–1760.

Li, H., and R. Durbin, 2011 Inference of human population history from individual whole-genome sequences. Nature **475**: 493–496.

Li, S., B. Li, C. Cheng, Z. Xiong, Q. Liu, *et al.*, 2014 Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. Genome Biology **15**: 557.

Lindblad-Toh, K., C. M. Wade, T. S. Mikkelsen, E. K. Karlsson, D. B. Jaffe, *et al.*, 2005 Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature **438**: 803–819.

Lohmueller, K., 2014a The distribution of deleterious genetic variation in human populations. Current opinion in genetics & development **29**: 139–146.

Lohmueller, K. E., 2014b The Impact of Population Demography and Selection on the Genetic Architecture of Complex Traits. PLoS Genetics **10**: 1004379.

Lohmueller, K. E., A. Albrechtsen, Y. Li, S. Y. Kim, T. Korneliussen, *et al.*, 2011 Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. PLoS Genetics **7**: 1002326.

Lohmueller, K. E., A. R. Indap, S. Schmidt, A. R. Boyko, R. D. Hernandez, *et al.*, 2008 Proportionally more deleterious genetic variation in European than in African populations. Nature **451**: 994–997.

Lohse, K., R. J. Harrison, and N. H. Barton, 2011 A general method for calculating likelihoods under the coalescent process. Genetics **189**: 977–87.

Lynch, M., 1989 Phylogenetic Hypotheses Under the Assumption of Neutral Quantitative-Genetic Variation. Evolution **43**: 1–17.

Lynch, M., 2011 The lower bound to the evolution of mutation rates. Genome Biology and Evolution **3**: 1107–1118.

Lynch, M., M. S. Ackerman, J.-F. Gout, H. Long, W. Sung, *et al.*, 2016 Genetic drift, selection and the evolution of the mutation rate. Nature Reviews Genetics **17**: 704–714.

Lynch, M., and W. G. Hill, 1986 Phenotypic Evolution by Neutral Mutation. Evolution **40**: 915–935.

Marsden, C. D., D. Ortega-Del Vecchyo, D. P. O'Brien, J. F. Taylor, O. Ramirez, *et al.*, 2016 Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proceedings of the National Academy of Sciences **113**: 152–157.

Marth, G. T., E. Czabarka, J. Murvai, and S. T. Sherry, 2004 The Allele Frequency Spectrum in Genome-Wide Human Variation Three Large World Populations. Genetics **372**: 351–372.

Martin, A. P., and S. R. Palumbi, 1993 Body size, metabolic rate, generation time, and the molecular clock. Proceedings of the National Academy of Sciences of the United States of America **90**: 4087–4091.

Maruyama, T., and P. A. Fuerst, 1984 Population bottlenecks and nonequilibrium models in population genetics. i. allele numbers when populations evolve from zero variability. Genetics **108**: 745–763.

Maruyama, T., and P. A. Fuerst, 1985a Population bottlenecks and nonequilibrium models in population genetics. ii. number of alleles in a small population that was formed by a recent bottleneck. Genetics **111**: 675–689.

Maruyama, T., and P. A. Fuerst, 1985b Population bottlenecks and nonequilibrium models in population genetics. iii. genic homozygosity in populations which experience periodic bottlenecks. Genetics **111**: 691–703.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the Adh locus in Drosophila. Nature **351**: 652–654.

McVicker, G., D. Gordon, C. Davis, and P. Green, 2009 Widespread genomic signatures of natural selection in hominid evolution. PLoS Genetics **5**: 1000471.

Mendes, F. K., J. A. Fuentes-González, J. G. Schraiber, and M. W. Hahn, 2018 Evolutionary inferences about quantitative traits are affected by underlying genealogical discordance. bioRxiv .

Metzger, B. P., F. Duveau, D. C. Yuan, S. Tryban, B. Yang, *et al.*, 2016 Contrasting Frequencies and Effects of cis- and trans-Regulatory Mutations Affecting Gene Expression. Molecular Biology and Evolution **33**: 1131–1146.

Meurer, A., C. P. Smith, M. Paprocki, O. Čertík, S. B. Kirpichev, *et al.*, 2017 Sympy: symbolic computing in python. PeerJ Computer Science **3**: e103.

Moorjani, P., C. E. G. Amorim, P. F. Arndt, and M. Przeworski, 2016 Variation in the molecular clock of primates. Proceedings of the National Academy of Sciences : 1–39.

Nadachowska-Brzyska, K., C. Li, L. Smeds, G. Zhang, and H. Ellegren, 2015 Temporal dynamics of avian populations during pleistocene revealed by whole-genome sequences. Current Biology **25**: 1375–1380.

Napierala, H., and H. P. Uerpmann, 2012 A 'new' palaeolithic dog from central Europe. International Journal of Osteoarchaeology **22**: 127–137.

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

Nelson, M. R., D. Wegmann, M. G. Ehm, D. Kessner, P. St Jean, *et al.*, 2012 An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science **337**: 100–104.

Nielsen, R., J. M. Akey, M. Jakobsson, J. K. Pritchard, S. Tishkoff, *et al.*, 2017 Tracing the peopling of the world through genomics. Nature **541**: 302–310.

Ohta, T., 1972 Population size and rate of evolution. Journal of Molecular Evolution **1**: 305–314.

Ohta, T., 1973 Slightly deleterious mutant substitutions in evolution. Nature **246**: 96–98.

Otto, S. P., and M. C. Whitlock, 1997 The probability of fixation in populations of changing size. Genetics **146**: 723–733.

Ovaskainen, O., M. Karhunen, C. Zheng, J. M. C. Arias, and J. Merilä, 2011 A new method to uncover signatures of divergent and stabilizing selection in quantitative traits. Genetics **189**: 621–632.

Ovodov, N. D., S. J. Crockford, Y. V. Kuzmin, T. F. Higham, G. W. Hodgins, *et al.*, 2011 A 33,000-Year-Old incipient dog from the Altai Mountains of Siberia: Evidence of the earliest domestication disrupted by the last Glacial Maximum. PLoS ONE **6**: 4–10.

Peischl, S., I. Dupanloup, A. Foucal, M. Jomphe, V. Bruat, *et al.*, 2017 Relaxed selection during a recent human expansion. Genetics .

Peischl, S., I. Dupanloup, M. Kirkpatrick, and L. Excoffier, 2013 On the accumulation of deleterious mutations during range expansions. Molecular Ecology **22**: 5972–5982.

Pfeifer, S. P., 2017 Direct estimate of the spontaneous germ line mutation rate in African green monkeys. Evolution **71**: 2858–2870.

Pionnier-Capitan, M., C. Bemilli, P. Bodu, G. Célérier, J. G. Ferrié, *et al.*, 2011 New evidence for Upper Palaeolithic small domestic dogs in South-Western Europe. Journal of Archaeological Science **38**: 2123–2140.

Polanski, a., and M. Kimmel, 2003 New explicit expressions for relative frequencies of single-nucleotide polymorphisms with application to statistical inference on population growth. Genetics **165**: 427–436.

Praebel, K., R. Knudsen, A. Siwertsson, M. Karhunen, K. K. Kahilainen, *et al.*, 2013 Ecological speciation in postglacial European whitefish: rapid adaptive radiations into the littoral, pelagic, and profundal lake habitats. Ecology and Evolution **3**: 4970–4986.

Prakash, S., 1972 Origin of reproductive isolation in the absence of apparent genic differentiation in a geographic isolate of Drosophila pseudoobscura. Genetics **72**: 143–155.

Prakash, S., R. C. Lewontin, and J. L. Hubby, 1969 A molecular approach to the study of genic heterozygosity in natural populations. IV. Patterns of genic variation in central, marginal and isolated populations of Drosophila pseudoobscura. Genetics **61**: 841–858.

Racimo, F., and J. Schraiber, 2014 Approximation to the Distribution of Fitness Effects across Functional Categories in Human Segregating Polymorphisms. PLoS Genetics **10**: 1004697.

Rahbari, R., A. Wuster, S. J. Lindsay, R. J. Hardwick, L. B. Alexandrov, *et al.*, 2016 Timing, rates and spectra of human germline mutation. Nature Genetics **48**: 126–133.

Ramu, A., M. J. Noordam, R. S. Schwartz, A. Wuster, M. E. Hurles, *et al.*, 2013 DeNovoGear: de novo indel and point mutation discovery and phasing. Nature methods **10**: 985–987.

Rannala, B., and Z. Yang, 2003 Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics **1656**: 1645–1656.

Reich, D. E., and D. B. Goldstein, 1998 Genetic evidence for a Paleolithic human population expansion in Africa. Proceedings of the National Academy of Sciences of the United States of America **95**: 8119–8123.

Rice, S., 2004 *Evolutionary Theory: Mathematical and Conceptual Foundations*. Sinauer.

Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, *et al.*, 2011 Integrative genomics viewer. Nature Biotechnology **29**: 24–26.

Robinson, M. R., G. Hemani, C. Medina-Gomez, M. Mezzavilla, T. Esko, *et al.*, 2015 Population genetic differentiation of height and body mass index across Europe. Nature Genetics **47**: 1357–1362.

Rogers, A., and H. Harpending, 1983 Population Structure and Quantitative Characters. Genetics **105**: 985–1002.

Rogers, A. R., 1995 Genetic Evidence for a Pleistocene Population Explosion. Evolution **49**: 608–615.

Sankararaman, S., S. Mallick, M. Dannemann, K. Prüfer, J. Kelso, *et al.*, 2014 The genomic landscape of Neanderthal ancestry in present-day humans. Nature **507**: 354–357.

Sawyer, S. A., and D. L. Hartl, 1992 Population genetics of polymorphism and divergence. Genetics **132**: 1161–1176.

Sayres, M. A. W., C. Venditti, M. Pagel, and K. D. Makova, 2011 Do variations in substitution rates and male mutation bias correlate with life-history traits? a study of 32 mammalian genomes. Evolution **65**: 2800–2815.

Scally, A., and R. Durbin, 2012 Revising the human mutation rate: implications for understanding human evolution. Nature reviews. Genetics **13**: 745–753.

Schiffels, S., and R. Durbin, 2014 Inferring human population size and separation history from multiple genome sequences. Nature genetics .

Schraiber, J. G., and J. M. Akey, 2015 Methods and models for unravelling human evolutionary history. Nat Rev Genet **16**: 727–740.

Schraiber, J. G., and M. J. Landis, 2015 Sensitivity of quantitative traits to mutational effects and number of loci. Theoretical population biology **102**: 85–93.

Schubert, M., H. Jónsson, D. Chang, C. Der Sarkissian, L. Ermini, *et al.*, 2014 Prehistoric genomes reveal the genetic foundation and cost of horse domestication. Proceedings of the National Academy of Sciences **111**: E5661–E5669.

Ségurel, L., M. J. Wyman, and M. Przeworski, 2014 Determinants of Mutation Rate Variation in the Human Germline. Annual Review of Genomics and Human Genetics **15**: 47–70.

Shannon, L. M., R. H. Boyko, M. Castelhano, E. Corey, J. J. Hayward, *et al.*, 2015 Genetic structure in village dogs reveals a Central Asian domestication origin. Proceedings of the National Academy of Sciences **112**: 13639–13644.

Simons, Y. B., and G. Sella, 2016 The impact of recent population history on the deleterious mutation load in humans and close evolutionary relatives.

Simons, Y. B., M. C. Turchin, J. K. Pritchard, and G. Sella, 2014 The deleterious mutation load is insensitive to recent population history. Nature Genetics **46**: 220–224.

Skoglund, P., E. Ersmark, E. Palkopoulou, and L. Dalén, 2015 Ancient Wolf Genome Reveals an Early Divergence of Domestic Dog Ancestors and Admixture into High-Latitude Breeds. Current Biology **25**: 1515–1519.

Skoglund, P., A. Götherström, and M. Jakobsson, 2011 Estimation of population divergence times from non-overlapping genomic sequences: Examples from dogs and wolves. Molecular Biology and Evolution **28**: 1505–1517.

Slatkin, M., 1987 Gene flow and the geographic structure of natural populations. Science **236**: 787–792.

Slatkin, M., and R. R. Hudson, 1991 Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics **129**: 555–562.

Smeds, L., A. Qvarnström, and H. Ellegren, 2016 Direct estimate of the rate of germline mutation in a bird. Genome Research **26**: 1211–1218.

Smit, A., R. Hubeley, and P. Green, 2013-2015 Repeatmasker open-4.0.

Spitze, K., 1993 Population structure in Daphnia obtusa: Quantitative genetic and allozymic variation. Genetics **135**: 367–374.

Stahler, D. R., D. R. Macnulty, R. K. Wayne, B. VonHoldt, and D. W. Smith, 2013 The adaptive value of morphological, behavioural and life-history traits in reproductive female wolves. Journal of Animal Ecology **82**: 222–234.

Stan Development Team, 2016 RStan: the R interface to Stan. R package version 2.14.1.

Subramanian, S., 2012 The abundance of deleterious polymorphisms in humans. Genetics **190**: 1579–83.

Sun, J. X., A. Helgason, G. Masson, S. S. Ebenesersdóttir, H. Li, *et al.*, 2012 A direct characterization of human mutation based on microsatellites. Nature Genetics **44**: 1161–1165.

Sung, W., M. S. Ackerman, S. F. Miller, T. G. Doak, and M. Lynch, 2012 Drift-barrier hypothesis and mutation-rate evolution. Proceedings of the National Academy of Sciences **109**: 18488–18492.

Tajima, F., 1989 The effect of change in population size on DNA polymorphism. Genetics **123**: 597–601.

Templeton, A. R., 1980 The theory of speciation via the founder principle.

Tennessen, J. a., A. W. Bigham, T. D. O'Connor, W. Fu, E. E. Kenny, *et al.*, 2012 Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science **337**: 64–9.

Thalmann, O., B. Shapiro, P. Cui, V. J. Schuenemann, S. K. Sawyer, *et al.*, 2013 Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. Science **342**: 871–874.

Thorvaldsdóttir, H., J. T. Robinson, and J. P. Mesirov, 2013 Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Briefings in Bioinformatics **14**: 178–192.

Torkamani, A., P. Pham, O. Libiger, V. Bansal, G. Zhang, *et al.*, 2012 Clinical implications of human population differences in genome-wide rates of functional genotypes. Frontiers in Genetics **3**: 211.

Turelli, M., 2017 Commentary: Fisher's infinitesimal model: A story for the ages. Theoretical Population Biology **118**: 46–49.

Turelli, M., and N. H. Barton, 1990 Dynamics of polygenic characters under selection. Theoretical Population Biology **38**: 1–57.

Turelli, M., J. H. Gillespie, and R. Lande, 1988 Rate Tests for Selection on Quantitative Characters During Macroevolution and Microevolution. Evolution **42**: 1085–1089.

Uchimura, A., M. Higuchi, Y. Minakuchi, M. Ohno, A. Toyoda, *et al.*, 2015 Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. Genome Research **25**: 1125–1134.

Venn, O., I. Turner, I. Mathieson, N. De Groot, R. Bontrop, *et al.*, 2014 Strong male bias drives germline mutation in chimpanzees. Science **344**: 1272–1275.

Vilà, C., P. Savolainen, J. E. Maldonado, I. R. Amorim, J. E. Rice, *et al.*, 1997 Multiple and ancient origins of the domestic dog. Science **276**: 1687–1689.

Voight, B. F., A. M. Adams, L. a. Frisse, Y. Qian, R. R. Hudson, *et al.*, 2005 Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. Proceedings of the National Academy of Sciences of the United States of America **102**: 18508–18513.

Wakeley, J., 2008 *Coalescent Theory: An Introduction*. W. H. Freeman.

Wall, J. D., M. a. Yang, F. Jay, S. K. Kim, E. Y. Durand, *et al.*, 2013 Higher levels of Neanderthal ancestry in east Asians than in Europeans. Genetics **194**: 199–209.

Wang, G.-D., M.-S. Peng, H.-C. Yang, P. Savolainen, and Y.-P. Zhang, 2016a Questioning the evidence for a Central Asian domestication origin of dogs. Proceedings of the National Academy of Sciences **113**: E2554–E2555.

Wang, G. D., W. Zhai, H. C. Yang, R. X. Fan, X. Cao, *et al.*, 2013 The genomics of selection in dogs and the parallel evolution between dogs and humans. Nature Communications **4**: 1860–1869.

Wang, G.-D., W. Zhai, H.-C. Yang, L. Wang, L. Zhong, *et al.*, 2016b Out of southern East Asia: the natural history of domestic dogs across the world. Cell Research **26**: 21–33.

Watterson, G. A., 1984 Allele frequencies after a bottleneck. Theoretical Population Biology **26**: 387–407.

Watterson, G. A., 1986 The homozygosity test after a change in population size. Genetics **112**: 899–907.

Wayne, R. K., and S. J. O'Brien, 1987 ALLOZYME DIVERGENCE WITHIN THE CANIDAE. Systematic Zoology **36**: 339–355.

Wegmann, D., C. Leuenberger, and L. Excoffier, 2009 Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. Genetics **182**: 1207–1218.

Weisburd, B., 2017 `igv_plotter`. https://github.com/macarthur-lab/igv_utils.

Westfall, P. H., 2014 Kurtosis as Peakedness, 1905-2014. R.I.P. The American Statistician **68**: 191–195.

Wheeler, H. E., K. P. Shah, J. Brenner, T. Garcia, K. Aquino-Michaels, *et al.*, 2016 Survey of the Heritability and Sparse Architecture of Gene Expression Traits across Human Tissues. PLOS Genetics **12**: e1006423.

Whitehead, A., and D. L. Crawford, 2006 Neutral and adaptive variation in gene expression. Proceedings of the National Academy of Sciences **103**: 5425–5430.

Whitlock, M. C., 1999 Neutral additive genetic variance in a metapopulation. Genetical Research **74**: 215–221.

Whitlock, M. C., 2008 Evolutionary inference from QST. Molecular Ecology **17**: 1885–1896.

Whitlock, M. C., and F. Guillaume, 2009 Testing for spatially divergent selection: Comparing QST to FST. Genetics **183**: 1055–1063.

Williamson, S. H., R. Hernandez, A. Fledel-Alon, L. Zhu, R. Nielsen, *et al.*, 2005 Simultaneous inference of selection and population growth from patterns of variation in the human genome. Proceedings of the National Academy of Sciences of the United States of America **102**: 7882–7887.

Wong, A., 2014 Covariance between testes size and substitution rates in primates. Molecular Biology and Evolution **31**: 1432–1436.

Wright, S., 1931 Evolution in mendelian populations. Genetics **16**: 97–159.

Wu, C. I., and W. H. Li, 1985 Evidence for higher rates of nucleotide substitution in rodents than in man. Proceedings of the National Academy of Sciences **82**: 1741–1745.

Xue, Y., J. Prado-Martinez, P. H. Sudmant, V. Narasimhan, Q. Ayub, *et al.*, 2015 Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science **348**: 242–245.

Yang, J.-R., C. J. Maclean, C. Park, H. Zhao, and J. Zhang, 2017 Intra and Interspecific Variations of Gene Expression Levels in Yeast Are Largely Neutral: (Nei Lecture, SMBE 2016, Gold Coast). Molecular Biology and Evolution **34**: 2125–2139.

Yang, S., L. Wang, J. Huang, X. Zhang, and Y. Yuan, 2015 Parent-progeny sequencing indicates higher mutation rates in heterozygotes. Nature **523**: 463–467.

Yu, N., F. C. Chen, S. Ota, L. B. Jorde, P. Pamilo, *et al.*, 2002 Larger genetic differences within Africans than between Africans and Eurasians. Genetics **161**: 269–274.

Zuckerkandl, E., and L. Pauling, 1965 Molecules as documents of evolutionary history. Journal of Theoretical Biology **8**: 357–366.