MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 1575–1581 doi: 10.1111/j.1365-294X.2011.05046.x

OPINION

The $n = 1$ constraint in population genomics

C. A. BUERKLE, Z. GOMPERT and T. L. PARCHMAN

Department of Botany, University of Wyoming, Laramie, WY 82071, USA

Abstract

A key objective of population genomics is to identify portions of the genome that have been shaped by natural selection rather than by neutral divergence. A previously recognized but underappreciated challenge to this objective is that observations of allele frequencies across genomes in natural populations often correspond to a single, unreplicated instance of the outcome of evolution. This is because the composition of each individual genomic region and population is expected to be the outcome of a unique array of evolutionary processes. Given a single observation, inference of the evolutionary processes that led to the observed state of a locus is associated with considerable uncertainty. This constraint on inference can be ameliorated by utilizing multi-allelic (e.g. DNA haplotypes) rather than bi-allelic markers, by analysing two or more populations with certain models and by utilizing studies of replicated experimental evolution. Future progress in population genomics will follow from research that recognizes the '$n = 1$ constraint' and that utilizes appropriate and explicit evolutionary models for analysis.

Keywords: Bayesian inference, F_{ST}, genome scan, outlier detection, population genomics

Received 28 October 2010; revision received 24 January 2011; accepted 31 January 2011

Over the last several decades, many evolutionary biologists have sought to use genetic variation in extant populations to make inferences about the evolutionary processes that have operated in the populations (reviewed in Luikart *et al.* 2003; Beaumont 2005; Nielsen 2005; Storz 2005; Butlin 2010). With increasing access to molecular markers and DNA sequences for a growing number of taxa, many researchers have investigated how evolutionary and demographic processes shape variation across the genome. This field of inquiry is commonly referred to as population genomics. A core aim of population genomics is to discriminate between regions of the genome that have experienced selection or have evolved neutrally. Our ability to discriminate among these classes of genetic regions is affected by whether evolutionary processes (drift, gene conversion, migration, mutation, etc.) other than selection vary across the genome and by the extent to which the form and intensity of selection vary across the genome. Although it is possible that most regions of the genome experience a single, common set of evolutionary forces (the same rates of recombination, mutation, gene conversion, gene flow, drift, etc.), there are good reasons to think this unlikely (Noor & Feder 2006; Noor & Bennett 2009; Bazin *et al.* 2010; Turner & Hahn 2010). Instead, we conclude that it is much more likely that each genetic region is affected by a unique array of evolutionary forces (see below and Table 1). If this is true, typically very little information will be available for inferring the evolutionary processes that have shaped variation at any point in the genome. More specifically, in many cases, there will be a single appropriate observation of the outcome of evolution at a locus in a population, with no clear evolutionary or genomic replication in other populations or at other loci. As a result, considerable uncertainty will be associated with the inferred amount of evolution and potential selection at each locus in a population, which will limit opportunities to detect true differences between selected and neutral loci. We refer to this underappreciated limitation as the '$n = 1$ constraint' (previously recognized in Balding 2003; Beaumont & Balding 2004; Riebler *et al.* 2008; Holsinger & Weir 2009). In this manuscript, we describe the '$n = 1$ constraint' and why we think it is likely to hold, evaluate its significance for various

Correspondence: C. Alex Buerkle, Fax: 307 766 2851;
E-mail: buerkle@uwyo.edu

Table 1 Several phenomena are known to vary across the genome and affect molecular evolution and population genomic divergence. For each phenomenon, a supporting example observation is given, along with relevant citations. The following phenomena are known to vary within genomes but also vary across populations (for example, recombination rates vary across the genome, among individuals and among populations, McMullen *et al.* 2009; N_e varies between autosomes and sex chromosomes Vicoso & Charlesworth, 2006, Charlesworth 2009 but also among populations Whitlock & McCauley 1999). Whereas this is a limited and non-exhaustive list of examples, the observations are thought to be typical rather than special cases

Phenomenon	Observation	Citation
Recombination	Rates vary ~30× across regions in human, mice and corn	Kauppi *et al.* (2007); Coop *et al.* (2008); McMullen *et al.* (2009)
	Segregating inversions suppress recombination in heterozygotes	McMullen *et al.* (2009)
Gene conversion	Higher rates in regions of high recombination	Andolfatto & Nordborg (1998); Marais (2003); Jeffreys & May (2004)
	Rates can be 1000× greater than mutation rate	Marais 2003
N_e	Smaller on X vs. autosomes	Vicoso & Charlesworth (2006); Charlesworth (2009)
Mutation rate	Hypervariable regions exist in human DNA	Jeffreys *et al.* (1988); Ellegren *et al.* (2003)
	Higher in regions experiencing high recombination	Lercher & Hurst (2002)
	Covaries with transposition rate and GC content	Hardison *et al.* (2003)
	Lower mutation rate on X chromosome	Vicoso & Charlesworth (2006)
Selection	Human genome contains evidence of negative and positive selection of varying intensity	Bustamante *et al.* (2005); Andrés *et al.* (2009)
	Natural selection causes variation in apparent N_e along the genome of *Caenorhabditis elegans*	Rockman *et al.* (2010)

models and approaches of population genomics and conclude by discussing possible strategies to ameliorate this problem.

Experimental, observational and theoretical results suggest that a single evolutionary model is unlikely to hold across the genome, or across populations, for plausible histories and genetic architectures of population divergence (Table 1). These include expected variance across the genome and across populations of effective population size (N_e; e.g. sex chromosome vs. autosomes, Vicoso & Charlesworth 2006; differences among populations, Whitlock & McCauley 1999; Charlesworth 2009), forms and strength of selection (Bustamante *et al.* 2005; Rockman *et al.* 2010), interference because of the Hill–Robertson effect (Hill & Robertson 1968), mutation rates (Ellegren *et al.* 2003), and recombination and biased gene conversion (Marais 2003; Jeffreys & May 2004; McMullen *et al.* 2009; Noor & Bennett 2009), all of which will shape allele frequency shifts over time. Similarly, even constant gene flow (Nm) among populations will lead to heterogeneous effects across the genome, depending on the difference in allele frequencies between the immigrant and resident individuals in the focal population, which will certainly vary among loci. Furthermore, even if a single evolutionary model held for most of the genome, it would certainly not hold for the regions of the genome affected by selection, unless they all experienced the exact same magnitude and form (e.g. directional, diversifying or disruptive, and background or purifying selection) of selection. Taken together these observations (Table 1) indicate that in many cases each genetic region can be expected to evolve under a unique set of forces, and thus, each should be modelled with an individual evolutionary model. Given that potentially distinct evolution will have occurred in each genetic region, with a single observation of its outcome, inference of the evolutionary processes that operated will be challenging and we refer to this as the '$n = 1$ constraint'.

The constraint for inference that results from having a single observation of an evolutionary process can be illustrated simply. Suppose that in a system we knew the allele frequencies in an ancestral population ($\boldsymbol{p}_0$) and also in a descendant population following some number of generations ($\boldsymbol{p}_1$) during which some evolutionary processes operated. The known allele frequencies allow us to avoid any uncertainty from sampling the population or the genome. With this perfect knowledge of the composition of the population at two points in time, if we were to examine any region of the genome, we would be able to observe directly the change in allele frequencies that occurred as a result of the evolutionary processes that operated at that locus. The '$n = 1$ constraint' is that we have a single observation of a result of the evolutionary process that operated on the focal locus. Consequently, even with perfect knowledge of the state of populations, based on a single outcome of the process, there will be considerable

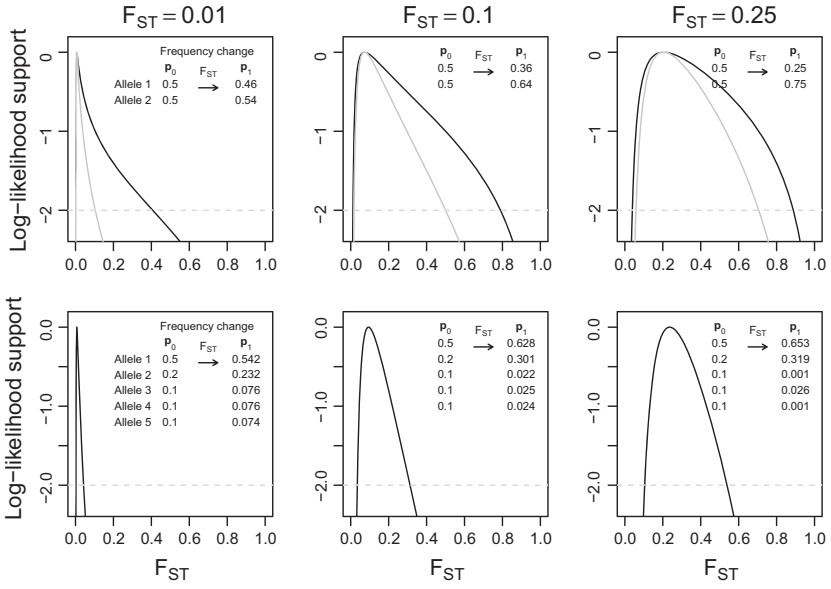

Fig. 1 The likelihood for locus-specific F_{ST} is based on the Dirichlet distribution below (related to eqn 1 in Beaumont & Balding 2004) and known allele frequencies in a population before (the vector of allele frequencies $\boldsymbol{p}_0$) and after ($\boldsymbol{p}_1$) a certain unknown amount of evolution quantified by the parameter F_{ST}. In each plot, curves indicate the likelihood function, which is scaled relative to maximum-likelihood estimate (MLE(h) = 0) and on a $\log_e$ scale. Based on the known allele frequencies, considerable uncertainty regarding the parameter F_{ST} exists if allele frequencies in the initial and final populations differ (dashed grey line indicates two log-likelihood units below the maximum estimate and defines the confidence interval for F_{ST}). In the upper three plots, the likelihood function based on a locus with two alleles is given for a single population ($n = 1$, solid black line) after evolution and for two known populations ($n = 2$, solid grey line). For each plot, allele frequencies change ($\boldsymbol{p}_0 \rightarrow \boldsymbol{p}_1$) and the specified amount of evolution is quantified by the parameter F_{ST}. In the lower three plots, the likelihood function is given for a locus with five alleles with the same three values for F_{ST}. Larger numbers of alleles narrow the confidence intervals for F_{ST} and increase the degrees of freedom.

$$P\left(\boldsymbol{p}_1 \middle| \alpha = \boldsymbol{p}_0\left(\frac{1}{F_{ST}} - 1\right)\right) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_k)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_k)} p_{11}^{\alpha_1 - 1} \cdots p_{1k}^{\alpha_k - 1}$$

uncertainty in any attempt to estimate population genetic structure (e.g. based on F_{ST}) and the underlying evolutionary processes that operated (Fig. 1; e.g. effective population size and the associated expectation for genetic drift, and the operation or strength of selection).

The constraint on inference can be further illustrated with a specific type of population genomic analysis, in this case a linear model for the evolutionary parameter $F_{ST_{ij}}$ for a genetic locus (i) nested within a population (j): $\log\left(\frac{F_{ST_{ij}}}{1-F_{ST_{ij}}}\right) = \mu + \beta_j + \alpha_{j(i)}$ (which is the context in which the constraint became apparent to us; similar models have been utilized previously, e.g. eqn 5 in Beaumont & Balding 2004). This type of model forms the potential basis for identifying locus-specific $F_{ST_{ij}}$ within populations that stand out relative to the distribution across loci (locus-specific deviations from $F_{ST_j} = \mu + \beta_j$). When we used a Bayesian approach to estimating the parameters of this model for a large SNP data set from human populations (Jakobsson *et al.* 2008), we found that point estimates of locus-specific effects ($\alpha_{j(i)}$ and the resulting $F_{ST_{ij}}$) varied over a small range and

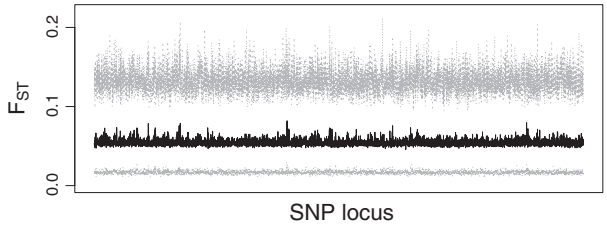

Fig. 2 Bayesian estimates of population and locus-specific $F_{ST_{ij}}$ across 7763 single nucleotide polymorphisms (SNP) on chromosome 22 and based on 14 individuals from the Adygei population in Russia (a representative population and chromosome were chosen arbitrarily; data from Jakobsson *et al.* 2008) relative to an inferred common ancestral population for human diversity. The black line connects the mean of the posterior distribution for each SNP locus, whereas the grey lines delimit the 95% credible interval for each posterior distribution for $F_{ST_{ij}}$. The credible intervals for each locus include the means for all other loci.

were associated with large credible intervals (Fig. 2). Likewise, in previous research, estimates of a similar locus- and population-specific parameter in a linear

model were thought to lack sufficient information (Beaumont & Balding 2004; Foll & Gaggiotti 2008; Riebler *et al.* 2008), which led to the simplifying assumption that loci have the same evolutionary histories across populations and that populations could be used for replication (Foll & Gaggiotti 2008; Riebler *et al.* 2008; Guo *et al.* 2009). The breadth of the credible intervals for the locus- and population-specific evolutionary parameter $F_{ST_{ij}}$ (Fig. 2) should have been expected based on the observation of only a single outcome of an evolutionary process, even with perfect knowledge of population allele frequencies (upper panels in Fig. 1 correspond to SNP data). We emphasize that the '$n = 1$ constraint' does not necessarily lead to incorrect inference (although it can when truly selected and neutral loci cannot be distinguished) but instead will typically result in substantial uncertainty in inference.

Most modelling approaches and analyses explicitly or implicitly assume that the '$n = 1$ constraint' can be circumvented by assuming that most of the genome evolves according to a single evolutionary model (with the exception of selected, 'outlier' loci) and that consequently individual loci are replicate samples from the distribution of outcomes that could result from the single underlying process (e.g. Beaumont & Nichols 1996). Likewise, as noted earlier, some modelling approaches attempt to circumvent this problem by assuming the evolutionary model for each locus is constant across populations, with multiple populations serving as replicate samples (Foll & Gaggiotti 2008; Riebler *et al.* 2008; Guo *et al.* 2009). Most population genomic studies of locus-specific evolutionary histories have used coalescent simulations as a basis for identifying exceptional, 'outlier' loci (see citations in Pérez-Figueroa *et al.* 2010). This method commonly utilizes an analytical approach implemented in software (*fdist* and *fdist2*; Beaumont & Nichols 1996; Beaumont & Balding 2004) that utilizes an empirical estimate of F_{ST} based on the weighted average across loci of the estimator θ (Weir & Cockerham 1984). As a first step in the analysis, the θ estimate of F_{ST} (after removing potentially non-neutral loci) is equated to an evolutionary parameter, the number of migrants (Nm) exchanged between populations in an island model of populations. This inference procedure assumes that all loci and populations used to estimate Nm have experienced a single, common evolutionary process (Whitlock & McCauley 1999). This is analogous to the assumption made in the linear models discussed previously that estimate marginal population or locus effects (Foll & Gaggiotti 2008; Riebler *et al.* 2008; Guo *et al.* 2009). Given that the assumption of homogeneity or constancy of evolutionary history across the genome and populations might not hold (Table 1), it becomes an open question as to whether the commonly used methods based on coalescent simulations and other analytical approaches are robust or sensitive to violations of this assumption. Despite some previous analyses of robustness under certain demographic histories, it is clear that violations of the homogeneity assumption can result in misleading or biased results (e.g. misidentification of outlier loci; Beaumont 2005; Excoffier *et al.* 2009). We think that this issue deserves further attention and that meantime inferences based on allele frequency differences should reflect our caution about this modelling assumption.

A related, but distinct, concern exists about the fundamental importance of the evolutionary models that lead to the detection of unlikely, outlier loci. When a single evolutionary model has been used to apply across the genome, or across populations, outlier loci have been routinely equated with loci that have experienced selection (Beaumont & Balding 2004; Beaumont 2005; Foll & Gaggiotti 2008; Riebler *et al.* 2008; Butlin 2010; Gautier *et al.* 2010). There is the potential that the outlier loci have been selected and therefore are unlikely to have come from the estimated distribution. However, clearly their low probability simply means that they are unlikely to have been derived from the modelled distribution, regardless of the causes. The predominant focus on selection, rather than locus-specific deviations in recombination rate or any other evolutionary process that could have led to the deviation, requires justification, which in many cases will only come from further evidence. Models that utilize a common genomic distribution to identify outliers require that most of the genome is not subject to selection and otherwise evolves according to equal conditions. Thus, future analytical approaches that directly model selection and other evolutionary processes hold promise for the direct estimation of the role of different forces that shape population genomic divergence.

Further progress in population genomics and the detection of the effects of selection on the genome will require viable means to address the '$n = 1$ constraint' and various alternatives warrant further consideration. In the context of model-based analyses of population differentiation (Balding 2003), including the aforementioned linear model (and related models; Beaumont & Balding 2004; Foll & Gaggiotti 2008; Riebler *et al.* 2008; Guo *et al.* 2009), there is the possibility of considering only the marginal effects of population (β_j) or locus (α_i, not nested within population) and to disregard the estimates of locus-by-population interaction (e.g. γ_{ij} in eqn 5 of Beaumont & Balding 2004) or nested locus terms ($\alpha_{j(i)}$ above) that will suffer from the '$n = 1$ constraint'. However, given the high potential for heterogeneity among populations and loci, and that certain alleles are unlikely to have the same beneficial effect on fitness in

all populations, this approach might not always be satisfactory, even if it leads to otherwise sound statistical estimates of the marginal parameters ('main effects'). A possible solution arises from assuming that local windows of the genome have a single evolutionary history and that loci within the region can be used for replication (e.g. sliding window analysis of F_{ST} in Weir *et al.* 2005) or at least provide some mutual information (Gautier *et al.* 2009; Guo *et al.* 2009). Depending on the extent to which this assumption holds for a particular set of data, this might be a viable solution, but the same reasons for scepticism of homogeneity across the whole genome apply to subsets of it. This issue will require further study as population genomics moves towards whole-genome resequencing and the concept of a locus loses definition. In the case of the F_{ST} model analysed here (and other similar models, but not those based on genealogies), analysis of independent pairs or larger numbers of populations might be an improvement (e.g. a pair offers $n = 2$; see upper panels Fig. 1; e.g. Wilding *et al.* 2001; Nosil & Yukilevich 2008; de Carvalho *et al.* 2010), although one that could add considerable complexity to the analysis if many pairs of populations are studied (i.e. when not looking at replicate pairs of populations).

In addition, confidence in parameter estimates is increased if they are based on multi-allelic rather than bi-allelic genetic markers (Fig. 1; Weir & Hill 2002). This statistical observation arises because of the greater degrees of freedom with multiple alleles (i.e. $1 < n < k–1$, where k is the number of alleles), and the increased confidence in parameter estimates can be expected to lead to more powerful analyses as long as the sampling of individuals is sufficient to accurately estimate the frequency of all alleles. There is good reason to expect that researchers will increasingly turn to high-throughput sequencing technologies to produce multi-allelic DNA sequence data rather than relying on bi-allelic SNPs and AFLPs for population genomics (e.g. Gompert *et al.* 2010). Methods for inference that utilize information in addition to allele frequencies at a locus are also promising; these include methods based on linkage disequilibrium as quantified by extended haplotype homozygosity (Sabeti *et al.* 2002). The use of increasingly accurate models of population- and locus-specific evolutionary histories in coalescent-based analyses, as well as the capacity to model and search among a set of plausible, alternative histories, should continue to advance population genomics (e.g. Bazin *et al.* 2010; Beaumont 2010; Peter *et al.* 2010) but will still be subject to the '$n = 1$ constraint'. A substantially different approach to identifying potentially selected loci in population genomics avoids the '$n = 1$ constraint' by analysing individuals as the unit of observation rather than populations and by contrasting locus-specific ancestry with ancestry across the genome (e.g. identifying potentially selected loci in hybrid or admixed individuals; Tang *et al.* 2007; Gompert & Buerkle 2009; Teeter *et al.* 2010). Finally, the study of the effects of selection at the genomic level is likely to be advanced by experimental populations, where careful replication is possible (e.g. Burke *et al.* 2010; Michel *et al.* 2010). Experimental studies of evolution are likely to reveal the diversity of patterns of genomic divergence and adaptive molecular evolution, but depending on the experimental manipulation may or may not be useful for understanding what patterns can be expected in natural, unmanipulated populations.

Recognition of the '$n = 1$ constraint' and the evolutionary model that it represents should affect how we design observational and experimental studies and pose questions in population genomics. The constraint suggests that inferences can be made with greater confidence if they are based on well-justified replication within the genome or across populations, on models that allow for locus- and population-specific deviations, and utilize multi-allelic DNA sequences rather than bi-allelic markers. It is clear that there are good reasons to be cautious in utilizing some existing approaches to analysis and that under some circumstances they are expected to be positively misleading and to misidentify outlier loci (e.g. Excoffier *et al.* 2009). There are now unprecedented opportunities to characterize the genomic consequences of evolutionary processes. Progress in empirical population genomics will be facilitated by researchers who address the '$n = 1$ constraint' and apply appropriate and explicit evolutionary models to data.

Acknowledgements

We appreciate the thoughtful and constructive input we received on these ideas from Jim Fordyce, Christian Lexer, Patrik Nosil, two anonymous reviews and the editor.

References

Andolfatto P, Nordborg M (1998) The effect of gene conversion on intralocus associations. *Genetics*, **148**, 1397–1399.

Andrés AM, Hubisz MJ, Indap A, *et al.* (2009) Targets of balancing selection in the human genome. *Molecular Biology and Evolution*, **26**, 2755–2764.

Balding DJ (2003) Likelihood-based inference for genetic correlation coefficients. *Theoretical Population Biology*, **63**, 221–230.

Bazin E, Dawson KJ, Beaumont MA (2010) Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. *Genetics*, **185**, 587–602.

Beaumont MA (2005) Adaptation and speciation: what can FST tell us?. *Trends in Ecology & Evolution*, **20**, 435–440.

Beaumont MA (2010) Approximate Bayesian Computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 379–406.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B – Biological Sciences*, **263**, 1619–1626.

Burke MK, Dunham JP, Shahrestani P, Thornton KR, Rose MR, Long AD (2010) Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature*, **467**, 587–590.

Bustamante CD, Fledel-Alon A, Williamson S, *et al.* (2005) Natural selection on protein- coding genes in the human genome. *Nature*, **437**, 1153–1157.

Butlin R (2010) Population genomics and speciation. *Genetica*, **138**, 409–418.

de Carvalho D, Ingvarsson PK, Joseph J, *et al.* (2010) Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula L.*), a widespread forest tree. *Molecular Ecology*, **19**, 1638–1650.

Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. *Nature Reviews. Genetics*, **10**, 95–205.

Coop G, Wen X, Ober C, Pritchard JK, Przeworski M (2008) High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science*, **319**, 1395–1398.

Ellegren H, Smith NGC, Webster MT (2003) Mutation rate variation in the mammalian genome. *Current Opinion in Genetics & Development* , **13**, 562–568.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Gautier M, Flori L, Riebler A, *et al.* (2009) A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics*, **10**, 550.

Gautier M, Hocking TD, Foulley JL (2010) A Bayesian outlier criterion to detect SNPs under selection in large data sets. *PLoS ONE*, **5**, e11913.

Gompert Z, Buerkle CA (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology*, **18**, 1207–1224.

Gompert Z, Forister ML, Fordyce JA, Nice CC, Williamson R, Buerkle CA (2010) Bayesian analysis of molecular variance in pyrosequences quantifies population genetic structure across the genome of Lycaeides butterflies. *Molecular Ecology*, **19**, 2455–2473.

Guo F, Dey DK, Holsinger KE (2009) A Bayesian hierarchical model for analysis of single- nucleotide polymorphisms diversity in multilocus, multipopulation samples. *Journal of the American Statistical Association*, **104**, 142–154.

Hardison RC, Roskin KM, Yang S, *et al.* (2003) Covariation in frequencies of substitution, deletion, transposition, and recombination during Eutherian evolution. *Genome Research*, **13**, 13–26.

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, **38**, 226–231.

Holsinger KE, Weir BS (2009) Fundamental concepts in genetics: genetics in geographically structured populations: defining, estimating and interpreting FST. *Nature Reviews. Genetics*, **10**, 639–650.

Jakobsson M, Scholz SW, Scheet P, *et al.* (2008) Genotype, haplotype and copy-number variation in worldwide human populations. *Nature*, **451**, 998–1003.

Jeffreys A, May C (2004) Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nature Genetics*, **36**, 151–156.

Jeffreys AL, Royle NJ, Wilson V, Wong Z (1988) Spontaneous mutation rates to new length alleles at tendem-repetitive hypervariable loci in human dna. *Nature*, **332**, 278–281.

Kauppi L, Jasin M, Keeney S (2007) Meiotic crossover hotspots contained in haplotype block boundaries of the mouse genome. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **104**, 13396–13401.

Lercher MJ, Hurst LD (2002) Human snp variability and mutation rate are higher in regions of high recombination. *Trends in Genetics*, **18**, 337–340.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews. Genetics*, **4**, 981–994.

Marais G (2003) Biased gene conversion: implications for genome and sex evolution. *Trends In Genetics*, **19**, 330–338.

McMullen MD, Kresovich S, Villeda HS, *et al.* (2009) Genetic properties of the maize nested association mapping population. *Science*, **325**, 737–740.

Michel AP, Sim S, Powell THQ, Taylor MS, Nosil P, Feder JL (2010) Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 9828–9832.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.

Noor MAF, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, **103**, 439–444.

Noor MAF, Feder JL (2006) Speciation genetics: evolving approaches. *Nature Reviews. Genetics* , **7**, 851–861.

Nosil P, Yukilevich R (2008) Mechanisms of reinforcement in natural and simulated polymorphic populations. *Biological Journal of the Linnean Society*, **95**, 305–319.

Pérez-Figueroa A, Garcýa-Pereira MJ, Saura M, Rolán-Alvarez E, Caballero A (2010) Com- paring three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, **23**, 2267–2276.

Peter BM, Wegmann D, Excoffier L (2010) Distinguishing between population bottleneck and population subdivision by a bayesian model choice procedure. *Molecular Ecology*, **19**, 4648–4660.

Riebler A, Held L, Stephan W (2008) Bayesian variable selection for detecting adaptive genomic differences among populations. *Genetics*, **178**, 1817–1829.

Rockman MV, Skrovanek SS, Kruglyak L (2010) Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science*, **330**, 372–376.

Sabeti PC, Reich DE, Higgins JM, *et al.* (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature*, **419**, 832–837.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.

Tang H, Choudhry S, Mei R, *et al*. (2007) Recent genetic selection in the ancestral admixture of Puerto Ricans. *The American Journal of Human Genetics*, **81**, 626–633.

Teeter KC, Thibodeau LM, Gompert Z, Buerkle CA, Nachman MW, Tucker PK (2010) The variable genomic architecture of isolation between hybridizing species of house mouse. *Evolution*, **64**, 472–485.

Turner TL, Hahn MW (2010) Genomic islands of speciation or genomic islands and speciation? *Molecular Ecology*, **19**, 848–850.

Vicoso B, Charlesworth B (2006) Evolution on the X chromosome: unusual patterns and processes. *Nature Reviews. Genetics*, **7**, 645–653.

Weir BS, Cardon LR, Anderson AD, Nielsen D, Hill W (2005) Measures of human population structure show heterogeneity among genomic regions. *Genome Research*, **15**, 1468–1476.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population-structure. *Evolution*, **38**, 1358–1370.

Weir BS, Hill WG (2002) Estimating F-statistics. *Annual Review of Genetics*, **36**, 721–750.

Whitlock M, McCauley D (1999) Indirect measures of gene flow and migration: F-ST not equal 1/(4Nm+1). *Heredity*, **82**, 117–125.

Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using aflp markers. *Journal of Evolution Biology*, **14**, 611–619.

The authors are broadly interested in the genetics of adaptation and speciation and develop and use modeling approaches to analyze their own and collaborator's empirical data. Z.G. is a Ph.D. student, T.L.P. is a postdoctoral researcher, and C.A.B. is an associate professor at the University of Wyoming.
