

# Speciation as a sieve for ancestral polymorphism

Rafael F. Guerrero<sup>1</sup>  | Matthew W. Hahn<sup>1,2</sup> 

<sup>1</sup>Department of Biology, Indiana University, Bloomington, IN, USA

<sup>2</sup>Department of Computer Science, Indiana University, Bloomington, IN, USA

## Correspondence

Rafael F. Guerrero, Department of Biology, Indiana University, Bloomington, IN, USA.  
Email: rafguerr@indiana.edu

## Funding information

Division of Molecular and Cellular Biosciences, Grant/Award Number: MCB-1127059; National Science Foundation, Grant/Award Number: MCB-1127059

## Abstract

Because they are considered rare, balanced polymorphisms are often discounted as crucial constituents of genome-wide variation in sequence diversity. Despite its perceived rarity, however, long-term balancing selection can elevate genetic diversity and significantly affect observed divergence between species. Here, we discuss how ancestral balanced polymorphisms can be “sieved” by the speciation process, which sorts them unequally across descendant lineages. After speciation, ancestral balancing selection is revealed by genomic regions of high divergence between species. This signature, which resembles that of other evolutionary processes, can potentially confound genomic studies of population divergence and inferences of “islands of speciation.”

## KEYWORDS

allopatric speciation, balancing selection, differential gene flow, genomic divergence

## 1 | INTRODUCTION

Studying the process of speciation using patterns of genomic divergence between species requires that we understand the determinants of genetic diversity within species. Because sequence diversity in an ancestral population determines the starting point from which divergent populations accumulate differences (Gillespie & Langley, 1979), any evolutionary forces that shape diversity within species can have a large impact on measures of divergence between species. These forces include those both decreasing (e.g., selective sweeps: Begun et al., 2007; Cruickshank & Hahn, 2014; or background selection: Phung, Huber, & Lohmueller, 2016) and increasing variation (e.g., balancing selection: Charlesworth, 2006). Selection can increase diversity by favouring the maintenance of polymorphism via overdominance, frequency dependence and heterogeneous selection. Nevertheless, balanced polymorphisms are considered rare in nature, and such loci are often overlooked as major contributors to genome-wide variation in levels of sequence diversity and divergence.

Here, we argue that speciation can act as a “sieve” that will reveal otherwise elusive balancing selection by sorting ancestral balanced polymorphisms unequally across descendant lineages. By sorting alternative alleles between different species, this process uncovers the existence of balancing selection in the ancestral population as regions of higher-than-expected divergence. Sorted

ancestral polymorphism may also be responsible for many of the observed peaks of genomic divergence between closely related taxa, mimicking the patterns produced by other processes and potentially confounding population genomic studies (including those of differential gene flow and “islands of speciation”).

## 2 | THE EFFECT OF BALANCING SELECTION ON GENETIC DIVERSITY

Balancing selection encompasses various, rather disparate, mechanisms that favour the maintenance of polymorphism. Most forms of balancing selection involve heterogeneous or variable selective forces: polymorphism is favoured when selection varies across space or time (reviewed in Felsenstein, 1976), between the sexes (reviewed in Otto et al., 2011), or as a function of allele frequency (negative frequency-dependent selection; reviewed in Ayala & Campbell, 1974). Balanced polymorphisms can also occur under constant selective pressures when heterozygotes have a fitness advantage over homozygotes (overdominance, Wright, 1931; see also Connallon & Clark, 2013).

Despite clear differences in their underlying mechanisms, all forms of balancing selection have qualitatively similar long-term effects on linked neutral variation (reviewed in Charlesworth, 2006;

Fijarczyk & Babik, 2015). Genomic regions closely linked to loci under balancing selection are expected to show increased divergence between the allelic classes defined by the balanced polymorphism (Barton & Bengtsson, 1986; Charlesworth, Nordborg, & Charlesworth, 1997). This increase in divergence—the extent of which depends on population size, recombination rate and strength of selection—results from a type of population subdivision imposed by the balanced polymorphism. When two (or more) balanced alleles persist for a long time in a population, closely linked regions tend to accumulate differences between the allelic classes. Thus, the genealogies of samples linked to balanced polymorphisms resemble those of structured populations, with allelic classes acting as subpopulations and recombination allowing “migration” between these subpopulations. Migration itself can, of course, play an important role in shaping the genealogies of these regions, for example, when the polymorphism is being balanced by spatially heterogeneous selection. However, interallelic genetic divergence by balancing selection can arise within a fully panmictic population.

The genomic signal of balancing selection can be difficult to detect when the balanced alleles at a locus are not known, because samples cannot be partitioned by allelic class. This makes it difficult to quantify the increased divergence between allelic classes directly (for example, using the statistic  $F_{ST}$ ). Instead, the only observable signals of balanced polymorphism are an excess of intermediate frequency alleles (detected using Tajima's  $D$ ), or overall increases in diversity (detected using the average number of pairwise differences, or  $\pi$ ). Unfortunately, the large amount of variance in both  $D$  and  $\pi$  associated with even neutrally evolving loci means that loci under balancing selection are hard to identify (Simonsen, Churchill, & Aquadro, 1995).

The potential difficulty in detecting balanced polymorphism is somewhat alleviated in cases of local adaptation. Because alternative balanced alleles differ in frequency between the populations where they are individually advantageous, measures of differentiation between populations can be used as a proxy for divergence between allelic classes. Moreover, actual population structure can lower the effective recombination rate between allelic classes, exacerbating their divergence. Perhaps due to its increased detectability, spatially varying selection is reported in natural populations much more frequently than other forms of balancing selection (Asthana, Schmidt, & Sunyaev, 2005; Charlesworth, 2006; Delph & Kelly, 2014; Fan, Hansen, Lo, & Tishkoff, 2016). Indeed, while overdominance and negative frequency dependence continue to be considered rare, local adaptation is considered pervasive and is even a compulsory first step in some models of speciation (Nosil, 2012).

### 3 | THE SIEVE: ANCESTRAL LINEAGE SORTING AFTER SPECIATION

Consider a simple case of allopatric speciation: a single population is split in two by vicariance (e.g., the rise of a mountain range or the construction of a thousand-mile wall). For neutral biallelic polymorphisms in the ancestral population, drift will fix alternative alleles in

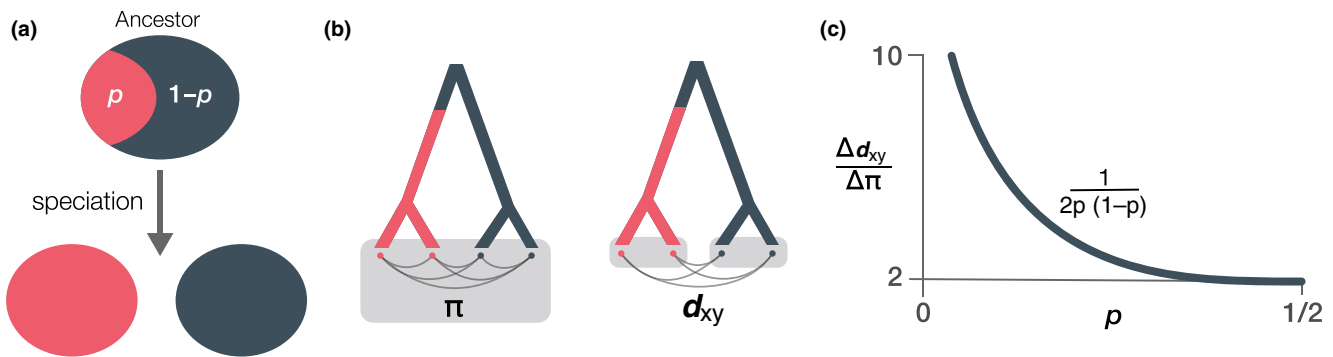
the two nascent species at half of all such loci (the rest of the time both species will fix the same allele). In the presence of balancing selection, expectations can differ. Balancing selection may favour the maintenance of polymorphism in both nascent species (resulting in “trans-specific polymorphism”; Muirhead, Glass, & Slatkin, 2002), or increase the chance that alternative alleles are fixed. For instance, when selection varies across space, a geographic barrier is likely to create two unequal ranges (i.e., areas with different proportions of habitats driving local adaptation) that may favour dramatically different equilibrium frequencies at the locally adapted loci in the nascent species. This could result in selection favouring the fixation of opposite alleles in each species, sieving ancestral alleles in the descendant lineages (Figure 1a). On the other hand, balancing selection may disappear after speciation, nonetheless leaving behind highly diverged (and selectively neutral) haplotypes that could be sieved by drift.

Sieved balanced polymorphisms carry the signature of selection in the form of increased sequence divergence between the descendant species. Immediately after the split, the level of genetic divergence is largely determined by the diversity present in the ancestral population. Formally, this can be seen in the expectation of absolute divergence:  $E(d_{XY}) = 2\mu t + \theta_{Anc}$  (Gillespie & Langley, 1979). As the time since the species split ( $t$ ) approaches zero,  $E(d_{XY})$  becomes approximately equal to the ancestral level of diversity,  $\theta_{Anc}$  ( $=4N_e\mu$  for diploids, where  $N_e$  is the effective population size and  $\mu$  is the neutral mutation rate). Initial levels of divergence are therefore strongly affected by forces that affect levels of ancestral neutral diversity, such as balancing selection. In regions linked to sieved balanced polymorphism, measures of divergence between species (such as  $d_{XY}$  or  $F_{ST}$ ) reflect the divergence accumulated between allelic classes both in the ancestor and since the lineages split, so sieved polymorphisms maintained in the ancestor for a long time can appear as regions of elevated divergence between nascent species.

Interestingly, this implies that balancing selection could be more readily detected after speciation (as a sieved polymorphism with elevated  $d_{XY}$ ) than in the ancestor when the causal alleles are unknown (using  $D$  or  $\pi$ ; Figure 1b). Indeed, the signature of balancing selection is expected to be twice as strong on  $d_{XY}$  than on  $\pi$  at loci with alleles maintained at equal frequencies in the ancestor (Figure 1c; Appendix S1). As either allele becomes rarer, the effect is more severe: when the minor allele frequency was 10% in the ancestor, the increase in  $d_{XY}$  is expected to be almost six times larger than the increase in  $\pi$ . Therefore, by separating the relevant haplotypes, speciation can dramatically increase our power to find loci under balancing selection.

### 4 | RELEVANCE OF SIEVED POLYMORPHISMS DURING RECENT GENOMIC DIVERGENCE

The prevalence of sieved polymorphisms in nature is unknown (and is probably low), but the importance of this phenomenon for



**FIGURE 1** (a) Schematic of a sieved polymorphism. In the ancestor, two alleles are balanced at frequencies  $p$  and  $1-p$ , and after speciation different alleles fix in descendant lineages. (b) The increased power to detect balancing selection stems from the partitioning of the ancestral population imposed by the sieve. While  $\pi$  is a measure of all pairwise distances in a population (left),  $d_{XY}$  compares only samples from different allelic classes (right). (c) The ratio of the effect of a balanced polymorphism on diversity ( $\Delta\pi = \pi - \pi_o$ , where  $\pi_o$  is the baseline diversity) and on divergence ( $\Delta d_{XY} = d_{XY} - d_o$ , and  $d_o = \pi_o$ ) increases as the minor allele becomes rare. As the time since the origin of the polymorphism increases, the ratio converges to the inverse of the expected heterozygosity,  $1/2p(1-p)$

observed patterns of genomic divergence does not stem from its frequency. The amount of balanced polymorphism sieved by speciation is proportional to the fraction of loci under balancing selection in the ancestor and the probability of fixing alternative alleles at those loci (which depends on the mode of selection operating at each locus). If balancing selection is as rare as usually assumed (Charlesworth, 2006), sieved polymorphisms are likely to be uncommon and would not drastically elevate average levels of divergence. Instead, the few instances of sieved polymorphisms will have subtle but significant repercussions: these regions will tend to appear at the top of the distribution of divergence across the genome, “fattening” its upper tail and potentially affecting further inferences. The potential consequences for the distribution of divergence depend not only on the fraction of loci sieved but also on parameters specific to each polymorphism (namely, its age, strength of selection and recombination rate). If, for instance, a genome carries only one sieved polymorphism, it will likely appear as a divergence outlier. On the other hand, a higher fraction of sieved regions—caused, for instance, by numerous locally adapted alleles differentiated between populations prior to speciation—will considerably fatten the upper tail in the divergence distribution. Such an observation could be interpreted as evidence of a period of differential gene flow following an initial split (cf. Yang, He, Shi, & Wu, 2017). This latter case highlights the fact that sieved polymorphism can mimic the signature of other evolutionary processes, and distinguishing among these may be challenging without additional pieces of evidence (see below).

Recent findings suggest that sieved polymorphisms do play an important role in shaping patterns of genomic divergence. Multiple studies have found regions with levels of divergence so high that differentiation at these loci almost certainly started before speciation, consistent with ancestral balanced polymorphism. In these cases, the timing of speciation—or at least a bound on the timing—can be independently estimated, highlighting the mismatch between species divergence and genetic divergence.

In the radiation of Darwin’s finches, speciation events happened approximately between 50 and 500 thousand years ago (Lamichaney et al., 2015), yet some genomic regions seem to have started differentiating well before then (up to one million years ago; Han et al., 2017). At least two of these genomic regions, linked to loci associated with beak shape and size (genes *ALX1* and *HMGA2*), are likely sieved polymorphisms. Across the nine species of tree and ground finches studied, these loci have two distinct haplotype classes (i.e., there is high divergence between classes and reduced divergence within class, even between species) that are responsible for marked phenotypic differences (blunt vs. pointed beaks and small vs. large beaks). As a result, species pairs that have fixed different haplotype classes show “islands of divergence” at these loci. As hypothesized by Han et al. (2017), however, this beak polymorphism was probably balanced in the ancestor (perhaps under negative frequency-dependent selection) and was later sieved across the Galapagos. These observations highlight an additional implication of the speciation sieve: ancestral polymorphism may persist through multiple speciation events, especially rapid radiations, before it is sieved. As a result, multiple species pairs will share sieved regions and may appear as if they had undergone repeated local adaptation (e.g., Campagna et al., 2017).

In the freshwater threespine sticklebacks of western North America, the *Eda* locus represents a clear example of how ancestral polymorphisms can emerge as conspicuous peaks of genomic divergence. Polymorphism at *Eda* has been maintained in marine populations for approximately two million years, and the minor allele has been selected repeatedly during colonization events of glacial lakes around ten thousand years ago (Colosimo et al., 2005). Expectedly, *Eda* shows dramatic differentiation between marine and freshwater populations, but most of this divergence happened before the invasion of the glacial lakes and is unrelated to recent processes.

Other types of polymorphism can also be sieved. Among these, chromosome rearrangements (e.g., inversions, fusions) are of special interest for their role during local adaptation. Rearrangements can

evolve by capturing locally adapted alleles (Guerrero & Kirkpatrick, 2014; Kirkpatrick & Barton, 2006; Yeaman, 2013), and established rearrangements can promote further local adaptation (Navarro & Barton, 2003). Moreover, balanced rearrangements (especially inversions) are usually conspicuous in population genomic data (e.g., Cheng et al., 2012; Kapun, Fabian, Goudet, & Flatt, 2016), as they typically cause a dramatic reduction in recombination, which in turn leads to much stronger population subdivision compared to other balanced polymorphisms (Guerrero & Kirkpatrick, 2014; Guerrero, Rousset, & Kirkpatrick, 2012). Due to their role in local adaptation and their large genomic footprint, rearrangements are thought to be key players in the build-up of differentiation that can lead to speciation. Some chromosome inversions have in fact been linked to speciation processes (e.g., in cactophilic *Drosophila*, Lohse, Clarke, Ritchie, & Etges, 2015; *Mimulus*, Fishman, Stathos, Beardsley, Williams, & Hill, 2013). In other cases, however, locally adapted inversions are maintained as polymorphisms within a species or as trans-specific polymorphisms—without being involved in speciation. In the *Anopheles gambiae* species complex, inversion *2La* arose in the ancestor of six species (well before the most recent speciation events), and it is still polymorphic in two of these (Fontaine et al., 2015). This inversion has been sieved at least two times, such that comparisons between species fixed for alternative arrangements show increased divergence across many megabases of sequence.

## 5 | IMPLICATIONS FOR INFERENCES FROM POPULATION GENOMIC STUDIES

Clearly, the patterns described above are not unique to sieved polymorphisms. Several evolutionary processes can produce regions of high divergence, potentially confounding inferences. After speciation, for instance, introgression from a third lineage into one of the nascent species will result in regions of high divergence between sister taxa (e.g., Stankowski & Streisfeld, 2015). Introgression events may be hard to distinguish from ongoing balancing selection (Fijarczyk & Babik, 2015) and could cause patterns similar to sieved polymorphisms if the introgressed regions fix. The duplication of a region and reciprocal loss in two daughter lineages can lead to the misassignment of paralogs as orthologs (“pseudoorthologs”; Koonin, 2005). As a result, divergence at the focal locus will appear much higher than the genome-wide average when duplication happened much before speciation.

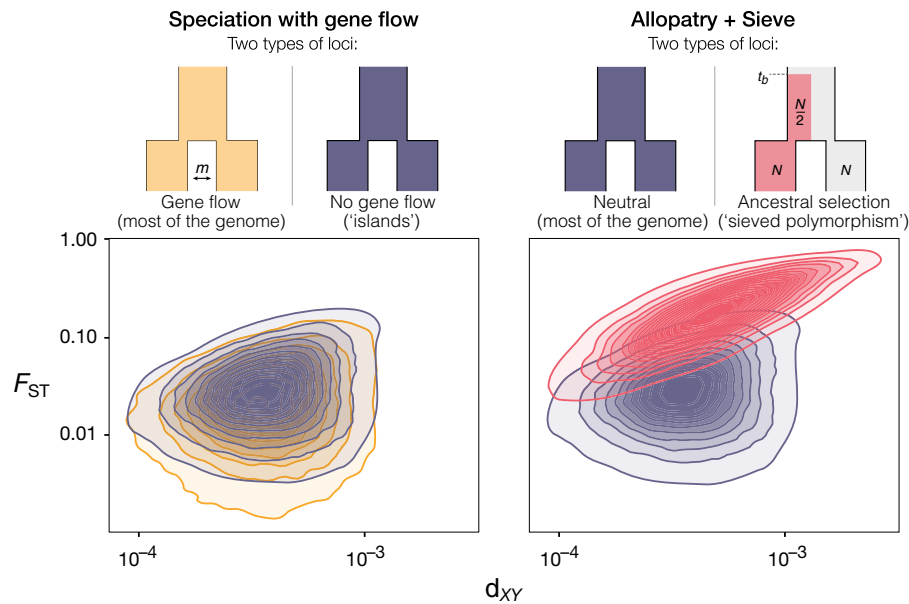
Sieved polymorphisms mimic predictions of models of speciation with gene flow. In the most common version of this model (which is conceptually similar to sympatric and parapatric speciation models; reviewed in Bush, 1975; Via, 2001), populations have uninterrupted exchange of migrants, but achieve total reproductive isolation gradually by the accumulation of locally adapted loci that limit effective gene flow (cf. Charlesworth et al., 1997). At the genomic level, differential gene flow is expected to leave a clear signature: regions linked to loci under divergent selection accumulate higher differentiation compared to the rest of the genome, appearing as “genomic

islands of speciation” (Turner, Hahn, & Nuzhdin, 2005). If the speciation process continues (with uninterrupted gene flow), an island may expand by incorporating new divergent loci (Aeschbacher & Burger, 2014; Yeaman, Aeschbacher, & Burger, 2016). At its origin, however, an island of speciation is essentially a locally adapted locus—and therefore has the same genomic signature.

Variation in divergence across the genome has been attributed to differential gene flow among loci, with individual loci showing much higher levels of divergence implicated as being causal in the speciation process. Several studies initially reported finding such islands using relative measures of divergence (such as  $F_{ST}$ ), which can be affected by selection in the sampled populations (e.g., Ellegren et al., 2012; Geraldine, Basset, Smith, & Nachman, 2011; Turner et al., 2005). Because absolute measures of divergence (such as  $d_{XY}$ ) are unaffected by current levels of polymorphism, it was suggested that these would be preferred in identifying regions that are truly resistant to introgression (Cruickshank & Hahn, 2014). It has therefore become more common to search for islands using  $d_{XY}$  and related statistics, and some researchers have reported finding these important loci (e.g., Malinsky et al., 2015; Marques et al., 2016). Using absolute measures of divergence, however, does not obviate the problem of variation in levels of diversity in the ancestral population. As discussed above, regions of elevated  $d_{XY}$  can be produced by other processes—including ancestral balanced polymorphism—and variance in levels of divergence across the genome can be driven by variance in diversity in ancestral populations. It is simply not true that all loci start out equally diverged at speciation or that (ongoing) differential gene flow is the only force that can produce variation in  $d_{XY}$  beyond that expected from neutral coalescent variation in the ancestor.

How would one distinguish between true islands and sieved balanced polymorphisms? One commonality shared by the clearest examples of sieved polymorphisms given above is that independent estimates exist for the earliest time when speciation could have started. Glacial lakes that could not have existed prior to the retreat of the glaciers, radiations onto geological features (such as oceanic islands) that recently appeared on the landscape, or simply the date of an earlier divergence from a more distantly related species, all limit the maximum time pairs of focal species could have been separated. Given such limits, we can then contrast hypotheses of speciation with gene flow with those involving sieved polymorphisms (Figure 2). In fact, it takes quite a long time for loci resistant to gene flow to appear as divergence outliers under models of speciation with gene flow (Figure 2a; Figure S1; also see fig. B1 in Cruickshank & Hahn, 2014). By contrast, sieved balanced polymorphisms are expected to be detected by both relative and absolute measures of divergence immediately after a vicariance event (Figure 2b; alternative alleles are assumed to fix instantly after vicariance by selection or drift).

We can apply these ideas to assess the likely causes of “islands of divergence” in two cases for which there are independent data on genomic divergence and the timing of speciation. In the cichlids of Lake Massoko, levels of divergence might stem from ancestral



**FIGURE 2** Distributions of absolute and relative divergence ( $d_{XY}$  and  $F_{ST}$ ) for genomes under two scenarios of recent speciation (1,000 generations ago, population size  $N = 10^4$  for each species). In each scenario, genomes have two types of regions. On the left, regions experience differential gene flow since the split: while in most of the genome (in light orange,  $m = 0.0001$ ) gene flow prevents divergence, in “speciation islands” (in purple,  $m = 0$ ) there is a slight increase in  $F_{ST}$ . On the right, there is no gene flow after speciation, but some regions are tightly linked to a sieved polymorphism (in pink; balanced locus is at  $r = 10^{-5}$  from the simulated region, originated  $t_b = 10^4$  generations ago, stable at frequency of  $\frac{1}{2}$  in ancestor, alternative alleles are fixed in descendants). We simulated the genealogy for a sample of 20 chromosomes drawn from each species ( $10^5$  coalescent simulations for each type of genomic region, a 10 Kb non-recombining segment with  $\mu = 10^{-8}$ )

polymorphism (Malinsky et al., 2015). The speciation process in the focal pair started some 3,500 generations ago (Malinsky et al., 2015), which would allow enough time for the accumulation of divergence on the order of  $d_{XY} = 1.05 \times 10^{-4}$  (assuming  $\mu = 1.5 \times 10^{-8}$  and  $\theta_{Anc} = 0$ ). However,  $d_{XY}$  observed in the most highly diverged regions is considerably higher than this expectation (mean  $d_{XY} = 9 \times 10^{-4}$ ; Malinsky et al., 2015; <https://twitter.com/milane1/status/758209899964862465>), suggesting that a large fraction of the observed differentiation is due to ancestral diversity. This rough calculation ignores many factors, such as variation in mutation rate, that can contribute to the observed patterns. However, it allows us to emphasize that the genomes of extant populations give us a glimpse into ancestral processes that transcend the most recent speciation event.

Similarly, in the threespine stickleback of Lake Constance, subspecies show several regions of divergence that most likely predate the current process of local adaptation (which started about 150 generations ago), and for which standing variation has been invoked as a probable source (Marques et al., 2016). In this case, high differentiation was inferred in 37 genomic regions based on allele frequency differences among populations (using SNPs obtained via RAD-seq). The observed diversity levels in these regions are not significantly reduced, suggesting that—while current selection may be driving allele frequency divergence—the accumulation of divergent SNPs is not the result of a recent sweep. Rather, many of these “islands” are likely ancestral balanced haplotypes currently being sorted.

Confounding sieved polymorphism, or other sources of variance in genomic divergence, with islands of speciation can lead to an additional erroneous inference: that large amounts of variation in divergence imply divergence in the face of gene flow. To some extent, these inferences follow from the observation of islands—if there are loci resistant to gene flow, then it follows that there must have been gene flow. But this false signal can also affect methods for inferring gene flow that assume that there is no selection, and therefore interpret the excess variance observed as due to migration. It has recently been recognized that modelling the effects of selection on the levels of sampled polymorphism is important in controlling such false positives (Roux et al., 2016). The implication here is that variation in levels of ancestral polymorphism must also be considered, as it can lead to false inference of recent gene flow and current selection (i.e., labelling sieved regions, which may be neutral now, as resistant to gene flow).

## 6 | CONCLUSIONS

Sieved polymorphism—in conjunction with factors such as population structure, assortative mating, introgression, background selection or variation in mutation and recombination rates—contributes to heterogeneity in genomic divergence levels. Due to the complexity of the divergence distribution, inferences that rely solely on its outliers (e.g., taking an arbitrary upper quantile of  $d_{XY}$  as speciation islands) can yield



misleading results by selecting regions, such as sieved polymorphisms, that are unrelated to speciation. Model-based analyses are necessary but not sufficient, because extremely similar patterns of genomic divergence can be generated by alternative models of speciation. In fact, biologically significant differences between speciation models are occasionally irrelevant from a theoretical standpoint (e.g., cessation of gene flow is modelled identically [ $m = 0$ ] regardless of the mechanism behind it, whether due to hybrid inviability, a geographic barrier, or other). For this reason, independent lines of evidence are critical to disentangle the multiple forces at play. For instance, having a lower bound on the time since speciation can allow us to determine how much ancestral polymorphism is expected, as its effect is strongest in recent speciation events (i.e.,  $t < 2N_e$ , when  $\theta_{Anc}$  accounts for more than half of  $E(d_{XY})$ ). Using current levels of polymorphism as a proxy for ancestral levels in model-based analyses will also be a useful starting point in trying to understand the causes of variation in divergence levels. Finally, judicious analysis of many new population genomic data sets being produced may unearth numerous sieved polymorphisms, which could cause us to reconsider the presumed rarity of balancing selection beyond local adaptation.

## ACKNOWLEDGEMENTS

We thank Nolan Kane, Sam Yeaman and an anonymous reviewer for valuable comments. Support comes from National Science Foundation grant MCB-1127059.

## AUTHOR CONTRIBUTIONS

R.F.G. and M.W.H. conceived and wrote this paper.

## ORCID

Rafael F. Guerrero  <http://orcid.org/0000-0002-8451-3609>  
 Matthew W. Hahn  <http://orcid.org/0000-0002-5731-8808>

## REFERENCES

- Aeschbacher, S., & Burger, R. (2014). The effect of linkage on establishment and survival of locally beneficial mutations. *Genetics*, *197*, 317–336.
- Asthana, S., Schmidt, S., & Sunyaev, S. (2005). A limited role for balancing selection. *Trends in Genetics*, *21*, 30–32.
- Ayala, F. J., & Campbell, C. A. (1974). Frequency-dependent selection. *Annual Review of Ecology and Systematics*, *5*, 115–138.
- Barton, N., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity (Edinb)*, *57*(Pt 3), 357–376.
- Begun, D. J., Holloway, A. K., Stevens, K., Hillier, L. W., Poh, Y. P., Hahn, M. W., ... Langley, C. H. (2007). Population genomics: Whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biology*, *5*, e310.
- Bush, G. L. (1975). Modes of animal speciation. *Annual Review of Ecology and Systematics*, *6*, 339–364.
- Campagna, L., Repenning, M., Silveira, L. F., Fontana, C. S., Tubaro, P. L., & Lovette, I. J. (2017). Repeated divergent selection on pigmentation genes in a rapid finch radiation. *Science Advances*, *3*, e1602404.
- Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, *2*, 379–384.
- Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, *70*, 155–174.
- Cheng, C. D., White, B. J., Kamdem, C., Mockaitis, K., Costantini, C., Hahn, M. W., ... Besansky, N. J. (2012). Ecological genomics of *Anopheles gambiae* along a latitudinal cline: A population-resequencing approach. *Genetics*, *190*, 1417–1432.
- Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G. Jr., Dickson, M., Grimwood, J., ... Kingsley, D. M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science*, *307*, 1928–1933.
- Connallon, T., & Clark, A. G. (2013). Antagonistic versus nonantagonistic models of balancing selection: Characterizing the relative timescales and hitchhiking effects of partial selective sweeps. *Evolution*, *67*, 908–917.
- Cruikshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*, 3133–3157.
- Delph, L. F., & Kelly, J. K. (2014). On the importance of balancing selection in plants. *New Phytologist*, *201*, 45–56.
- Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., ... Wolf, J. B. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, *491*, 756–760.
- Fan, S., Hansen, M. E., Lo, Y., & Tishkoff, S. A. (2016). Going global by adapting local: A review of recent human adaptation. *Science*, *354*, 54–59.
- Felsenstein, J. (1976). The theoretical population genetics of variable selection and migration. *Annual Review of Genetics*, *10*, 253–280.
- Fijarczyk, A., & Babik, W. (2015). Detecting balancing selection in genomes: Limits and prospects. *Molecular Ecology*, *24*, 3529–3545.
- Fishman, L., Stathos, A., Beardsley, P. M., Williams, C. F., & Hill, J. P. (2013). Chromosomal rearrangements and the genetics of reproductive barriers in *Mimulus* (monkey flowers). *Evolution*, *67*, 2547–2560.
- Fontaine, M. C., Pease, J. B., Steele, A., Waterhouse, R. M., Neafsey, D. E., Sharakhov, I. V., ... Besansky, N. J. (2015). Mosquito genomics. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, *347*, 1258524.
- Geraldes, A., Basset, P., Smith, K. L., & Nachman, M. W. (2011). Higher differentiation among subspecies of the house mouse (*Mus musculus*) in genomic regions with low recombination. *Molecular Ecology*, *20*, 4722–4736.
- Gillespie, J. H., & Langley, C. H. (1979). Are evolutionary rates really variable? *Journal of Molecular Evolution*, *13*, 27–34.
- Guerrero, R. F., & Kirkpatrick, M. (2014). Local adaptation and the evolution of chromosome fusions. *Evolution*, *68*, 2747–2756.
- Guerrero, R. F., Rousset, F., & Kirkpatrick, M. (2012). Coalescent patterns for chromosomal inversions in divergent populations. *Philosophical Transactions of the Royal Society London B*, *367*, 430–438.
- Han, F., Lamichhaney, S., Grant, B. R., Grant, P. R., Andersson, L., & Webster, M. T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Research*, *27*, 1004–1015.
- Kapun, M., Fabian, D. K., Goudet, J., & Flatt, T. (2016). Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. *Molecular Biology and Evolution*, *33*, 1317–1336.
- Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, *173*, 419–434.
- Koonin, E. V. (2005). Orthologs, paralog, and evolutionary genomics. *Annual Review of Genetics*, *39*, 309–338.
- Lamichhaney, S., Berglund, J., Almen, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., ... Andersson, L. (2015). Evolution of Darwin's

- finches and their beaks revealed by genome sequencing. *Nature*, 518, 371–375.
- Lohse, K., Clarke, M., Ritchie, M. G., & Etges, W. J. (2015). Genome-wide tests for introgression between cactophilic *Drosophila* implicate a role of inversions during speciation. *Evolution*, 69, 1178–1190.
- Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., ... Turner, G. F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, 350, 1493–1498.
- Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2016). Genomics of rapid incipient speciation in sympatric Threespine Stickleback. *PLoS Genetics*, 12, e1005887.
- Muirhead, C. A., Glass, N. L., & Slatkin, M. (2002). Multilocus self-recognition systems in fungi as a cause of trans-species polymorphism. *Genetics*, 161, 633–641.
- Navarro, A., & Barton, N. H. (2003). Chromosomal speciation and molecular divergence—accelerated evolution in rearranged chromosomes. *Science*, 300, 321–324.
- Nosil, P. (2012). *Ecological speciation* Oxford University Press. New York, NY: Oxford.
- Otto, S. P., Pannell, J. R., Peichel, C. L., Ashman, T. L., Charlesworth, D., Chippindale, A. K., ... McAllister, B. F. (2011). About PAR: The distinct evolutionary dynamics of the pseudoautosomal region. *Trends in Genetics*, 27, 358–367.
- Phung, T. N., Huber, C. D., & Lohmueller, K. E. (2016). Determining the effect of natural selection on linked neutral divergence across species. *PLoS Genetics*, 12, e1006199.
- Roux, C., Fraisse, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology*, 14, e2000234.
- Simonsen, K. L., Churchill, G. A., & Aquadro, C. F. (1995). Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics*, 141, 413–429.
- Stankowski, S., & Streisfeld, M. A. (2015). Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. *Proceedings of the Royal Society B: Biological Sciences*, 282, 154–162.
- Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, 3, 1572–1578.
- Via, S. (2001). Sympatric speciation in animals: The ugly duckling grows up. *Trends in Ecology & Evolution*, 16, 381–390.
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16, 97–159.
- Yang, M., He, Z., Shi, S., & Wu, C. I. (2017). Can genomic data alone tell us whether speciation happened with gene flow? *Molecular Ecology*, 26, 2845–2849.
- Yeaman, S. (2013). Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences of the USA*, 110, E1743–E1751.
- Yeaman, S., Aeschbacher, S., & Burger, R. (2016). The evolution of genomic islands by increased establishment probability of linked alleles. *Molecular Ecology*, 25, 2542–2558.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Guerrero RF, Hahn MW. Speciation as a sieve for ancestral polymorphism. *Mol Ecol*. 2017;26:5362–5368. <https://doi.org/10.1111/mec.14290>