Posted on Authorea 3 Jan 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.160968166.65928724/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

The population genomics of invasive species

Henry North[1], Angela McGaughran[2], and Chris Jiggins[1]

[1]University of Cambridge
[2]University of Waikato

January 3, 2021

Abstract

By studying invasive species, evolutionary geneticists have been able to simultaneously inform management strategies and quantify rapid evolution in the wild. The role of genomics in invasion science is increasingly recognised, and the growing availability of reference genomes for invasive species is paving the way for whole-genome resequencing studies in a wide range of systems. Here, we survey the literature to assess the application of whole-genome resequencing data in invasion biology. For some applications, such as the reconstruction of invasion routes in time and space, sequencing the whole genome of many individuals simply increases the accuracy of existing methods. In other cases, population genomic approaches such as haplotype analysis can permit entirely new questions to be addressed and new technologies to be applied. To date whole-genome resequencing has only been applied to a handful of invasive systems, but these studies have highlighted important roles for processes such as balancing selection and hybridization that allow invasive species to reuse existing adaptations and rapidly overcome the challenges of a foreign ecosystem. The use of genomic data does not constitute a paradigm shift *per se*, but by leveraging new theory, tools, and technologies, population genomics can provide unprecedented insight into basic and applied aspects of invasion science.

Introduction

As an unintended consequence of global commerce and climate change, biodiversity is being redistributed at an unprecedented rate (Ding, Mack, Lu, Ren, & Huang, 2008; Muirhead, Minton, Miller, & Ruiz, 2015; Ricciardi, 2007; Sardain, Sardain, & Leung, 2019; Seebens et al., 2015). Many introductions fail to form viable populations on foreign soil, but those that go on to establish and spread — invasive species — are a dominant cause of biodiversity declines and a major threat to global food security (Clavero, Brotons, Pons, & Sol, 2009; Clavero & García-Berthou, 2005; Maxwell, Fuller, Brooks, & Watson, 2016; Oerke, 2006). Invasion biology is an interdisciplinary field that aims to understand the transport, establishment and spread of invasive species and inform management strategies that mitigate their impact. Invasion genetics has proven to be an essential part of this effort (Barrett, 2015). From a purely biological perspective, many invasive species are ideal natural experiments that enable the observation of rapid adaptation, parallel evolution, inter- and intra-specific hybridization and speciation in the wild (C. E. Lee, 2002; Prentis, Wilson, Dormontt, Richardson, & Lowe, 2008; Vallejo-Marin & Hiscock, 2016). Quantifying such phenomena in invasive species will inevitably shed light on the factors that facilitate their transport, establishment and spread. A cursory examination of past issues of *Molecular Ecology* will attest to the long history of invasion geneticists working on pure and applied aspects of invasion biology in equal measure.

High-throughput sequencing is now broadly recognised as an important tool for monitoring, managing and mitigating the impact of invasive species (Chown et al., 2015; Hamelin & Roe, 2020; Rius, Bourne, Hornsby, & Chapman, 2015; Tay & Gordon, 2019). As a result, there has been a recent increase in the availability of reference genomes for invasive species, laying the groundwork for population resequencing projects (McCartney, Mallez, & Gohl, 2019). With the ease of transferable skills between study systems in molecular

Posted on Authorea 3 Jan 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.160968166.65928724/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

population genetics, the cost of sequencing continuing to decline, general bioinformatics literacy continuing to increase, and a greater recognition of the value of DNA sequence data in invasion science, we anticipate that whole-genome resequencing (WGR) will become a crucial tool in invasion genetics.

[BOX 1]

What can population genomics add to the field of invasion genetics? The decision to sequence multiple whole genomes is not straightforward. This is primarily because progress in the field of invasion genetics is limited by a lack of manipulative experiments at least as much as it is limited by a lack of genome sequence data (Bock et al., 2015). Additionally, WGR is still a non-trivial cost in many systems and reduced-representation sequencing is in some cases sufficient to address key questions in invasion genetics (see Box 1). Here, we review the extent to which WGR has been adopted in the field of invasion genetics, assessing its existing and potential impact beyond other sequencing technologies (*e.g.* , transcriptomics, reduced-representation sequencing or comparative genomics of single reference genomes). To achieve this, we assessed 1,614 publications that appeared in a Web of Science search using the term "invasive species" [?] "genom*" ! "cancer". By combining the results of this search with recent pre-prints at the time of writing, a total of 31 studies that used WGR to study invasive species were identified (Supplementary Table 1, examples shown in Figure 1). We highlight key case studies form this list, summarise theoretical considerations relevant to the population genomics of invasive species, and highlight newly developed technologies and analyses that enable novel insights though the use of WGR. Population genomic studies of native species evolving in response to invasive species, and studies of pathogens, are outside the scope of this review.

[FIGURE 1]

We discuss five broad themes in invasion biology research that have involved, or could benefit from, high-throughput sequencing. These are 1) the role of pre-invasion adaptation in enabling subsequent spread; 2) tools to reconstruct invasion routes in space; 3) demographic inference to reconstruct the timing of invasion events, which also sheds light on the role of population bottlenecks during invasion; 4) post-introduction adaptation as a driver of spread in novel bioregions; and finally 5) the role of hybridization and introgression during invasion, which brings together all four of the preceding themes. These key themes reflect the focus of existing WGR papers in invasion biology and span the temporal range of the invasion sequence (Figure 2).

[FIGURE 2]

1: Measuring the contribution of pre-invasion selection to invasion success

1.1: Time-series data can distinguish between species destined for invasion and those that adapt in situ

One of the oldest postulations in the field of invasion biology is that some species are "predisposed" to invasiveness (Baker, 1965). In other words, some traits that facilitate spread in a new environment are not adaptations that arise following colonization (see Part 4), but instead exist in native populations prior to transport (Figure 2). For example, the kudzu bug (*Megacopta cribraria*) was introduced from Asia to North America where it initially grew on kudzu, but within 9 months of detection had begun exploiting soy crops. Studies in the native range had shown that the genotype of the symbiont *Candidatus*Ishikawaella *capsulata* ('Ishikawaella' hereafter) mediates*M. cribraria* 's ability to grow on soy (Hosokawa, Kikuchi, Nikoh, Shimada, & Fukatsu, 2006). Therefore, it was initially unclear whether the switch to soy was enabled by the evolution of Ishikawaella in the invasive range. Brown *et al.* (2014) inferred the evolution of the symbiont Ishikawaella by sequencing its genome at various locations, including the founding population in the year it was first detected, and at different time points since. Their analysis revealed that the founding population closely resembled native Japanese Ishikawaella samples known to enable growth on soy, with little evidence of allele frequency changes during invasion. This suggests that *M. cribraria* and its symbiont Ishikawaella had arrived in the US already able to spread on soy plantations. Though the 750kbp Ishikawaella genome is sequenced at scale with relative ease, this study exemplifies the significant and under-exploited benefits of WGR time-series data as a means of tracking genome-wide shifts in allele frequency during biological

Posted on Authorea 3 Jan 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.160968166.65928724/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

invasion. This approach could in principle be applied to any invasive species to distinguish between pre- and post-introduction adaptation.

1.2: Balancing selection maintains adaptive diversity that can facilitate invasion

Unlike the example above, pre-existing adaptations that increase invasive spread in the invaded range can sometimes experience a different selective regime in the native range. Lee and Gelembuik (2008) argue that fluctuating selection pressures in the native range (*e.g.* , through regular disturbance events) can maintain either genetic variation or phenotypic plasticity, which can then be acted on by positive selection in novel environments. This prediction has been tested in the copepod species complex *Eurytemora affinis* (Stern & Lee, 2020). Several native populations from high-salinity environments have independently invaded freshwater ecosystems in North America. Seasonal fluctuations in salinity in the native environment, combined with several life history traits of *E. affinis* including overlapping generations, create favourable conditions for balancing selection. In the invaded range, a genome-wide scan showed parallel signatures of directional selection at ion transport genes in replicate populations (Figure 3, upper panel). Furthermore, the loci identified as being under positive selection in the introduced range also showed signatures of long-term balancing selection in the native range (Figure 3, lower panel), consistent with the hypothesis that fluctuating selection had maintained genetic variation that could enable invasive spread in a novel environment. Through the use of genomic data, localised signatures of both directional and balancing selection could be detected without any prior knowledge of the genes or traits underlying invasion success.

[FIGURE 3]

2: *Reconstructing invasion routes in space*

2.1: A phylogeographic perspective on invasive species

A useful application of population genetics is to quantify patterns of isolation-by-distance to identify genetically distinct management units, infer the source population(s) of invasive species, and estimate the timing of introduction(s) (Cristescu, 2015). Both phylogenetic and assignment-based population genetic methods, typically using mitochondrial DNA or microsatellite samples from native and invasive populations, have been applied to hundreds of invasive species over several decades (reviewed by Estoup & Guillemaud, 2010). The same methods can be used with WGR data, which can add resolution especially in systems with little population structure (*e.g.* , recent introductions or high inter-population connectivity). For example, WGR data were used to trace incursions of *Spodoptera frugiperda* (fall armyworm) into Africa and the Asia-Pacific region with unprecedented resolution (Tay et al., 2020); to show that the cosmopolitan crop pest *Plutella xylostella* (diamondback moth) most likely originated in South America (You et al., 2020); and to infer multiple independent introductions of *Aedes aegypti* into California (Lee et al., 2019).

2.2: Making the most of genomic data

Population genomic data are best suited to analytical tools designed to work efficiently with large datasets and make the most of the available information. To this end, a number of new analytical approaches have been developed to infer the geographic origin of a genomic sample using continuous spatial models (*e.g.* , Battey, Ralph, & Kern, 2020; Guillot, Jonsson, Hinge, Manchih, & Orlando, 2016). Due to their computational efficiency, such measures can also be used to estimate the geographic origin of a sample in chromosomal windows. This feature is particularly useful when tracing the geographic origin of a candidate locus (*e.g.,* a haplotype containing a pesticide resistance gene or a QTL known to be associated with invasiveness) or when investigating the contribution of different source populations across the genome. For example, *Locator* (Battey et al., 2020) has been used to identify the geographic origin of *Anopheles* samples (Figure 4).*Locator* is one of several new analytical tools in population genomics that make use of machine learning (reviewed by Schrider & Kern, 2018; see also Flagel, Brandvain, & Schrider, 2019). Modern tools for geographic inference have been taken up more readily in other fields to date (*e.g.* , forensic investigations into illegal wildlife trade), though they hold great potential in invasion science as a means of biomonitoring — for instance, determining the origin of invasive species intercepted at ports at a fine spatial scale.

Posted on Authorea 3 Jan 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.160968166.65928724/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

[FIGURE 4]

WGR increases our ability to infer the geographic origin of invasive species, though an associated increase in precision depends on the analytical tools used. Many analyses optimised for genomics can be used with reduced-representation datasets as well, so the application of WGR will depend on the nature of the system and the question being asked (see Box 1). For example, analysis of RAD was used to quantify ongoing migration rates of *Aedes albopictus* in Australia, and to identify the source of the incursions using *Locator* (Schmidt et al., 2020). WGR can also add value to geographic inference where there is a weak signature of population structure, or in studies of admixed invasive populations aiming to infer the geographic origin of specific loci (see Part 5).

[BOX 2]

3: *Inferring demographic change during invasion and its quantifying its impact on invasion success*

In the simplest model of a successful invasion, some small fraction of a native population is transported to a new environment where a viable population forms and increases in size over time. This leads to a demographic bottleneck (a decrease in population size followed by an increase), which can have wide-ranging implications for both practical and theoretical aspects of invasion genetics (see Box 2). From a pragmatic perspective, the extreme demographic dynamics of colonization can be used to reconstruct invasion events in time. From a biological perspective, invasion geneticists are tasked with explaining how so many invasive species form viable populations, let alone dominate foreign ecosystems, in the face of a population bottleneck that reduces genetic diversity and increases the risk of inbreeding depression (Estoup et al., 2016). We discuss these two perspectives separately.

3.1 Population bottlenecks as a timestamp in the genome

Population bottlenecks, detected using demographic inference methods, can help to infer the number and timing of independent colonization events. For example, there was a clear signature of population bottleneck in the invasive fall webworm (*Hyphantria cunea*), although this predated introduction to China in 1979 (Wu et al., 2019). In contrast, there was no signal of a population bottleneck in invasive North American populations of the common carp, which instead shared a similar demographic history to putative ancestral populations in Europe (Yuan et al., 2018). A criticism of these studies is that they both used a sequentially Markovian coalescent (SMC), which is most appropriate for inferring demography in deep time (McVean & Cardin, 2005; Patton et al., 2019). Methods that use the site frequency spectrum, on the other hand, tend to perform better when inferring recent demographic change (Patton et al., 2019). A further complexity is that invasions often involve sequential introductions and admixture among differentiated source populations or even species, making their origins much harder to identify. Nonetheless, the timing and magnitude of admixture can be inferred by analysing the size distribution of haplotypes (Harris & Nielsen, 2013; see Box 3). In particular, Bayesian estimates of complex demographic parameters can be improved with the use of whole-genome sequence data compared to reduced-representation sequencing, largely due to the information added by haplotype statistics (Smith & Flaxman, 2020).

Whole genome sequence data should, in general, give more accurate demographic inference. This is because demographic inference in the recent past relies on the shallow coalescence times of low-frequency alleles. Assuming that sequencing is conducted at sufficient depth to accurately call low-frequency alleles (see Box 1), whole genome sequence data are therefore more likely to capture rare alleles required to time recent bottlenecks (Hahn, 2019). For example, Puckett *et al.*(2020) set out to test the hypothesis that a 1768 shipwreck introduced the brown rat (*Rattus norvegicus*) to the Faroe Islands using reduced representation sequencing. Although three introduction events could be inferred, the authors were unable to estimate the timing of each event due to a lack of rare alleles, which were removed through the very bottlenecks they were attempting to date. Thus, in some cases the detection of recent bottlenecks is an empirically intractable problem. However, because whole genome sequence data are much more likely to capture the rare allele frequency data required to time recent bottlenecks, WGR may give better resolution in examples

Posted on Authorea 3 Jan 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.160968166.65928724/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

such as this. In summary, although other sequencing technologies, such as transcriptome sequencing and microsatellite markers, have shown considerable success in reconstructing invasion histories (*e.g.* , Fontaine et al., 2020; Popovic, Matias, Bierne, & Riginos, 2020), WGR data can increase resolution and will be especially useful in estimating the time of a recent invasion event.

3.2 Population bottlenecks as a paradox to resolve

A central aim in invasion genetics is to understand the general impact of genetic drift on invasion success beyond individual case studies (Bock et al., 2015). Debates about the role of genetic drift in invasion success are as old as the field of invasion genetics (Baker & Stebbins, 1965; Barrett, 2015). Even today researchers seek to resolve the "paradox of biological invasion" by explaining how invasive species rapidly adapt to new environments despite a loss in genetic diversity, a reduction in the efficiency of natural selection and an increased risk of inbreeding depression (Estoup et al., 2016). Additionally, if an introduced population can overcome (or avoid) these challenges, it must then somehow endure expansion load (see Box 2). WGR data are well suited to examining this apparent paradox.

First, WGR can be used to reliably test whether an introduced population has experienced a population bottleneck, and to distinguish among different demographic scenarios that that explain the observed level of genetic diversity (Smith & Flaxman, 2020; Welles & Dlugosch, 2018). For example, higher genetic diversity was seen in invasive populations of *P. xylostella* compared to their native range, despite clear evidence of a population bottleneck, apparently as a result of admixture among introduced populations (You et al., 2020). Second, WGR allows quantification of the genomic landscape of genetic diversity, which can be used to better understand the processes that mediate diversity. Comeault *et al.* (2020) show that, although introduced populations of *Zaprionus indianus* (African fig fly) have lower genetic diversity than native populations, diversity is proportionally lower in regions of low recombination. This result indicates the effect of linked selection and suggests that the reduction in diversity in the introduced range is not solely due to demography. Finally, WGR can be used to observe subtle detrimental effects of population bottlenecks, such as expansion load (see Box 2). Thus, population genetic analyses applied to WGR data can shed light on longstanding fundamental questions in invasion biology.

[BOX 3]

4: Identifying post-introduction adaptation

4.1 Approaches to measuring post-introduction adaptation

Perhaps the most exciting application of genomics in invasion biology is in the identification of loci underlying post-introduction adaptation. Although there are relatively few examples to date, it seems likely that this will be the major contribution of genomic approaches over the coming decade. In terms of the invasion sequence, adaptive change in the invaded range can either end the lag phase by facilitating spread or accelerate spread in an already-invasive population (Prentis et al., 2008) (Figure 2). Although *in situ* adaptation is thought to play an important role in many systems, it has often been difficult to quantify the contribution of an adaptive trait to the rate of spread (Bock et al., 2015). Addressing these questions will not only contribute to a general understanding of biological invasion but will also provide information for integrated management strategies.

4.2 Forward-genetics in the wild

"Top-down" forward-genetic approaches start with a particular trait and dissect its genetic basis. There are few examples of this approach using WGR data in invasion biology. Forward genetic approaches have historically involved QTL mapping in invasive species reared under controlled conditions. However, if an invasiveness trait is easy to score in wild individuals, genome-wide association studies can be used with WGR. For example, the genetic basis of wing length was investigated with admixture mapping using WGR from field-collected samples of introduced honey bees (*Apis mellifera*), though no major effect loci were identified (Calfee, Agra, Palacio, Ramirez, & Coop, 2020). In another WGR study, on *Aedes aegypti* throughout the native sub-Saharan range, a handful of major effect loci underlying preference for human odour were

Posted on Authorea 3 Jan 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.160968166.65928724/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

identified (Rose et al., 2020). Although the latter study was conducted within the native range of *A. aegypti* , the trait of human preference appears to contribute to its spread into urban habitats.

4.3 Scans for the genomic signature of selection

WGR has more commonly contributed to reverse-genetic approaches, where whole-genome scans are used to identify loci that have been subject to selection without directly knowing the traits involved. In this way, inferences can be made about the genetic basis and evolutionary history of adaptation even when the ecology and life history of an invasive species is poorly understood. There are various ways to identify the signature of natural selection from genomic datasets. When studying a single invasive population, the footprint of a selective sweep can be identified from the site frequency spectrum of genetic variation (DeGiorgio, Huber, Hubisz, Hellmann, & Nielsen, 2016; Nielsen et al., 2005). Alternatively, comparisons between populations (*e.g.* , between different timepoints during invasion or between native and invasive populations) can be used to identify regions of high divergence, using summary statistics such as F_{ST}or the population branch statistic (Yi et al., 2010). Another rarely exploited approach to measuring adaptation in invasive species is the use of sequence data collected in a time series – analogous to 'evolve-and-resequence' experiments carried out in laboratory populations (Long, Liti, Luptak, & Tenaillon, 2015; Schlotterer, Kofler, Versace, Tobler, & Franssen, 2015). Not only can this approach be used to rule out pre-invasion adaptation (see Part 1.1), it also provides a powerful framework in which to identify allele frequency shifts resulting from simple or polygenic adaptation (Buffalo & Coop, 2020; Otte & Schlotterer, 2020). Where samples are not readily available from early timepoints in the invasion, historical museum or herbarium samples can be used to infer past allele frequencies (Bi et al., 2019; McGaughran, 2020). All the approaches mentioned above can be used with SNPs, transposable elements or structural variants, which are readily detectable using WGR data and difficult to measure using other sequencing technologies (Bertolotti et al., 2020).

There are already a handful of examples of selection scans being used in invasion biology. For example, a genome-wide scan for association with invasiveness in 16 invasive and 6 native populations of *Drosophila suzukii* identified SNPs in two genes associated with independent invasion routes (Olazcuaga et al., 2020). Using a similar approach that controlled for population structure, genome scans across the global distribution of *P. xylostella* identified three potentially novel insecticide resistance alleles (You et al., 2020), and signatures of positive selection were associated with sugar receptor genes in in*Hyphantria cunea* (mulberry moth) (Wu et al., 2019). Other studies have made use of WGR data by identifying structural variants and transposable elements, investigating their effect on fitness in invasive populations. For example, again in *Drosophila suzukii* , fifteen putative adaptive transposable elements were identified, one of which was 399bp from a SNP previously associated with invasion success in this species (Merel et al., 2020). In this way, WGR can identify an otherwise invisible dimension of genetic variation.

It is has long been realised that genome scans for selection need to account for background genomic processes that can lead to false positives for adaptive loci. These can include genetic drift caused by demographic changes and selective processes such as background selection. In some cases, the peculiar biology of invasive species makes them especially prone to such problems, as genetic bottlenecks can lead to signatures of reduced variation similar to those caused by selection (see Box 2). Furthermore, any summary statistic capturing the coalescent process will be influenced by variation in recombination rate (c) and changes in the effective population size (N_e) (Barton & Etheridge, 2004; Booker, Yeaman, & Whitlock, 2020; Brandvain & Wright, 2016). N_e and c can be estimated empirically with WGR data. Changes inN_e can be inferred using demographic inference methods (see Part 3.1), while recombination rate variation along the genome can be estimated by constructing a linkage map or with phased WGR data (Chan, Jenkins, & Song, 2012). User-friendly modelling tools, such as *SLiM* , can be used to explore the expected distribution of summary statistics under various combinations ofN_e and c (Haller, Galloway, Kelleher, Messer, & Ralph, 2019; Haller & Messer, 2019). Tests for selection that explicitly incorporate demography and recombination can also be used (*e.g.* , Luqman, Widmer, Fior, & Wegmann, 2020). Therefore, despite the confounding effects of recombination rate variation and demographic history on the summary statistics used in genome scans, it is now easier than ever to identify and account for these effects.

Posted on Authorea 3 Jan 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.160968166.65928724/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

5: Quantifying Inter- and Intra-specific hybridization during invasion

5.1 Mapping introgression during invasion: new motivations, new tools

Hybridization within and between species has long been recognised as a potentially important process mediating invasion success (Bock et al., 2015). However, in part because of the increasing use of genomic data, there is now a much greater appreciation of the true extent of intra- and interspecific hybridization during invasions. Hybridization has become a central part of invasion genetics (Grabenstein & Taylor, 2018; McFarlane & Pemberton, 2019; Todesco et al., 2016; Viard, Riginos, & Bierne, 2020). For example, newly updated models of the expected spatial distribution of introgression following invasive hybridization can be used to reconstruct invasion routes (Quilodran, Tsoupas, & Currat, 2020). There are a wealth of different strategies for detecting introgression (Malinsky, 2019), and adaptive introgression in particular, that to date have not been applied to invasive species (Gower, Picazo, Fumagalli, & Racimo, 2020; Malinsky, 2019; Setter et al., 2020; Svedberg, Shchur, Reinman, Nielsen, & Corbett-Detig, 2020).

5.2 Intraspecific hybridization

As discussed in Part 3.2, a longstanding challenge in invasion biology is to explain how invasive species overcome or avoid the deleterious consequences of a demographic bottleneck. One solution to this challenge is seen where invasions involve admixture among multiple genetically differentiated source populations (Cristescu, 2015; Smith et al., 2020). For example, population genomic studies of the invasive fungus *Cryphonectria parasitica* (causing chestnut blight) and the fall armyworm (*Spodoptera frugiperda*) show that gene flow among invasive lineages maintains genetic diversity (Demene et al., 2019; Tay et al., 2020; Yainna et al., 2020). Admixture not only alleviates the effects of inbreeding depression but can lead to the sorting of adaptive alleles into beneficial combinations. This may often explain the 'bridgehead effect', where an initially successful invasion acts as a source of colonists for subsequent invasions (Lombaert et al., 2010). Rispe *et al.* (2020) provided a recent example, showing that multiple native North American populations of the viticultural pest *Daktulosphaira vitifoliae* were introduced to France. The invasive hybrid French population then acted as a bridgehead for the subsequent invasion of vineyards throughout the rest of Europe, and possibly throughout the Southern Hemisphere.

Strong genomic evidence for the sorting of adaptive alleles following the hybridization of genetically differentiated populations has also come from replicate studies of geographic clines. In introduced Australian and North American populations of *Drosophila melanogaster* , an F_{ST} outlier scan was used to identify polymorphisms responsible for parallel latitudinal clines in both continents (Bergland, Tobler, Gonzalez, Schmidt, & Petrov, 2016). In both cases, invasive populations were the result of hybridization between African and European populations. High-latitude populations in both North America and Australia share more ancestry with native European populations, in contrast to low-latitude populations that share more African ancestry.

[FIGURE 5]

Similarly, African honey bees (*Apis mellifera scutellata*) were introduced to the Americas in 1957 where they hybridize with, and often outcompete, conspecifics with European ancestry (Figure 5A) (Calfee, Agra, Palacio, Ramirez, et al., 2020). Genome-wide geographic clines in African ancestry in both California and Argentina (Figure 5B) are concordant with phenotypic clines in wing length, consistent with a model of highly polygenic divergent selection in response to differences in climate. Using this framework of migration-selection balance, loci that introgressed further along the transect than expected (*i.e.* , those loci important for spread, subjected to different spatially

varying selection pressure) could be identified (Figure 5C). In both of these examples, spatially varying selection in hybrid populations has sorted standing genetic variation along environmental gradients, facilitating the rapid adaptation of these invasive species.

5.3 Interspecific hybridization

Interspecific hybridization is now recognised as being reasonably frequent, occurring in some 10% of ani-

Posted on Authorea 3 Jan 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.160968166.65928724/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

mal and 25-30% of plant species (Mallet, 2005; Rieseberg, Wood, & Baack, 2006). Increasingly, through unintended anthropogenic dispersal, pairs of species are interacting for the first time since they last shared a common ancestor (Grabenstein & Taylor, 2018; McFarlane & Pemberton, 2019; Muirhead et al., 2015; Seebens et al., 2015). This opens the possibility of adaptive gene exchange that could contribute to invasion success in much the same way it does within species (Hovick & Whitney, 2014). However, interspecific hybridization will more commonly produce unfit offspring compared to intraspecific admixture, as reproductive incompatibilities are more likely to have accumulated. The trade-off between the cost of hybridization and the benefit of adaptive introgression creates ideal conditions for the study of speciation.

One example of interspecific introgression during invasion is the Iberian hare, *Lepus granatensis* , which replaced the now-extinct Arctic species, *L. timidus* , in its northern range. *IL12B* , a gene implicated in the inflammatory process and immune response to viruses in rabbits, underwent adaptive introgression from *L. timidus* to *L. granatensis* , potentially contributing to its northern range expansion following the last glacial maximum (Seixas, Boursot, & Melo-Ferreira, 2018). Similarly, some introduced populations of the three-spined stickleback (*Gasterosteus aculeatus*) have higher genetic diversity as a result of introgression from *G. nipponicus* (Yoshida et al., 2016).

In some cases, hybridization might contribute to increased fitness of native species. For example, the crop pests *Helicoverpa armigera*and *H. zea* developed strong prezygotic barriers to hybridization following more than one million years of divergence in allopatry (Laster & Sheng, 1995). *H. armigera* was introduced to Brazil within the past decade, where it encountered the native *H. zea* . Whole genome resequencing has shown that the pesticide resistance allele*CYP337B3* was subsequently introduced from *H. armigera* to*H. zea* , increasing its ability to evade previously effective control measures (Valencia-Montoya et al., 2020).

Another related phenomenon is hybridization among multiple co-invading species. Examples include fishes in the genera *Hypophthalmichthys*(carp) and *Cottus* (freshwater sculpins), and invasive fungi in the genus *Ophiostoma* (the cause of Dutch elm disease) (Dennenmoser et al., 2017, 2019; Hessenauer et al., 2020; Wang et al., 2020). Notably, interspecific introgression among *Ophiostoma*species increased genetic diversity and was associated with individual growth rate (Hessenauer et al., 2020). Moreover, introgression on chromosome 1 was positively associated with virulence, apparently as a consequence of adaptive introgression (Hessenauer et al., 2020). These observations suggest that interspecific hybridization can create novel combinations of adaptive variants that enhance spread, in addition to mitigating the negative impacts of population bottlenecks by maintaining genetic diversity.

Regardless of whether it increases the spread of invasive species, interspecific hybridization can threaten local biodiversity as a result of genetic swamping (where local genotypes are replaced by hybrids) or demographic swamping (where local population growth rates are reduced via outbreeding depression) (Todesco et al., 2016). Genetic swamping is both a potentially cryptic mode of extinction (Todesco et al., 2016) and a mechanism by which genetic material from introduced domesticated species can dominate populations of wild relatives (Haygood, Ives, & Andow, 2003). The high resolution of WGR means that it can be a powerful tool for monitoring and quantifying genetic swamping.

Conclusion

Many of the key questions in invasion genetics highlighted by Bock*et al.* (2015) remain unanswered, though our ability to obtain and interpret genome sequence data has matured substantially in the past five years. Though WGR data are certainly not a singular solution to outstanding questions about biological invasion, we are increasingly appreciating their potential; of the studies we assessed, over one third were published in the past year and more than two thirds were published within the past two years (Supplementary Table 1). Whilst some research questions have more to gain from WGR than others (*e.g.* , quantifying hybridization versus spatial inference), appropriately designed population genomics studies can address multiple questions about invasions simultaneously. Indeed, almost without exception, the examples we have highlighted addressed hypotheses from many areas of interest.

Based on existing genomic data, processes that maintain adaptive genetic diversity (*i.e.* , balancing selection,

Posted on Authorea 3 Jan 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.160968166.65928724/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

admixture and adaptive interspecific introgression) are often key to the success of invasive species (*e.g.* , Calfee et al., 2020; Hessenauer et al., 2020; Stern & Lee, 2020; Valencia-Montoya et al., 2020; Yainna et al., 2020). In other words, standing genetic diversity that has already been shaped by natural selection is often repurposed to rapidly overcome adaptive challenges; invasive species do not have time to re-invent the wheel. This is not a new observation. Several authors have argued that the same 'combinatorial' evolutionary processes known to facilitate major ecological transitions and adaptive radiations can also enable biological invasion (Hegarty, 2012; Marques, Meier, & Seehausen, 2019; Prentis et al., 2008; Rieseberg et al., 2003). Genomic data have revealed the frequency of this phenomenon among invasive species. It is now clear that the 'paradox' of biological invasion is often explained not only by the fact that many invasive species avoid the negative effects of demographic bottlenecks, but also because they avoid the need for *de novo* mutation followed by *in situ* adaptation. This implies a genic view of biological invasion in which the primary aim of management strategies should be to minimise the spread of alleles known to confer invasive success through introduced populations, and potentially through reproductively compatible native populations.

In the studies we surveyed, the most substantial individual contributions to our understanding of invasive species did not come from the largest datasets, but from studies that associated phenotypic or spatial information with genomic data in a hypothesis-driven design that incorporated appropriate modelling (*e.g.* , Calfee, Agra, Palacio, Ramirez, et al., 2020; Olazcuaga et al., 2020; Stern & Lee, 2020). Future genomic studies will contribute considerably to our understanding of pre- and post-introduction adaptation if they adopt such an approach. Useful inferences will also require development of tailored non-equilibrium models that can incorporate the full complexity of invasive species. In particular, an exciting area is understanding the evolutionary history of loci that contribute to adaptive spread. This will be easiest when studying recent invasions with samples taken over a time series. The second, more challenging step will be to quantify the marginal contribution of positively selected loci to invasive spread. Increasing use of forward-genetic approaches to dissect the genetic basis of invasiveness traits will likely make this task easier. Together, these approaches can be used to test whether the 'combinatorial' view of invasion success holds up as a general trend.

Given the pervasive role of hybridization in invasion success, alongside the declining cost of sequencing, we anticipate that whole genome sequence data will become a standardised approach for monitoring the ongoing global redistribution of biodiversity. Comprehensive genomic datasets will eventually allow invasion events to be consistently reconstructed at a resolution that is useful for informing management plans and they will put us in a better position to quantify the contribution of specific mechanisms to overall invasion success.

FIGURES AND BOXES

Hosted file

`image1.emf` available at https://authorea.com/users/386791/articles/502257-the-population-genomics-of-invasive-species

Figure 1: Examples of invasive species that have been analysed using whole-genome resequencing data. Population genomic analyses have been conducted in invasive species from all kingdoms, though the phylum Arthropoda is over-represented, constituting half of all the studies we assessed (Supplementary Table 1). **A)** *Eurytemora affinis*species complex, Pacific clade (see Stern & Lee, 2020) photographed by Carol Eunmi Lee. **B)** Diamondback Moth (*Plutella xylostella*) (see Wei et al., 2017) photographed by Jessa Thurman.**C)** Yellow Fever Mosquito (*Aedes aegypti*) (see Lee et al., 2019) photographed by Chen Wu. **D)** European honey bee (*Apis mellifera*) (see Calfee, Agra, Palacio, Ramirez, & Coop, 2020) photograph from Bees of Australia: a Photographic Exploration (CSIRO Publishing), used with permission from the author.

Posted on Authorea 3 Jan 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.160968166.65928724/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

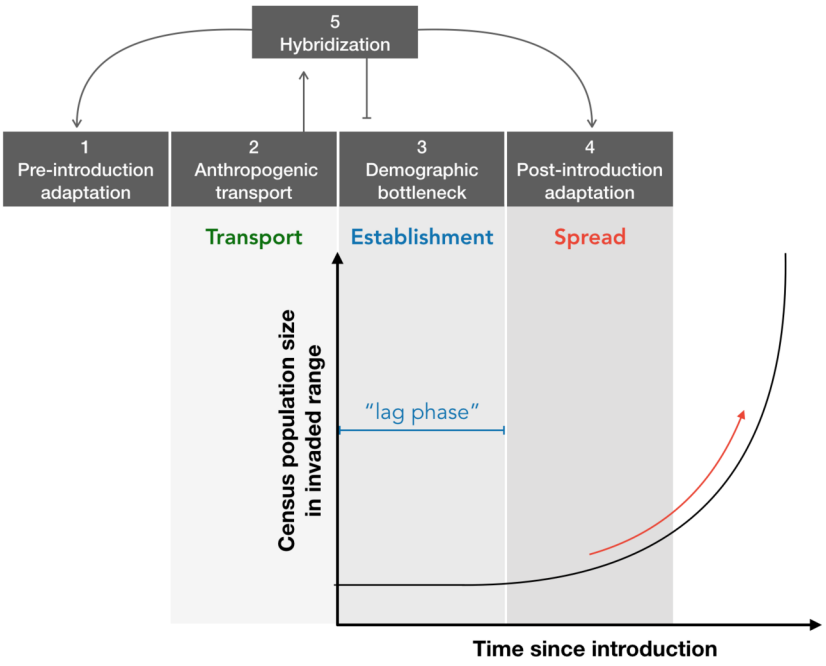


Figure 2: Through anthropogenic dispersal, many species are increasingly being transported to new bioregions. Only some of these species are able to establish a viable population in the foreign environment and, of these, a smaller fraction will grow exponentially to become 'invasive'. The time between initial colonization and rapid population growth is known as the lag phase (Sakai et al., 2001). The simplicity of this model of biological invasion, known as the invasion sequence, means that its components can be parameterised in terms of ecology and evolutionary biology (Kolar & Lodge, 2001; Lodge, 1993; Sakai et al., 2001). For example, it is critical to understand the eco-evolutionary dynamics that underlie the transition from establishment to spread, or the traits that allow some species but not others to establish small yet viable populations. Across the temporal extent of the invasion sequence, we discuss five research themes in invasion genetics that whole-genome resequencing can shed light on.

Posted on Authorea 3 Jan 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160906816.65928724/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

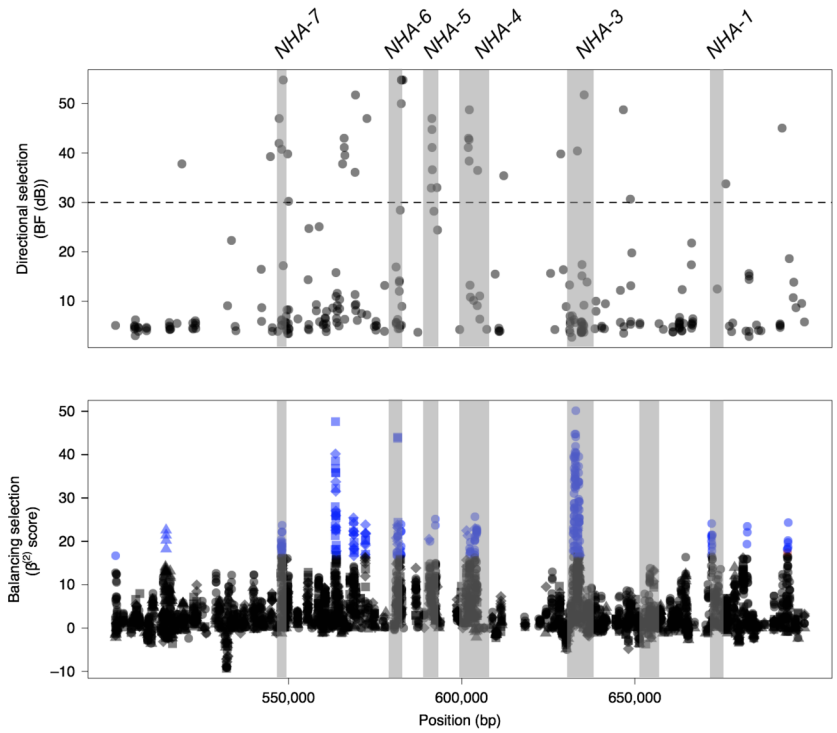


Figure 3 : Stern and Lee (2020) show that balancing selection in native saltwater populations of the *Eurytemora affinis* species complex acted on the same loci repeatedly subject to directional selection in the invasive freshwater range. **Upper panel:** scan for the footprint of directional selection using *Bayescan 3*(Foll, Gaggiotti, Daub, Vatsiou, & Excoffier, 2014). Dashed line indicates significance threshold for test of directional selection.**Lower panel:** Signatures of balancing selection were quantified in four native populations (each shape corresponds to a population) using the summary statistic $\beta^{(2)}$ (Siewert & Voight, 2020). Points in blue are within the top 1% of$\beta^{(2)}$ scores calculated within each population. The chromosomal coordinates of genes in the NHA family are highlighted in grey in both panels. Figure adapted from Stern and Lee (2020).

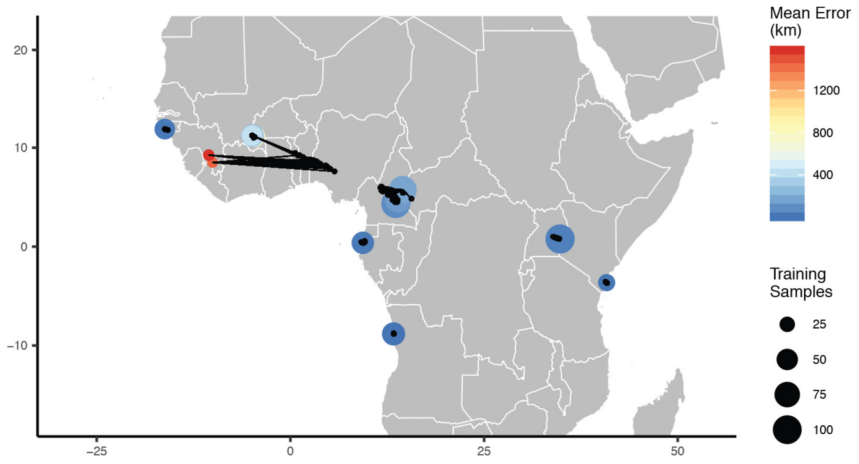


Figure 4 : Battey, Ralph and Kern (2020) developed machine learning software, *Locator* , to estimate

Posted on Authorea 3 Jan 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160968166.65928724/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

the geographic origin of a genetic sample. **A)** True and predicted sample locations for 153 *Anopheles gambiae/coluzzii* samples in sub-Saharan Africa using a total of 612 training samples and a 2Mb window size. Each inferred sample location (geographic centroid of per-window estimates) is a black point connected by a line to the true location of the sample. True sample locality point sizes are scaled by the number of training samples used for the estimate and coloured by average test error. See Battey, Ralph and Kern (2020) for further details.

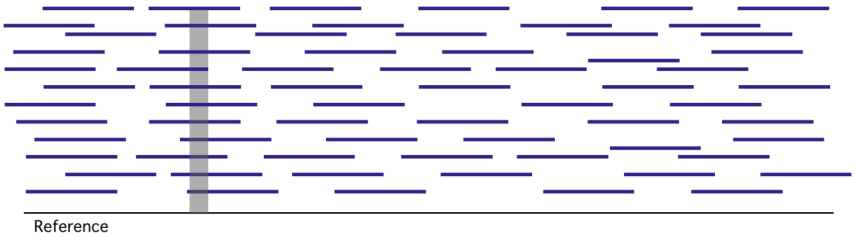
Hosted file

`image5.emf` available at https://authorea.com/users/386791/articles/502257-the-population-genomics-of-invasive-species

Figure 5 : Calfee *et al.* (2020) studied replicate geographic ancestry clines in invasive honey bees (*Apis mellifera*). **A)** In 1957, western honey bees derived from African populations (*A. m. scutellata* ; '*scutellata'*hereafter) escaped from a captive breeding programme in Rio Claro, São Paulo. *Scutellata* populations both outcompeted and hybridized with honey bee populations from Europe (previously introduced to the Americas), forming invasive hybrid *scutellata* -European populations. Replicate North and South American invasion routes (indicated by arrows) of hybrid populations are shown with pink arrows, with dates of first occurrence indicated. **B)** Geographic clines in genome-wide ancestry. Curves are logistic cline models of ancestry predicted by latitude, with dotted horizontal lines indicating the latitude at which the model predicts 50% *scutellata* ancestry.**C)** Genomic location of *scutellata* ancestry in the two hybrid zones shown in (B). Dashed line indicates mean *scutellata*ancestry. Parallel peaks in excess *scutellata* ancestry, where selection has outweighed migration to 'push' *scutellata* alleles to high frequency hundreds of kilometres past the expected cline centre, occur in chromosomes 1 and 11. Figure adapted from Calfee *et al.*(2020).

Box 1: Sequencing strategies in population genomics

Whole-genome resequencing (WGR) is one of several sequencing technologies that can be used for population genomic analysis. Currently the most common family of technologies for WGR is short-read sequencing, where reads are aligned to an already-available reference genome. The per-base pair error rate is approximately 0.31% for Illumina reads (Schirmer, D'Amore, Ijaz, Hall, & Quince, 2016). Therefore, if rare variants (those that occur in few individuals, and which differ from the reference genome) need to be identified with high accuracy, high sequence depth (*i.e.* , a relatively large average number of reads that cover each base pair in the genome for each individual) may be required, although this can also be achieved by sequencing large population samples at low per-individual coverage. This may often be the case when using demographic inference to estimate the timing of an invasion event (see Part 3.1). In the example below, the sequence depth at the locus highlighted in grey is 11X.

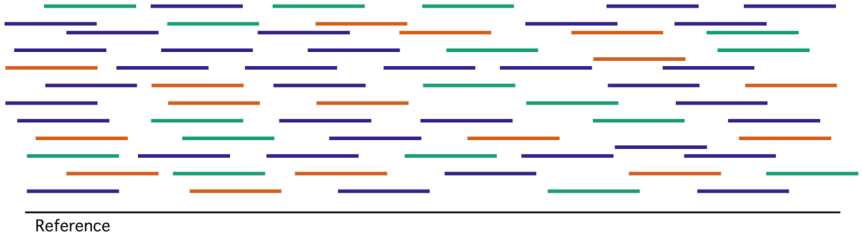


With limited resources, there exists a trade-off between the number of individuals sequenced and the sequence depth. Depending on the biological question, higher individual sample sizes may be more valuable than high read depth (see Fumagalli, 2013; Lou, Jacobs, Wilder, & Therkildsen, 2020). For example, when inferring the geographic source of an invasive population, many reference individuals are required (see Part 2). In some cases, such as when analysing historical museum or herbarium samples, low sequencing depth may be unavoidable (McGaughran, 2020), necessitating analytical pipelines specifically designed for low-coverage genomic data (*e.g.* , Korneliussen, Albrechtsen, & Nielsen, 2014). Linked-read technologies such

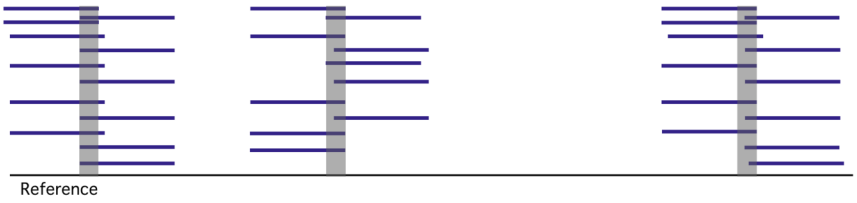
Posted on Authorea 3 Jan 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160968166.65928724/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

as haplotagging allow variants to be imputed with high accuracy, which means that many individuals can be sequenced at the cost of low-depth sequencing with less of a compromise in terms of effective read depth (Meier et al., 2020; see Box 3).

An alternative WGR sequencing strategy is PoolSeq (see Hivert, Leblois, Petit, Gautier, & Vitalis, 2018). If a given analysis requires allele frequencies from separate populations (*e.g* ., detecting directional or balancing selection; see Figure 3), genomic DNA from many individuals of the same population can be pooled in equimolar proportion and sequenced together. The concept of PoolSeq is shown below, where read colours correspond to three different (but unlabelled) individuals sequenced together. PoolSeq may be largely outdated now that methods for individual barcoding of large numbers of individuals for sequencing have become more affordable.



Reduced-complexity sequencing is an alternative to the WGR strategies discussed above (see Andrews, Good, Miller, Luikart, & Hohenlohe, 2016; Deschamps, Llaca, & May, 2012). Reduced-representation technologies produce sequence data from a small fraction of the genome, at restriction endonuclease cut sites (shown in grey below).



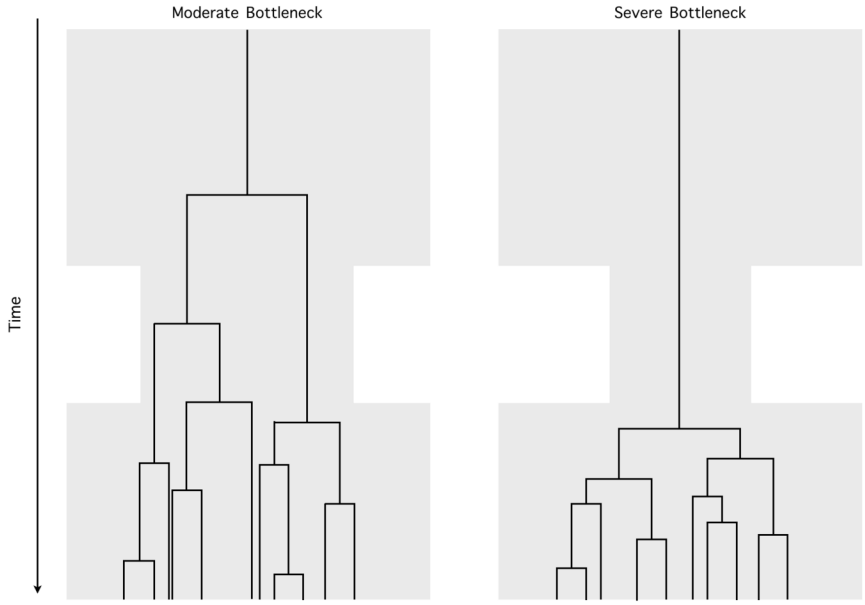
Reduced-representation sequencing can be an affordable alternative to WGR, especially where many individual samples are required, where the species of interest has a large reference genome, or when a reference genome is unavailable. Successful applications of reduced-representation sequencing (or other approaches that do not use whole genome sequence data, such as transcriptomics) have been particularly useful when reconstructing the evolutionary history of biological invasions (*e.g.* , Gibson, de Lourdes Torres, Brandvain, & Moyle, 2020; Schmidt et al., 2020). However, reduced-representation sequencing does not allow haplotype data to be used, (see Box 3), cannot identify structural variants (see Part 4), may not provide sufficient resolution to estimate the timing of recent biological invasions (Part 3), and provides limited resolution when detecting adaptation or introgression throughout the genome (Parts 4 and 5). In addition, library preparation can be more time-consuming (and therefore expensive) than WGS approaches. Figures in this box are adapted from Hahn (2019).

Box 2: The demography of biological invasion

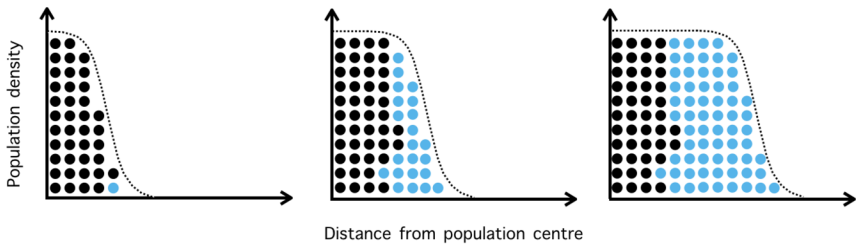
Demographic bottlenecks (in which the census population size decreases and then increases again) are expected to occur under simple models of biological invasion and may be common among invasive species. The consequences of demographic bottlenecks are well-established and far reaching in terms of population genetic models (reviewed by Gattepaille, Jakobsson, & Blum, 2013). Tajima's D will depend on the magnitude and

Posted on Authorea 3 Jan 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160968166.65928724/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

number of generations since the invasion, and will ultimately show greater than expected variance among loci (Stajich & Hahn, 2005). After a sufficiently extreme bottleneck, all lineages will coalesce to form a star-like genealogy with an excess of rare variants; however if the bottleneck is more moderate and multiple lineages persist, an excess of intermediate-frequency may be observed (Depaulis, Mousset, & Veuille, 2003). These scenarios are depicted below (adapted from Gattepaille et al., 2013).



Various ramifications become apparent when thinking about simple invasion scenarios not just in time, but in space. For example, during invasion, individuals at the range edge are presumably more likely to disperse into unexploited habitat than those at the centre of the metapopulation. Over time, this would create strong drift at the leading edge of an expanding population. As a consequence, a subset of low-frequency mutations that arise at the range edge will propagate over space and reach high frequencies simply as a consequence of population expansion (Edmonds, Lillie, & Cavalli-Sforza, 2004). This phenomenon is known as genetic surfing (or 'allele surfing'), and is more likely to occur in small, fast-expanding populations, which could include many invasive species (Klopfstein, Currat, & Excoffier, 2006). Below, a mutation at the expanding range edge spreads to high frequency over a large area through genetic surfing (adapted from Foutel-Rodier & Etheridge, 2020).



Genetic surfing creates fitness costs at the range edge (reviewed by Angert, Bontrager, & Ågren, 2020) in two ways. First, at least in one dimensional simulations, surfing causes genetic diversity to decline over space away from the population centre (Hallatschek & Nelson, 2008). Second, deleterious mutations can surf on the

Posted on Authorea 3 Jan 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160968166.65928724/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

wave of advance to reach high frequencies over a large range (Peischl, Dupanloup, Kirkpatrick, & Excoffier, 2013; Travis et al., 2007) – a phenomenon known as expansion load (Peischl & Excoffier, 2015). These costs make the success of invasive species seem even more paradoxical (Estoup et al., 2016). However, a number of solutions have been proposed to the cost of range expansion. First, long range dispersal can ameliorate the loss of genetic diversity through surfing under some conditions (Paulose & Hallatschek, 2020). Second, the spatial sorting of dispersal traits that results from superior dispersers finding mates more often at the range edge (Shine, Brown, & Phillips, 2011) can rescue populations from expansion load (Peischl & Gilbert, 2020).

Genetic Surfing can also create geographic clines in allele frequency in the direction of range expansion (Klopfstein et al., 2006), clusters of low genetic diversity, and sweeps of random loci in different regions of the metapopulation (Hallatschek, Hersen, Ramanathan, & Nelson, 2007). These allele frequency patterns may be falsely interpreted as a footprint of selection (Excoffier & Ray, 2008).

Thus, when using genomic data to detect post-introduction adaptation in an invasive species known to have undergone a population bottleneck, modelling approaches should be used to rule out potentially confounding demographic and spatial effects (*e.g.* , Currat et al., 2006). Moreover, Peischl and Excoffier's (2015) model of expansion load provides clear expectations in terms of the expected shape of the site frequency spectrum at the range front. Invasive species are therefore ideal systems in which to validate or reject these expectations.

Box 3: Maximising the advantage of whole genome sequencing with haplotype data

All sequencing technologies allow allele frequencies to be measured. One of the key advantages of whole-genome resequencing over other technologies is the opportunity to exploit additional information, such as the haplotypes on which physically linked alleles are coinherited. Haplotype data enable the use of several powerful analytical methods (reviewed by Leitwein, Duranton, Rougemont, Gagnaire, & Bernatchez, 2020) that are relevant to invasion genomics.

Because recombination and mutation reconfigure haplotypes over time, the size and frequency of haplotypes convey evolutionary information – a phenomenon that Moorjani *et al.* (2016) refer to as the 'recombination clock'. For example, a haplotype on which a beneficial allele arises is swept to fixation faster than recombination can break it down to its expected size under neutrality. Therefore a signature of selection is left by unusually large stretches of haplotype homozygosity (*i.e.* , linkage extends further from the selected locus than expected), and by the unexpectedly high frequency of a core haplotype (Sabeti et al., 2002). This is the basis for tests of extended haplotype homozygosity, used to scan the genome for signatures of selection (see Parts 1 and 4). Haplotype data are also useful for reconstructing population size change through time (Part 3). By analysing long haplotypes identical by descent (that have not yet been broken down by recombination), Browning and Browning (2015) were able to accurately reconstruct changes in human population size in the recent past (4 to 50 generations before present). This approach holds great potential for invasion genetics, where it is often difficult to reconstruct recent demography (see Part 3.1).

Haplotype data show most promise in recently admixed populations (see Part 5). Any analysis of hybridization using haplotype data will require the ancestry of an introgressed haplotype ('ancestry tract' or 'ancestry block') to be inferred (for a review of approaches to ancestry assignment see Leitwein et al., 2020). Duranton *et al.* (2019) studied the introgression of Atlantic sea bass (*Dicentrarchus labrax*) into Mediterranean populations of the same species. By modelling the diffusion of introgressed haplotypes through space (by gene flow) as they are broken down over time (by recombination), the average per-generation dispersal distance could then be estimated. This approach is likely to be useful for reconstructing the spatial extent of introgression in invasive species (See Parts 2 and 5). Finally, adaptive introgression can be accurately detected using haplotype data (see Shchur, Svedberg, Medina, Corbett-Detig, & Nielsen, 2020). In summary, haplotype data open many possibilities in invasion genetics research, representing one of the key advantages of using WGR to study invasive species.

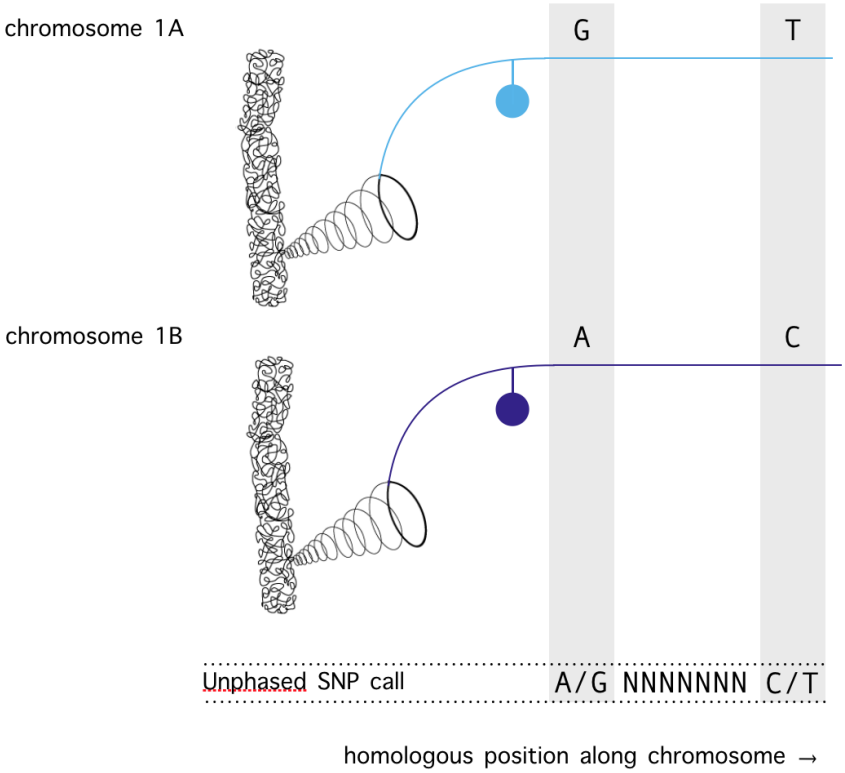
However, haplotype information cannot be directly extracted from WGR data generated using short reads. Therefore, until long-read sequencing becomes scalable, direct or indirect methods for inferring gametic phase

Posted on Authorea 3 Jan 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160908166.65928724/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

(*i.e.* , the two DNA sequences on which alleles occur, in the case of diploids) need to be used to leverage haplotype information from WGR data.

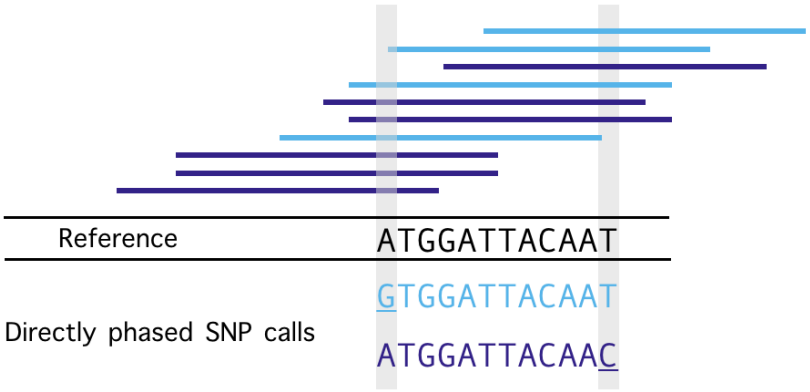
Indirect or statistical phasing methods can be applied to whole-genome datasets obtained with short-read sequencing technology (reviewed by Rhee et al., 2016). The accuracy of these methods depend on factors such as the number of samples and the density of nucleotide polymorphisms (Browning & Browning, 2007). Phasing errors can affect the downstream biological interpretations made by analysing haplotypes. Direct phasing methods, on the other hand, record chromosomal haplotypes during the generation of sequence data. Linked-read sequencing is a newly developed family of direct phasing technologies that results in fewer errors than indirect statistical approaches (Amini et al., 2014; Choi, Chan, Kirkness, Telenti, & Schork, 2018).

Though linked-read sequencing approaches show great promise in population genomics (*e.g* ., Lutgen et al., 2020), many platforms are currently prohibitively expensive. One notable exception is haplotagging, a recent low-cost linked-read sequencing method (Meier et al., 2020). Through haplotagging, kilobase-length DNA fragments are tagged with unique barcodes as they wrap around unique microbeads in solution.



Standard short-read sequencing can then proceed with long-range haplotype information retained as unique barcodes. This method also allows individuals to be sequenced at lower depth, because missing genotypes can be imputed using haplotype information, and structural variants to be more readily identified.

Posted on Authorea 3 Jan 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160968166.65928724/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.



References

Amini, S., Pushkarev, D., Christiansen, L., Kostem, E., Royce, T., Turk, C., ... others. (2014). Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. *Nature Genetics* , *46* (12), 1343–1349.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* , *17* (2), 81.

Angert, A. L., Bontrager, M. G., & Ågren, J. (2020). What Do We Really Know About Adaptation at Range Edges? *Annual Review of Ecology, Evolution, and Systematics* , *51* , 341–361.

Baker, H. G. (1965). Characteristics and modes of origin of weeds. In H. G. Baker & G. Stebbins (Eds.), *The genetics of colonizing species: Proceedings of the First International Union of Biological Sciences Symposia on General Biology* (pp. 147–172). ACADEMIC PRESS INC.

Baker, H., & Stebbins, G. (1965). *The Genetics of Colonizing Species* (1st ed.). New York: ACADEMIC PRESS INC.

Barrett, S. C. H. (2015). Foundations of invasion genetics: the Baker and Stebbins legacy. *Molecular Ecology* , *24* (9), 1927–1941.

Barton, N. H., & Etheridge, A. M. (2004). The effect of selection on genealogies. *Genetics* , *166* (2), 1115–1131.

Battey, C. J., Ralph, P. L., & Kern, A. D. (2020). Predicting geographic location from genetic variation with deep neural networks. *ELife* , *9* , e54507.

Bergland, A. O., Tobler, R., Gonzalez, J., Schmidt, P., & Petrov, D. (2016). Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in Drosophila melanogaster. *MOLECULAR ECOLOGY* , *25* (5), 1157–1174. doi:10.1111/mec.13455

Bertolotti, A. C., Layer, R. M., Gundappa, M. K., Gallagher, M. D., Pehlivanoglu, E., Nome, T., ... others. (2020). The structural variation landscape in 492 Atlantic salmon genomes. *BioRxiv* .

Posted on Authorea 3 Jan 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160968166.65928724/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Bi, K., Linderoth, T., Singhal, S., Vanderpool, D., Patton, J. L., Nielsen, R., ... Good, J. M. (2019). Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during recent climate change. *PLoS Genetics* , *15* (5), e1008119.

Bock, D. G., Caseys, C., Cousens, R. D., Hahn, M. A., Heredia, S. M., Hübner, S., ... Rieseberg, L. H. (2015). What we still don't know about invasion genetics. *Molecular Ecology* , *24* (9), 2277–2297.

Booker, T. R., Yeaman, S., & Whitlock, M. (2020). Variation in recombination rate affects detection of FST outliers under neutrality.*BioRxiv* .

Brandvain, Y., & Wright, S. I. (2016). The limits of natural selection in a nonequilibrium world. *Trends in Genetics* , *32* (4), 201–210.

Brown, A. M. V, Huynh, L. Y., Bolender, C. M., Nelson, K. G., & McCutcheon, J. P. (2014). Population genomics of a symbiont in the early stages of a pest invasion. *MOLECULAR ECOLOGY* , *23* (6, SI), 1516–1530. doi:10.1111/mec.12366

Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics* , *81* (5), 1084–1097.

Browning, S. R., & Browning, B. L. (2015). Accurate non-parametric estimation of recent effective population size from segments of identity by descent. *The American Journal of Human Genetics* , *97* (3), 404–418.

Buffalo, V., & Coop, G. (2020). Estimating the genome-wide contribution of selection to temporal allele frequency change. *Proceedings of the National Academy of Sciences* , *117* (34), 20672–20680. doi:10.1073/pnas.1919039117

Calfee, E., Agra, M. N., Palacio, M. A., Ramirez, S. R., & Coop, G. (2020). Selection and hybridization shaped the Africanized honey bee invasion of the Americas. *BioRxiv* .

Calfee, E., Agra, M. N., Palacio, M. A., Ramírez, S. R., & Coop, G. (2020). Selection and hybridization shaped the rapid spread of African honey bee ancestry in the Americas. *PLoS Genetics* , *16* (10), e1009038.

Chan, A. H., Jenkins, P. A., & Song, Y. S. (2012). Genome-wide fine-scale recombination rate variation in Drosophila melanogaster.*PLoS Genetics* , *8* (12), e1003090.

Choi, Y., Chan, A. P., Kirkness, E., Telenti, A., & Schork, N. J. (2018). Comparison of phasing strategies for whole human genomes.*PLoS Genetics* , *14* (4), e1007308.

Chown, S. L., Hodgins, K. A., Griffin, P. C., Oakeshott, J. G., Byrne, M., & Hoffmann, A. A. (2015). Biological invasions, climate change and genomics. *EVOLUTIONARY APPLICATIONS* , *8* (1), 23–46. doi:10.1111/eva.12234

Clavero, M., Brotons, L., Pons, P., & Sol, D. (2009). Prominent role of invasive species in avian biodiversity loss. *Biological Conservation* , *142* (10), 2043–2049.

Clavero, M., & García-Berthou, E. (2005). Invasive species are a leading cause of animal extinctions. *Trends in Ecology & Evolution* , *20* (3), 110.

Comeault, A. A., Wang, J., Tittes, S., Isbell, K., Ingley, S., Hurlbert, A. H., & Matute, D. R. (2020). Genetic Diversity and Thermal Performance in Invasive and Native Populations of African Fig Flies.*MOLECULAR BIOLOGY AND EVOLUTION* , *37* (7), 1893–1906. doi:10.1093/molbev/msaa050

Cristescu, M. E. (2015). Genetic reconstructions of invasion history.*MOLECULAR ECOLOGY* , *24* (9, SI), 2212–2225. doi:10.1111/mec.13117

Currat, M., Excoffier, L., Maddison, W., Otto, S. P., Ray, N., Whitlock, M. C., & Yeaman, S. (2006). Comment on" Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens" and"

Posted on Authorea 3 Jan 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160968166.65928724/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans".*Science* , *313* (5784), 172.

DeGiorgio, M., Huber, C. D., Hubisz, M. J., Hellmann, I., & Nielsen, R. (2016). SweepFinder2: increased sensitivity, robustness and flexibility.*Bioinformatics* , *32* (12), 1895–1897.

Demené, A., Legrand, L., Gouzy, J., Debuchy, R., Saint-Jean, G., Fabreguettes, O., & Dutech, C. (2019). Whole-genome sequencing reveals recent and frequent genetic recombination between clonal lineages of Cryphonectria parasitica in western Europe. *FUNGAL GENETICS AND BIOLOGY* , *130* , 122–133. doi:10.1016/j.fgb.2019.06.002

Dennenmoser, S., Sedlazeck, F. J., Iwaszkiewicz, E., Li, X.-Y., Altmueller, J., & Nolte, A. W. (2017). Copy number increases of transposable elements and protein-coding genes in an invasive fish of hybrid origin. *MOLECULAR ECOLOGY* , *26* (18), 4712–4724. doi:10.1111/mec.14134

Dennenmoser, S., Sedlazeck, F. J., Schatz, M. C., Altmueller, J., Zytnicki, M., & Nolte, A. W. (2019). Genome-wide patterns of transposon proliferation in an evolutionary young hybrid fish. *MOLECULAR ECOLOGY* , *28* (6, SI), 1491–1505. doi:10.1111/mec.14969

Depaulis, F., Mousset, S., & Veuille, M. (2003). Power of neutrality tests to detect bottlenecks and hitchhiking. *Journal of Molecular Evolution* , *57* (1), S190–S200.

Deschamps, S., Llaca, V., & May, G. D. (2012). Genotyping-by-sequencing in plants. *Biology* , *1* (3), 460–483.

Ding, J., Mack, R. N., Lu, P., Ren, M., & Huang, H. (2008). China's booming economy is sparking and accelerating biological invasions.*Bioscience* , *58* (4), 317–324.

Duranton, M., Bonhomme, F., & Gagnaire, P.-A. (2019). The spatial scale of dispersal revealed by admixture tracts. *Evolutionary Applications* , *12* (9), 1743–1756.

Edmonds, C. A., Lillie, A. S., & Cavalli-Sforza, L. L. (2004). Mutations arising in the wave front of an expanding population.*Proceedings of the National Academy of Sciences* , *101* (4), 975–979.

Estoup, A., & Guillemaud, T. (2010). Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology* ,*19* (19), 4113–4130.

Estoup, A., Ravigné, V., Hufbauer, R., Vitalis, R., Gautier, M., & Facon, B. (2016). Is there a genetic paradox of biological invasion?*Annual Review of Ecology, Evolution, and Systematics* , *47* , 51–72.

Excoffier, L., & Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution* , *23* (7), 347–351.

Flagel, L., Brandvain, Y., & Schrider, D. R. (2019). The unreasonable effectiveness of convolutional neural networks in population genetic inference. *Molecular Biology and Evolution* , *36* (2), 220–238.

Foll, M., Gaggiotti, O. E., Daub, J. T., Vatsiou, A., & Excoffier, L. (2014). Widespread signals of convergent adaptation to high altitude in Asia and America. *The American Journal of Human Genetics* ,*95* (4), 394–407.

Fontaine, M. C., Labbe, F., Dussert, Y., Deliere, L., Richart-Cervera, S., Giraud, T., & Delmotte, F. (2020). Europe as a bridgehead in the worldwide invasion history of grapevine downy mildew, Plasmopara viticola. *BioRxiv* . doi:https://doi.org/10.1101/2020.09.22

Foutel-Rodier, F., & Etheridge, A. M. (2020). The spatial Muller's ratchet: surfing of deleterious mutations during range expansion.*Theoretical Population Biology* , *135* , 19–31.

Fumagalli, M. (2013). Assessing the effect of sequencing depth and sample size in population genetics inferences. *PLoS One* ,*8* (11), e79667.

Gattepaille, L. M., Jakobsson, M., & Blum, M. G. B. (2013). Inferring population size changes with sequence and SNP data: lessons from human bottlenecks. *Heredity* , *110* (5), 409–419.

Posted on Authorea 3 Jan 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160968166.65928724/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Gibson, M. J. S., de Lourdes Torres, M., Brandvain, Y., & Moyle, L. C. (2020). Reconstructing the history and biological consequences of a plant invasion on the Galápagos islands. *BioRxiv* . doi:https://doi.org/10.1101/2020.09.26.313627;

Gower, G. R., Picazo, P. I., Fumagalli, M., & Racimo, F. (2020). Detecting adaptive introgression in human evolution using convolutional neural networks. *BioRxiv* . doi:https://doi.org/10.1101/2020.09.18.301069

Grabenstein, K. C., & Taylor, S. A. (2018). Breaking barriers: causes, consequences, and experimental utility of human-mediated hybridization. *Trends in Ecology & Evolution* , *33* (3), 198–212.

Guillot, G., Jónsson, H., Hinge, A., Manchih, N., & Orlando, L. (2016). Accurate continuous geographic assignment from low-to high-density SNP data. *Bioinformatics* , *32* (7), 1106–1108.

Hahn, M. W. (2019). *Molecular population genetics* . Sinauer Associates New York.

Hallatschek, O., Hersen, P., Ramanathan, S., & Nelson, D. R. (2007). Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences* , *104* (50), 19926–19930.

Hallatschek, O., & Nelson, D. R. (2008). Gene surfing in expanding populations. *Theoretical Population Biology* , *73* (1), 158–170.

Haller, B. C., Galloway, J., Kelleher, J., Messer, P. W., & Ralph, P. L. (2019). Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. *Molecular Ecology Resources* , *19* (2), 552–566.

Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Molecular Biology and Evolution* , *36* (3), 632–637.

Hamelin, R. C., & Roe, A. D. (2020). Genomic biosurveillance of forest invasive alien enemies: A story written in code. *EVOLUTIONARY APPLICATIONS* , *13* (1, SI), 95–115. doi:10.1111/eva.12853

Harris, K., & Nielsen, R. (2013). Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genet* , *9* (6), e1003521.

Haygood, R., Ives, A. R., & Andow, D. A. (2003). Consequences of recurrent gene flow from crops to wild relatives. *Proceedings of the Royal Society of London. Series B: Biological Sciences* ,*270* (1527), 1879–1886.

Hegarty, M. J. (2012). Invasion of the hybrids. *MOLECULAR ECOLOGY* , *21* (19), 4669–4671. doi:10.1111/j.1365-294X.2012.05720.x

Hessenauer, P., Fijarczyk, A., Martin, H., Prunier, J., Charron, G., Chapuis, J., ... Landry, C. R. (2020). Hybridization and introgression drive genome evolution of Dutch elm disease pathogens. *NATURE ECOLOGY & EVOLUTION* , *4* (4), 626–638. doi:10.1038/s41559-020-1133-6

Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring genetic differentiation from pool-seq data. *Genetics* , *210* (1), 315–330.

Hosokawa, T., Kikuchi, Y., Nikoh, N., Shimada, M., & Fukatsu, T. (2006). Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. *PLoS Biol* , *4* (10), e337.

Hovick, S. M., & Whitney, K. D. (2014). Hybridisation is associated with increased fecundity and size in invasive taxa: meta-analytic support for the hybridisation-invasion hypothesis. *Ecology Letters* , *17* (11), 1464–1477.

Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution* , *23* (3), 482–490.

Kolar, C. S., & Lodge, D. M. (2001). Progress in invasion biology: predicting invaders. *Trends in Ecology & Evolution* ,*16* (4), 199–204.

Posted on Authorea 3 Jan 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160968166.65928724/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* ,*15* (1), 356.

Laster, M. L., & Sheng, C. F. (1995). Search for hybrid sterility for Helicoverpa zea in crosses between the North American H. zea and H. armigera (Lepidoptera: Noctuidae) from China. *Journal of Economic Entomology* , *88* (5), 1288–1291.

Lee, C. E. (2002). Evolutionary genetics of invasive species. *Trends in Ecology & Evolution* , *17* (8), 386–391.

Lee, C. E., & Gelembiuk, G. W. (2008). Evolutionary origins of invasive populations. *Evolutionary Applications* , *1* (3), 427–448.

Lee, Y., Schmidt, H., Collier, T. C., Conner, W. R., Hanemaaijer, M. J., Slatkin, M., ... Cornel, A. J. (2019). Genome-wide divergence among invasive populations of Aedes aegypti in California. *BMC GENOMICS* , *20* . doi:10.1186/s12864-019-5586-4

Leitwein, M., Duranton, M., Rougemont, Q., Gagnaire, P.-A., & Bernatchez, L. (2020). Using haplotype information for conservation genomics. *Trends in Ecology & Evolution* , *35* (3), 245–258.

Lodge, D. M. (1993). Species invasions and deletions: community effects and responses to climate and habitat change. In *Species invasions and deletions: community effects and responses to climate and habitat change.* (pp. 367–387). Sinauer Associates Inc.

Lombaert, E., Guillemaud, T., Cornuet, J.-M., Malausa, T., Facon, B., & Estoup, A. (2010). Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *PloS One* , *5* (3), e9743.

Long, A., Liti, G., Luptak, A., & Tenaillon, O. (2015). Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nature Reviews Genetics* , *16* (10), 567–582.

Lou, R. N., Jacobs, A., Wilder, A., & Therkildsen, N. O. (2020). A beginner's guide to low-coverage whole genome sequencing for population genomics. *Authorea Preprints* .

Luqman, H., Widmer, A., Fior, S., & Wegmann, D. (2020). Identifying loci under selection via explicit demographic models. *BioRxiv* . doi:https://doi.org/10.1101/2020.07.20.211581

Lutgen, D., Ritter, R., Olsen, R.-A., Schielzeth, H., Gruselius, J., Ewels, P. A., ... others. (2020). Linked-read sequencing enables haplotype-resolved resequencing at population scale. *BioRxiv* .

Malinsky, M. (2019). Dsuite-fast D-statistics and related admixture evidence from VCF files. *MOLECULAR ECOLOGY RESOURCES* . doi:10.1111/1755-0998.13265

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* , *20* (5), 229–237. doi:http://dx.doi.org/10.1016/j.tree.2005.02.010

Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A combinatorial view on speciation and adaptive radiation. *Trends in Ecology & Evolution* .

Maxwell, S. L., Fuller, R. A., Brooks, T. M., & Watson, J. E. M. (2016). Biodiversity: The ravages of guns, nets and bulldozers. *Nature News* , *536* (7615), 143.

McCartney, M. A., Mallez, S., & Gohl, D. M. (2019). Genome projects in invasion biology. *Conservation Genetics* , 1–22.

McFarlane, S. E., & Pemberton, J. M. (2019). Detecting the true extent of introgression during anthropogenic hybridization. *Trends in Ecology & Evolution* , *34* (4), 315–326.

McGaughran, A. (2020). Effects of sample age on data quality from targeted sequencing of museum specimens: what are we capturing in time? *BMC Genomics* , *21* (1), 1–10.

McVean, G. A. T., & Cardin, N. J. (2005). Approximating the coalescent with recombination. *Philosophical Transactions of the Royal Society B: Biological Sciences* , *360* (1459), 1387–1393.

Posted on Authorea 3 Jan 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160968166.65928724/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Meier, J. I., Salazar, P. A., Kučka, M., Davies, R. W., Dréau, A., Aldás, I., ... others. (2020). Haplotype tagging reveals parallel formation of hybrid races in two butterfly species. *BioRxiv* .

Mérel, V., Gibert, P., Buch, I., Rada, V. R., Estoup, A., Gautier, M., ... Vieira, C. (2020). The worldwide invasion of Drosophila suzukii is accompanied by a large increase of transposable element load and a small number of putatively adaptive insertions. *BioRxiv* . doi:https://doi.org/10.1101/2020.11.06.370932

Moorjani, P., Sankararaman, S., Fu, Q., Przeworski, M., Patterson, N., & Reich, D. (2016). A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. *Proceedings of the National Academy of Sciences* ,*113* (20), 5652–5657.

Muirhead, J. R., Minton, M. S., Miller, W. A., & Ruiz, G. M. (2015). Projected effects of the Panama Canal expansion on shipping traffic and biological invasions. *Diversity and Distributions* , *21* (1), 75–87.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Research* , *15* (11), 1566–1575.

Oerke, E. C. (2006). Crop losses to pests. *The Journal of Agricultural Science* , *144* , 31.

Olazcuaga, L., Loiseau, A., Parrinello, H., Paris, M., Fraimout, A., Guedot, C., ... others. (2020). A whole-genome scan for association with invasion success in the fruit fly Drosophila suzukii using contrasts of allele frequencies corrected for population structure.*Molecular Biology and Evolution* .

Otte, K. A., & Schlötterer, C. (2020). Detecting selected haplotype blocks in evolve and resequence experiments. *Molecular Ecology Resources* .

Patton, A. H., Margres, M. J., Stahlke, A. R., Hendricks, S., Lewallen, K., Hamede, R. K., ... others. (2019). Contemporary demographic reconstruction methods are robust to genome assembly quality: A case study in Tasmanian Devils. *Molecular Biology and Evolution* ,*36* (12), 2906–2921.

Paulose, J., & Hallatschek, O. (2020). The impact of long-range dispersal on gene surfing. *Proceedings of the National Academy of Sciences* , *117* (14), 7584–7593.

Peischl, S., Dupanloup, I., Kirkpatrick, M., & Excoffier, L. (2013). On the accumulation of deleterious mutations during range expansions.*Molecular Ecology* , *22* (24), 5972–5982.

Peischl, S., & Excoffier, L. (2015). Expansion load: recessive mutations and the role of standing genetic variation. *Molecular Ecology* , *24* (9), 2084–2094.

Peischl, S., & Gilbert, K. J. (2020). Evolution of dispersal can rescue populations from expansion load. *The American Naturalist* ,*195* (2), 349–360.

Popovic, I., Matias, A. M. A., Bierne, N., & Riginos, C. (2020). Twin introductions by independent invader mussel lineages are both associated with recent admixture with a native congener in Australia.*EVOLUTIONARY APPLICATIONS* , *13* (3, SI). doi:10.1111/eva.12857

Prentis, P. J., Wilson, J. R. U., Dormontt, E. E., Richardson, D. M., & Lowe, A. J. (2008). Adaptive evolution in invasive species. *TRENDS IN PLANT SCIENCE* , *13* (6), 288–294. doi:10.1016/j.tplants.2008.03.004

Puckett, E. E., Magnussen, E., Khlyap, L. A., Strand, T. M., Lundkvist, A., & Munshi-South, J. (2020). Genomic analyses reveal three independent introductions of the invasive brown rat (Rattus norvegicus) to the Faroe Islands. *HEREDITY* , *124* (1), 15–27. doi:10.1038/s41437-019-0255-6

Quilodrán, C. S., Tsoupas, A., & Currat, M. (2020). The Spatial Signature of Introgression After a Biological Invasion With Hybridization. *Frontiers in Ecology and Evolution* , *8* , 321.

Rhee, J.-K., Li, H., Joung, J.-G., Hwang, K.-B., Zhang, B.-T., & Shin, S.-Y. (2016). Survey of computational haplotype determination methods for single individual. *Genes & Genomics* , *38* (1), 1–12.

Posted on Authorea 3 Jan 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160968166.65928724/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Ricciardi, A. (2007). Are modern biological invasions an unprecedented form of global change? *Conservation Biology* , *21* (2), 329–336.

Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K., Nakazato, T., ... Lexer, C. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization.*Science* , *301* (5637), 1211–1216.

Rieseberg, L. H., Wood, T. E., & Baack, E. J. (2006). The nature of plant species. *Nature* , *440* (7083), 524.

Rispe, C., Legeai, F., Nabity, P. D., Fernandez, R., Arora, A. K., Baa-Puyoulet, P., ... Tagu, D. (2020). The genome sequence of the grape phylloxera provides insights into the evolution, adaptation, and invasion routes of an iconic pest. *BMC BIOLOGY* , *18* (1). doi:10.1186/s12915-020-00820-5

Rius, M., Bourne, S., Hornsby, H. G., & Chapman, M. A. (2015). Applications of next-generation sequencing to the study of biological invasions. *CURRENT ZOOLOGY* , *61* (3), 488+. doi:10.1093/czoolo/61.3.488

Rose, N. H., Sylla, M., Badolo, A., Lutomiah, J., Ayala, D., Aribodor, O. B., ... McBride, C. S. (2020). Climate and Urbanization Drive Mosquito Preference for Humans. *Current Biology* , *30* (18), 3570–3579.e6. doi:https://doi.org/10.1016/j.cub.2020.06.092

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., ... others. (2002). Detecting recent positive selection in the human genome from haplotype structure.*Nature* , *419* (6909), 832–837.

Sakai, A. K., Allendorf, F. W., Holt, J. S., Lodge, D. M., Molofsky, J., With, K. A., ... others. (2001). The population biology of invasive species. *Annual Review of Ecology and Systematics* , *32* (1), 305–332.

Sardain, A., Sardain, E., & Leung, B. (2019). Global forecasts of shipping traffic and biological invasions to 2050. Nat Sustain 2: 274–282.

Schirmer, M., D'Amore, R., Ijaz, U. Z., Hall, N., & Quince, C. (2016). Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics* , *17* (1), 125.

Schlötterer, C., Kofler, R., Versace, E., Tobler, R., & Franssen, S. U. (2015). Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. *Heredity* , *114* (5), 431–440.

Schmidt, T. L., Swan, T., Chung, J., Karl, S., Demok, S., Yang, Q., ... Hoffmann, A. A. (2020). Spatial population genomics of a recent mosquito invasion. *BioRxiv* . doi:10.1101/2020.11.30.405191

Schrider, D. R., & Kern, A. D. (2018). Supervised machine learning for population genetics: a new paradigm. *Trends in Genetics* ,*34* (4), 301–312.

Seebens, H., Essl, F., Dawson, W., Fuentes, N., Moser, D., Pergl, J., ... others. (2015). Global trade will accelerate plant invasions in emerging economies under climate change. *Global Change Biology* ,*21* (11), 4128–4140.

Seixas, F. A., Boursot, P., & Melo-Ferreira, J. (2018). The genomic impact of historical hybridization with massive mitochondria DNA introgression. *GENOME BIOLOGY* , *19* . doi:10.1186/s13059-018-1471-8

Setter, D., Mousset, S., Cheng, X., Nielsen, R., DeGiorgio, M., & Hermisson, J. (2020). VolcanoFinder: genomic scans for adaptive introgression. *PLoS Genetics* , *16* (6), e1008867.

Shchur, V., Svedberg, J., Medina, P., Corbett-Detig, R., & Nielsen, R. (2020). On the distribution of tract lengths during adaptive introgression. *G3: Genes, Genomes, Genetics* , *10* (10), 3663–3673.

Shine, R., Brown, G. P., & Phillips, B. L. (2011). An evolutionary process that assembles phenotypes through space rather than through time. *Proceedings of the National Academy of Sciences* ,*108* (14), 5708–5711.

Siewert, K. M., & Voight, B. F. (2020). BetaScan2: Standardized statistics to detect balancing selection utilizing substitution data.*Genome Biology and Evolution* , *12* (2), 3873–3877.

Posted on Authorea 3 Jan 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160960816.65928724/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Smith, A. L., Hodkinson, T. R., Villellas, J., Catford, J. A., Csergo, A. M., Blomberg, S. P., ... Buckley, Y. M. (2020). Global gene flow releases invasive plants from environmental constraints on genetic diversity. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA* , *117* (8), 4218–4227. doi:10.1073/pnas.1915848117

Smith, C. C. R., & Flaxman, S. M. (2020). Leveraging whole genome sequencing data for demographic inference with approximate bayesian computation. *Molecular Ecology Resources* , *20* (1), 125–139.

Stajich, J. E., & Hahn, M. W. (2005). Disentangling the effects of demography and selection in human history. *Molecular Biology and Evolution* , *22* (1), 63–73.

Stern, D. Ben, & Lee, C. E. (2020). Evolutionary origins of genomic adaptations in an invasive copepod. *NATURE ECOLOGY & EVOLUTION* ,*4* (8), 1084–1094. doi:10.1038/s41559-020-1201-y

Svedberg, J., Shchur, V., Reinman, S., Nielsen, R., & Corbett-Detig, R. (2020). Inferring Adaptive Introgression Using Hidden Markov Models.*BioRxiv* . doi:https://doi.org/10.1101/2020.08.02.232934;

Tay, W. T., & Gordon, K. H. J. (2019). Going global - genomic insights into insect invasions. *CURRENT OPINION IN INSECT SCIENCE* ,*31* , 123–130. doi:10.1016/j.cois.2018.12.002

Tay, W. T., Rane, R., Padovan, A., Walsh, T., Elfekih, S., Downes, S., ... others. (2020). Global FAW population genomic signature supports complex introduction events across the Old World.*BioRxiv* . doi:https://doi.org/10.1101/2020.06.12.147660

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., ... others. (2016). Hybridization and extinction.*Evolutionary Applications* , *9* (7), 892–908.

Travis, J. M. J., Münkemüller, T., Burton, O. J., Best, A., Dytham, C., & Johst, K. (2007). Deleterious mutations can surf to high densities on the wave front of an expanding population. *Molecular Biology and Evolution* , *24* (10), 2334–2343.

Valencia-Montoya, W. A., Elfekih, S., North, H. L., Meier, J. I., Warren, I. A., Tay, W. T., ... Jiggins, C. D. (2020). Adaptive introgression across semipermeable species boundaries between local Helicoverpa zea and invasive Helicoverpa armigera moths. *Molecular Biology and Evolution* . doi:10.1093/molbev/msaa108

Vallejo-Marin, M., & Hiscock, S. J. (2016). Hybridization and hybrid speciation under global change. *NEW PHYTOLOGIST* , *211* (4), 1170–1187. doi:10.1111/nph.14004

Viard, F., Riginos, C., & Bierne, N. (2020). Anthropogenic hybridization at sea: three evolutionary questions relevant to invasive species management. *PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES* , *375* (1806, SI). doi:10.1098/rstb.2019.0547

Wang, J., Gaughan, S., Lamer, J. T., Deng, C., Hu, W., Wachholtz, M., ... Lu, G. (2020). Resolving the genetic paradox of invasions: Preadapted genomes and postintroduction hybridization of bigheaded carps in the Mississippi River Basin. *EVOLUTIONARY APPLICATIONS* ,*13* (2), 263–277. doi:10.1111/eva.12863

Wei, S.-J., Zhou, Y., Fan, X.-L., Hoffmann, A. A., Cao, L.-J., Chen, X.-X., & Xu, Z.-F. (2017). Different genetic structures revealed resident populations of a specialist parasitoid wasp in contrast to its migratory host. *ECOLOGY AND EVOLUTION* , *7* (14), 5400–5409. doi:10.1002/ece3.3097

Welles, S. R., & Dlugosch, K. M. (2018). Population genomics of colonization and invasion. In *Population genomics* (pp. 655–683). Springer.

Wu, N., Zhang, S., Li, X., Cao, Y., Liu, X., Wang, Q., ... Zhan, S. (2019). Fall webworm genomes yield insights into rapid adaptation of invasive species. *NATURE ECOLOGY & EVOLUTION* , *3* (1), 105–115. doi:10.1038/s41559-018-0746-5

Yainna, S., Tay, W. T., Fiteni, E., Legeai, F., Clamens, A.-L., Gimenez, S., ... others. (2020). Genomic balancing selection is key to the invasive success of the fall armyworm. *BioRxiv* .

Posted on Authorea 3 Jan 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160968166.65928724/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

doi:https://doi.org/10.1101/2020.06.17.154880

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., ... others. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* , *329* (5987), 75–78.

Yoshida, K., Miyagi, R., Mori, S., Takahashi, A., Makino, T., Toyoda, A., ... Kitano, J. (2016). Whole-genome sequencing reveals small genomic regions of introgression in an introduced crater lake population of threespine stickleback. *ECOLOGY AND EVOLUTION* , *6* (7), 2190–2204. doi:10.1002/ece3.2047

You, M., Ke, F., You, S., Wu, Z., Liu, Q., He, W., ... others. (2020). Variation among 532 genomes unveils the origin and evolutionary history of a global insect herbivore. *Nature Communications* ,*11* (1), 1–8.

Yuan, Z., Huang, W., Liu, S., Xu, P., Dunham, R., & Liu, Z. (2018). Historical demography of common carp estimated from individuals collected from various parts of the world using the pairwise sequentially markovian coalescent approach. *GENETICA* ,*146* (2), 235–241. doi:10.1007/s10709-017-0006-7

Data Accessibility

The list of 31 papers identified in our literature review are uploaded as supplementary material (Supplementary Table 1).

Acknowledgements

We thank Cynthia Riginos, Iva Popovic and Sinan-Saleh Kassam for their input in early discussions on this topic. All members of the Butterfly Genetics group at the University of Cambridge provided useful feedback, especially Joana Meier and Gabriela Montejo-Kovacevich. We also extend our thanks to the photographers who shared their work free of charge to produce Figure 1: Carol Eunmi Lee, Jessa Thurman, Chen Wu, and James Dorey.