Evolutionary Applications ISSN 1752-4571

ORIGINAL ARTICLE

What, if anything, are hybrids: enduring truths and challenges associated with population structure and gene flow

Zachariah Gompert[1] and C. Alex Buerkle[2]

1 Department of Biology, Utah State University, Logan, UT, USA
2 Department of Botany, University of Wyoming, Laramie, WY, USA

Keywords
admixture, conservation biology, genetic ancestry, hybridization, population genetics.

Correspondence
Zachariah Gompert, Department of Biology, Utah State University, Logan, UT 84322, USA.
Tel.: +1 435 797 9463;
e-mail: zach.gompert@usu.edu

Received: 18 November 2015
Accepted: 27 February 2016

doi:10.1111/eva.12380

Abstract

Hybridization is a potent evolutionary process that can affect the origin, maintenance, and loss of biodiversity. Because of its ecological and evolutionary consequences, an understanding of hybridization is important for basic and applied sciences, including conservation biology and agriculture. Herein, we review and discuss ideas that are relevant to the recognition of hybrids and hybridization. We supplement this discussion with simulations. The ideas we present have a long history, particularly in botany, and clarifying them should have practical consequences for managing hybridization and gene flow in plants. One of our primary goals is to illustrate what we can and cannot infer about hybrids and hybridization from molecular data; in other words, we ask when genetic analyses commonly used to study hybridization might mislead us about the history or nature of gene flow and selection. We focus on patterns of variation when hybridization is recent and populations are polymorphic, which are particularly informative for applied issues, such as contemporary hybridization following recent ecological change. We show that hybridization is not a singular process, but instead a collection of related processes with variable outcomes and consequences. Thus, it will often be inappropriate to generalize about the threats or benefits of hybridization from individual studies, and at minimum, it will be important to avoid categorical thinking about what hybridization and hybrids are. We recommend potential sampling and analytical approaches that should help us confront these complexities of hybridization.

Introduction

Sexual reproduction that involves mating with other individuals (outcrossing rather than selfing) and meiotic recombination mix alleles among different genomic backgrounds. Physical dispersal of individuals before reproduction moves alleles farther from where they originated by mutation and is referred to as gene flow. At some point, crosses can occur between individuals that are unrelated enough that we refer to these as hybrids. Although hybridization has sometimes been viewed as an unimportant dead end, there is a long history of interest in hybridization as a potent creative and destructive evolutionary process (e.g. Stebbins 1950; Ellstrand 1992; Rieseberg and Wendel 1993; Buerkle et al. 2003; Arnold 2006). Numerous cases where hybridization and introgression have had substantial ecological or evolutionary consequences in plants are known. For example, hybridization between the sunflower species *Helinathus annuus* and *Helinathus petiolaris* resulted in multiple distinct hybrid species (Rieseberg et al. 1990, 1995, 2003a), and hybridization in *Populus* affects community composition and ecosystem processes (Driebe and Whitham 2000; Martinsen et al. 2000; Whitham et al. 2006; Floate et al. 2016). Hybridization is particularly common among oak species, where it may spread or generate adaptive genetic variation and where it has been proposed as a key component of natural and human-induced invasions (Petit et al. 2004; Moran et al. 2012).

The consequences of hybridization are directly relevant to aspects of conservation biology and agriculture. Hybridiza-

© 2016 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

tion, whether natural or human induced, can affect the origin, maintenance, and loss of biodiversity (Rhymer and Simberloff 1996; Wolf et al. 2001; Buerkle et al. 2003; Zalapa et al. 2010; Muhlfeld et al. 2014). Hybridization in plants could help endemic species survive periods of climate change (Becker et al. 2013) or result in extinction, when, for example, native species are assimilated by non-native species or experience demographic decline due to outbreeding depression (Ellstrand 1992; Levin et al. 1996; Balao et al. 2015; Gómez et al. 2015). Introgressive hybridization also occurs between crops and their wild relatives, and this too can have beneficial or detrimental consequences for biodiversity (Linder et al. 1998; Ellstrand et al. 2013; Hufford et al. 2013; Warschefsky et al. 2014). Of particular interest is the potential for crop–wild hybridization to allow modified or engineered genes to escape into the wild, which could negatively affect native species or increase public distrust of genetically modified crops (Ellstrand 2001; Stewart et al. 2003; Chapman and Burke 2006; Garnier et al. 2014). Another practical issue is whether and under what conditions hybrid populations or taxa warrant conservation efforts. Hybrids were not granted protection under the US Endangered Species Act, but this was questioned in a federal rule proposed in 1996 (this rule was never adopted; Allendorf et al. 2001, 2013). The proposed federal rule used the term 'intercross' rather than 'hybrid' to avoid a negative connotation of the latter (Allendorf et al. 2013) and we suspect that some people would view even natural hybrids as less worthy of protection than 'pure' species (e.g. the decision to conserve eastern wolves has in part been based on species or hybrid status; Rutledge et al. 2015). Clearly, the potential outcomes and practical consequences of hybridization are multifarious, and thus, different cases of hybridization will need to be treated differently.

Confronting this complexity requires careful consideration of what hybridization is, and when distinguishing among different processes is necessary and possible. The recognition of hybrids between named taxa is relatively uncontroversial, but it is somewhat poorly resolved as to what distance of a cross constitutes hybridization, and what therefore qualifies as a hybrid (Harrison 1993; Arnold 2006; Allendorf et al. 2013). Similarly, different histories of gene flow and selection, such as primary divergence versus secondary contact, have been referred to as hybridization (Barton and Hewitt 1985). However, discriminating among these different histories could be necessary from a management perspective, if, for example, we are to treat cases of natural and human-induced hybridization differently as suggested by Allendorf et al. (2001). Unfortunately, different histories of hybridization can generate very similar or identical patterns of genetic and phenotypic variation (e.g. Barton and Hewitt 1985; Kruuk et al. 1999; Barton and de Cara 2009). This means we might not always be able to distinguish different histories even when doing so would be useful.

In this article, we review and discuss ideas that are relevant to recognition of hybrids and supplement these with simulations to illustrate important contrasts. We acknowledge that is atypical to have a paper contain review, synthesis of concepts and novel simulations, but we think the combination can be useful. The issues we address have a relatively long history, some of which is underappreciated, and clarifying these ideas should have practical consequences for managing hybridization and gene flow in plants. A reexamination of some of these points is worthwhile too because recent population genomic studies have led to a greater appreciation of variation within species and genomic heterogeneity in differentiation between species or populations (e.g. Martin and Orgogozo 2013; Gompert et al. 2014; Mandeville et al. 2015). Additionally, we have learned more about models and approaches that can be used to describe patterns of variation in hybrids (Patterson et al. 2012; Gompert and Buerkle 2013). Along these lines, it is important to recognize what we can and cannot infer about hybrids and hybridization from molecular data; in other words, we must be aware that genetic data provide incomplete information about hybridization. Our simulations and discussion focus on patterns of variation when hybridization is recent and populations are polymorphic; this contrasts with the bulk of theoretical work that concerns long-term equilibrium outcomes of hybridization and often is most applicable when hybridizing taxa exhibit fixed differences. This distinction increases the novelty of our results and makes them particularly informative for applied issues and contemporary hybridization following recent ecological change. In the following, we first address the question of what constitutes hybridization and then turn to the definition of hybrids. We combine literature review and new simulations to answer these questions and conclude each section with recommendations for applied studies of hybridization and gene flow in plants.

What, if anything, is hybridization

Hybridization has been variously defined as interbreeding between different species or subspecies, distinct populations or cultivars, or any individuals with heritable phenotypic differences (Stebbins 1950; Barton and Hewitt 1985; Harrison 1993; Allendorf et al. 2001; Arnold 2006). However, such distinctions downplay the continuous nature of genetic and phenotypic differentiation and distract from the fact that gene flow can have similar consequences anywhere along this continuum (Mayr 1963; Mallet et al. 2007; Martin and Orgogozo 2013). For example, because of population genetic structure and local adaptation within species, intraspecific gene flow can have positive, negative,

 © 2016 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd

or negligible effects on populations that are similar to those of interspecific gene flow (e.g. Ellstrand 1992; Kremer et al. 2012; Nosil et al. 2012; Roe et al. 2014). Moreover, the consequences of interspecific gene flow frequently depend on the specific individuals involved, because of polymorphisms within and among conspecific populations (Sweigart et al. 2007; Escobar et al. 2008; Good et al. 2008; Gompert et al. 2013). In other words, it is the evolutionary and ecological consequences of gene flow that should be considered when defining hybridization. Importantly, the consequences of gene flow do not depend on taxonomy or a specific definition of species, but rather on the nature of differences between groups. Of course, such differences also represent a continuum, and thus, an unambiguous and objective definition of hybridization as something distinct from gene flow is not likely possible. With that said, we think it is useful to reserve the term hybridization for cases where outcrossing and gene flow occur between populations that differ, at least quantitatively, at multiple heritable characters or genetic loci that affect fitness. Thus, we argue that the distinction between gene flow and hybridization is fuzzy and quantitative, rather than discrete and qualitative. While such a view could complicate management decisions, we think it more accurately captures patterns of variation in nature.

Different histories or geographies of gene flow and selection have often been referred to as hybridization. For example, several authors have argued that both primary divergence with gene flow and gene flow following secondary contact (i.e. gene flow after a prolonged period of geographic separation with very little or no gene flow) constitute hybridization (Barton and Hewitt 1985). We think that the case for secondary contact is uncontroversial, but that informed opinions might differ about whether primary divergence includes hybridization. Certainly, primary divergence is not the common conception of hybridization in conservation biology (Allendorf et al. 2001, 2013). Likewise, hybrid zones maintained primarily by exogenous (environment dependent) versus endogenous (environment-independent) selection have been classified and treated similarly. However, management efforts could benefit from distinguishing among these different histories and processes. We might be more inclined to intervene when secondary contact occurs after an anthropogenic disturbance than when primary divergence occurs, even if the latter takes place in a disturbed area.

An equally important question is whether and under what conditions we can in fact discriminate among these different cases. On the one hand, theory shows that over the long-term, primary divergence and secondary contact with exogenous or endogenous selection have similar equilibrium conditions and result in similar geographic patterns of genetic and phenotypic variation (Endler 1977; Barton and Hewitt 1985; Kruuk et al. 1999; Navarro and Barton 2003; Barton and de Vladar 2009; Barton 2013; Flaxman et al. 2014). However, it is also true that well-documented examples of these different cases are known. For example, convergent clines in flowering time in sunflowers are best explained by primary divergence driven by exogenous selection (Blackman et al. 2011; Kawakami et al. 2011), whereas hybridization between *H. annuus* and *H. petiolaris*, which are not sister species, can be attributed to secondary contact (Rieseberg 1991). Additionally, the bulk of evidence suggests that many classic hybrid zones are tension zones maintained by endogenous selection (reviewed in Barton and Hewitt 1985). Consistent with this, Dobzhansky–Muller incompatibilities (DMIs) have been documented in several plant taxa, such as *Mimulus* and *Solanum* (Sweigart et al. 2007; Moyle and Nakazato 2010).

Here, we ask when genetic analyses commonly used to study hybridization might mislead us about the history or nature of gene flow and selection. We are particularly interested in cases where being misled could affect decisions in applied science. We consider primary divergence versus secondary contact, and neutral evolution versus selection on a quantitative trait along an environmental gradient or reduced hybrid fitness due to intrinsic epistatic incompatibilities (i.e. DMIs). We simulate genetic data under each of these conditions and then summarize the results by (i) examining allele frequency and trait clines, (ii) summarizing genetic variation with principal component analysis (PCA), and (iii) estimating admixture proportions. Our goal is not an exhaustive evaluation of these methods, but rather to provide illustrative examples of the potential to be misled by genetic data. We then turn to the related problem of finite sampling. In particular, we show that sparse population sampling when organisms are continuously distributed can lead to false inferences about population structure. That is to say, clinal variation can appear more demic and even suggestive of hybrid speciation. Importantly, and in contrast to most theoretical work on hybridization or hybrid zones, our simulations incorporate shared polymorphism across populations (or species), rather than focusing on genetic markers with fixed differences. This is realistic in general and better reflects the current generation of molecular data (e.g. SNPs identified and scored through genotyping-by-sequencing or exome sequencing).

Simulations and analyses

We used individual-based, genetically explicit simulations to generate pseudo-data under different demographic and evolutionary histories. Simulations were conducted using the program `nemo` version 2.3.44 (Guillaume and Rougemont 2006). Generations were discrete, and each

generation consisted of the following ordered events: breeding, dispersal, viability selection (some histories), and aging. Patches were arranged according to a 1-D stepping-stone model with dispersal allowed only between adjacent patches (dispersal off the outer-edges of the patch vector was allowed). We assumed logistic growth within each patch with a carrying capacity of 5000 individuals and a mean fecundity of two. Genomes consisted of a single chromosome with a recombinational map length of one Morgan. We tracked 200 neutral bi-allelic SNPs in all simulations, and 10 quantitative trait SNPs or DMI SNPs in relevant subsets of the simulations. In all cases, mutation rates were 0.0001 per locus per generation and SNPs were distributed according to a random uniform distribution along the recombinational map of the chromosome (this included neutral and non-neutral SNPs). Simulations lasted 2000 generations.

Starting allele frequencies were generated for neutral markers, quantitative trait SNPs and DMI SNPs to mimic secondary contact or primary divergence (Figure S1). Ancestral allele frequencies were first generated for neutral SNPs by sampling from a beta distribution with α and β equal to 20 (this distribution has a mean of 0.5 and a standard deviation of 0.08). We then obtained initial allele frequencies for the two taxa experiencing secondary contact by sampling from beta$(\alpha = \pi\frac{1-F}{F}, \beta = (1-\pi)\frac{1-F}{F})$, where π is the ancestral allele frequency for the SNP and F corresponds to F_{ST} (Balding and Nichols 1995; Falush et al. 2003), which was set to 0.3 (i.e. substantial population genetic differentiation). We assigned one set of allele frequencies to patches 1–5 and a different set of allele frequencies to patches 6–10. We used the same procedure to generate initial neutral allele frequencies for primary divergence, except the same allele frequencies were assigned to all 10 patches. We initialized quantitative SNPs by assuming the two taxa were perfectly adapted to alternative ends of the patch vector (secondary contact; mean phenotypes of −0.5 and 0.5 were used for patches 1–5 and 6–10, respectively), or by setting the mean phenotype in each patch equal to 0 (primary divergence). We initialized DMI SNPs with different taxa fixed for different sets of derived alleles, such that no fitness reduction occurred within taxa but hybrids would experience reduced fitness (secondary contact), or with all populations fixed for the ancestral allele.

We then simulated five replicate data sets with the following conditions: neutral evolution following secondary contact (no DMIs and no effect of the quantitative trait on fitness), exogenous selection along an environmental gradient with primary divergence, exogenous selection along an environmental gradient following secondary contact, exogenous selection at a sharp ecotone with primary divergence, exogenous selection at a sharp ecotone following secondary contact, endogenous selection caused by DMIs with primary divergence, and endogenous selection caused by DMIs following secondary contact (summarized in Table 1). We repeated all simulations with migration rates of 0.01 and 0.001. Exogenous selection was based on a single quantitative trait that was under stabilizing selection in each patch; we used a Gaussian fitness function with mean μ and variance 0.5. μ varied from −0.5 to 0.5 in steps of 0.1 (most patches) or 0.2 (patches 5 and 6) between patches for the environmental gradient, and was set to −0.5 (patches 1–5) or 0.5 (patches 6–10) for the sharp ecotone. This means that an individual perfectly adapted to one end of the patch vector would have relative fitness of 0.37 at the other end. DMIs were modeled as negative fitness effects between derived alleles at pairs of SNPs. Considering a single locus pair, we assumed the double homozygote for different derived alleles had a fitness of 0.6, and an individual heterozygous at one locus and homozygous for derived alleles at the other had a fitness of 0.8; all other genotypes had a fitness of 1.0. We assumed fitness was absolute (not relative) and multiplicative across DMIs.

Additional data were simulated to evaluate the effect of limited sampling on inference. Our primary motivations were to determine whether sampling gaps would provide false evidence of discrete population clusters or a lack of hybrids when the underlying population structure was continuous (i.e. with isolation by distance). Here, we assumed neutral primary divergence in a 1-D stepping-stone model with 50 patches, each with a carrying capacity of 2500 individuals and a dispersal rate between neighboring patches of 0.001 (our focus on neutral primary divergence reflects our interest in isolation by distance). We initialized neutral allele frequencies as described above. We analyzed either samples from all 50 patches (50 or 5 individuals each),

Table 1. . Summary of conditions for simulations conducted with `nemo` (five replicates each).

Geography	Selection	Migration rate
Secondary contact	None	0.001
Primary divergence	Exogeneous, smooth gradient	0.001
Secondary contact	Exogeneous, smooth gradient	0.001
Primary divergence	Exogeneous, sharp ecotone	0.001
Secondary contact	Exogeneous, sharp ecotone	0.001
Primary divergence	Endogenous (DMIs)	0.001
Secondary contact	Endogenous (DMIs)	0.001
Secondary contact	None	0.01
Primary divergence	Exogeneous, smooth gradient	0.01
Secondary contact	Exogeneous, smooth gradient	0.01
Primary divergence	Exogeneous, sharp ecotone	0.01
Secondary contact	Exogeneous, sharp ecotone	0.01
Primary divergence	Endogenous (DMIs)	0.01
Secondary contact	Endogenous (DMIs)	0.01

DMI, Dobzhansky–Muller incompatibilities.

 © 2016 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd

from sets of four patches at the edges and center of the patch vector (50 individuals each), and from the 12 center patches (50 individuals each).

We used three common analytical approaches to quantify and summarize patterns of genetic variation from the simulations: (i) character and allele frequency clines, (ii) ordination via PCA, and (iii) inference of admixture proportions. We plotted geographic clines in allele frequencies at all neutral SNPs and for the quantitative trait (with the exceptions of DMI simulations, which did not include a quantitative trait). Allele frequencies were polarized such that the rarer allele in the first patch was shown. We conducted PCA on the centered genotype data from 50 individuals from each patch in each simulation (i.e. 500 individuals total for most simulations) using the `prcomp` function in R (R Development Core Team 2015). Genotypes were coded as 0, 1, or 2 copies of one allele at each locus. We estimated admixture proportions using these same genetic data. We used the program `admixture` version 1.23 (Alexander et al. 2009) for this, which fits the same model as the admixture model in `structure` Pritchard et al. (2000), but uses maximum likelihood rather than Bayesian inference. We used the block-relaxation method for parameter estimation with a tolerance of 0.0001 and the Quasi-Newton algorithm for convergence acceleration.

Our analyses show that time since the onset of secondary contact or primary divergence has a profound effect on patterns of genetic variation (Figs 1–3), even over the relatively short temporal scale of our simulations (2000 generations). By the end of the simulations, allele frequency clines were somewhat similar for both histories (i.e. secondary contact and primary divergence), despite clear differences earlier on. PCA and admixture proportions gave similar results. Thus, there may be a relatively narrow window of time during which can distinguish between these histories based on patterns of genetic or phenotypic data. With that said, time here is measured in generations, which could represent vastly different amounts of absolute time for species with different life histories and reproductive strategies (e.g. annual plants versus long-lived, clonal trees).

At the end of the simulations (2000 generations), allele frequency clines and population structure were weak overall, particularly when the migration rate was 0.01 (Figures S2–S4). Phenotypic clines were much more pronounced and followed the environmental gradient or ecotone when exogenous selection occurred. This contrast is not surprising even though the neutral SNPs and quantitative trait SNPs were linked on a single chromosome, because without greater allele frequency differences among populations, limited linkage disequilibrium (LD) is expected. A lower migration rate slowed the decay of differences following secondary contact, but also resulted

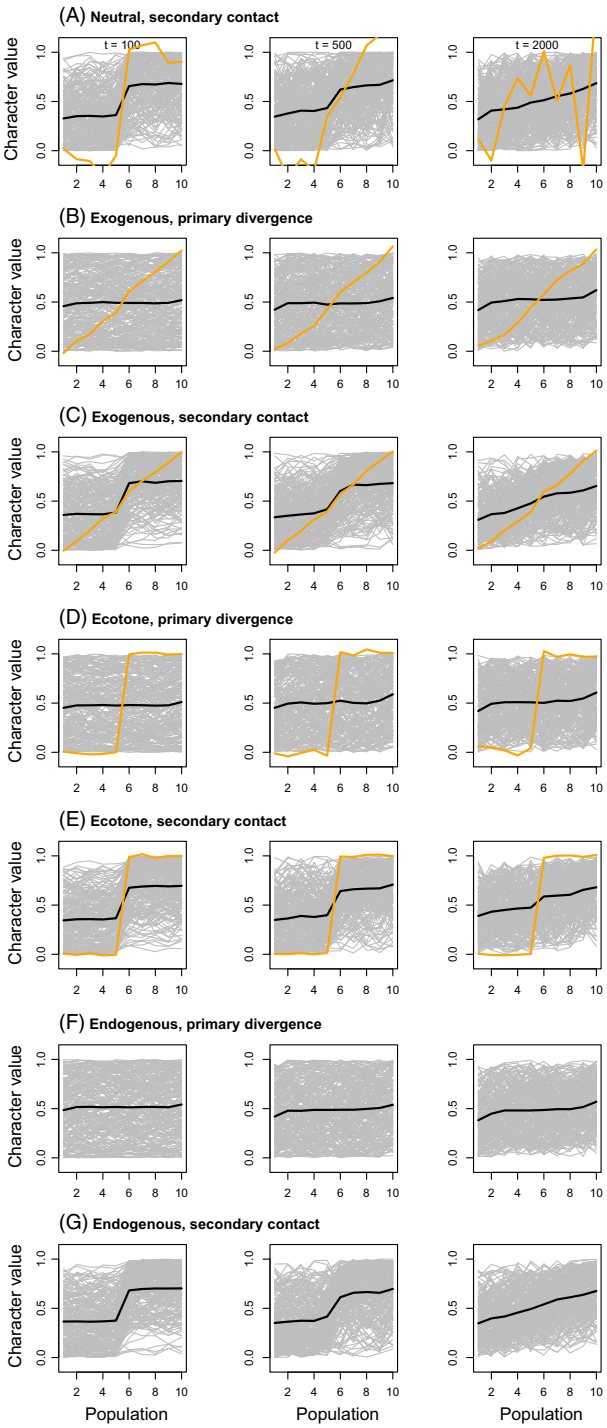

Figure 1 Plots show neutral allele frequency (gray) and quantitative trait (orange) clines from simulated data with a migration rate of 0.001. The mean allele frequency cline with SNPs polarized such that the allele plotted was rarer in patch 1 than patch 10 is depicted with a black line. Clines after 100, 500, and 2000 generations are shown. Results from a single simulation are shown, but replicate simulations produced qualitatively similar results. Clines from simulations with a higher migration rate of 0.01 are shown in Figure S2.

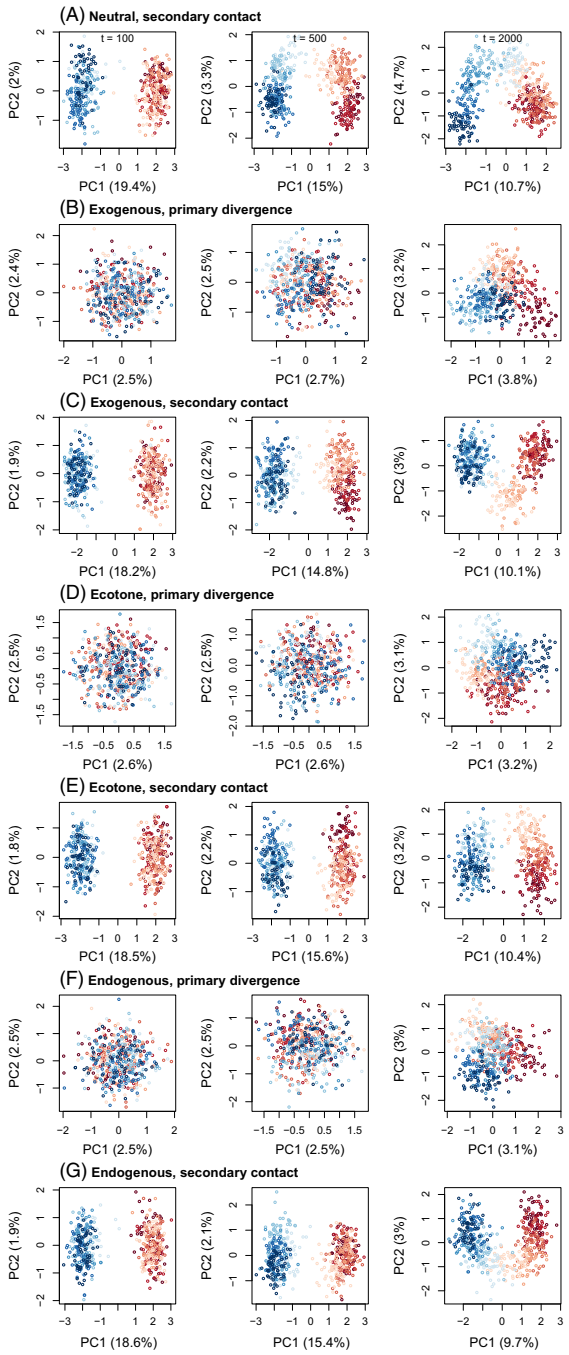

Figure 2 Scatterplots summarize patterns of genotypic variation for simulated data based on principal component analysis (PCA). Points denote individuals and are colored based on patch (dark red and dark blue for patches 1 and 10, with lighter shades indicating patches closer to the center). Results are shown for a migration rate of 0.001 and 100, 500, or 2000 generations. Results from a single simulation are shown, but replicate simulations produced qualitatively similar results. Clines from simulations with a higher migration rate of 0.01 are shown in Figure S3.

in smaller-scale isolation by distance, including sharp phenotypic clines under the neutral secondary contact model, which could be incorrectly attributed to selection. Consistent with previous studies focused on equilibrium dynamics (Kruuk et al. 1999), we found that patterns of variation generated by exogenous and endogenous selection can also be difficult to distinguish earlier in the evolutionary process.

Neutral simulations that included 50 patches resulted in weak population structure overall, and this pattern was robust to sampling a smaller number of individuals per patch (5 vs 50; Fig. 4). However, other sampling approaches resulted in greater distortions of the true population structure. Sampling only center and edge patches resulted in three distinct genotypic clusters, which could be incorrectly interpreted as evidence of an isolated hybrid lineage or even hybrid species (e.g. Gompert et al. 2014). Even sampling only the central patches exaggerates levels of population structure. Together these results highlight the importance of broad geographic sampling to accurately recover clinal variation (also see Witherspoon et al. 2006; Schwartz and McKelvey 2009), as opposed to more limited sampling of putative hybrids and isolated 'pure' parental populations.

Recommendations

Our illustrative simulations are consistent with other theoretical work on hybridization (e.g. Barton and Hewitt 1985; Kruuk et al. 1999; Barton and de Vladar 2009) and show that it will often be difficult to discriminate among different histories of selection and gene flow from genetic data. However, we show that even though primary divergence and secondary contact are thought of as hybridization and result in similar long-term or equilibrium patterns of genetic variation (Barton and Hewitt 1985), recent primary divergence and secondary contact generate different patterns of variation. These differences occur because time is required for LD to buildup between neutral and selected variants with primary divergence (Barton and de Vladar 2009; Flaxman et al. 2014), whereas allele frequency differences between geographically isolated populations will generate LD upon secondary contact. This also means that, during the early stages of hybridization, secondary contact might often lead to segregation of greater functional (and nonfunctional) variation than primary divergence. On the other hand, early stages of primary divergence might be limited to sharp phenotypic and genetic differences for strongly selected characters (e.g. Poelstra et al. 2014; Soria-Carrasco et al. 2014), with less segregating variation for other traits or genes in hybrids. We thus recommend that conservation and management practitioners treat recent primary divergence and secondary contact distinctly, as

 © 2016 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd

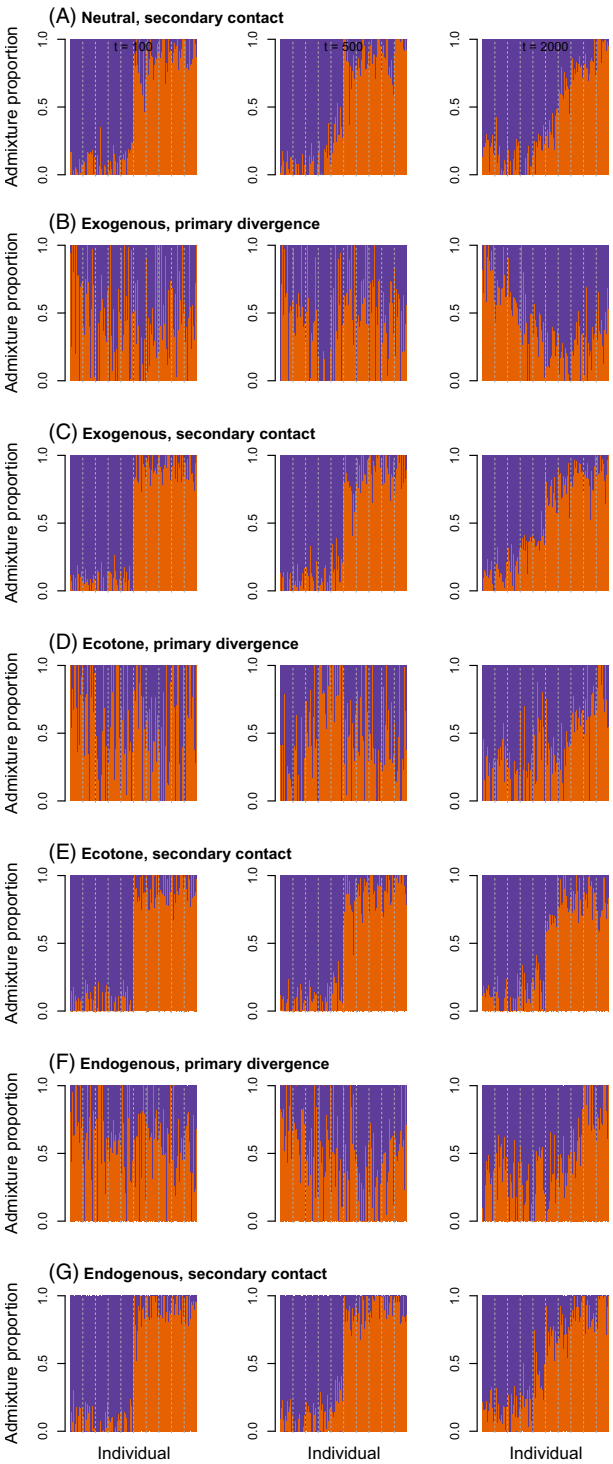

Figure 3 Barplots show maximum likelihood estimates of admixture proportions. Different colors denote ancestry from different hypothetical source populations. Here, we give results for a migration rate of 0.001 and 100, 500, or 2000 generations from a single set of simulations. Replicate simulations produced qualitatively similar results. Admixture from simulations with a higher migration rate of 0.01 is shown in Figure S4.

these processes can be distinguished and have different consequences. Once hybridization has occurred for a greater amount of time, patterns will become similar, and additional data, such as the phylogenetic relationship between or geographic distribution of hybridizing species, will be needed to parse these histories. Our results also show that widespread geographic sampling is important to accurately describe population structure and patterns of hybridization. As argued by practitioners of landscape genetics (e.g. Schwartz and McKelvey 2009), this means that structured, sensible sampling is preferable to sparse opportunistic sampling, or sampling focused on ends of a continuum.

Additional information will likely be gained from studies of hybridization that parse different types of genetic variants rather than treating them all in a single analysis (e.g. Gompert et al. 2014). For example, our simulations and discussion have considered genetic polymorphism, but we have focused on common rather than rare genetic variants. Rare variants, that is genetic variants with minor allele frequencies <1%, have become more accessible with current sequencing methods and could further help discriminate among different histories and provide information about recent evolutionary dynamics (Gravel et al. 2011; Mathieson and McVean 2012; Nelson et al. 2012). In particular, rare variants are often spatially restricted and can be informative about the dispersal of individuals from neighboring populations (Slatkin 1985; Barton and Bengtsson 1986; Gompert et al. 2014) and thus might provide better measures of contemporary gene flow among plant populations of conservation concern. Although more difficult to identify, genetic variants affecting important phenotypes or those linked to such variants could provide additional information if they are strongly structured by the environment (e.g. contrast phenotypic and neutral clines in Fig. 1). When one or a few genes of large effect determine functional phenotypes, it might be useful to examine patterns of genetic variation at these loci. However, when phenotypic variation is due to many variants with smaller effects, statistical approaches that combine information across genetic loci will be more useful (Berg and Coop 2014). Complementary methods that attempt to identify genetic variants potentially affected by selection in hybrids could also be used (e.g. Payseur et al. 2004; Gompert and Buerkle 2009, 2011). Thus, studies of hybridization between crops and wild species or native and non-native plants, as well as gene flow in plants with fragmented populations, would benefit from an increased emphasis on the spread of functional genetic variation via hybridization (e.g. Rieseberg and Willis 2007; Hufford et al. 2013). Such information is needed to determine the fitness consequences of hybridization and thus to decide when hybridization should be valued, allowed or prevented.

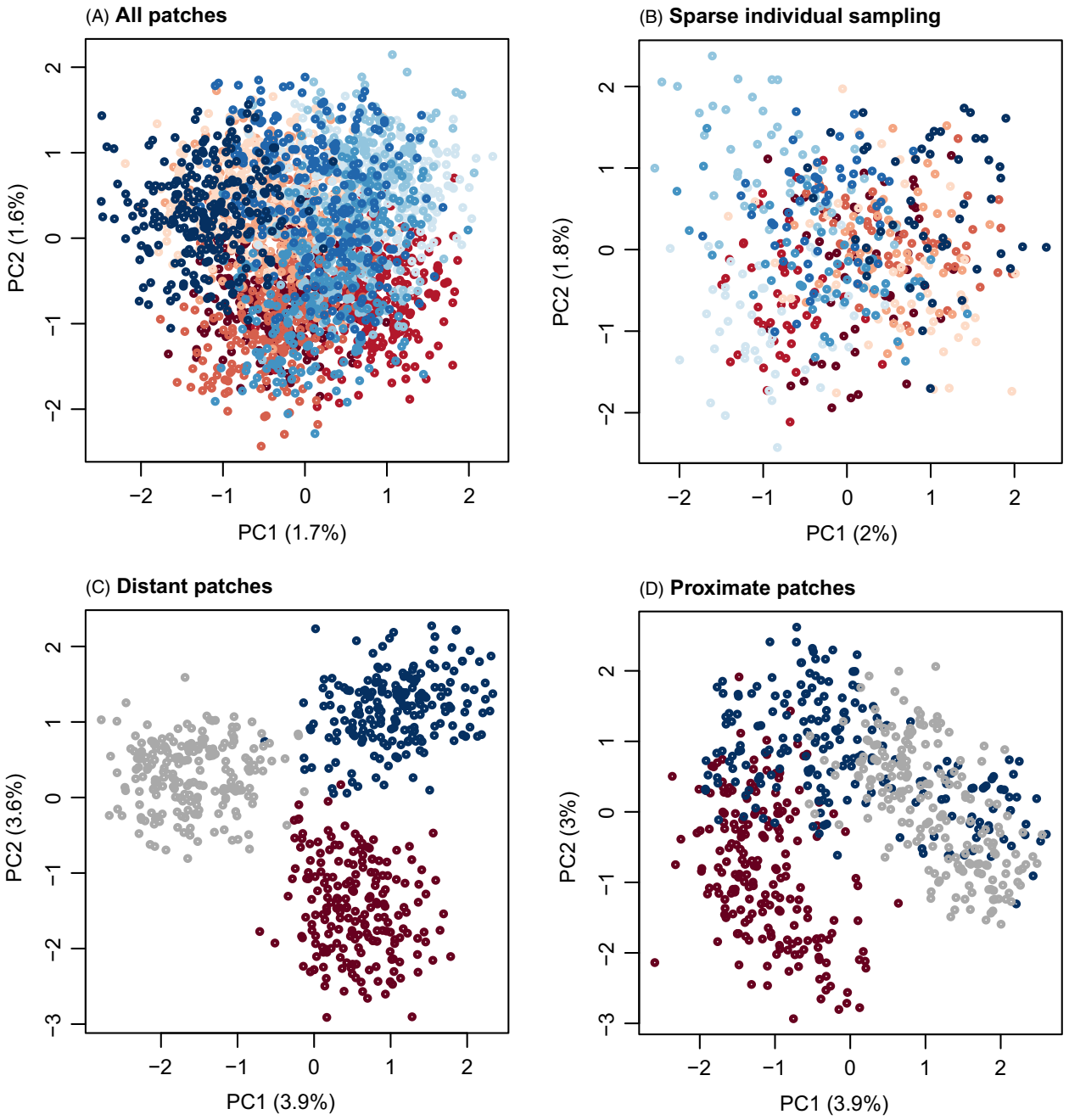

Figure 4 Principal component analysis (PCA) plots illustrate the effect of subsampling on summaries of genetic variation. Points denote individuals and are colored based on patch. Dark red, dark blue, and gray are used to denote peripheral and central patches when a subset of patches were sampled; otherwise dark red and blue indicate patches on opposite ends, with lighter colors used for more central patches. In panes (A) and (B), 50 or 5 individuals were included from each patch. In pane (C), 50 individuals were included from patches 1–4, 24–27 and 47–50, and in pane (D), 50 individuals were sampled from patches 20–31. Results are shown for a migration rate of 0.001 and 100, 500, or 2000 generations.

What, if anything, are hybrids

As noted above, there is a long history of recognizing phenotypically intermediate individuals as putative hybrids between differentiated parental populations or species, including the use of multivariate phenotypic analysis (Alston and Turner 1962; Hatheway 1962; Freeman et al. 1991) and gaining understanding of evolutionary relationships through crossing studies (e.g. Heiser 1947, 1956; Rieseberg 2000). The advent of molecular markers gave rise

 © 2016 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd

to the use of genetic information as the basis of inference of ancestry and the recognition of hybrids (e.g. Harrison and Arnold 1982; Vanlerberghe et al. 1986; Barton and Gale 1993). A variety of statistical models exist to support the recognition of hybrids and their distinction from individuals from parental populations (including species), both using population genetic (Boecklen and Howard 1997; Barton 2000; Pritchard et al. 2000; Anderson and Thompson 2002; Falush et al. 2003) and tree-based models (Durand et al. 2011; Patterson et al. 2012). These various models and their implementation in software allow quantitative, model-based recognition of hybrids, given sufficient, informative genetic data (Anderson and Thompson 2002; Falush et al. 2003; Vaha and Primmer 2006). Yet, the papers that describe these models often include explicit cautionary statements regarding the difficulty of distinguishing among different hybrid genealogies, as well as distinguishing hybrids from parentals (e.g. Barton 2000; Anderson and Thompson 2002). Aside from the problem of alleles shared between parental taxa and the resulting imperfect information about ancestry from allelic state, hybrids can be difficult to recognize simply because genetic recombination and sexual reproduction in different genealogies can lead to the same, ambiguous combination of alleles in genotypes. While the genetic variation that results from hybridization is known, it is not clear that as biologists we appreciate the extent to which different hybrid genealogies can lead to the same genetic composition. To illustrate the overlapping expectations for ancestry and genotypic composition of hybrids, we present a simple set of simulations in this section (reprising related simulations and results in Fitzpatrick 2012; Gompert et al. 2014; Lindtke et al. 2014), and their continuous variation along multiple dimensions of hybridization. These illustrations lead to the conclusions that it can be misleading to think about ancestry categories of hybrids and that hybrids will often be genetically and functionally diverse.

The fractional contribution of two (or more) parental taxa to the ancestry of hybrids is a common measure of hybridity and ancestry and is typically referred to as a hybrid index (Barton and Gale 1993; Boecklen and Howard 1997; Buerkle 2005) or admixture proportion (Pritchard et al. 2000; Falush et al. 2003). In the simple case of putative hybridization between two parental taxa, the hybrid index or admixture proportion (q) corresponds to variation along a single axis, with parental ancestry at each end and hybrids intermediate. Summarizing admixture in this way is very common, but it also disregards important information about the history of admixture (Barton 2000; Anderson and Thompson 2002; Fitzpatrick 2012; Lindtke et al. 2012; Gompert et al. 2014; Lindtke et al. 2014). For example, F_1 individuals will have a hybrid index of 0.5, but this is also the expected (mean) hybrid index of any $F_2 \cdots F_n$ hybrid individuals, which do not have one of the parental taxa as a parent after the first generation of hybridization (i.e. they have experienced no backcrossing). Consequently, whereas a hybrid index does quantify a continuum of genetic hybridity and is preferable to a categorical analysis, it cannot discriminate among very different genealogies, including the differences in ancestry between an F_2 and an F_{20}. Additional information can be obtained from a second dimension of admixture, the fraction of loci that combine ancestry from the two parental taxa, which has been referred to as interspecific heterozygosity or interpopulation ancestry (denoted Q_{12} here; Barton 2000; Fitzpatrick and Shaffer 2007; Fitzpatrick 2012; Lindtke et al. 2012; Gompert et al. 2014; Lindtke et al. 2014). Some software models this parameter explicitly from genetic data in hybrids and source populations (e.g. `HIest` and `entropy`; models for interpopulation ancestry are described in Fitzpatrick 2012; Gompert et al. 2014), but the most commonly used software for admixture analysis does not (`structure`; Pritchard et al. 2000; Falush et al. 2003). The combination of admixture proportion (q) and interpopulation ancestry (Q_{12}) contains additional information about admixture histories and thus is a general tool for summarizing the genomic composition of hybrids. For one, it allows identification of individuals that had a parental taxon as an immediate parent (including F_1 and any backcrossed hybrids), as these have maximal Q_{12} for a given q.

Simulations and analyses

As has been done in previous studies (Fitzpatrick 2012; Gompert et al. 2014; Lindtke et al. 2014), we performed individual-based simulations of hybridization. In the first set, we repeatedly modeled two generations of hybridization that included parental, F_1, F_2, and backcross (BC) individuals. In a second set, we used replicates to generate expectations for the ancestry of F_2, F_5 and F_{20} individuals. The simulations were of finite populations of 50 individuals that contribute to the parentage of any set of progeny (F_1, F_2, etc.). Diploid meiotic recombination and segregation were modeled, with 1000 marker loci distributed across 10 chromosomes, and a single, randomly located crossover per chromosome in each gamete. Thus, we were able to track ancestry with complete knowledge. To superimpose allelic states (including shared alleles between parental taxa and polymorphism within), we utilized an F-model for shared ancestry of parental taxa and the genetic drift they experienced relative to the common ancestor (as above, and in Balding and Nichols 1995; Falush et al. 2003), with a beta distribution of allele frequencies in the ancestral population with parameters α and β equal to 0.8 (this distribution has a mean allele frequency of 0.5 and a

standard deviation of 0.31). We set $F_{ST} = 0.5$ and only considered a random subset of 1000 marker loci with a minor allele frequency >0.05 in the sampled individuals (i.e. what are typically referred to as 'polymorphic' loci or common variants). We arbitrarily sampled 20 individuals of each of the parental taxa, 20 F_2, and 10 each F_5, F_{20}, and BC to each parental taxon. The simulations were performed in R (version 3.2.2; R Development Core Team 2015) and the script to perform the simulations is in the Supporting information.

Our simulations tracked both the ancestry and allelic state of loci, and we present summaries of both (Fig. 5). Because we simulated admixture, we had perfect knowledge of the admixture proportions and interpopulation ancestry rather than needing to infer them. If one were to infer ancestries based on models and software (e.g. Gompert et al. 2014), there would be more uncertainty and variance around the true values shown here (uncertainty in ancestry is inversely proportional to allele frequency differences between the parental taxa, that is, to the extent that allelic state is informative about ancestry). With the level of allele frequency difference between our parental populations ($F_{ST} = 0.5$), recognition of parental individuals and distinguishing them from all hybrids was unambiguous with PCA (Fig. 5, PCA performed in R; R Development Core Team 2015; it would be more difficult to distinguish parental and hybrid individuals based on allelic state if parental populations were more similar genetically). In terms of ancestry, F_1 individuals were distinguishable from more advanced generation F_n hybrids (F_2, F_5, and F_{20}) and backcrossed individuals on the basis of their maximal interpopulation ancestry. Likewise, BC individuals are recognizable on the basis of their maximal interpopulation ancestry for a given admixture proportion. Distinguishing among different generations of backcrossing (e.g. whether F_1, or F_2 was hybrid parent) would not be possible based only on the information contained in Q_{12} and q (knowledge of chromosomal blocks of ancestry would be helpful; Gompert and Buerkle 2013). Segregation even in the F_1 is highly variable and ancestry in later generation hybrid F_n parents is expected to overlap with that of the F_1. More generally and as noted in previous research, discriminating between genealogies beyond the first two generations of admixture is difficult (Barton 2000; Anderson and Thompson 2002) without additional information. This is illustrated in these simulations by the overlapping expectations for Q_{12} and q across the individuals in the F_2, F_5, and F_{20} generations. While drift would cause Q_{12} to decline over further

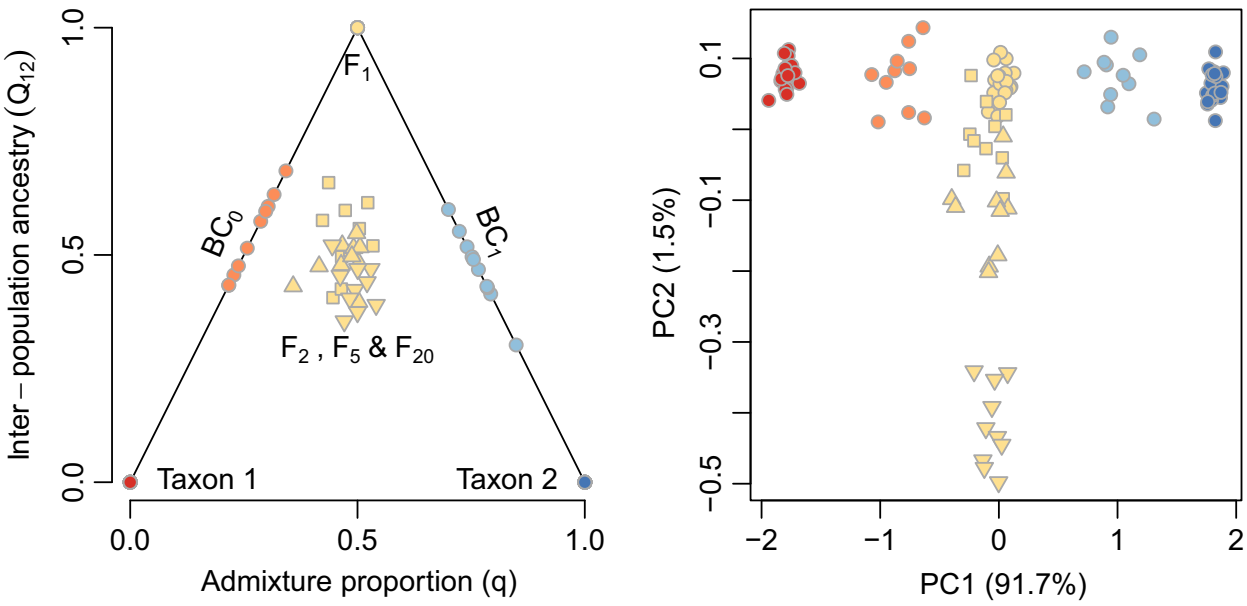

Figure 5 Ancestry for simulated individuals from parental taxa (Taxon 1 and 2) and hybrids vary in admixture proportion (q) and the fraction of loci at which individuals have ancestry from both parental taxa (Q_{12}, interpopulation ancestry; left pane of plot). Hybrids that are progeny from a cross involving one (BC) or both (F_1) parental taxa have maximal interpopulation ancestry for a given admixture proportion (on the edges of the triangle). In contrast, progeny from crosses between hybrid individuals ($F_2 \cdots F_n$) has less than maximal interpopulation ancestry for a given admixture proportion. Principal component analysis (PCA) of genetic covariances among individuals in the simulated population (right pane) shows that genetic differences between the parental species (ancestry variation) constitute the dominant axis of genetic variation (colors as in left pane). $F_1 \cdots F_n$ are genetically intermediate on PC1, and across all hybrids, PC1 mirrors the admixture proportion. F_{20} individuals (downward-pointing triangles) are distinguishable genetically from earlier F_n hybrids and in general PC2 is associated with genetic variation among F_n generations.

 © 2016 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd

generations and ultimately lead to the fixation of ancestry states in finite populations over time (Stam 1980; Chapman and Thompson 2002, 2003; MacLeod et al. 2005; Buerkle and Rieseberg 2008), 20 generations are insufficient to have a detectable effect in a simulated population of 50 individuals.

If hybridization is restricted to two generations, with sufficient genomic sampling, it can be possible to identify different parental, F_1, F_2, and BC categories. Here, we have considered perfect knowledge of ancestry and previous work has addressed sources of uncertainty that would lessen the prospects for clear expectations for ancestry (Anderson and Thompson 2002; Vaha and Primmer 2006; Burgarella et al. 2009; Fitzpatrick 2012). But beyond the issues of whether hybridization involves only two generations of hybridization and uncertainty in empirical ancestry estimates, analysis of ancestry categories is problematic because these classes mask the fact that they will contain genetic and functional phenotypic variation. Perhaps we stand to be the most misled in the case of F_1 individuals, where allelic polymorphisms in the parental populations will result in genetically variable F_1 individuals, contrary to the expectation for a single genotype of F_1 resulting from typical experimental crosses between homozygous parents. This genetic variation is evident among F_1 individuals in our simulations (Fig. 5). This variation would be even greater if one considered F_1 individuals from different geographic locations, where allele frequencies in parental taxa are likely to differ even more (e.g. Gompert et al. 2014; Mandeville et al. 2015). Somewhat similarly, categorical treatment of ancestry in hybrids leads to overlapping expectations for Q_{12} and q in F_2, F_5, and F_{20} individuals, but analysis of their allelic states shows that F_{20} individuals differ genetically from early F_n generations (PCA in Fig. 5). These genetic differences could be responsible for functional phenotypic differences and comparable functional differences might arise from more subtle genotypic differences between F_n generations.

Overall, the use of the term *hybrid classes* or *categories*, and methods for their inference, could obscure important variation that exists within classes. Instead, Rieseberg and Carney (1998) suggested it is worthwhile to focus on the fitness of individual genotypes, rather than hybrid classes. Certainly, our simple model illustrates that it can be nonsensical to refer to the 'fitness of hybrids', as the genetic and ancestry composition of hybrids can be highly variable and hybrids would be expected to vary substantially for phenotypes. Given the expected genetic and phenotypic variability within hybrids, and the potential for transgressive phenotypes (Rieseberg et al. 1999a, 2003b), discussion of hybrid fitness should be in the context of the typical complexity of tying phenotype (including fitness) to genotype in natural populations, which is particularly difficult in variable environments and in variable genetic backgrounds (Weiss 2008; Rockman 2012).

Recommendations

In both applied and basic science settings, knowledge of the existence and attributes of hybrids can provide a foundation for learning about species interactions and maintenance (Arnold 2006; Allendorf et al. 2013). For example, a predominance of BC hybrids would lead to genetic exchange between parental taxa and a potential local erosion of species differences, whereas if hybrids are restricted to relatively abundant F_n individuals, these will affect the demography of parental taxa through wasted reproductive effort on F_1 hybrid progeny and possibly through competition. Our simple model reflects our understanding that the ancestry and genetic composition of hybrids vary along multiple axes and treatment of hybrids as a singular entity would disregard potentially important variation. Thus, management decisions might need to consider the types of hybrids generated and could even accommodate different actions for different hybrids within the same biological system.

Furthermore, hybrids beyond the F_1 will also vary in ancestry along their chromosomes, both in tracts of ancestry that have not yet recombined, and as a result of drift and selection leading individual loci to deviate from the average ancestry in the genome (reviewed in Gompert and Buerkle 2013). For this reason, hybrids have been of interest to evolutionary biologists who are interested in the genetics of species boundaries (Rieseberg et al. 1999b; Rieseberg and Buerkle 2002; Buerkle and Lexer 2008; Payseur 2010; Gompert et al. 2012). Incomplete reproductive isolation and hybridization have provided support for the 'genic view' of speciation and species boundaries (Wu 2001; Abbott et al. 2013). Additionally, recent studies in a variety of taxa have drawn attention to variability in the genetic outcomes of hybridization that followed secondary contact between the same pairs of species in multiple locations or contexts (Rieseberg 2006; Nolte et al. 2009; Teeter et al. 2010; Lepais and Gerber 2011; Lagache et al. 2013; Gompert et al. 2014; Mandeville et al. 2015). For both applied and basic evolutionary biology, this variability in outcomes means that it can be difficult to formulate categorical statements about the composition, importance, and likely conservation threats of hybrids. The empirical abundance of parental taxa and hybrids at one site may or may not be informative about other locations where the taxa co-occur (e.g. Aldridge and Campbell 2009; Mandeville et al. 2015). Likewise, as noted above, genetic variation in parents and hybrids makes it difficult to make categorical statements about the genotypes, phenotypes, and fitness of hybrids (e.g. Sweigart et al. 2007; Good et al. 2008). This

challenge is not a matter of uncertainty that arises from analytical approaches and software, but is inherent to the process of hybridization, as we have illustrated with the simulations in this paper.

Overall, these complexities mean it will be difficult to know the consequences of hybridization without detailed study (Allendorf et al. 2013), which could include estimation of multiple dimensions of ancestry (Gompert and Buerkle 2013), sampling multiple geographic locations and contexts (e.g. Hamilton et al. 2013; Haselhorst and Buerkle 2013; Gompert et al. 2014; Mandeville et al. 2015), and characterization of the demography of parental and hybrid individuals in populations (e.g. Carney et al. 2000; Fitzpatrick and Shaffer 2007).

Synthesis and conclusions

A common theme of our results and discussion is that hybridization is not a singular process, but rather a collection of related processes with variable outcomes and consequences. In support of this, as noted above, empirical studies have often documented variation in outcomes of hybridization in different locations or contexts, in terms of the genomic composition of hybrids, patterns of introgression, and the ecological consequences of hybridization (e.g. Yanchukov et al. 2006; Lepais et al. 2009; Nolte and Tautz 2010; Teeter et al. 2010; Nice et al. 2013; Gompert et al. 2014; Mandeville et al. 2015). Indeed, consistent outcomes of hybridization appear to be mostly limited to taxa that exhibit limited intraspecific variation for loci affecting fitness and where endogenous selection dominates (e.g. Buerkle and Rieseberg 2001). Such variability limits our ability to predict the outcome of specific instances of hybridization and thus is relevant for our understanding of evolutionary biology in general and has practical consequences for management. For example, invasion by a non-native species could result in extirpation of a native species in one area but not in another, or transgene escape from a crop could occur readily into some wild populations but not others. Thus, it might be difficult to make valid general statements about the threats or benefits of hybridization, even for individual species. Likewise, extrapolation from single empirical examples (i.e. studies or sites) could be problematic. Clearly, such problems will be exacerbated when species exhibit substantial isolation by distance or local adaptation and fail to function as cohesive entities. We conclude by noting that while we have focused on hybridization between pairs of diploid populations or species, the points we have made should also apply when hybridization generates polyploids or involves multiple species. However, in such cases, even more factors could affect the ecological and evolutionary dynamics, rendering the outcomes of these instances of hybridization even less predictable.

Acknowledgements

We thank Norman Ellstrand and Loren Rieseberg for the invitation to contribute to this special issue. Computing, storage, and other resources from the Division of Research Computing in the Office of Research and Graduate Studies at Utah State University are gratefully acknowledged.

Data archiving statement

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.g7g45.

Literature cited

Abbott, R., D. Albach, S. Ansell et al. 2013. Hybridization and speciation. Journal of Evolutionary Biology **26**:229–246.

Aldridge, G., and D. Campbell 2009. Genetic and morphological patterns show variation in frequency of hybrids between *Ipomopsis* (*Polemoniaceae*) zones of sympatry. Heredity **102**:257–265.

Alexander, D. H., J. Novembre, and K. Lange 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research **19**:1655–1664.

Allendorf, F., R. Leary, P. Spruell, and J. Wenburg 2001. The problems with hybrids: setting conservation guidelines. Trends in Ecology & Evolution **16**:613–622.

Allendorf, F. W., G. Luikart, and S. N. Aitken 2013. Conservation and the Genetics of Populations, 2nd edn. Wiley-Blackwell, Hoboken, NJ.

Alston, R. E., and B. L. Turner 1962. New techniques in analysis of complex natural hybridization. Proceedings of the National Academy of Sciences of the USA **48**:130–137.

Anderson, E. C., and E. A. Thompson 2002. A model-based method for identifying species hybrids using multilocus genetic data. Genetics **160**:1217–1229.

Arnold, M. L. 2006. Evolution Through Genetic Exchange. Oxford University Press, Oxford, UK.

Balao, F., R. Casimiro-Soriguer, J. L. García-Castaño, A. Terrab, and S. Talavera 2015. Big thistle eats the little thistle: does unidirectional introgressive hybridization endanger the conservation of *Onopordum hinojense*? The New Phytologist **206**:448–458.

Balding, D. J., and R. A. Nichols 1995. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. Genetica **96**:3–12.

Barton, N. H. 2000. Estimating multilocus linkage disequilibria. Heredity **84**:373–389.

Barton, N. 2013. Does hybridization influence speciation? Journal of Evolutionary Biology **26**:267–269.

Barton, N., and B. Bengtsson 1986. The barrier to genetic exchange between hybridizing populations. Heredity **57**:357–376.

Barton, N. H., and M. A. R. de Cara 2009. The evolution of strong reproductive isolation. Evolution **63**:1171–1190.

Barton, N. H., and K. S. Gale 1993. Genetic analysis of hybrid zones. In R. G. Harrison, ed. Hybrid Zones and the Evolutionary Process, pp. 13–45. Oxford University Press, New York, NY.

 © 2016 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd

Barton, N. H., and G. M. Hewitt 1985. Analysis of hybrid zones. Annual Review of Ecology and Systematics **16**:113–148.

Barton, N. H., and H. P. de Vladar 2009. Statistical mechanics and the evolution of polygenic quantitative traits. Genetics **181**:997–1011.

Becker, M., N. Gruenheit, M. Steel et al. 2013. Hybridization may facilitate in situ survival of endemic species through periods of climate change. Nature Climate Change **3**:1039–1043.

Berg, J. J., and G. Coop 2014. A population genetic signal of polygenic adaptation. PLoS Genetics **10**:e1004412.

Blackman, B. K., S. D. Michaels, and L. H. Rieseberg 2011. Connecting the sun to flowering in sunflower adaptation. Molecular Ecology **20**:3503–3512.

Boecklen, W. J., and D. J. Howard 1997. Genetic analysis of hybrid zones: numbers of markers and power of resolution. Ecology **78**:2611–2616.

Buerkle, C. A. 2005. Maximum-likelihood estimation of a hybrid index based on molecular markers. Molecular Ecology Notes **5**:684–687.

Buerkle, C. A., and C. Lexer 2008. Admixture as the basis for genetic mapping. Trends in Ecology & Evolution **23**:686–694.

Buerkle, C. A., and L. H. Rieseberg 2001. Low intraspecific variation for genomic isolation between hybridizing sunflower species. Evolution **55**:684–691.

Buerkle, C. A., and L. H. Rieseberg 2008. The rate of genome stabilization in homoploid hybrid species. Evolution **62**:266–275.

Buerkle, C. A., D. E. Wolf, and L. H. Rieseberg 2003. The origin and extinction of species through hybridization. In C. A. Brigham, and M. W. Schwartz, eds. Population Viability in Plants: Conservation, Management, and Modeling of Rare Plants, pp. 117–141. Springer Verlag, New York, NY.

Burgarella, C., Z. Lorenzo, R. Jabbour-Zahab et al. 2009. Detection of hybrids in nature: application to oaks (*Quercus suber* and *Q. ilex*). Heredity **102**:442–452.

Carney, S. E., K. A. Gardner, and L. H. Rieseberg 2000. Evolutionary changes over the fifty-year history of a hybrid population of sunflowers (*Helianthus*). Evolution **54**:462–474.

Chapman, M. A., and J. M. Burke 2006. Letting the gene out of the bottle: the population genetics of genetically modified crops. The New Phytologist **170**:429–443.

Chapman, N. H., and E. A. Thompson 2002. The effect of population history on the lengths of ancestral chromosome segments. Genetics **162**:449–458.

Chapman, N. H., and E. A. Thompson 2003. A model for the length of tracts of identity by descent in finite random mating populations. Theoretical Population Biology **64**:141–150.

Driebe, E., and T. Whitham 2000. Cottonwood hybridization affects tannin and nitrogen content of leaf litter and alters decomposition. Oecologia **123**:99–107.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin 2011. Testing for ancient admixture between closely related populations. Molecular Biology and Evolution **28**:2239–2252.

Ellstrand, N. C. 1992. Gene flow by pollen: implications for plant conservation genetics. Oikos **63**:77–86.

Ellstrand, N. C. 2001. When transgenes wander, should we worry? Plant Physiology **125**:1543–1545.

Ellstrand, N. C., P. Meirmans, J. Rong et al. 2013. Introgression of crop alleles into wild or weedy populations. Annual Review of Ecology, Evolution, and Systematics **44**:325–345.

Endler, J. A. 1977. Geographic Variation, Speciation, and Clines. Princeton University Press, Princeton, NJ.

Escobar, J. S., A. Nicot, and P. David 2008. The different sources of variation in inbreeding depression, heterosis and outbreeding depression in a metapopulation of *Physa acuta*. Genetics **180**:1593–1608.

Falush, D., M. Stephens, and J. K. Pritchard 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **164**:1567–1587.

Fitzpatrick, B. 2012. Estimating ancestry and heterozygosity of hybrids using molecular markers. BMC Evolutionary Biology **12**:131.

Fitzpatrick, B. M., and H. B. Shaffer 2007. Hybrid vigor between native and introduced salamanders raises new challenges for conservation. Proceedings of the National Academy of Sciences of the USA **104**:15793–15798.

Flaxman, S. M., A. C. Wacholder, J. L. Feder, and P. Nosil 2014. Theoretical models of the influence of genomic architecture on the dynamics of speciation. Molecular Ecology **23**:4074–4088.

Floate, K. D., J. Godbout, M. K. Lau, N. Isabel, and T. G. Whitham 2016. Plantherbivore interactions in a trispecific hybrid swarm of *Populus*: assessing support for hypotheses of hybrid bridges, evolutionary novelty and genetic similarity. The New Phytologist **209**:832–844, 2015-19480.

Freeman, D. C., W. A. Turner, E. D. McArthur, and J. H. Graham 1991. Characterization of a narrow hybrid zone between two subspecies of big sagebrush (*Artemisia tridentata*: Asteraceae). American Journal of Botany **78**:805–815.

Garnier, A., H. Darmency, Y. Tricault, A. M. Chèvre, and J. Lecomte 2014. A stochastic cellular model with uncertainty analysis to assess the risk of transgene invasion after crop-wild hybridization: oilseed rape and wild radish as a case study. Ecological Modelling **276**:85–94.

Gómez, J. M., A. González-Megías, J. Lorite, M. Abdelaziz, and F. Perfectti 2015. The silent extinction: climate change and the potential hybridization-mediated extinction of endemic high-mountain plants. Biodiversity and Conservation **24**:1843–1857.

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

Gompert, Z., and C. A. Buerkle 2011. Bayesian estimation of genomic clines. Molecular Ecology **20**:2111–2127.

Gompert, Z., and C. A. Buerkle 2013. Analyses of genetic ancestry enable key insights for molecular ecology. Molecular Ecology **22**:5278–5294.

Gompert, Z., T. L. Parchman, and C. A. Buerkle 2012. Genomics of isolation in hybrids. Philosophical Transactions of the Royal Society of London B: Biological Sciences **367**:439–450.

Gompert, Z., L. K. Lucas, C. C. Nice, J. A. Fordyce, C. A. Buerkle, and M. L. Forister 2013. Geographically multifarious phenotypic divergence during speciation. Ecology and Evolution **3**:595–613.

Gompert, Z., L. K. Lucas, C. A. Buerkle, M. L. Forister, J. A. Fordyce, and C. C. Nice 2014. Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. Molecular Ecology **23**:4555–4573.

Good, J. M., M. A. Handel, and M. W. Nachman 2008. Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. Evolution **62**:50–65.

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

Guillaume, F., and J. Rougemont 2006. Nemo: an evolutionary and population genetics programming framework. Bioinformatics **22**:2556–2557.

Hamilton, J. A., C. Lexer, and S. N. Aitken 2013. Genomic and phenotypic architecture of a spruce hybrid zone (*Picea sitchensis* × *P. glauca*). Molecular Ecology **22**:827–841.

Harrison, R. G. 1993. Hybrids and Hybrid Zones: Historical Perspective, Chap. 1, pp. 3–12. Oxford University Press, New York, NY.

Harrison, R. G., and J. Arnold 1982. A narrow hybrid zone between closely related cricket species. Evolution **36**:535–552.

Haselhorst, M. S. H., and C. A. Buerkle 2013. Population genetic structure of *Picea engelmannii*, *P. glauca* and their previously unrecognized hybrids in the central Rocky Mountains. Tree Genetics & Genomes **9**:669–681.

Hatheway, W. H. 1962. A weighted hybrid index. Evolution **16**:1–10.

Heiser, C. B. Jr 1947. Hybridization between the sunflower species *Helianthus annuus* and *H. petiolaris*. Evolution **1**:249–262.

Heiser, C. B. 1956. Biosystematics of *Helianthus debilis*. Madroño **13**:145–167.

Hufford, M. B., P. Lubinksy, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand, and J. Ross-Ibarra 2013. The genomic signature of crop-wild introgression in maize. PLoS Genetics **9**:e1003477.

Kawakami, T., T. J. Morgan, J. B. Nippert et al. 2011. Natural selection drives clinal life history patterns in the perennial sunflower species, *Helianthus maximiliani*. Molecular Ecology **20**:2318–2328.

Kremer, A., O. Ronce, J. J. Robledo-Arnuncio et al. 2012. Long-distance gene flow and adaptation of forest trees to rapid climate change. Ecology Letters **15**:378–392.

Kruuk, L. E. B., S. J. E. Baird, K. S. Gale, and N. H. Barton 1999. A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. Genetics **153**:1959–1971.

Lagache, L., E. K. Klein, E. Guichoux, and R. J. Petit 2013. Fine-scale environmental control of hybridization in oaks. Molecular Ecology **22**:423–436.

Lepais, O., and S. Gerber 2011. Reproductive patterns shape introgression dynamics and species succession within the European white oak species complex. Evolution **65**:156–170.

Lepais, O., R. Petit, E. Guichoux et al. 2009. Species relative abundance and direction of introgression in oaks. Molecular Ecology **18**:2228–2242.

Levin, D. A., J. K. Francisco-Ortega, and R. K. Jansen 1996. Hybridization and the extinction of rare plant species. Conservation Biology **10**:10–16.

Linder, C. R., I. Taha, G. J. Seiler, A. A. Snow, and L. H. Rieseberg 1998. Long-term introgression of crop genes into wild sunflower populations. Theoretical and Applied Genetics **96**:339–347.

Lindtke, D., C. A. Buerkle, T. Barbará et al. 2012. Recombinant hybrids retain heterozygosity at many loci: new insights into the genomics of reproductive isolation in *Populus*. Molecular Ecology **21**:5042–5058.

Lindtke, D., Z. Gompert, C. Lexer, and C. A. Buerkle 2014. Unexpected ancestry of *Populus* seedlings from a hybrid zone implies a large role for postzygotic selection in the maintenance of species. Molecular Ecology **23**:4316–4330.

MacLeod, A. K., C. S. Haley, J. A. Woolliams, and P. Stam 2005. Marker densities and the mapping of ancestral junctions. Genetical Research **85**:69–79.

Mallet, J., M. Beltran, W. Neukirchen, and M. Linares 2007. Natural hybridization in Heliconiine butterflies: the species boundary as a continuum. BMC Evolutionary Biology **7**:28.

Mandeville, E. G., T. L. Parchman, D. B. McDonald, and C. A. Buerkle 2015. Highly variable reproductive isolation among pairs of catostomus species. Molecular Ecology **24**:1856–1872.

Martin, A., and V. Orgogozo 2013. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. Evolution **67**:1235–1250.

Martinsen, G. D., K. D. Floate, A. M. Waltz, G. M. Wimp, and T. G. Whitham 2000. Positive interactions between leafrollers and other arthropods enhance biodiversity on hybrid cottonwoods. Oecologia **123**:82–89.

Mathieson, I., and G. McVean 2012. Differential confounding of rare and common variants in spatially structured populations. Nature Genetics **44**:243–246.

Mayr, E. 1963. Animal Species and Evolution. Harvard University Press, Cambridge, MA.

Moran, E. V., J. Willis, and J. S. Clark 2012. Genetic evidence for hybridization in red oaks (*Quercus* sect. *Lobatae*, *Fagaceae*). American Journal of Botany **99**:92–100.

Moyle, L. C., and T. Nakazato 2010. Hybrid incompatibility "snowballs" between *Solanum* species. Science **329**:1521–1523.

Muhlfeld, C. C., R. P. Kovach, L. A. Jones et al. 2014. Invasive hybridization in a threatened species is accelerated by climate change. Nature Climate Change **4**:620–624.

Navarro, A., and N. H. Barton 2003. Chromosomal speciation and molecular divergence–accelerated evolution in rearranged chromosomes. Science **300**:321–324.

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

Nice, C. C., Z. Gompert, J. A. Fordyce, M. L. Forister, L. K. Lucas, and C. A. Buerkle 2013. Hybrid speciation and independent evolution in lineages of alpine butterflies. Evolution **67**:1055–1068.

Nolte, A. W., and D. Tautz 2010. Understanding the onset of hybrid speciation. Trends in Genetics **26**:54–58.

Nolte, A. W., Z. Gompert, and C. A. Buerkle 2009. Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. Molecular Ecology **18**:2615–2627.

Nosil, P., Z. Gompert, T. E. Farkas et al. 2012. Genomic consequences of multiple speciation processes in a stick insect. Proceedings of the Royal Society of London B: Biological Sciences **279**:5058–5065.

Patterson, N., P. Moorjani, Y. Luo et al. 2012. Ancient admixture in human history. Genetics **192**:1065–1093.

Payseur, B. A. 2010. Using differential introgression in hybrid zones to identify genomic regions involved in speciation. Molecular Ecology Resources **10**:806–820.

Payseur, B. A., J. G. Krenz, and M. W. Nachman 2004. Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. Evolution **58**:2064–2078.

Petit, R. J., C. Bodénès, A. Ducousso, G. Roussel, and A. Kremer 2004. Hybridization as a mechanism of invasion in oaks. The New Phytologist **161**:151–164.

Poelstra, J. W., N. Vijay, C. M. Bossu et al. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science **344**:1410–1414.

Pritchard, J. K., M. Stephens, and P. Donnelly 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945–959.

R Development Core Team 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Rhymer, J. M., and D. Simberloff 1996. Extinction by hybridization and introgression. Annual Review of Ecology and Systematics **27**:83–109.

 © 2016 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd

Rieseberg, L. H. 1991. Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. American Journal of Botany **78**:1218–1237.

Rieseberg, L. H. 2000. Crossing relationships among ancient and experimental sunflower hybrid lineages. Evolution **54**:859–865.

Rieseberg, L. H. 2006. Hybrid speciation in wild sunflowers. Annals of the Missouri Botanical Garden **93**:34–48.

Rieseberg, L. H., and C. A. Buerkle 2002. Genetic mapping in hybrid zones. The American Naturalist **159**:S36–S50.

Rieseberg, L. H., and S. E. Carney 1998. Tansley review no. 102 plant hybridization. The New Phytologist **140**:599–624.

Rieseberg, L. H., and J. F. Wendel 1993. Introgression and its consequences in plants. In R. G. Harrison, ed. Hybrid Zones and the Evolutionary Process, pp. 70–109. Oxford University Press, New York, NY.

Rieseberg, L. H., and J. H. Willis 2007. Plant speciation. Science **317**:910–914.

Rieseberg, L. H., R. Carter, and S. Zona 1990. Molecular tests of the hypothesized hybrid origin of two diploid *Helianthus* species (Asteraceae). Evolution **44**:1498–1511.

Rieseberg, L. H., C. Van Fossen, and A. Desrochers 1995. Hybrid speciation accompanied by genomic reorganization in wild sunflowers. Nature **375**:313–316.

Rieseberg, L. H., M. A. Archer, and R. K. Wayne 1999a. Transgressive segregation, adaptation, and speciation. Heredity **83**:363–372.

Rieseberg, L. H., J. Whitton, and K. Gardner 1999b. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics **152**:713–727.

Rieseberg, L. H., O. Raymond, D. M. Rosenthal et al. 2003a. Major ecological transitions in wild sunflowers facilitated by hybridization. Science **301**:1211–1216.

Rieseberg, L. H., A. Widmer, A. M. Arntz, and J. M. Burke 2003b. The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. Philosophical Transactions of the Royal Society of London B: Biological Sciences **358**:1141–1147.

Rockman, M. V. 2012. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution **66**:1–17.

Roe, A. D., C. J. K. MacQuarrie, M. C. Gros-Louis et al. 2014. Fitness dynamics within a poplar hybrid zone: I. Prezygotic and postzygotic barriers impacting a native poplar hybrid stand. Ecology and Evolution **4**:1629–1647.

Rutledge, L. Y., S. Devillard, J. Q. Boone, P. A. Hohenlohe, and B. N. White 2015. RAD sequencing and genomic simulations resolve hybrid origins within North American Canis. Biology Letters **11**:20150303.

Schwartz, M., and K. McKelvey 2009. Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. Conservation Genetics **10**:441–452.

Slatkin, M. 1985. Rare alleles as indicators of gene flow. Evolution **39**:53–65.

Soria-Carrasco, V., Z. Gompert, A. A. Comeault et al. 2014. Stick insect genomes reveal natural selection's role in parallel speciation. Science **344**:738–742.

Stam, P. 1980. The distribution of the fraction of the genome identical by descent in finite random mating populations. Genetical Research **35**:131–155.

Stebbins, G. L. 1950. Variation and Evolution in Plants. Columbia University Press, New York, NY.

Stewart, C. N., M. D. Halfhill, and S. I. Warwick 2003. Transgene introgression from genetically modified crops to their wild relatives. Nature Reviews Genetics **4**:806–817.

Sweigart, A. L., A. R. Mason, and J. H. Willis 2007. Natural variation for a hybrid incompatibility between two species of *Mimulus*. Evolution **61**:141–151.

Teeter, K. C., L. M. Thibodeau, Z. Gompert, C. A. Buerkle, M. W. Nachman, and P. K. Tucker 2010. The variable genomic architecture of isolation between hybridizing species of house mouse. Evolution **64**:472–485.

Vaha, J., and C. Primmer 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. Molecular Ecology **15**:63–72.

Vanlerberghe, F., B. Dod, P. Boursot, M. Bellis, and F. Bonhomme 1986. Absence of Y-chromosome introgression across the hybrid zone between *Mus musculus domesticus* and *Mus musculus musculus*. Genetics Research **48**:191–197.

Warschefsky, E., R. V. Penmetsa, D. R. Cook, and E. J. von Wettberg 2014. Back to the wilds: tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. American Journal of Botany **101**:1791–1800.

Weiss, K. M. 2008. Tilting at quixotic trait loci (QTL): an evolutionary perspective on genetic causation. Genetics **179**:1741–1756.

Whitham, T., J. Bailey, J. Schweitzer et al. 2006. A framework for community and ecosystem genetics: from genes to ecosystems. Nature Reviews Genetics **7**:510–523.

Witherspoon, D., E. Marchani, W. Watkins et al. 2006. Human population genetic structure and diversity inferred from polymorphic L1 (LINE-1) and Alu insertions. Human Heredity **62**:30–46.

Wolf, D. E., N. Takebayashi, and L. H. Rieseberg 2001. Predicting the risk of extinction through hybridization. Conservation Biology **15**:1039–1053.

Wu, C. I. 2001. The genic view of the process of speciation. Journal of Evolutionary Biology **14**:851–865.

Yanchukov, A., S. Hofman, J. M. Szymura, and S. V. Mezhzherin 2006. Hybridization of *Bombina bombina* and *B. variegata* (Anura, Discoglossidae) at a sharp ecotone in Western Ukraine: comparisons across transects and over time. Evolution **60**:583–600.

Zalapa, J. E., J. Brunet, and R. P. Guries 2010. The extent of hybridization and its impact on the genetic diversity and population structure of an invasive tree, *Ulmus pumila* (Ulmaceae). Evolutionary Applications **3**:157–168.

Supporting Information

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

Figure S1. Diagram summarizing the spatial layout and sampling of populations for the 'What, if anything, is hybridization' simulations.

Figure S2. Plots show neutral allele frequency (gray) and quantitative trait (orange) clines from simulated data with a migration rate of 0.01.

Figure S3. Scatterplots summarize patterns of genotypic variation for simulated data based on PCA.

Figure S4. Barplots show maximum likelihood estimates of admixture proportions.

Data S1. R code used to simulate and analyze hybrids.