History, chance and adaptation during biological invasion: separating stochastic phenotypic evolution from response to selection

Abstract
Introduced species often exhibit changes in genetic variation, population structure, selection regime and phenotypic traits as they colonize and expand into new ranges. For these reasons, species invasions are increasingly recognized as promising systems for studying adaptive evolution over contemporary time scales. However, changes in phenotypic traits during invasion occur under non-equilibrium demographic conditions and may reflect the influences of prior evolutionary history and chance events, as well as selection. We briefly review the evidence for phenotypic evolution and the role of selection during invasion. While there is ample evidence for evolutionary change, it is less clear if selection is the primary mechanism. We then discuss the likelihood that stochastic events shift phenotypic distributions during invasion, and argue that hypotheses of adaptation should be tested against appropriate null models. We suggest two experimental frameworks for separating stochastic evolution from adaptation: statistically accounting for phenotypic variation among putative invasion sources identified by using phylogenetic or assignment methods and by comparing estimates of differentiation within and among ranges for both traits and neutral markers ($Q_{ST}$ vs. $F_{ST}$). Designs that incorporate a null expectation can reveal the role of history and chance in the evolutionary process, and provide greater insights into evolution during species invasions.

Keywords
Admixture, assignment tests, colonization, drift, founder effect, $F_{ST}$, invasion, $Q_{ST}$.
phenotypic divergence to adaptive evolution, this inference is too often made in the absence of an appropriate null model (e.g. Gould & Lewontin 1979). By analogy to community ecology, the neutral theory of biodiversity suggests communities may differ in species diversity because of the deterministic match between resident species and the available ecological niches, as well as the historical and stochastic nuances of dispersal and sampling processes. In this sense, the forces structuring genetic diversity and species diversity are similar, and we must expect the joint influences of stochasticity and determinism to affect their outcomes (Antonovics 1976a).

In this paper, we explore the likelihood that chance events, such as founder effects, interact with prior evolutionary history as major factors driving phenotypic evolution during species invasions. We briefly review the potential role of selection and summarize studies that demonstrate phenotypic divergence at some phase of the invasion process. Although empirical evidence for phenotypic evolution is strong, in many cases the experimental design does not permit an unambiguous, or even a probabilistic, assessment of whether chance sampling of evolutionary history or adaptive evolution has influenced the outcome. We highlight reasons why neutral phenotypic change is a probable outcome during invasion and offer suggestions for experimental frameworks that test hypotheses of adaptation against null models of neutral phenotypic evolution. Designs that statistically incorporate a null expectation can reveal the role of history and chance in the evolutionary process, and allow a more precise understanding of the contemporary evolution of adaptations.

A PROMINENT ROLE FOR SELECTION DURING INVASION

The spread of non-native species involves several phases including dispersal, colonization, establishment of self-perpetuating populations (naturalization) and range expansion from the point of introduction. The ability of a species to respond to selection is thought to be an important determinant of geographic range (Antonovics 1976b; Holt et al. 2005). Hence, natural selection operating during each of these phases could be critical to invasion success (Sakai et al. 2001; Lee 2002). The colonizing phase represents an interesting challenge as long-distance (often trans-oceanic) dispersal causes colonizing genotypes to abruptly experience an environment that may differ dramatically from their place of origin. Thus, the transition to establishment and range expansion could be constrained or delayed if colonizing genotypes are maladapted to their environment (Holt et al. 2005). For invasions whose historical dynamics have been documented, the spread of introduced populations often starts out slowly before undergoing a rapid increase,
resulting in a lag-phase between colonization and eventual range expansion. While the lag phase may have multiple explanations, including basic exponential population growth, several studies suggest it may result from the waiting time for pre-adapted genotypes to colonize or the time required for the in situ evolution of adaptations to the new selective environment (Sakai et al. 2001; Lee 2002; Holt et al. 2005). Strong selection on the colonizing propagule pool is likely to truncate the phenotypic distribution of the establishing population (Simons 2003), yielding not only a shift in trait mean but also a sharp reduction in effective population size ($N_e$). Indeed, a lag time following introduction may actually be caused by intense selection because the few selected survivors begin the process of exponential growth from a smaller population size. Thus selection during establishment may generate different genotypes as well as different numerical dynamics, relative to the native range (Antonovics 1976a).

Initial establishment may be facilitated by phenotypic plasticity (Richards et al. 2006), with selection gaining importance after populations reorganize their genetic variance through repeated introductions. Multiple introductions from the native range can create genetic admixture within introduced populations, which may influence the process of adaptation following establishment. Admixture occurs when individuals from genetically divergent sources are brought together in new populations, and is an increasingly common feature of many species invasions (e.g. Gaskin & Schaal 2002; Kolbe et al. 2004; Wares et al. 2005). Mating between previously isolated gene pools can produce recombinant genotypes that may be entirely absent from the native range (de la Vega et al. 1991; Ellstrand & Schierenbeck 2000; Lavergne & Molofsky 2007). Such admixture can elevate the genetic variance in phenotypic traits and enhance the response to local selection, setting the stage for adaptive diversification within the introduced range (Lee 2002; Novak & Mack 2005).

What traits are likely to be under selection during invasion, and what is the empirical evidence for phenotypic divergence? Adaptation for invasiveness should be operationally definable (Antonovics 1976a), yet there are few generalizations to be made from empirical studies. The introduction of species to new ranges and their subsequent spatial expansions suggest selection may act directly on dispersal mechanisms (Holt et al. 2005). For example, Bufo marinus (cane toads) introduced to Australia possess longer legs at the invasion front relative to older established populations, indicating that dispersal ability may be evolving under selection (Phillips et al. 2006). Many invaders are also ecological opportunists, and changes in disturbance regime may select for shifts towards a ‘faster’ or more r-selected life history (Lewontin 1965; Sakai et al. 2001). Early reproduction and allocation for increased offspring number is predicted for colonizing species experiencing high levels of disturbance or environmental unpredictability (Lewontin 1965; Grime 1977; Simons 2007). Genetically based shifts in offspring size, reproductive capacity and the ability to reproduce across heterogeneous environments has been demonstrated in several plant taxa (e.g. Siemann & Rogers 2001; Blair & Wolfe 2004; DeWalt et al. 2004; Brown & Eckert 2005). Demographic uncertainty during the founding of new populations also predicts invasion will select for self-compatibility, asexuality, or other means of reproductive assurance (i.e. Baker’s Law: Baker 1955; Taylor et al. 1999; Kolar & Lodge 2001; Barrett et al. 2008). Finally, biotic interactions such as predation, pathogen attack, or mutualisms may also drive phenotypic evolution in the introduced range. In particular, the impact of natural enemies often differs systematically between the native and introduced ranges (reviewed in Colautti et al. 2004). The relaxation of selection from enemies may favour the evolution of traits conferring increased growth, competitive ability, or reproduction at the expense of defence (the ‘EICA’ hypothesis: Blossey & Notzold 1995). While support for the evolutionary consequences of enemy escape remains mixed (Daehler & Strong 1997; Willis et al. 2000; Siemann & Rogers 2001; Blair & Wolfe 2004; Bossdorf et al. 2004; DeWalt et al. 2004; Wolfe et al. 2004; Genton et al. 2005; Joshi & Vriezing 2005; Meyer et al. 2005), few definitive tests have been conducted (Colautti et al. 2004, Bossdorf et al. 2005).

Forces of selection operating during invasion may also relate to changes in the abiotic components of the species’ physiological niche, such as temperature, precipitation, soils, or growing season length (Kaufman & Smouse 2001; Holt et al. 2005; Broennimann et al. 2007). A response to physiological selection may be evident in clinal patterns of quantitative traits across the introduced range. Latitudinal clines in body size have been observed among introduced populations of sparrows (Johnston & Selander 1964) and in chromosomal inversion frequency and wing size in fruit flies (Prevosti et al. 1988; Huey et al. 2000). In invasive plants, latitudinal clines have been reported for biomass, height, number of shoots, flowering time, fecundity and several physiological traits (Weber & Schmid 1998; Kollmann & Banuelos 2004; Maron et al. 2004, 2007; Leger & Rice 2007; Montague et al. 2007). Because of the covariance between latitude and many aspects of the abiotic environment (e.g. temperature, growing season length), clines of traits with latitude suggest physiological adaptation during the course of invasion, although this is rarely tested against alternative explanations (but see Maron et al. 2004).

Taken together, there is good reason to believe that responding to selection may be a common or even necessary component of species that successfully establish self-perpetuating populations and undergo range expansion. However, the role of selection has rarely been
explicitly tested against null models of phenotypic evolution.

THE PROBLEM: $\Delta \xi \neq h^2S$

To understand adaptive evolution during invasion, we must study changes in the distribution of genetically based phenotypic traits associated with fitness during establishment and range expansion. Traits affecting fitness often have a polygenic basis and can be analysed by using the methods of quantitative genetics (Lynch & Walsh 1998). Quantitative genetic designs parse out genetic and environmental influences on phenotypic traits by raising pedigreed individuals, usually full or half-sib families, under controlled conditions or in common garden experiments. Quantitative genetic theory gives an inferential framework for analysing the response of a trait ($\zeta$) to selection, known as the breeder’s equation: $R = h^2S$ where $R$ is the response to selection (the cross-generational change in mean phenotype, $\Delta \xi$, before and after selection), $h^2$ the narrow sense heritability (the proportion of total phenotypic variance attributable to additive genetic effects), and $S$ the selection differential (the covariance between a trait and relative fitness). The implication of the breeder’s equation is straightforward: if both $h^2$ and $S$ are non-zero, the phenotypic distribution will shift in response to selection on the trait (assuming no countering effects from genetic correlations). For example, suppose that for an invasive plant, competition with neighbours causes individuals with greater stem height to produce more seeds, thereby generating a positive covariance between height and relative fitness ($S > 0$). If height is at least partially heritable ($h^2 > 0$), and is not negatively genetically correlated with another trait under selection, then the mean height of the invasive population will increase in the next generation through the process of adaptive evolution. By extension, populations that currently occupy different selective environments and are observed to differ genetically for a trait ($\Delta \xi > 0$) are often interpreted as having diverged in response to selection, although this may leave the agent(s) of selection unspecified.

The quantitative genetics of population divergence is naturally finding applications in invasion biology. Many studies use families collected from across one or both ranges, rear them in a common environment, and demonstrate significant genetically based divergence in trait means (reviewed in Bossdorf et al. 2005). In some cases, populations are phenotypically divergent but with no obvious connection to a putative force of selection (Parker et al. 2003; DeWalt et al. 2004). In other cases, a history of adaptive evolution is reasonable, but has not been validated experimentally. Thus, it is difficult to interpret phenotypic divergence as a response to selection without additional information on the history of the sample, especially the native range sources of introduced genotypes.

The problem with adopting an adaptationist view of evolutionary change is that invasions are inherently non-equilibrium demographic situations where the influences of historical events and stochastic processes are expected to be prominent. Founder effects (Eckert et al. 1996; Cabe 1998; Tsutsui et al. 2000; Kliber & Eckert 2005; Taylor & Keller 2007), multiple introductions and admixture (de la Vega et al. 1991; Collins et al. 2002; Gaskin & Schaal 2002; Kolbe et al. 2004; Wares et al. 2005; Lavergne & Molofsky 2007; Taylor & Keller 2007), and metapopulation dynamics (McCayley et al. 2003) all suggest demographic perturbations may be responsible for changes in genetic diversity during invasions (see also reviews in Bossdorf et al. 2005; Novak & Mack 2005; Wares et al. 2005; Roman & Darling 2007; Dlugosch & Parker 2008).

Chance demographic events can impact invasions through the sampling, introduction, and redistribution of alleles from the native range with diverse evolutionary histories (Fig. 1). The role of chance is frequently investigated by testing for genetic bottlenecks within introduced populations, as demographic reductions of sufficient magnitude and duration are expected to reduce genetic variation. The available evidence suggests bottlenecks may commonly reduce allelic diversity and heterozygosity within introduced populations, but also that these effects can be mitigated through high propagule pressure and multiple introductions (Bossdorf et al. 2005; Novak & Mack 2005; Wares et al. 2005; Roman & Darling 2007; Dlugosch & Parker 2008). However, it is important to note that a change in the average genetic diversity within populations is only one type of founder effect possible during invasion. Founder effects may also arise from biased sampling among different native range sources, which may not be detected by tests of within-population diversity. The frequency of different sources in the inocula, their geographic points of introduction, and their genetic similarity to each other, are all potential components of the founder effect that may be common during invasion. Founder effects among sources can have important consequences for evolution in the introduced range, as well as bias our inference of what forces have shaped the current population structure. This is because differentiation among newly founded populations will often reflect the sampling of source diversity, rather than an equilibrium between currently acting deterministic forces ( Whitlock & McCayley 1999).

A pair of recent studies on the invasive aquatic plant, Butomus umbellatus, highlight the importance of founder effects for phenotypic evolution during invasion. In Europe where it is native, B. umbellatus occurs as both asexual triploid genotypes that reproduce via clonal bulbils and sexual diploid genotypes that reproduce via both bulbils and...
seeds. In Europe, 16% of populations are diploid, while in North America, the frequency of diploids has increased dramatically to 71% (Brown & Eckert 2005; Kliber & Eckert 2005). One explanation is that natural selection is acting on the greater genetic variability and dispersal potential of sexually produced offspring (Brown & Eckert 2005). Consistent with this, when plants from North America and Europe are compared in a common greenhouse environment, introduced genotypes show a greater proportional allocation to reproduction via sexual inflorescences, as well as asexual bulbils (Brown & Eckert 2005). However, population genetic analyses provide two important pieces of information to suggest that these differences probably result from a founder effect. First, there has been a dramatic overall reduction in the number of unique genotypes in North America compared with Europe, suggesting a small number of founders and a biased sampling of diploid and triploid lineages. Second, across the 38 introduced diploid populations surveyed, 95% of plants share a single heterozygous genotype (Kliber & Eckert 2005). This suggests that despite higher allocation to sexual inflorescences and seeds, the actual recruitment of offspring in North America is almost exclusively asexual. Thus, what appears to be the adaptive evolution of increased allocation to sexual reproduction may actually be a founder effect, which skewed the introduced sample towards genotypes that invest heavily in seed production but with no realized fitness benefit (Brown & Eckert 2005; Kliber & Eckert 2005).

Founder effects can also generate clinal patterns via isolation by distance. This occurs when genetically distinct sources establish at different points of introduction, creating a pattern of genotypes in close proximity being more closely related than those farther away. This has clearly occurred during the invasion of North America by Silene latifolia, where haplotypes from divergent European clades established at geographically separated sites (Taylor & Keller 2007; Fig. 2). A similar pattern is apparent from the invasion of the European green crab along the Atlantic coast of North America, where multiple introductions of haplotypes from distinct sources generated a cline in genetic differentiation among introduced populations (Roman 2006). Demographic processes such as these are capable of generating ‘clines’ in both neutral markers and phenotypic traits, especially when the respective source regions differ for both types of variation. Therefore, even phenotypic clines should be interpreted cautiously, especially when they occur over geographic gradients as well as environmental ones (i.e. latitude).

Among the growing number of studies testing hypotheses about phenotypic evolution during invasion, very few incorporate a null expectation that accounts for chance sampling of evolutionary history from the native range (but see Maron et al. 2004; Kliber & Eckert 2005; Kolbe et al. 2007; Lavergne & Molofsky 2007). This approach is essential, since chance events can affect quantitative traits as well as neutral loci (Merila & Crnokrak 2001; McKay & Latta 2002), especially when founder effects produce a

![Figure 2](https://example.com/figure2.png)

Figure 2 Isolation by distance among chloroplast sequence haplotypes in Silene latifolia generated by the spatial pattern of colonization by divergent native range haplotypes. In Europe(a), haplotypes show significant phylogeographic structure (AMOVA: $\Phi_{ST} = 0.27; P < 0.0001$). In North America (b), colonization of western European haplotypes (blue symbols) occurred primarily in the southeastern USA, while haplotypes from eastern Europe (orange symbols) colonized elsewhere. Given the genealogical relationships among haplotypes (c), this pattern lead to a signature of isolation by distance in eastern North America (d) (Mantel’s test: $r = 0.63, P < 0.0001$). Figure 2a-c adapted from Taylor & Keller (2007).
biased sample of source populations. To see this, consider a scenario of stochastic phenotypic evolution that is probably quite common during invasion (Fig. 3). In the first example (Fig. 3a), a species in its native range is structured into several discrete demes that have diverged for some quantitative trait. If the process of colonization samples only a subset of the native demes, then the mean phenotype of the colonists will be shifted, relative to the mean across the native range. In the second example (Fig. 3b), we assume the same native range structure, but consider that sampling during invasion may change the relative frequencies of the demes, again resulting in a shift in the phenotypic mean of the colonists relative to the native range. Thus, stochastic events may determine which individuals contribute to the invasion, with the phenotypes of those individuals reflecting a complex history of selection and drift in the native range. In these cases, a quantitative genetic comparison of the native and introduced ranges may reveal phenotypic divergence, but $\Delta z \neq b^2 S$. One way to account for stochastic divergence owing to sampling effects is to estimate phenotypic divergence directly between ancestral lineages in the native range and their descendents in the introduced range (Fig. 3c). Fortunately, with the advent of high resolution molecular markers and analytical techniques, establishing ancestor–descendent relationships at the intraspecific level is increasingly feasible. This presents an opportunity to integrate knowledge about the demographics of invasion, using neutral markers, with hypotheses about adaptive evolution of the phenotype, using measurements on traits.

**TESTING PHENOTYPIC EVOLUTION AGAINST NEUTRAL EXPECTATIONS**

Separating the effects of history and chance from adaptation is a significant challenge, but is experimentally tractable. The key observation is that neutral loci are subject to the demographic and genetic forces of founder effect, drift, and gene flow, while loci contributing to quantitative traits are subject to these same forces, plus the action of natural selection. Thus, the genetic contributions of history and chance can be jointly controlled for by incorporating neutral molecular variation into experimental designs that test for divergence in phenotypes. To address this issue, we offer two experimental designs that combine neutral and quantitative genetic information to help parse the relative roles of selection following introduction vs. other evolutionary forces during species invasion.

**Conceptual design 1: ancestor–descendent comparisons**

History and chance contribute to phenotypic evolution during invasion when native range sources are divergent for traits and the invasion process samples these sources in a biased way (Fig. 3). In the simplest case, the native range is comprised of a single ancestral deme from which the invasion sampled descendents representatively. However, when there are multiple ancestral demes that have diverged phenotypically and are sampled at biased frequencies, then differences between ranges (native vs. invasive) arise from the chance manner in which they were sampled.

One way to deal with this biological reality is to first pair descendents (invaders) with their ancestral lineages in the native range using molecular methods such as neutrally evolving DNA sequences or marker loci (e.g. allozymes,
microsatellites, AFLPs, SNPs). Phenotypic measurements on a sample of genotypes from these sources then sets up a neutral expectation for how founder effects could have resulted in phenotypic divergence between ranges. Divergence in excess of this neutral expectation would implicate the action of selection driving evolution. Thus, a quantitative comparison of phenotypic variation between invasive and native range genotypes that are descendants from the same ancestral deme provides crucial insight into the causes of phenotypic divergence during invasion (Fig. 3c).

The approach

Two molecular approaches show promise for generating ancestor-descendent comparisons: phylogenetics and multi-locus assignment methods. In the phylogenetic approach, DNA sequencing of individuals from native and introduced ranges is used to identify haplotypes and their relatedness. If sufficient phylogeographic structure exists in the haplotype network, then invaders can be traced back to their ancestral sources by searching for shared haplotypes between ranges (Collins et al. 2002; Gaskin & Schaal 2002; Saltonstall 2002; Kolbe et al. 2004; Taylor & Keller 2007). Phylogenies from sequence data provide the most complete genealogical information for resolving ancestor-descendent relationships because they produce historically ordered alleles and are generally less prone to homoplasy than marker loci. However, they may lack resolution depending on the rate of mutation. As it is a priority to connect introduced genotypes as closely as possible to their native range ancestors, anticipated phylogenetic resolution should be carefully considered.

As an alternative to phylogenetic methods, marker loci can be used to connect introduced genotypes to their ancestral gene pools. While most marker loci are not preferred for inferring phylogenetic relationships, they are ideal for generating multi-locus genotypes for use in assigning sources under non-equilibrium conditions (Davies et al. 1999). Assignment methods are a growing class of analyses that share the general feature of using multi-locus genotypes to assign individuals a probability of membership to different putative sources. The approach is similar to forensic or parentage analyses (Manel et al. 2005), and is based on the idea that at any given locus, an individual has a probability of matching one or more sources. By combining information from many independent loci, these probabilities can be calculated with enough precision to exclude all but the most likely source(s).

Assignment methods can be divided into two types, distinguished in part by the assumptions they make about the source populations. First, assignment tests group individuals with their most likely population, chosen by the investigator from an a priori group of putative source populations (Rannala & Mountain 1997). Simulations have shown assignment tests have considerable statistical power for producing correct assignments, provided the genetic structure among the set of source populations is not too low (Waples & Gaggiotti 2006) and of course that representatives of the source population are included in the sample. A second parallel approach uses Bayesian model-based clustering to identify genetic structure and genotype membership while making relatively few assumptions about the source populations (Pritchard et al. 2000; Corander et al. 2003). Clustering methods work by using information on the allelic associations (i.e. statistical linkage disequilibria) that develop within isolated demes. These methods first solve for the most likely number of genetically distinct demes, given a data set of multilocus genotypes, and then assign individuals to demes probabilistically based on the estimated allele frequencies (Pritchard et al. 2000). Clustering methods usually perform well when population structure is moderate to high and with mixed results when structure is low (Pritchard et al. 2000; Manel et al. 2005; Waples & Gaggiotti 2006). In simulation tests, two of the most frequently used Bayesian clustering methods (BAPS and STRUCTURE) were found to perform well when differentiation (FST) was above 0.03, suggesting that detecting populations and conducting assignments is possible even when structure is subtle (Latch et al. 2006).

The ability to detect signals of weak genetic structure and assign genotypes to sources is a powerful tool for empirical studies of introduced species, since high dispersal ability (either intrinsic or human-mediated) is a feature of many successful invaders. For example, consider the historically weedy plant *Arabidopsis thaliana*, which has attained a global distribution as a result of its dispersal abilities and association with humans. Previous tree-based analyses gave no indication of genetic structure in Europe and Asia where it is native (Miyashita et al. 1999), or in North America where it is introduced (Jorgensen & Mauricio 2004). In contrast, a recent analysis based on Bayesian clustering has uncovered clear genetic structure that was previously undetected (Beck et al. 2008), and allowed for introduced genotypes to be assigned to native range source demes.

Although phylogenetic surveys and multilocus genotyping are routinely being used to examine the introduction of genetic diversity during invasion, neither has been well integrated with studies of phenotypic evolution (but see Maron et al. 2004; Kliber & Eckert 2005; Kolbe et al. 2007). Once invasive genotypes are assigned to putative sources, this ancestry information can be incorporated into quantitative genetic designs to test for adaptation. To illustrate this, consider a straightforward test for phenotypic change during invasion by sampling individuals from multiple populations distributed across the native and introduced
ranges and measuring their traits (e.g. Blair & Wolfe 2004; Leger & Rice 2007). It is important to note that traits measurements should be made under a common set of environmental conditions, preferably on individuals with known pedigree, to avoid confounding genetic and environmental effects. A history of natural selection operating in the same direction across invasive populations, for example, favouring life history shifts towards a colonizing strategy, would cause divergence in the mean phenotype between ranges. Including the ancestral deme as a fixed effect in ANOVA (with i levels corresponding to the number of lineages or demes common to both ranges) permits the phenotypic divergence caused by shifts in deme frequencies during invasion to be controlled for statistically. The principal test of the fixed effect of range then takes on a new interpretation. Phenotypic divergence between ranges that persists after controlling for divergence among demes lends strong support to selection driving the change, or at least phenotypic change that has occurred within the introduced range following colonization. This design is flexible to a variety of biological and statistical outcomes (Fig. 4). For example, not all demes included will experience selection following introduction, or the strength of selection and magnitude of response may vary among demes. This may occur because demes differ in how preadapted they are to their new environment, because demes are introduced to locations that differ in the strength or direction of selection, or because demes contain different amounts of genetic variability and hence differ in the potential to respond to selection. These types of outcomes are captured by the range*deme interaction effect, indicating that not all demes experienced the same amount of phenotypic evolution (Fig. 4).

A slightly different approach is appropriate when testing the hypothesis of selection generating clinal patterns. As above, trait measurements should be taken on individuals raised in a controlled environment to ensure variation is genetic and not environmental acclimation. Next, if ancestry to discrete sources can be assigned, then a test of clinal evolution can be made with analysis of covariance (ANCOVA), testing for trend differences among sources (the fixed factor), an environmental gradient (the covariate), and their interaction. Alternatively, the ancestry of multilocus genotypes can be evaluated on a more continuous scale using distance-based ordination methods such as principal coordinates analysis (PCO). Relatedness among genotypes based on their neutral markers can be described by the relative positions of invasive and native range genotypes in multilocus genetic space. The ancestry information summarized by one or more axes from the PCO could be used as independent variables in a multiple regression model along with the putative environmental gradients influencing trait evolution. Significant covariance between a trait and an environmental gradient, while holding the effects of relatedness constant, would be compelling evidence for the action of clinal selection rather than a correlated effect of spatial structure among different introduced sources (Fig. 5).

Figure 4 Hypothetical 'norm of reaction' showing two possible outcomes of conceptual experimental design 1. In (a), plants from different demes posses divergent phenotypes in the native range prior to invasion (e.g. a significant effect of 'Deme' but not 'Range' in an ANOVA model). If stochastic sampling has changed the frequencies of demes during invasion, then phenotypic evolution occurs but is attributable to neutral processes. In contrast, (b) shows that invaders from some demes have evolved new phenotypic means, after controlling for differences due to common ancestry (e.g. a significant 'range' or range × deme' effect in ANOVA). This suggests invaders have evolved toward new phenotypic optima in response to selection during or since the invasion.

Dealing with admixture

When multiple introductions among differentiated sources produce admixture within introduced populations, some additional considerations are necessary. The evolutionary consequences of admixture and the complications it introduces to the analysis of phenotypic evolution depend in part on whether it is purely demographic (genotypes from different sources co-occurring) or genomic (mating and recombination between sources creating novel genotypes). The first step is determining the type of admixture that has
occurred. Demographic admixture is straightforward to detect, for example by differences in the number of haplotypes (or demes) and their genetic relatedness within introduced populations relative to native range populations. Several methods are available to detect if mating has produced recombinant genotypes. If few loci are used, then a direct comparison of genotype classes can be made between ranges (de la Vega et al. 1991; Novak & Mack 2001; Lavergne & Molofsky 2007). Admixture can also be inferred by computing an estimate of linkage disequilibrium among loci, such as $D$ or $D'$ (Hartl & Clark 1989). Recent admixture between source populations with differing allelic frequencies should leave a temporary signal of elevated linkage disequilibrium within populations that eventually decays with generations of random mating and recombination. Finally, detection of genomic admixture can be made by using Bayesian clustering, which can provide estimates of the fraction of loci in an admixed genotype that belongs to each source (Pritchard et al. 2000).

Fortunately, admixture need not prevent tests for whether phenotypic evolution exceeds neutral expectations. Under purely demographic admixture, an analysis of phenotypic variance that incorporates source assignments would still control for history and chance by removing the variance attributable to different native range sources. In a similar approach, Kolbe et al. (2007) used matrix correlations between morphological differences and source differences (identified from a mtDNA phylogeny) to test for phenotypic divergence among introduced populations in excess of neutral expectations. While introduced populations were morphologically variable, the matrix correlations showed that these differences could be explained by the number and frequency of different haplotypes within populations (Kolbe et al. 2007). Thus, neutral demographic admixture was sufficient to explain population morphological differences without evoking the action of selection. When admixture is genomic and not just demographic, then care must be taken to not confound selection with phenotypic variance arising from novel interactions among alleles. In this case, a comparison of differentiation in neutral markers and the additive genetic variance in traits would help control for admixture and novel allelic interactions and identify phenotypic divergence that exceeds neutral expectations (see Conceptual design 2 below).

**Issues and considerations**

Several considerations should be kept in mind regarding these analyses, some minor and others more substantial. First, what constitutes a sufficient sampling strategy to connect invaders with their native range sources? This will largely depend on the scale of variation present in the native range, the precision desired for identifying the sources, and the type of method being employed. Naturally, identifying sources will be easier when strong genetic structure is present (Kolbe et al. 2004). If using phylogeographic methods and tracing invaders to biogeographical regions (and not to specific point locations) provides sufficient precision, then relatively few individuals need to be sampled from within each population (e.g. Collins et al. 2002). Indeed, it may often be more desirable to maximize the number of locations sampled across the entire range at the expense of replicating within populations. Alternatively, if structure is weak or assignment methods are used that depend on accurate assessments of allele frequencies, then it is
necessary to sample more intensively within populations, or at least within the unit of variation (defined at an appropriate scale). Prior knowledge on the biology of the species and the method being used should be considered carefully in light of these sampling tradeoffs before embarking on any intensive survey.

A second consideration pertains to phylogenetic studies of invaders, the majority of which preferentially use organelle genomes (mitochondria in animals and chloroplasts in plants) because of their high mutation rates and lack of frequent recombination. While mitochondrial or chloroplast DNA sequences may be excellent for inferring certain aspects of demographic history, these histories may not always reflect the history of the nuclear genome (which is presumably responsible for the majority of quantitative trait variation). To the extent that the nuclear and organelle genomes have experienced different histories either prior to or during the invasion, phylogenetic information from the organelles may be inappropriate neutral controls for phenotypic evolution. Admixture during invasion may exacerbate this issue, as hybridization weakens the associations between nuclear and cytoplasmic loci. In the worst case scenario, extensive admixture can completely homogenize nuclear loci across different cytoplasmic backgrounds, rendering haplotypes derived from organelles uninformative about phenotypic differences among sources. Advances in nuclear gene phylogeography may offer solutions to this problem (Gaskin & Schaal 2002). Finally, gene coalescence is a stochastic process, which makes inferences from a single locus subject to considerable variation around the ‘mean’ demographic history of the species. Therefore, to obtain robust inferences of demographic history, it is preferable to use multiple loci.

For the assignment method approach to produce an effective neutral expectation, it is assumed that introduced genotypes are a representative and unbiased sample of the source deme in the native range from which they descended. When founder effects produce colonists that are a biased sample of their source deme, then ancestor–descendent comparisons may no longer provide a valid neutral expectation for phenotypic divergence. This assumption can be assessed by first testing for reductions in diversity at neutral marker loci for each introduced deme relative to its native range source. Evidence of strong bottlenecks in neutral marker loci for each introduced deme provides a null model for adaptation, rejecting the null does not reveal the phase of the invasion during which the response to selection occurred. Selection during establishment and selection to local conditions during expansion would both cause the phenotypic mean of invaders to deviate from their native range ancestors, leaving the precise timing of the selective events open to further experimentation.

Conceptual design 2: \(Q_{ST} \) vs. \(F_{ST} \)

Another way to partition history and chance from selection is to make the population the unit of comparison and compare the genetic variance at neutral loci (\(F_{ST} \)) relative to the variance in traits measured on pedigreed progeny (\(Q_{ST} \)). This approach is appealing because it is directly tied to the methodologies of traditional population genetic surveys, which analyze hierarchical genetic structure within and between populations. As it is frequency-based and not dependent on individual genotypes, it is appropriate for any demographic scenario including bottlenecks or admixture among genetically distinct sources (the demographic consequences of founder effects and admixture are captured by \(F_{ST} \)). This approach is also well suited towards studies that explicitly want to follow the process of populations locally adapting to conditions within the introduced range. Finally, the hierarchical design allows an overall test for adaptation between ranges (i.e. selection for ‘invasiveness’), as well as adaptive differentiation among populations within ranges.

The approach

The neutral theory of phenotypic evolution poses that the additive genetic variance for a trait, \(\sigma^2_g \), can be partitioned into within \(\sigma^2_g(b) \) and between \(\sigma^2_g(b) \) population variance components in a manner analogous to single locus population genetics (Wright 1951). Ignoring for a moment the effects of mutation and selection, the balance between drift and gene flow will result in the hierarchical partitioning of \(\sigma^2_g \) in proportion to Wright’s fixation coefficient for neutral loci, \(F_{ST} \) (Wright 1951). From the results of the neutral phenotypic theory, the analogous fixation coefficient for quantitative traits is \(Q_{ST} = \sigma^2_g(b)/\left(\sigma^2_g(b) + 2\sigma^2_g(w)\right) \) (Spitze 1993). The important result is that, in the absence of selection, \(Q_{ST} \) and \(F_{ST} \) estimated from a set of populations are expected to have closely similar values (Merila & Crnokrak 2001; McKay & Latta 2002). When \(F_{ST} \) is estimated from neutral loci distributed across the genome, it estimates the sum of the demographic processes that contribute to divergence, such as founder effects and genetic drift, but is less affected by the force of selection. In contrast, \(Q_{ST} \) summarizes divergence at loci affecting a phenotypic trait and will be affected by the same demographic forces as neutral loci, as well as potentially influenced by a history of selection on the phenotype. Therefore, \(F_{ST} \) provides a null expectation for divergence.
caused by chance and drift, against which divergence at putatively selected traits \( (Q_{ST}) \) can be compared. If quantitative traits are evolving neutrally, their divergence will approximate that for neutral loci, and \( Q_{ST} = F_{ST} \). Therefore, \( Q_{ST} > F_{ST} \) is evidence of a history of adaptive divergence, while \( Q_{ST} < F_{ST} \) indicates a history of stabilizing selection (Merila & Crnokrak 2001; McKay & Latta 2002; Leinonen et al. 2008).

By assaying both types of genetic variation, it is possible to make this comparison at different hierarchical levels (e.g. among ranges, among populations within ranges) to test hypotheses regarding the history of selection and adaptive evolution. There are a number of historical scenarios relevant to species invasions that can be disentangled by this method (Table 1). For example, invasion may have involved selection for traits that promote productivity early in the life history. Our expectation would then be for strong phenotypic divergence between ranges for these life history traits (a high between-range \( Q_{ST} \)). However, this must be tested against the neutral divergence between ranges as a result of the sampling process \( (F_{ST}) \). Thus, a result in which \( Q_{ST} > F_{ST} \) for the invasive vs. native range lends strong support to the action of selection driving invasiveness. A second possibility is that over evolutionary time, native populations have become locally adapted to the environments within the native range \( (Q_{ST} > F_{ST} \) within the native range). During invasion, a colonization process involving multiple introductions and admixture may reduce \( Q_{ST} \) and \( F_{ST} \) to near zero. Or alternatively, invasive populations may show some evidence of structure due to colonization-extinction dynamics (McCauley et al. 2003), but with traits and markers affected similarly. In these cases, \( Q_{ST} = F_{ST} \), and any divergence detected is attributable to founder effects and drift. As before, this conceptual design should be robust to a wide variety of possible outcomes (Table 1), making it a powerful means to decouple neutral phenotypic evolution from adaptation during invasion.

Comparisons of \( Q_{ST} - F_{ST} \) have become a popular approach among ecological geneticists since the publication of two prominent reviews (Merila & Crnokrak 2001; McKay & Latta 2002). Among invasion biologists, however, there are few applications to understanding the mechanisms of phenotypic evolution. One recent study examined divergence in the invasive grass, *Phalaris arundinacea*, using a hierarchical sampling design of three to four populations within each of two regions in Europe (native) and North America (invasive; Lavergne & Molofsky 2007). Divergence was quantified using neutral allozyme loci \( (F_{ST}) \) and several phenotypic traits measured in a common greenhouse environment \( (Q_{ST}) \). The results showed that \( Q_{ST} \) generally exceeded \( F_{ST} \) among populations within regions, pointing to selection operating locally within the native and introduced ranges (Lavergne & Molofsky 2007). In other analyses, the native and introduced ranges were divergent for traits related to invasiveness in *P. arundinacea*, such as emergence rate, biomass and clonal expansion; however, these tests did

---

**Table 1** Hypothetical outcomes from a hierarchical comparison of genetic variance in quantitative traits \( (Q_{ST}) \) and neutral loci \( (F_{ST}) \)

<table>
<thead>
<tr>
<th>Traits: ( Q_{ST} ) vs. ( F_{ST} ) between NR vs. IR†</th>
<th>Traits: ( Q_{ST} ) vs. ( F_{ST} ) among NR populations†</th>
<th>Traits: ( Q_{ST} ) vs. ( F_{ST} ) among NR populations†</th>
<th>Traits: ( Q_{ST} ) vs. ( F_{ST} ) among IR populations†</th>
<th>Biological interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>*** &gt;</td>
<td>*** =</td>
<td>*** &gt;</td>
<td>*** &gt;</td>
<td>Selection during establishment or selection in the IR after establishment promotes adaptive divergence between ranges</td>
</tr>
<tr>
<td>*** =</td>
<td>*** =</td>
<td>*** &gt;</td>
<td>*** &gt;</td>
<td>Unrepresentative sampling of traits during invasion. Evolution driven by stochastic events</td>
</tr>
<tr>
<td>*** &gt;</td>
<td>*** =</td>
<td>*** &gt;</td>
<td>*** &gt;</td>
<td>Local adaptation in the NR; disrupted by stochastic processes during invasion</td>
</tr>
<tr>
<td>n.s. &lt;</td>
<td>n.s. &lt;</td>
<td>n.s. &lt;</td>
<td>n.s. &lt;</td>
<td>Diversifying selection drives local adaptation within both ranges</td>
</tr>
<tr>
<td>n.s. &lt;</td>
<td>*** &gt;</td>
<td>Stabilizing selection maintains similar trait means within and between ranges</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n.s. &lt;</td>
<td>*** &gt;</td>
<td>Stabilizing selection within the NR, but release from selective constraint promotes adaptive radiation in the IR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The outcomes and their interpretation are a non-exhaustive list assuming a nested experimental design involving comparisons of the native range (NR) and introduced range (IR) and comparisons among populations within each range.

†Outcomes from an analysis of phenotypic divergence measured in a common environment:

**Significant difference in trait means, n.s. no significant difference in trait means. Entries left blank reflect where the specific outcome does not qualitatively affect the overall interpretation**
not account for neutral divergence. It would be interesting to conduct the $Q_{ST}$ vs $F_{ST}$ comparison at each hierarchical level, including between ranges, to understand if selection is the primary force responsible for the evolution of invasiveness.

**Issues and considerations**

Comparisons of $Q_{ST}$ and $F_{ST}$ are currently an active area of research in population genetics, from both empirical and theoretical perspectives. While the behaviour of the estimators and the sensitivity of their assumptions still receive attention, several recent insights are relevant to the conceptual design proposed here. First, $Q_{ST}$ is formally a partitioning of additive genetic variance, which requires intensive breeding designs capable of isolating just the additive effects of genes (Lynch & Walsh 1998). Less complicated breeding designs may produce variance components that include some amount of non-additive genetic effects such as dominance or epistasis. A recent simulation study suggests the general effect of dominance variance may be to lower $Q_{ST}$ and therefore avoid type I errors (Goudet & Buchi 2006), while meta-analysis suggests no significant bias in $Q_{ST}$ when estimated from designs that include dominance variance (Leinonen et al. 2008); however, this area of research warrants additional attention.

An important assumption of the $F_{ST} - Q_{ST}$ comparison is that marker loci behave neutrally and are independent of each other and selected regions of the genome. However, for highly selfing or asexual species, limited recombination can generate linkage disequilibrium between markers and gene regions. If selection acts on a gene region, then marker loci in linkage with it will also increase or decrease in frequency through a process of genetic hitchhiking. This tends to make inference of adaptive differentiation conservative in highly selfing species, as differentiation in both markers ($F_{ST}$) and traits ($Q_{ST}$) may increase under heterogeneous selection (Porcher et al. 2006). This is an important consideration for invaders which are highly selfing or clonal.

Another assumption that is infrequently discussed is that the rate of mutation is small relative to the migration rate and that the model of evolution for neutral and quantitative traits is similar (Hendry 2002). The mutation rate of hypervariable markers like microsatellites may be high relative to the migration rate, leading to lower estimates of $F_{ST}$ (Hedrick 2005) and the potential for type I errors of falsely rejecting the null hypothesis $F_{ST} = Q_{ST}$ (Hendry 2002; Leinonen et al. 2008). This problem principally arises when there are a large number of alleles at each locus, leading to high expected heterozygosities within populations even when divergence among populations is pronounced. One solution might be to standardize $F_{ST}$ by its maximum possible value given the gene diversity within populations (Hedrick 2005), but it is unclear if a standardized $F_{ST}$ can be validly compared with an unstandardized $Q_{ST}$. While a recent meta-analysis indicates no significant differences in $F_{ST}$ estimated from microsatellites vs. other common marker types, it remains unclear how the $Q_{ST} - F_{ST}$ comparison may be affected by high mutation rates (Leinonen et al. 2008).

A final issue relates to statistical power. The power of $Q_{ST}$ estimates is affected by the number of populations sampled; some simulations suggest fewer than 20 populations may compromise the ability to detect the signature of selection (Goudet & Buchi 2006). Similarly, studies of invasions that wish to make statistical statements about phenotypic divergence between the native vs. introduced ranges also require a large and unbiased sample of populations for reliable inference (R.I. Colautti, J.L. Maron, and S.C.H. Barrett, unpublished manuscript). The need for many populations when estimating $Q_{ST}$ makes for potentially large experimental designs, though Goudet & Buchi (2006) suggest replication within populations can be somewhat relaxed. For example, if we replicated the native and introduced ranges with 20 populations each, and performed a modest sized paternal half-sib design within each population (ex. 10 sires each mated to two dams and raising five offspring from each family), the experimental design would involve phenotyping 4000 individuals. Less replication intensive designs exist (eg. full-sib families), but will involve some amount of non-additive genetic variance.

**Conclusions**

Invasive species have great potential to reveal the process of adaptive evolution, but evidence for selection should be evaluated relative to null expectations based on neutral phenotypic evolution. The experimental designs presented here are meant to further this goal. The conceptual advance is that by incorporating insights gained from neutral molecular markers, experiments can be designed that isolate the sometimes complex demographic history of an invasion from the history of selection on the phenotype. We do not regard these as the only methods of accounting for demographic effects when studying adaptation, but rather view them as promising examples of the more general approach of incorporating neutral expectations for phenotypic change.

It is also important to keep in mind that these approaches are necessarily statistical in nature (i.e. they do not reveal the agents of selection), and are best used as a first step in the study of adaptation during invasion. Field experiments such as reciprocal transplants among the putative selective environments, coupled with direct measurements of contemporary selection on the traits, would complement the experiments described here. Nevertheless, some of the most
interesting questions in biological invasions involve inferences of past selection shaping phenotypic distributions. Tests of adaptation against null models of neutral evolution make this possible.

ACKNOWLEDGEMENTS
We appreciate the thoughtful comments of Janis Antonovics, Laura Galloway, the Taylor lab group, and three anonymous referees. This paper was supported by NSF DEB-0608358 and NSF DEB-444176.

REFERENCES

© 2008 Blackwell Publishing Ltd/CNRS


Editor, Rebecca Irwin

Manuscript received 15 February 2008

Manuscript accepted 12 March 2008