SPECIAL ISSUE: INVASION GENETICS: THE BAKER AND STEBBINS LEGACY
INVITED REVIEWS AND SYNTHESSES

Genetic reconstructions of invasion history

MELANIA E. CRISTESCU
Department of Biology, McGill University, Montreal, QC H3A 1B1, Montreal, Canada

Abstract

A diverse array of molecular markers and constantly evolving analytical approaches have been employed to reconstruct the invasion histories of the most notorious invasions. Detailed information on the source(s) of introduction, invasion route, type of vectors, number of independent introductions and pathways of secondary spread has been corroborated for a large number of biological invasions. In this review, I present the promises and limitations of current techniques while discussing future directions. Broad phylogeographic surveys of native and introduced populations have traced back invasion routes with surprising precision. These approaches often further clarify species boundaries and reveal complex patterns of genetic relationships with noninvasive relatives. Moreover, fine-scale analyses of population genetics or genomics allow deep inferences on the colonization dynamics across invaded ranges and can reveal the extent of gene flow among populations across various geographical scales, major demographic events such as genetic bottlenecks as well as other important evolutionary events such as hybridization with native taxa, inbreeding and selective sweeps. Genetic data have been often corroborated successfully with historical, geographical and ecological data to enable a comprehensive reconstruction of the invasion process. The advent of next-generation sequencing, along with the availability of extensive databases of repository sequences generated by barcoding projects opens the opportunity to broadly monitor biodiversity, to identify early invasions and to quantify failed invasions that would otherwise remain inconspicuous to the human eye.

Keywords: aquatic invasions, invasion history, invasion routes, invasive species, molecular markers, range dynamics

Received 23 November 2014; revision received 11 February 2015; accepted 12 February 2015

Introduction

Reconstructing the complex history of biological invasions represents an important step in understanding the invasion process. Species that rapidly expand their distributitional ranges far beyond their native ranges (experience dramatic range expansions) or outside their original ecological spaces (experience major habitat transitions) provide, in many ways, large-scale natural experiments and the opportunity to investigate dynamic evolutionary processes in real time (Lee 2002; Sax et al. 2007). Often, a deep understanding of invasion history is required for building predictive models of secondary spread and ecological or economic impact, as well as for implementing sound management measures of prevention and rapid intervention (Kulhanex et al. 2011), such as efficient methods of biological control (Barrett 1992). It is not surprising that over the last few decades, the reconstruction of invasion history has become the foundation of many long-term investigations of biological invasions (Estoup & Guillemaud 2010). In this review, I evaluate recent progress and future trends in reconstructing the history of invasive species. I contrast natural and human-mediated invasions and focus on well-documented aquatic invasions, particularly zooplankton species that are considered to be dispersalists par excellence. I discuss the most common genetic approaches and molecular markers that have been used to determine source populations, invasion routes and vectors, as well as the genetic, evolutionary and ecological changes that invasive species or populations undergo during the process of invasion.
Historical perspectives on reconstructing invasion histories

Early natural historians compiled extensive records of natural and human-mediated colonization events, including major distributional shifts, faunal turnovers and extinctions associated with biological invasions. Many of these events were triggered by intentional introductions or habitat alterations that accompanied the successful expansion of modern humans ‘out of Africa’ about 70 000–90 000 years ago (Brahic 2012). The early observations on biological invasions were based on detailed historical and comparative biogeographical data and were often focused on particular regions (reviewed in Di Castri 1989; Simberloff 2013). A century after Darwin’s influential work, the birth of invasion biology was prompted by two memorable moments. Elton’s (1958) book on The Ecology of Invasions by Animals and Plants provided the first comprehensive work on biological invasions. Elton found his inspiration in the early faunal history and in the field of ecology that he was advancing at the time. He described with eloquence the significant faunal and floral turnover associated with the continual destruction of natural dispersal barriers due to increased human trade and travel. Although Elton’s work brought biological invasions into the spotlight, the effort of synthesizing the knowledge on biological invasions was firmly advanced by the influential volume on The Genetics of Colonizing Species edited by Hebert Baker and Ledyard G. Stebbins in Baker & Stebbins 1965 (Barrett this issue). The broad synthesis inspired by the Asilomar meeting 50 years ago greatly stimulated the research field by placing biological invasions into an evolutionary context and by setting up a conceptual framework for studying the evolution of invasiveness (weediness, persistence, dispersal). The Genetics of Colonizing Species covered many aspects on the phenotypic properties that facilitate the invasion process. The properties of the gene pools and the special genetic architecture of invasive species also received close attention. Dobzansky (1965) and Carson (1965) took advantage of the well-understood genetics of Drosophila at that time to make one of the first attempts to compare levels of genetic diversity in invasive and noninvasive species. The last fifty years has been marked heavily by the genetic revolution that greatly facilitated the solidification of the field of invasion genetics. The genetic reconstruction of invasion history has been central to many research programmes aimed at understanding the causes and consequences of invasiveness. Determining the geographical origin of introduced populations often provided essential reference points for ecological and evolutionary studies that involved contrasting native and introduced populations (Milne & Abbott 2000; Hierro et al. 2005; Dlugosch & Parker 2008).

What can molecular markers tell us about the invasion process?

The common use of molecular markers in ecological and evolutionary studies has enabled us to reconstruct the evolutionary history of biological invasions, sometime with surprising precision (Handley et al. 2011). Molecular approaches have become commonplace and have been used to ask important questions about the invasion process (Sakai et al. 2001). What is the taxonomic identity of invasive species? What are their native areas and their sources of introduction? How do these species manage to travel around the world? What are the vectors of primary introduction and vectors of secondary spread? Are we dealing with single or multiple introduction events? How are these serendipitous voyages shaping the genetic landscape of the introduced populations? Are demographic bottlenecks experienced during the invasion process? Are invasive populations depleted of standing genetic variation?

Ultimately, molecular approaches proved to be effective when exploring invasion histories (Estoup & Guillemaud 2010) and opened the door for extensive investigations of the evolutionary and ecological forces that shape invasiveness (Sakai et al. 2001; Lee 2002). Although technical challenges are still associated with next-generation sequencing, rapid progress is being made on the use of population genomics approaches and metabarcoding techniques. For example, genotype-by-sequencing (GBS) techniques, such as RAD-seq (Baird et al. 2008), that involve the use of large numbers of markers randomly distributed across genomes are being used increasingly for organisms for which few genomic resources presently exist. These methods enable the transition from coarse molecular approaches to dense population genomics surveys. Such high molecular resolution is often necessary to accurately resolve the invasion history of species and identify severe bottleneck events associated with the invasion process. The power of next-generation sequencing will probably be combined with new computational methods and will allow researchers to test complex invasion scenarios (Estoup et al. 2010; Fitzpatrick et al. 2012). Moreover, metabarcoding approaches open the possibility of applying molecular identification techniques to very complex communities. While this method is commonly used to infer basic biodiversity estimates in healthy or disturbed ecosystems, it also offers the possibility of rapid and relatively inexpensive methods for the detection of invasive species. Such advanced molecular approaches will soon enable the application of
‘forensic’ methods to invasion biology by allowing the detection of traces of extracellular DNA when invasive species are still at very low abundance and their presence cannot be identified by morphological methods.

**Contrasting natural and human-mediated invasions**

Changes in the geographical range of species are recognized to be natural processes (Elton 1958; Lodge 1993). The distribution of species constantly expands and contracts, and these changes can trigger dynamic ecological and evolutionary processes, leading to radiations and extinctions, which ultimately shape the Earth’s biodiversity. For example, over geological time, many marine species colonized estuaries, rivers and lakes and eventually colonized the land. It is easy to appreciate that some invasive waves occurred abruptly, over ecological time scales. This is in contrast with the major colonizations and recolonizations of particular regions that often involve geological times (e.g. after glaciations; Wilson et al. 2009; Estoup & Guillemaud 2010). The temporal and geographical scales that we consider when investigating natural invasions are often vast (Box 1). However, contemporaneous invasions have received much closer investigation than historical invasions. Inherently, the field of invasion biology has a strong anthropogenic focus due to the problem that we are currently facing: the homogenization of biodiversity. Natural barriers to dispersal and gene flow are continually being removed due to accelerated human trade and travel (Mooney & Cleland 2001). Moreover, human activities contribute to extensive worldwide homogenization of the environment through agriculture practices and urbanization (Sax & Brown 2000). By focusing on anthropogenic factors, we recognize humans as the main force facilitating the global hitchhiking of species (directly or indirectly). In addition, human welfare is often used to evaluate the impact of these invasions. As a consequence of this strong anthropogenic focus, the terminology associated with invasive species is constantly evolving and revised (reviewed by Colautti & Maclsaac 2004). Extensive efforts have been directed towards the implementation of specific and consistent terminology, particularly when focusing on human-mediated invasions. Terms such as colonizing, invasive and weedy have been replaced by nonindigenous, exotic, introduced and naturalized. In this study, I will use the term invasive in a very broad sense to refer to the species that rapidly spread (over ecological times as opposed to geological times), irrespective of their ecological or economic impact. I will discuss species that have an exceptionally high ability to establish populations in a geographical or ecological space not previously occupied. By doing this, I aim to facilitate the comparison and contrast between natural and human-mediated invasions and to partially depart from the anthropogenic perspective.

**Natural invasions: lessons from zooplankton species**

Zooplankton organisms can be studied in the context of invasion biology in ways that enable us to contrast natural and human-mediated invasions. This may come as a surprise, as few zooplankton species are considered to be notorious invaders and are on our radar when implementing preventive measures. In fact, zooplankton species have been regarded as cosmopolitan for nearly a century. Their apparent widespread distribution has been linked to their production of resting eggs, which allows zooplankton species to travel long distances and disperse between continents. This view of zooplankton cosmopolitanism has slowly been invalidated (reviewed in Hebert & Cristescu 2002). Detailed morphological and ecological studies have revealed that many species with a broad geographical distribution represent complexes of closely related species that often have restricted, allopatric distributions (Frey 1982; Havel & Shurin 2004).

Detailed molecular data have also revealed an unexpectedly high level of regional endemism and population-level differentiation, suggesting that geographical and/or ecological barriers to dispersal have been effective in maintaining a high level of continental and regional endemism even in species considered to be exceptional dispersalists (Hebert & Wilson 1994; Taylor et al. 1998). For example, detailed phylogeographic surveys of cladoceran species that were assumed to have Holarctic distributions, such as *Polyphemus pediculus* and *Leptodora kindtii* revealed unexpectedly high levels of cryptic endemism (Xu et al. 2009; Millette et al. 2011). Many species of *Daphnia* thought to have geographical distributions that span entire continents are known to have marked phylogeographic structure produced by restricted levels of gene flow among populations. The *Daphnia pulex* group (*sensu lato*) represents a rich species complex of about 12 species with distinct ecology (Adamowicz et al. 2009). The best investigated species of this complex, the North American *D. pulex* (*sensu stricto*) and *D. pulicaria*, are probably to be among the most subdivided species yet documented (Lynch et al. 1999) due to their unusually high level of among-population structure across small geographical scales, which suggests the existence of strong ecological barriers that restrict gene flow (Cristescu et al. 2012). The quest to identify true cosmopolitan species has also failed when investigating zooplankton groups with smaller body size and higher effective population sizes, such as...
rotifers (Gómez et al. 2002). Despite the growing evidence of strong regionalism and increased endemism, zooplankton species remain regarded as exceptionally good colonists. When barriers (either physical or ecological) that previously separated biota for millions of years are naturally passed or disrupted, zooplankton species can attain very long-distance colonization events (Adamowicz et al. 2009) or significant ecological transitions (Cristescu et al. 2012). The achieved regional distribution and the restriction of gene flow, despite exceptional dispersal abilities, are often considered to be the result of a strong priority effect shaped by invasion order (Robinson & Dickerson 1987), but also by the rates and timing of invasions (Robinson & Edgemont 1988). Propagule banks (dormant stages of plants and animals that remain viable for decades) are expected to

<table>
<thead>
<tr>
<th>Box 1: Contrasting the history of natural and human-mediated invasions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristics</td>
</tr>
<tr>
<td>Temporal scale and general considerations</td>
</tr>
<tr>
<td>Origin of invasions and invasion routes</td>
</tr>
<tr>
<td>Rates of invasions (number of primary introductions per unit of time)</td>
</tr>
<tr>
<td>Rates of secondary spread</td>
</tr>
<tr>
<td>Molecular approaches</td>
</tr>
<tr>
<td>Analytical approaches</td>
</tr>
</tbody>
</table>
also buffer resident populations against local extinction and to enhance the priority effect on subsequent immigrants (Mergen et al. 2007). These attributes make zooplankton species particularly suitable for elucidating prolonged invasion histories and contrasting natural and human-mediated invasions.

During the last 50 years, at least five cladoceran species invaded the Great Lakes region of North America alone, suggesting an intercontinental rate of human-mediated dispersal of at least one invasion per decade (Hebert & Cristescu 2002). Using phylogeographic data and genetic divergence between closely allied species of cladocerans, Hebert & Cristescu (2002) estimated that rates of human-mediated invasions are thousands of times higher than natural, intercontinental colonization events. Biological invasion is, of course, a natural process. Many species experience periods of dramatic range expansion. However, the rate and the geographical and ecological scale of human-mediated invasions are incredible today and dramatically different from those of the geological past.

Genetic data also allow us to investigate patterns of very rapid intracontinental range expansions, particularly postglacial recolonization routes. We can compare the genetic structure of populations and/or species that inhabit glaciated regions and experienced strong environmental disturbances, with populations that persisted in more stable, unglaciated regions. For example, demographic analysis using Tajima’s D (Tajima 1989), Fu’s Fs (Fu 1997) and mismatch distributions (Rogers & Harpending 1992) of lineages that show contrasting distributional patterns allows the detection of non-neutral evolution and rapid population expansion events. Such approaches have been used in conjunction with phylogeographic and haplotype network analyses to identify the centre of origin (equivalent to the source of introduction) of populations or species that inhabit previously glaciated regions or more stable, unglaciated regions. For example, the conventional belief is that many plant and animal species in Europe recolonized large parts of the continent from southern refugia that largely correspond to southern European peninsulas, which were less influenced by glaciations. However, for the predatory cladoceran Leptodora kindti, detailed phylogeographic and network analyses based on mitochondrial markers have revealed a pattern which sharply contrasts with this perception. Southern lineages remained localized in very narrow distributional ranges, sometimes in single habitats, while cryptic refugia situated in the central (Carpathian region) or northern parts of the continent have been responsible for the recolonization of large parts of the European continent and large parts of Eurasia (Millette et al. 2011). Similar cryptic, northern refugia, corresponding to the Russian Yaroslave region and Siberia, have been identified for other zooplankton species such as the onychopod cladoceran Polyphemus pediculus (Xu et al. 2009). Paleoecological reconstructions of northern Eurasia indicate that large ice-dammed paleolakes covered large parts of the White Sea basin and west Siberia (Mangerud et al. 2004). These extensive lakes were formed due to obstruction of the north-flowing rivers whose drainages were diverted towards the Caspian basin. This facilitated the southern dispersal of species trapped among the ice sheets. The persistence of zooplankton species in northern refugia was probably facilitated by their ability to produce diapause eggs known for their ability to withstand desiccation and freezing and to maintain viability for a prolonged period of time (sometimes hundreds of years). This adaptation allowed zooplankton species to persist in regions in which aquatic habitats underwent more dramatic geological ephemerality. What is particularly impressive is the high ‘invasive’ ability of those lineages that persisted under such harsh conditions. The phylogeography of L. kindti, reconstructed by Millette et al. (2011), revealed that sister clades/species within each of the major geographical regions analysed (North America, Europe and Far East Asia) contain clades that show marked distributional differences. Each of the very narrowly distributed clades situated in previously unglaciated areas of Oregon, Greece, Turkey and Japan has a sister clade characterized by a very widespread distribution that spans entire continents and often shows a clear signature of relatively recent population expansion. The widespread clades arguably contain populations that are more invasive than the clades that were identified in single habitats, which are often remnants of paleolakes. As expected, population genetics analyses conducted on freshwater zooplankton and fishes across both glaciated and unglaciated regions revealed marked genetic differences (Bernatchez & Wilson 1998). While glaciated regions are generally occupied by populations that show a marked signature of recent demographic expansion and generally low genetic diversity (low nucleotide diversity, but high haplotype diversity), unglaciated regions harbour populations with a higher level of intrapopulation genetic diversity (in particular nucleotide diversity due to the presence of divergent haplotypes), larger genetic splits between sister clades and no signature of recent demographic expansions. These dramatic differences in distributional ranges and genetic structure between closely related species that inhabit formally glaciated regions and those found in unglaciated regions are also relevant to invasion biology. These patterns provide support for a hypothesis, proposed by Lee & Gelembiuk (2008), suggesting that the evolutionary history of disturbances might select for...
life history traits that favour rapid colonization, such as rapid growth, persistence, high somatic plasticity and long-distance dispersal ability.

**Human-mediated invasions**

*Invasion forensics: identifying the source(s) of invasions and reconstructing the invasion routes*

When investigating human-mediated invasions, their relatively recent history (centuries and decades vs. millennia) facilitates approaches that mimic forensic techniques, with the goal of reconstructing the most likely invasion scenarios. However, the correct identification of the source population or route of invasion is not always straightforward. The native areas of invasive species can be unknown or very large and difficult to explore exhaustively (Lombaert et al. 2011). For many species with a long history of invasion (several decades) and nearly cosmopolitan distribution, the native ranges are an educated guess at best. This is often the case for marine invasive species such as the colonial or solitary tunicates. Carlton (1996) noted that the general approach when working with populations that are not demonstrably native or introduced is still to classify all populations without records of introduction as being ‘native’ rather than cryptogenic populations. For example, Carlton (1996) conservatively estimates that about 100 aquatic species in San Francisco Bay are cryptogenic. This conventional approach of classifying cryptogenic populations as native introduces significant noise to large-scale data sets and obscures the reconstruction of invasion history. Native populations may also be genetically homogenous, due to a recent homogenization effect induced by human activities, high levels of natural gene flow or a short evolutionary history of the lineage under investigation. However, given that genetic homogeneity is never fully achieved, such situations can be resolved by employing very fast evolving marker or genomic approaches that cover a large portion of the genome and a large number of individuals. Moreover, native populations are sometime not completely disconnected from the gene pool that encompasses populations of the invaded ranges. This is particularly common when investigating invasions that start as outbreaks, species that successfully enter new ecological spaces while only slowly expanding beyond the native geographical range. Furthermore, human-mediated invasions often involve complex routes with multiple sources of introduction and repeated introductions (Wilson et al. 2009). Such complex invasive scenarios which often achieve network patterns can greatly complicate the reconstruction of the invasion steps.

Despite these general difficulties, the genetic reconstruction of source populations and routes of invasion have proven to be a fruitful endeavour (reviewed in Estoup & Guillemaud 2010). Many studies identified the source introduction(s) and invasion routes with high precision. For example, the study of the Ponto-Caspian invader Cercopagis pengoi, the fishhook waterflea (Cristescu et al. 2001), demonstrated the usefulness of mitochondrial DNA (mtDNA) markers in locating the source of invasion as well as the corresponding invasion corridor. The lagoons of the Black Sea were identified as the source of the Baltic population of Cercopagis. This initial introduction resulted in a severe bottleneck that virtually wiped out the intrapopulation genetic variation in mtDNA. Moreover, the genetic similarity between populations in the Baltic Sea and those in the Great Lakes suggested a subsequent transfer of animals from the Baltic Sea to North America (Cristescu et al. 2001). A few other early studies conducted on the southern house mosquito, Culex quinquefasciatus, the vector of avian malaria, pointed to the complexity of the invasion process and the importance of identifying invasive populations that act as major sources or hubs of secondary spread (Fonseca et al. 2000; Estoup et al. 2001). Such transit populations have been termed ‘invasive bridgeheads’ (Estoup & Guillemaud 2010) and are considered to either foster favourable evolutionary shifts or to be situated in the path of major vectors. These early studies also demonstrated that fine-scale reconstruction of invasion history is possible when the native region harbours significant genetic structure, and both the invasive and native areas are sampled systematically. Moreover, these studies revealed regions and/or habitats that acted as donors for a larger number of species with shared evolutionary history but divergent life history attributes (Ricciardi & Maclsaac 2000; Cristescu et al. 2004). A large number of more recent studies combined historical and distributional data with genetic data (genetic patterns observed within and between populations across native and introduced ranges) to reconstruct complex scenarios of invasions for taxa with various life history attributes (brown algae, Voisin et al. 2005; crabs, Darling et al. 2008; clams, Hoos et al. 2010; toads, Estoup et al. 2010; insects, Lombaert et al. 2010). Collectively, these studies revealed the potential of genetic analyses to reveal detailed patterns of introduction, establishment and spread of invasive species.

Both nuclear and mitochondrial (or chloroplast) markers have been employed successfully with many studies making use of multiple markers with distinct levels of resolution (e.g. organelle markers coupled with nuclear microsatellite markers). The methods commonly used involve building dendrograms (such as UPGMA or neighbour joining, e.g. Saitou & Nei 1987) and parsimony networks often interpreted in a geographical
context. Detailed population genetic analyses frequently accompany conclusions derived from interpreting dendrograms and networks. Markers are generally tested for the level of resolution provided, for concordance with Hardy–Weinberg (HW) expectations and for linkage disequilibrium (e.g. nonrandom association of alleles between loci). Several traditional methods are often implemented to determine population structure and demographic history (Box 2). Population assignment methods that allow pairing of introduced populations with their corresponding source population(s) represent an effective tool for identifying population structure. For example, methods of clustering multilocus genotypes into genetically discrete groups such as those implemented in STRUCTURE (Pritchard et al. 2000), BAPS (Corander et al. 2003) or GENELAND (Guillot et al. 2005) are commonly used to infer donor populations and reconstruct invasion routes. More recently, the method based on approximate Bayesian computations (ABC) allows the evaluation of alternative invasion scenarios. ABC uses a model-based inference in a Bayesian setting (Beaumont et al. 2002) and provides probabilities (with confidence intervals) for all invasion scenarios considered. The array of population genetics approaches available have proved to be useful in reconstructing invasion routes and inferring the number of independent introductions despite their inherent limitations. However, these well-established methods are not always fully applicable to invasion studies. It is generally accepted that as human-mediated invasions have very restricted timescales, much smaller than the timescale of mutation and genetic drift, analyses based on neutral theory cannot always accurately resolve such dynamics and have to be used cautiously (Fitzpatrick et al. 2012).

Founding events

Theory predicts that founding events will drastically deplete genetic variation in invading populations (Nei et al. 1975; Roman & Darling 2007). Thus, invasive populations are expected to be less genetically diverse than the populations from which they are derived. However, invasive species persist and adapt to new conditions despite the likely reduction in genetic variation. While a severe reduction in genetic diversity could reduce adaptive potential when such reduction involves quantitative traits, other mechanisms of genomic reconfiguration, the fate of which is driven by genetic drift or a combination of genetic drift and selection, can provide populations with unique adaptive opportunities. Such serendipitous events could involve intraspecific and interspecific admixture which could provide invasive populations with the ability to explore diverse ecological spaces. Moreover, hybridization followed by fixation of large-scale inversions might also provide reproductive isolation from donor populations or from other introduced populations, generating unexpected barriers to gene flow (Hoffmann & Rieseberg 2008). Chromosomal inversions can spread when such inversions capture locally adapted alleles facilitating range expansions (Prevosti et al. 1988; Kirkpatrick & Barrett this issue). Furthermore, prolonged periods of isolation and low population size result in severe inbreeding that could potentially reduce the mutational load of invasive populations. Because inbreeding increases homozygosity, and hence the effectiveness of selection against detrimental alleles that are fully or partially recessive, severe inbreeding can result in a reduction of the frequency of such detrimental alleles and a significant fitness rebound (Crnokrak & Barrett 2002). Phenotypic changes that provide invasive populations the ability to explore divergent niches are also expected to occur due to random sampling of the genetic diversity of source populations (Keller & Taylor 2008).

One of the biggest surprises revealed by genetic studies has been the realization that invasive species are not always experiencing the expected reduction in genetic diversity. Many studies that compared neutral genetic diversity of invasive populations to that of native populations or related, noninvasive species reported comparable levels of genetic variation in native and introduced populations. Moreover, for many of the worldwide marine invaders, such as the European green crab, Carcinus maenas (Darling et al. 2008), the brown alga Undaria pinnatifida (Voisin et al. 2005), or the violet tunicate Botrylloides violaceus (Bock et al. 2011), invasive populations vary broadly in the degree to which genetic diversity is retained in invasive populations. These studies revealed no strong correlation between the genetic diversity retained by an introduced population and its invasive potential (Roman & Darling 2007; Dlugosch & Parker 2008). For example, one of the least diverse populations of C. maenas (from the western USA) has been the most successful (nearly 2000 km coastline expansion), while the highly diverse Cape Town population has remained static (Darling et al. 2008). The large comparative study conducted by Dlugosch & Parker (2008) on 80 species of animals, plants and fungi revealed that losses of neutral variation are sometime detectable but are certainly not ubiquitous. Fitzpatrick et al. (2012) suggests that this is probably explained by the fact that genetic drift is not efficient in removing large amounts of genetic variation from populations experiencing expansions soon after introduction. Simulations of isolated populations experiencing sudden decline with or without recovery show that rapid recovery and/or immigration can make severe

© 2015 John Wiley & Sons Ltd
### Box 2: Analytical methods employed when reconstructing invasion histories

#### Phylogenies

Phylogenetic approaches (both distance based, e.g., Saitou & Nei 1987; and character based) are often used to infer the taxonomic identity of invasive species by exploring their relationship with the noninvasive relatives. Moreover, these approaches are often used to infer genetic relationships among native and introduced populations and investigate patterns of range expansions on postglacial time scales. Due to the generally short evolutionary time frame covered, such data sets involve a low number of informative characters or short overall genetic distances. Moreover, phylogenetic relationships can be difficult to interpret when the source population(s) is not sampled. These limitations make the reconstruction of true phylogeny difficult (Estoup & Guillemard 2010).

#### Population genetics

Population genetics provides a solid mathematical framework for understanding evolutionary processes in natural populations. Markers are generally tested for their level of resolution, concordance with Hardy–Weinberg (HW) expectations and for linkage disequilibrium (i.e., nonrandom association of alleles between loci). Programs such as **GENEPOP** (Raymond & Rousset 1995), **FSTAT** (Goudet 1995), **ARLEQUIN** (Schneider et al. 2000) and **Baps** (Dieringer & Schlötterer 2003) are used often to investigate marker properties as well as population diversity (proportion of polymorphic loci, allelic richness, heterozygosity).

Recent bottlenecks are often detected by testing for reduced allele number and excess of heterozygosity compared to that expected for an equilibrium population. **BOOTTLENECK** (Cornuet & Luikart 1996) is the program commonly used for detecting recent effective population size reductions from allele frequency data. The disadvantage of this method is the limited period of time over which the test is able to show a bottleneck effect, the large bottleneck size, number of loci and sample sizes needed to generate high statistical power.

A good understanding of invasion history requires often the identification of population structure and migration rates across native and invasive ranges. Assignment tests based on clustering multilocus genotypes into groups represent one of the most robust ways of investigating population structure. The computer program **STRUCTURE** (Pritchard et al. 2000) is used often to estimate the most likely number of actual populations from the full genotype data set by comparing posterior probabilities across a range of possible population numbers. The programs (**BAPS**) developed by Corander et al. (2003) and **GENELAND** developed by Guillot et al. (2005) estimate the number of populations using information about the source populations.

A fundamental problem in invasion biology is distinguishing between ongoing gene flow (an equilibrium situation) and recent separation of populations with decreased or no gene flow (a nonequilibrium situation). One useful approach is to estimate $F_{ST}$ and gene flow between all pairs of populations and to look for the effect of isolation by distance (Slatkin 1993). Recent geographical isolation is not expected to show an inverse correlation between gene flow and geographical distance (Hutchison & Templeton 1999).

In general, the methods described take poorly into account the effect of demographic stochasticity or genetic history. Given that stochasticity plays an important role during the invasion process and exhaustive sampling is often not feasible, population genetics results need to be interpreted cautiously.

#### Using approximate Bayesian computation (ABC) to evaluate distinct invasion scenarios

The model-based method called approximate Bayesian computation (ABC, Beaumont et al. 2002) has been used to make inferences about the invasion routes. Posterior probabilities of different models and/or the posterior distribution of the demographic parameters under a given model are determined by measuring the similarity between the observed data set and a large number of simulated data sets. Several ABC programs such as **popABC** (Lopes et al. 2009), **DIYABC** (Cornuet et al. 2008) and **ABCtoolbox** (Wegmann et al. 2010) provide user-friendly interfaces that allow researchers to test complex invasion scenarios.

#### Population genomics

NGS technologies coupled with powerful software platforms enable researchers to analyse tens of thousands of genetic markers. Often genomes are subsampled at homologous locations using genotype by sequencing (GBD) or restriction site-associated DNA sequencing (RAD-seq; Davey et al. 2011). Several recent softwares such as **Stacks** (Catchen et al. 2013a), **SAMtools/BCFtools** (Le et al. 2009) and the Genome Analysis Toolkit (**GATK**, McKenna et al. 2010) produce core population genomic summary statistics and SNP-by-SNP statistical tests. These tests are often conducted across a reference genome. However, Stacks was developed specifically for projects that do not benefit from a well-characterized reference genome.

#### Phylogenomics

Phylogenetic studies using genome data (e.g., GBS markers) have the potential to elucidate complex evolutionary histories that are obscured by hybridization events or prolonged periods of gene flow after lineage splitting and incomplete lineage sorting (Catchen et al. 2013b). When applied to invasive species, such an approach can elucidate the phylogenetic relationship among native and invasive populations or among invasive lineages and their noninvasive relatives.
bottlenecks undetectable (Fitzpatrick et al. 2012). Population bottlenecks generally generate a pattern of ‘heterozygosity excess’ due to a deficiency of rare alleles (Nei et al. 1975). Recent bottlenecks are often detected by testing for reduced allele number and excess of heterozygosity compared to that expected for an equilibrium population.

**Single or multiple introduction events?**

Several recent studies on invasion histories have focused deeper on the role that single or multiple introduction events might play in shaping the genetic diversity of invasive populations (Novak & Mack 2005; Roman & Darling 2007; Dlugosch & Parker 2008). Single introduction events followed by subsequent introductions and bottlenecks are expected to result in a significant reduction of genetic diversity. On the other hand, recurrent gene flow due to repeated introductions is expected to largely mitigate a drastic reduction of genetic diversity (Dlugosch & Parker 2008). Introduced populations are expected to have higher probability of establishment and spread if the propagule pressure is maintained high, due to the high number of individuals being released at a particular location (Simberloff 2009; Blackburn et al. this issue). This theoretical prediction has been confirmed by empirical studies. Multiple introductions are often correlated with successful establishment (Barrett & Husband 1990) because repeated introductions have the combined benefit of maintaining high population size and increased genetic variation. The study conducted by Kolbe et al. (2004) on the brown anole revealed eight independent introductions in Florida from across the native range of the invasive lizard. The authors postulate that the infusion of genetic material from distinct native sources resulted in introduced populations that harbour substantially more genetic variation than the source populations. The case of the brown anole is certainly not an isolated case. When investigating invasions with a relatively long history, multiple introductions appear to be the norm rather than the exception (Facon et al. 2003; Kelly et al. 2006).

**Genetic reconstruction of invasion history: where next?**

Recent advancements driven by the genomics revolution have enabled rapid, deep and relatively inexpensive characterization of genetic patterns in natural populations, including the relationship between native and introduced populations, and the pathways that link these populations (Dlugosch et al. 2013; Puzey & Vallejo-Marin 2014; Box 3). Often, such approaches involve methods of complexity reduction such as transcriptome sequencing (Dlugosch et al. 2013; Hodgins et al. this issue) or RAD sequencing (Baird et al. 2008; Roda et al. 2013). However, for studies that benefit from a well-characterized reference genome, whole-genome resequencing is often employed for genotyping SNPs markers (Savolainen et al. 2013; Puzey & Vallejo-Marin 2014). Such high-throughput methods that involve gen-
nomewide markers provide a global view of the genome and enable deep screening of large numbers of populations. As a consequence, population genomics approaches greatly facilitate the identification of demographic events, the identification of loci that experience selection during the invasion process or other evolutionary events (Stapley et al. 2010; Puzey & Vallejo-Marin 2014). It is generally recognized that microsatellite-based bottleneck tests often do not detect bottlenecks in populations known to have experienced dramatic declines (Peery et al. 2012). This observation was followed by simulations that confirmed that bottleneck tests have limited statistical power to detect bottlenecks largely as a result of the limited sample sizes typically used in microsatellite-based studies and poorly estimated values for mutation model parameters. The application of population genomics approach to invasive species is probably to improve the characterization of demographic history by greatly increasing the number of loci analysed (Davey et al. 2011; Peery et al. 2012).

Several genomics studies indicate that population genomic approaches can provide a solid understanding on the genomic consequences of invasion. By analysing whole-genome sequences of 22 populations of *Mimulus guttatus*, Puzey & Vallejo-Marin (2014) demonstrated that introduced plants in the UK are characterized by a 50% reduction in neutral (synonymous) genetic diversity and suggested a common origin for non-native populations. Hohenlohe et al. (2010) used RAD tags data and confirmed the long-standing biogeographical hypothesis that the large panmictic oceanic populations of stickleback have repeatedly colonized freshwater habitats giving rise to phenotypically divergent populations. The genomic signature of both balancing and divergent selection was remarkably consistent across the populations investigated. Moreover, Catchen et al. (2013b) used a RAD tags data set to test the hypothesis of recent stickleback introduction into central Oregon, where this species was only recently documented. The authors documented a clear genetic division between coastal and inland populations as well as the role of introgressive hybridization in coastal populations and recent expansion in central Oregon. Although genomic approaches are still in their infancy and the pipelines necessary for analysing genomic data are only emerging, the great opportunity for exploring invasion histories and the underlying evolutionary consequences is overwhelming.

A suite of genetic tools and computational approaches have been developed to sequence and detect organisms from complex environmental samples, including those found at low abundance or which are partially degraded (Box 3). These provide a sensitive approach to monitoring biodiversity (Bik et al. 2012; Thomsen et al. 2012), detecting invasive species (Lodge et al. 2012; Fochon et al. 2013; Darling 2014) and inferring large-scale patterns of natural or human-mediated connectivity between populations. The method known as metabarcoding combines next-generation sequencing (NGS) with traditional barcoding techniques (Hebert et al. 2003) in order to successfully circumvent the laborious effort otherwise necessary to identify single organisms from complex mixtures (reviewed in Taberlet et al. 2012). Many technical limitations have been identified. Several solutions have been proposed to overcome them, and many are under development (Bik et al. 2012; Cristescu 2014). These metabarcoding techniques often need to be carefully calibrated against simple biological assemblages of well-known composition in order to assess the most effective methods of recovering biodiversity. The ultimate goal is that each biological or ecological species (reproductively isolated units or distinct units with characteristic ecological requirements) will be represented by a single genetic cluster or OTU. Comprehensive databases comprised of repository sequences would allow researchers to link their OTUs to corresponding Linnean species and access valuable ecological information (Cristescu 2014). Recent studies provide a framework for the application of metabarcoding to monitor biodiversity and for the early detection of invasive species in complex environmental samples (Talbot et al. 2014). Moreover, long-term surveillance of vulnerable habitats situated near entry points would provide valuable data on introduced, exotic species that fail to invade (Box 2). Such approaches can be also employed to quantify propagule pressure associated with the major vectors of introductions or secondary spread. For example, species assemblages commonly transported in ballast water or as hull fouling on transoceanic vessels can be also monitored efficiently. Genetic information on the source of invasions, invasion vectors and routes, the number of introductions, and the pathways of secondary spread can be also used by managers and other decision makers to support conservation projects and other regulations that minimize accidental introductions (Estoup & Guillemaud 2010; Darling 2014).

Conclusions

Given the complex evolutionary forces and ecological settings involved during the invasion process, the reconstruction of invasion history remains central to studies that aim to understand the mechanisms driving successful invasions. Molecular tools have provided effective ways of tracing the history of both natural and human-mediated invasions. Many attributes of historical and contemporaneous biological invasions are alike, but often the methods (specific molecular markers or
analytical approaches) applied to study them need careful consideration. Biological invasions are indeed natural processes; however, their rate of introduction and geographical scale are extraordinarily high today and dramatically different from the recent past. As one of the most successful colonizing species, humans have an inherent desire to understand the fundamental nature of the invasion process and the important historical events that shape invasion trajectories. Biological invasions have provided complex natural experiments that have stimulated deep reflections on the ecological footprint of humanity and our future as a species.

Acknowledgements

I would like to thank the organizers for the invitation to participate to the Genetics of Colonizing Species Symposium that inspired this volume. The manuscript benefited from valuable comments provided by the anonymous reviewers the editor, Dr. Loren Rieseberg, and laboratory members. Funding was provided by the Natural Science and Engineering Research Council (NSERC).

References


Hodgins KA, Bock DG, Hahn MA et al. (this issue) Comparative genomics in the Asteraceae reveals little evidence for parallel evolutionary change in invasive taxa. Molecular Ecology, this issue.


© 2015 John Wiley & Sons Ltd


Schneider S, Roessli D, Excoffier L (2000) Arlequin Ver. 2.00: A Software For Population Genetic Data Analysis. Genetics and Biometry Laboratory, University of Geneva, Bern, Switzerland.


M.C. wrote the manuscript.