The consequences of genomic architecture on ecological speciation in postglacial fishes

Sean M. ROGERS*, Jonathan A. MEE, Ella BOWLES
Department of Biological Sciences, University of Calgary, 2500 University Drive NW, Calgary, Alberta, T2N 1N4, Canada

Abstract The quest for the origin of species has entered the genomics era. Despite decades of evidence confirming the role of the environment in ecological speciation, an understanding of the genomics of ecological speciation is still in its infancy. In this review, we explore the role of genomic architecture in ecological speciation in postglacial fishes. Growing evidence for the number, location, effect size, and interactions among the genes underlying population persistence, adaptive trait divergence, and reproductive isolation in these fishes reinforces the importance of considering genomic architecture in studies of ecological speciation. Additionally, these populations likely adapt to new freshwater environments by selection on standing genetic variation, as de novo mutations are unlikely under such recent divergence times. We hypothesize that modular genomic architectures in postglacial fish taxa may be associated with the probability of population persistence. Empirical studies have confirmed the genic nature of ecological speciation, implicating surprisingly extensive linkage disequilibrium across the genome. An understanding of these genomic mosaics and how they contribute to reproductive barriers remains unclear, but migration rates and the strength of selection at these loci is predicted to influence the likelihood of population divergence. Altogether, understanding the role of genomic architecture is an important component of speciation research and postglacial fishes continue to provide excellent organisms to test these questions, both from the perspective of variation in architectures among taxa, and with respect to the distinct environments they have colonized. However, more empirical tests of ecological speciation predictions are needed [Current Zoology 59 (1): 53–71, 2013].

Keywords Adaptation, Ecological speciation, Genomic architecture, Modularity, Postglacial fishes, QTL

1 Introduction

The quest to understand the origin of species has never been more exciting and within our grasp. Traditional speciation models were built upon the assumption that geographic isolation (i.e., complete lack of gene flow) is required for the evolution of reproductive isolation (RI) during population divergence (Mayr, 1963). Yet, Darwin recognized that reproductive isolation evolves due to natural selection (Darwin 1859, Charlesworth and Charlesworth 2009) and it is increasingly clear that adaptive divergence (with or without geographic isolation) is indeed fundamental to the speciation process (Dobzhansky, 1951; Diekmann and Doebeli, 1999; Kondrashov and Kondrashov, 1999; Schluter, 2000; Mallet, 2001; Nosil and Feder, 2012). Not all species evolve by natural selection, but the evidence suggests that most of them do (Schluter, 2009). Thus, understanding how natural selection gives rise to species in close association to their environment remains important in efforts to understand the evolution of populations, species, and the nature of biodiversity.

Ecological speciation occurs when adaptive population divergence results in the formation of reproductive barriers. Divergent or disruptive natural selection favours different alleles advantageous for populations living in different environments, and these alleles are not favoured in alternative environments (Schluter, 2009). Adaptive genetic divergence is predicted to reduce gene flow and result in the formation of reproductive barriers at the genes underlying these adaptive phenotypes, or those genes genetically correlated with them (i.e., exhibiting gametic phase disequilibrium). Ecological speciation can therefore cause RI under any geographic scenario of population divergence, including “divergence-with-gene-flow” models (Smadja and Butlin, 2011). This process differs from “non-ecological” speciation models that invoke drift and fixation of different incompatible alleles in the evolution of RI between populations adapting to similar selection pressures (so-called mutation-order speciation, Mani and Clarke, 1990; Schluter, 2009). Whereas ecological speciation favors different alleles between populations,
mutation-order speciation favors the same alleles in diverging populations, but the probability of alleles arising and fixing in one population is due to chance (Schluter and Conte, 2009; Nosil and Flaxman, 2011). Contrasting these two speciation models highlights two processes that are required for ecological speciation to occur. First, following adaptive peak shift, population persistence is necessary prior to the evolution of RI. This is particularly relevant in the initial stages of population divergence, such as the colonization of new ecological environments or following sudden environmental change (Pavey et al., 2010c; Rogers et al., 2012).

Second, the evolution of adaptive genetic divergence in these new environments has to be associated with prezygotic or postzygotic RI (Rundle and Nosil, 2005; Nosil, 2012).

Tests of ecological speciation must therefore focus on the role of selection during the initial stages of population divergence (Rundle and Schluter, 2004). The most general prediction is that RI should evolve between populations adapting to contrasting environments, but not between populations adapting to similar environments (Rundle and Nosil, 2005). An extension of this prediction is that the extent of ecological divergence should be positively correlated with the strength of RI (Funk et al., 2006). A second prediction in that the phenotypic traits associated with adaptive divergence should cause RI via intrinsic (hybrid inviability and/or segregation distortion) or extrinsic (reduced hybrid fitness in parental environments) postzygotic barriers (Rogers and Bernatchez, 2006), or via the development of prezygotic barriers that cause individuals to discriminate against heterospecific matings (Rundle and Schluter, 2004). Finally, neutral gene flow should decrease as adaptive population divergence increases (Lu and Bernatchez, 1999; Ogden and Thorpe, 2002; Nosil et al., 2009; Smadja and Butlin, 2011).

Several empirical studies have now demonstrated the role of adaptive divergence in promoting ecological speciation in nature (e.g., McKinnon et al., 2004), leading some to claim that natural selection is the primary force leading to the origin of species (Schluter, 2001, 2009). One of the most compelling lines of evidence comes from numerous tests of parallel ecological speciation, whereby greater RI repeatedly evolves between independent populations adapting to contrasting environments than between independent populations adapting to similar environments (Rice and Hostert, 1993; Schluter and Nagel, 1995; Bernatchez et al., 1996; Rundle et al., 2000; Taylor and McPhail, 2000; McKinnon et al., 2004; Langerhans et al., 2007; Butlin et al., 2008; Nosil et al., 2008; Schluter, 2009; Deagle et al., 2012; De Busschere et al., 2012; Jones et al., 2012;). Yet, there are cases in which the formation of RI is inconclusive or even absent (Turgeon et al., 1999; Docker et al., 2012; Taylor et al., 2012b), reinvigorating discussion about the power and tendency of ecological divergence to form reproductive barriers between diverging populations in nature, as well as a reevaluation of the evidence required to support ecological speciation (Hendry, 2009; Lowry, 2012).

The lack of RI (or, at least, the lack of strong RI) in many cases of ecological divergence highlights an important gap in our understanding of ecological speciation. Despite increasing knowledge confirming the role of the environment in divergence, an understanding of the genomics of divergence and ecological speciation is still in its infancy (Via and Hawthorne, 2002; Schluter and Conte, 2009; Feder et al., 2012a; Nosil and Feder, 2012), precluding insight into the underlying mechanisms of RI. Distinct adaptive phenotypes might be apparent, but their underlying genomic architecture can vary, suggesting that selection may produce rapid changes in RI in some circumstances but not others (Butlin et al., 2012). We use the term “genomic architecture” to broadly describe the number of genes underlying adaptation and/or RI, their distribution across the genome, and the linkage disequilibrium among these genes and neutral loci throughout the genome. The gene is the elemental unit of adaptation (Rogers et al., 2012), and an understanding of the genetic basis of adaptation is known to be critical to understanding the action of selection and the evolution of RI in nature (Barrett et al., 2008; Linnen and Hoekstra, 2009).

A number of questions related to the genetic basis of ecological speciation, and the genomic architecture of ecological speciation, are outstanding. What are the genes associated with ecological speciation? Do they, like genes for adaptation, evolve in parallel? How are ecological speciation genes organized within the genome and how do genomes evolve during adaptive population divergence? What influence does phenotypic plasticity have on ecological speciation (Fitzpatrick, 2011)? How does natural selection lead to speciation and what causes variation in the rate of speciation among taxa? What is the contribution of adaptive genes underlying changes in behavior, physiology, or morphology to the formation of reproductive barriers? Are incompatible genotypes the result of a few large effect genes or the interaction of many small effect genes?
Does the level of modularity (i.e., linkage among functionally-related genes) affect the probability of RI forming? All of these questions can be addressed by studying the genomic architecture of divergent populations. Despite the advances we have made in understanding ecological speciation in nature, the importance of the role of genomic architecture during ecological speciation, namely how it facilitates or impedes population persistence, adaptive divergence, and reproductive isolation, is largely unknown. Hence, our understanding of the nature of biodiversity is limited, in part, by our poor understanding of genomic architecture.

In this paper we investigate the role of genomic architecture in promoting ecological speciation in postglacial fishes. We focus on genomic architecture because, in contrast to other factors associated with ecological speciation, the consequences of genomic architecture are far less understood (Nosil and Feder, 2012; Feder et al., 2012a). We focus on postglacial fishes because several aspects of genomic architecture have been elucidated for several taxa under various scenarios of environmental change and various stages of ecological speciation (Table 1). Our objective is to clarify the potential role for genomic architecture in promoting or constraining ecological speciation in postglacial fishes, and to draw some generalities from this body of research. Although we refer to the significance of biogeography and secondary contact, we do not directly discuss the role of genomic architecture in sympatric and allopatric speciation, but instead focus on the influence of genomic architecture on ecological and mutation-order speciation when there is gene flow. Specifically, we investigate whether genomic architecture can impede or facilitate adaptive divergence. We examine the contribution of genomic architecture to the genetic basis of adaptive trait divergence and the association with RI. We examine how genomic architecture under these conditions has contributed to ecological speciation in postglacial fishes. We ask whether the existing evidence in postglacial fishes supports ecologically-based RI in contrast to the buildup of RI by incompatible side-effects of selected loci that may also occur under mutation order speciation (Nosil and Flaxman, 2011). We comment on recent theoretical predictions and advances to propose avenues for future research in the genomic architecture of ecological speciation.

2 Postglacial Fishes

Postglacial fishes colonized north temperate environments following the last ice age and represent an excellent model to study the genomic architecture of ecological speciation. Evolutionary studies of postglacial fishes have contributed significantly to our understanding of speciation (for a comprehensive review see Taylor, 1999), including the evolution of reproductive isolation (Nelson, 1968; Aspinwall and McPhail, 1995; Lu and Bernatchez, 1998; Rogers and Bernatchez, 2006), hybridization (Dowling and Moore, 1985; Rogers et al., 2001), species concepts (e.g., Mayr, 1963; Mayden and Wood, 1995), and recently even reversal of the speciation process (e.g., Taylor et al., 2006; Vonlanthen et al., 2012). The latter examples also include several contributions specifically focused on gaining an understanding of ecological speciation in species pairs or species complexes of fishes. “Species pairs” or “complexes” are those fishes that are reproductively isolated in sympathy and exhibit adaptive differences (e.g., spawning times, distinct physiologies or size at maturity) in association with their environments (e.g., habitat use, diet, predation regimes) (Schluter, 1996; Taylor, 1999; Schluter and Conte, 2009; Bernatchez et al., 2010). These unique groups demonstrate both population persistence and rapid adaptive divergence and include some of the youngest species on earth, with ecological predictions continuing to lead to new discoveries of this phenomenon in the wild (Hendry et al., 2000; Gow et al., 2008; Goetz et al., 2010).

The most common examples stem from colonizations of postglacial lake and stream environments following the end of the Pleistocene glaciations between approximately 10K to 18K years ago (McPhail and Lindsey, 1970; Schluter, 1996; Taylor, 1999; Bernatchez et al., 1999). The melting and recession of glacier ice precipitated a terrestrial rebound of the earth resulting in the formation of several-hundred lake environments that were colonized in parallel by both marine and freshwater fish species at the end of the ice age. The evidence for distinct gene pools in species complexes that evolved within these new environments is widespread and includes Arctic charr (e.g., Hindar and Jonsson 1993, Corrigan et al. 2011), Atlantic salmon (e.g., Verspoor and Cole, 1989; Gilbey et al., 2006), sockeye salmon (e.g., Hamon and Foote, 2005; Pavey et al., 2010b), lake whitefish (e.g., Lindsey, 1963; Fenderson, 1964; Bodaly, 1979; Bernatchez and Dodson, 1990), rainbow smelt (e.g., Taylor and Bentzen, 1993; Shaw and Curry, 2011), stickleback (e.g., Lavin and McPhail, 1993; McPhail, 1993; Thompson et al., 1997; Taylor and McPhail, 1999) and some lampreys (e.g., Docker,
Table 1  Elements of genomic architecture in the ecological speciation of postglacial fishes

<table>
<thead>
<tr>
<th>Species, Genus, or Family</th>
<th>Peak shift</th>
<th>Adaptive trait divergence</th>
<th>Genetic architecture</th>
<th>Large effect genes?</th>
<th>Modular?</th>
<th>Evidence for reduced gene flow/RI at adaptive traits and/or genes?</th>
<th>Role for biogeography (secondary contact)?</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three-spined stickleback</td>
<td>Habitat, predation, physiology: Marine</td>
<td>Linkage maps, pQTL, genome scans</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Unknown</td>
<td>Colosimo et al., 2004; Albert et al., 2008; Jones et al., 2012; Rogers et al., 2012.</td>
<td></td>
</tr>
<tr>
<td>Gasterosteus aculeatus</td>
<td>Freshwater</td>
<td>Armour, body size, shape</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Assortative mating</td>
<td>Possibly</td>
<td>Peichl et al., 2001; Malek et al., in press.</td>
</tr>
<tr>
<td>Habitat, diet:</td>
<td>Limnetic</td>
<td>Linkage map, admixture mapping</td>
<td>Yes</td>
<td>Yes</td>
<td>Assortative mating</td>
<td>Possibly</td>
<td>Raeymaeker et al., 2010; Berner et al., 2011; Kaeuffler et al., 2012.</td>
<td></td>
</tr>
<tr>
<td>Habitat, diet:</td>
<td>North American Lake</td>
<td>Shape</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>No</td>
<td>No</td>
<td>Reusch et al., 2001a; Reusch et al., 2001b; Milinski et al., 2005; Eizaguirre et al., 2009; Eizaguirre et al., 2012a; Eizaguirre et al., 2012b.</td>
</tr>
<tr>
<td>Stream</td>
<td>North American Lake</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Habitat, diet:</td>
<td>European Lake</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limnetic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benthic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Habitat, diet, predation:</td>
<td>Habitat, diet:</td>
<td>Parasite resistance</td>
<td>MHC loci</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Unknown</td>
<td>Olafsdottir and Snorrason 2009, Olafsdottir et al., 2007, Olafsdottir et al., 2006</td>
</tr>
<tr>
<td>North American Lake Whitefish</td>
<td>Lava</td>
<td>Armor, body shape</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Assortative mating</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Coregonus clupeaformis</td>
<td>Mud</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diet: Benthic (invertebrate feeder)</td>
<td>Habitat, diet:</td>
<td>Swimming behavior, physiology, morphology, body shape</td>
<td>Linkage maps, pQTL, eQTL, genome scans, gene expression</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes and No</td>
<td>Rogers and Bernatchez, 2007; Bernatchez et al., 2010; Renaut et al., 2012</td>
</tr>
<tr>
<td>Coregonus lavaretus</td>
<td>Limnetic (zooplankton feeder)</td>
<td>Morphology</td>
<td>Genome scans, gene expression</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Yes</td>
<td>Yes</td>
<td>Østbye et al., 2005; Østbye et al., 2006; Siwertsson et al., 2010; Kahilainen et al., 2007; Jeukens et al., 2009</td>
</tr>
<tr>
<td>Diet: Small benthic</td>
<td>Arctic char</td>
<td>Body size, reproductive timing</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Hindar and Jonsson 1993, Snorras et al., 1994, Corrigan et al., 2011</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Species, Genus, or Family</th>
<th>Peak shift</th>
<th>Adaptive trait divergence</th>
<th>Genetic architecture</th>
<th>Large effect genes?</th>
<th>Modular?</th>
<th>Evidence for reduced gene flow/RI at adaptive traits and/or genes?</th>
<th>Role for biogeography (secondary contact)?</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Sockeye salmon (Oncorhynchus nerka)</em></td>
<td>Habitat: Anadromous ↑↓ Landlocked (Kokanee / Kunimasu)</td>
<td>Body size</td>
<td>Gene expression</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>No</td>
<td>Taylor et al., 1996; Wood and Foote, 1996; Pavey et al., 2007; Pavey et al., 2010b</td>
</tr>
<tr>
<td></td>
<td>Habitat, predation, diet: Lake rearing, River rearing</td>
<td>Swimming behaviour, body shape</td>
<td>Gene expression</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Yes</td>
<td>No</td>
<td>Pavey et al., 2007; Pavey et al., 2010a</td>
</tr>
<tr>
<td></td>
<td>Habitat: stream spawner, beach spawner</td>
<td>Body shape, reproductive timing</td>
<td>Gene expression</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>No</td>
<td>Blair et al., 1993; Hendry et al., 2000; Quinn et al., 2001; Pavey et al., 2010b</td>
</tr>
<tr>
<td><em>Bahamas mosquitofish (Gambusia spp.)</em></td>
<td>Predation: high predation (G. affinis and G. heterochir) ↑↓ low predation (G. geiseri and G. hurtadoi)</td>
<td>Body shape</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Assortative mating</td>
<td>No</td>
<td>Schug et al., 1998; Langerhans et al., 2007; Langerhans (this issue)</td>
</tr>
<tr>
<td></td>
<td>Habitat, diet: parasitic species in several genera ↑↓ non-parasitic species in several genera</td>
<td>Body Size, timing of reproduction</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Salewski, 2003; Docker, 2009.</td>
</tr>
<tr>
<td></td>
<td>Habitat: anadromous (Entosphenus tridenatus) ↑↓ freshwater (E. macrostomus)</td>
<td>Body size, physiology, spawning behaviour</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Taylor et al., 2012.</td>
</tr>
<tr>
<td><em>Smelt Osmerus mordax</em></td>
<td>Habitat, diet: dwarf ↑↓ normal</td>
<td>Body size, morphology</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Yes (Pst)</td>
<td>Yes</td>
<td>Taylor and Bentzen, 1993; Berntchev, 1997; Saint- Laurent et al., 2003; Bradbury et al., 2010; Shaw and Curry, 2011.</td>
</tr>
</tbody>
</table>
2009). In all of these species, parallel evolution in response to the environment has occurred in multiple environments (Table 1). In some systems, the potential for ecological transitions was also influenced by allopatric genetic divergence shaped by the geography of past glaciations (Taylor and McPhail, 2000; Rogers et al., 2001). The phylogeography of north temperate fish diversity, among other taxa, was significantly influenced by the Quaternary expansions and contractions of glacial ice sheets (Petit et al., 2003; Rowe et al., 2004; Fraser and Bernatchez, 2005), whereby postglacial range shifts and colonization routes were variable by region, direction, and frequently implicated secondary contact between previously allopatric populations that inhabited ice-free refugia (Taylor and McPhail, 2000; Lyons, 2003). Thus, both geography and the environment may have influenced the adaptation and divergence of postglacial fish populations. Although the exact consequences of past biogeography on speciation remain unknown in postglacial fishes, the occurrence of allopatry and divergence-with-gene-flow in these systems presents unique opportunities to test predictions about the genetics of ecological speciation (Table 1). Postglacial fish species complexes are ideal systems to test these questions because, in addition to the tremendous adaptive divergence, many species complexes are also amenable to genomic techniques. Many postglacial fishes can be reared to maturity and used to generate crosses under controlled conditions (e.g., Lu and Bernatchez, 1998; Peichel et al., 2001) Their high fecundity can yield dozens to thousands of progeny in a single cross. DNA and RNA are easily isolated, even from the embryonic stage, and genomic resources can now be rapidly developed for just about any postglacial fish species (e.g., Kingsley et al., 2004). Overall, given the utility of recent advances in genomics for ecology and evolution (Stapley et al., 2010), we are arguably in the midst of the greatest revolution since the modern synthesis with respect to our understanding the speciation process, but many questions remain (see above, as well as Butlin et al., 2012).

3 Genomic Architecture

An important contribution to our understanding of speciation has been the ability to isolate and genetically map molecular polymorphisms across the genome (Rieseberg, 1998). Genetic mapping can contribute to an understanding of speciation in a myriad of ways, including close examination of complex quantitative trait loci (QTL) (e.g., Doerge, 2002; Mackay et al., 2009) and QTL mapping of gene expression profiles (expression QTL, or eQTL) (Whiteley et al., 2008; Majewski and Pastinen, 2011; Cubillos et al., 2012). Comparative mapping among closely related species or diverging populations can be used to infer the genetic changes that accompany or facilitate speciation (Whitkus, 1998; Rogers and Bernatchez, 2007; Baxter et al., 2008; Slotté et al., 2012), including the detection of genomic regions resistant to introgression following population divergence or speciation (e.g., Rieseberg et al., 1999; Rogers et al., 2001; Lexer et al., 2010; Chamberlain et al., 2011; Koeverts et al., 2012). Ultimately, it should be possible to identify “speciation” genes in ecological speciation with linkage mapping (Barrett and Hoekstra, 2011; Nosil and Schluter, 2011; Shaw and Mullen, 2011).

Genetic architecture refers to the pattern of genetic effects that build and control a given phenotypic character and its variational properties (Hansen, 2006). Collectively, the integrative application of genetic mapping and quantitative genetics (QTL mapping, linkage disequilibrium) provides insight into this “genetic architecture” under a genic view of speciation. We define genetic architecture as the number, location, and effects of genes underlying population divergence and speciation (Rieseberg, 1998; Rieseberg et al., 1999; Turelli et al., 2001). Just as architecture is both the process and product of design and construction, the genetic architecture of ecological speciation can result from both the process of population divergence and the product of adaptation and reproductive isolation. Of course, the “design” relies on the influence of drift, migration, and selection, with each evolutionary force having the potential to leave a distinct signature on the genome. For this reason, we refer to “genomic architecture” to reinforce the study of genetic architecture at the genome level. Genomic architecture may either promote or constrain adaptive divergence (Via and West, 2008; Nosil et al., 2008; Feder and Nosil, 2010) and can additionally bias the molecular variation presented to selection (McDonald et al., 2009). As such, genome-wide perspectives are increasingly integral for a full understanding of the functional genomic response to the evolutionary processes incurred by populations as they diverge, including the development of reproductive barriers. Four principal questions can therefore be addressed with respect to the components of genomic architecture during ecological speciation: (1) how does the number of genes associated with adaptive divergence affect the development and probability of RI, (2) does the location of such genes affect the formation of RI, (3) does the
effect size of these genes affect RI, and (4) what are the consequences of these elements of genomic architecture to the sequence of events, including gene flow and recombination, during population persistence and adaptive divergence?

There are several possible approaches to the study of genomic architecture in speciation. These approaches include the localization and dissection of complex traits (QTL mapping), genome scans to characterize candidate adaptive loci (so-called outlier loci (see Barrett and Hoekstra, 2011), admixture mapping in the wild (Willing et al., 2010; Winkler et al., 2010), genomic tests of introgressive hybridization (Warren et al., 2012; Arnold et al., 2012), as well as direct studies of RI (Sambatti et al., 2012; Sanchez-Guillen et al., 2012). Several excellent papers have already reviewed the merit and application of these methods for understanding issues in population genomics (Stinchcombe and Hoekstra, 2008; Dalziel et al., 2009; Storz and Wheat, 2010; Blackman, 2010; Elmer and Meyer, 2011, Rice et al., 2011). Consequently, we do not discuss these methodological approaches here. However, a consensus that has emerged is the notion that an understanding of genomic architecture in adaptation and speciation requires the integration of methods across multiple levels of biological organization. Integrative methods that bridge phenotype and genotype, as well as their effect on fitness, have a better chance of differentiating actual targets of natural selection as opposed to the genetic changes that occurred independently of ecological divergence, such as endogenous genetic barriers that may evolve independent of the environment (Bierne et al., 2011). Overall, this approach decreases the likelihood of falling into the pitfalls of the “molecular adaptationist paradigm” (Barrett and Hoekstra, 2011), but, given the difficulties involved in determining the selective pressures acting on a particular phenotype for which the genetic basis is known, this approach has yet to be broadly applied.

Finally, we note that recent advances in genomics, especially the ability to characterize genome-wide molecular polymorphisms for virtually any species, will be invaluable for future speciation studies. The fine-scale genome-wide sampling that is now possible will allow researchers to explore the genic concept of RI (Wu and Ting, 2004). In this genic view, RI develops only for particular genomic regions under selection, and gene flow continues to occur in unlinked neutral genomic regions unless complete (i.e., genome-wide) RI evolves (Wu, 2001). A central question in ecological speciation research is therefore the extent to which these “genomic islands” of divergence occur in diverging populations, and the extent to which these genomic islands facilitate or impede adaptation and the evolution of RI (Wu and Ting, 2004).

4 Genomic Architecture, Gene Flow, and Population Persistence

Populations must first persist in new environments if they are to speciate. Genetic variation can be associated with colonization success, with genotypic richness and phenotypic dissimilarity enhancing persistence in new environments (Crawford and Whitney, 2010; Ellers et al., 2011). Population persistence may also be facilitated by phenotypic plasticity (Yeh and Price, 2004; Fierst, 2011; Muschick et al., 2011; Hughes, 2012), especially with respect to differential gene expression (Pavey et al., 2010c). Such variation in addition to the modulation of behavioral, morphological or physiological traits could occur before any adaptive genetic evolution occurs (Price et al., 2003, Fitzpatrick, 2011). In this section, we discuss how genomic architecture may influence the ability of populations to persist following an environmental change.

In postglacial fishes, a fundamental aspect of the range shifts and expansions following the ice age was the extent to which these range shifts and expansions involved the colonization of novel environments, and hence the extent to which postglacial fishes did or did not successfully establish persistent populations in these novel environments. In at least three well-known cases (sockeye salmon, threespine stickleback and lake whitefish), post-glacial range expansion was associated with the colonization of several novel freshwater environments. It appears unlikely, however, that the degree of postglacial evolutionary change observed in these taxa (Table 1) is typical when considering all postglacial taxa (that include around 200 freshwater fish species, Scott and Crossman 1973). This begs the question. Why did some populations of certain species undergo adaptive divergence during the process of postglacial range shift? How do post-glacial populations persist in novel environments? We suggest that there may be a distinct role for genomic architecture to influence the propensity for postglacial fish populations to persist following colonization of a novel environment from their ancestral counterparts. Of course, we do not deny the possibility that other traits unrelated to genomic architecture (e.g., life-history variation and plasticity associated with ana-
domy) could have contributed to the post-colonization persistence of these species.

The degree of genomic modularity may be one aspect of genomic architecture associated with the probability of population persistence. Genetic architecture has typically been treated as a specific set of parameters (e.g., number, location, and effect) rather than a set of dynamic variables (Hansen, 2006). However, advances in sequencing have given rise to approaches that characterize biological processes at a genomic or “systems” level (Ideker et al., 2001; Ideker et al., 2011). One of the main findings of such systems biology approaches is that genomes appear to be modular and consist of specific, independent, standardized genomic regions that can be used to construct complex structures or phenotypes. Thus, modularity implies a functional independence among multiple subunits (e.g., genomic neighbourhoods, Butlin et al., 2012) of a larger entity (e.g., the genome). Modularity has long been predicted to be an important component in the genomic architecture of almost any complex functional entity. Numerous mechanisms that create modularity have been discussed previously, including chromosomal rearrangements (via translocations, inversions, or transposable elements) (Kirkpatrick and Barton, 2006; Feder and Nosil, 2009), duplications (Yeaman and Whitlock, 2011), and divergence hitchhiking (Via and West, 2008; Feder et al., 2012b). Modular variation, whereby the relative degree of connectivity varies within a species, is thought to evolve over long time scales through a process of selection that favors the integration of certain traits and the decoupling of others. In this way, variation in modularity may influence evolvability. Alternatively, conserved patterns of modularity may act to influence the rate and direction of evolution (Hill and Zhang, 2012; Parsons et al., 2012; Sanger et al., 2012).

There are two kinds of modularity that may have an impact on population persistence. The first is modularity between genotypes and phenotypes, whereby different traits may be influenced by different sets of genes, such that evolution may proceed independently without epistatic interference (Wagner and Altenberg, 1996; Pepper, 2000; Pepper, 2003; Hansen, 2006). For this reason, parallel evolution need not implicate parallel genes. Second, there may be modularity in the physical organization of genes. A genome’s architecture is modular if sets of genes with epistatic effects on fitness are physically clustered on the chromosomes (Pepper, 2003). Unlike modularity of the genotype-phenotype map, genomic modules do not affect how a given mutation will alter an individual’s phenotype or fitness (Pepper, 2000). Its only significance is that it affects the response to recombination, because crossover events are more likely between than within sets of epistatically interacting genes. For this reason, the frequency of gene exchange should be lower in genomic regions of low recombination. These regions are predicted to be more differentiated between diverging populations (Butlin, 2005), although selection at linked sites also predicts greater differentiation in regions of low recombination simply as a result of faster sorting of ancestral alleles even in the absence of gene flow (Nachman and Payseur, 2012). In addition, genomic regions of significantly increased recombination could be an artifact of variable population size over time (Johnston and Cutler, 2012), reinforcing the importance of integration between inferences of linkage disequilibrium in the wild and estimates from meiotic maps.

The degree of flexibility of constructing phenotypes is associated with the degree of modularity (Pepper, 2000; Ideker et al., 2001). Modularity may arise and become conserved in all descendant species, or may underlie multiple functionalities among species. Wu and Ting (2004) reinforced the importance of the modular nature of the genome for speciation. For example, if plant genomes are more modular than animal genomes, this may explain the prevalence of introgression and hybrid speciation (as opposed to hybrid inviability, hybrid sterility, or complete reproductive isolation) in plants. For example, three *Helianthus* species have arisen via hybridization between *H. petiolaris* and *H. annuus* (approximately 60,000 to 200,000 years before present), and Rieseberg et al. (2003) recreated many of the transgressive (i.e., non-intermediate) traits found in these hybrid species by backcrossing offspring of *H. annuus × H. petiolaris* crosses for two generations. There was a significant amount of genetic correlation among the loci associated with these traits, which suggests that modularity likely facilitated the origin of the three ancient hybrid species. Interestingly, the two hybrid species with the highest degree of positive genetic correlation among transgressive traits (i.e., the highest modularity) have multiple hybrid origins in nature, while the hybrid species with the lowest degree of positive genetic correlation (i.e., the lowest modularity) has a single hybrid origin. Hence, hybrid speciation may indeed reveal the extent of modularity in a genome (Wu and Ting, 2004).
5 Genomic Architecture, Gene Flow, and Adaptive Divergence

There is an implication in the previous section that modularity is important for adaptive divergence (and ecological speciation) when functionally-related genes are more likely to segregate together as a physically-linked module. An increasingly modular genomic architecture should reduce the likelihood of intrinsic incompatibilities forming upon recombination, but may still be subject to extrinsic RI depending on the environmental context of divergence. For example, previous models have also shown that such modularity is expected to be more effective in temporally variable environments (Pepper, 2000). Al-Shahrour et al. (2010) demonstrated that eukaryotic taxa have variable numbers and types of modules or “functional neighbourhoods” in their genomes. These modules may be akin to communities of co-adapted gene complexes, or multilocus genotypes favoured by selection (Palopoli and Wu, 1996; Hittinger et al., 2010; McPhee et al., 2012). One implication of this perspective is that the degree of modularity in genomic architecture may determine the propensity for introgression, and RI during adaptive divergence. Genomic architectures that circumvent incompatibilities (e.g., high modularity) would be amenable to adaptive divergence with gene flow, but would also be less likely to become fully reproductively isolated. On the other hand, genomic architectures that are prone to incompatibilities (e.g., low modularity) would not diverge with gene flow, but could develop strong RI during allopatry (Figure 1). This hypothesis is consistent with the notion that there is variation in genomic architecture (e.g., different degrees of modularity) among taxa. This variation may influence rates of adaptive divergence and formation of reproductive barriers.

Recent empirical evidence from divergent populations in two postglacial fishes is consistent with the notion that genes underlying adaptive divergence may be present in modules. A recent global study of threespine stickleback involved the complete sequencing of multiple genomes, finding that most of the outlier regions underlying adaptation to freshwater environments contained clusters of genes (Jones et al., 2012). When examining sliding windows of 50 Kb there were over 240 genomic regions associated with adaptive divergence. Remarkably, many of these regions were modular, exhibiting higher gene density than the rest of the genome. For example, a 66 Kb region of reported low recombination on chromosome IV representing one of the most highly differentiated regions of the stickleback genome has nine embedded genes, 3X the average gene density (1 protein-coding gene per 22 Kb, Jones et al., 2012). These results are consistent with population genetic patterns of linkage disequilibrium and population structure in the species (Hohenlohe et al., 2012). Other postglacial fishes reveal similar patterns of modularity. Lake whitefish species pairs exhibit distributions of eQTL associated with adaptive divergence in modular “eQTL hotspots” (Whiteley et al., 2008), also maintaining population genetic patterns of linkage disequilibrium among species pairs in association with the degree of ecological specialization (Renaut et al., 2012). Differential gene expression patterns between ecotypes, evolving in parallel among lakes, implicated a role for modularity at the hotspots (Derome et al., 2006), indicating that genetic constraints may depend on the gene position in metabolic networks, as well as network functions themselves (Cork and Purugganan, 2004).

An alternative to epistatic selection between these modules is multifarious selection, whereby multiple sites in the genome may be selected against via extrinsic postzygotic RI, although these two processes do not necessarily need to be mutually exclusive (Nosil et al., 2009; Feder et al., 2012a). Overall, the increasingly detailed genomic resolution has provided some evidence that the genomic architecture of ecological speciation may be modular, involving unique physical organization of genes on chromosomes, but the influence of modularity on RI remains to be tested.

Migration is a key evolutionary force associated with the genomic architecture of adaptive divergence. The likelihood and spatial scale of speciation, including ecological speciation, are influenced by the timing and the strength of gene flow (Kirkpatrick and Ravigne, 2002; Kisel and Barraclough, 2010; Smadja and Butlin, 2011). Yeaman and Whitlock (2011) modeled the evolution of genetic architectures under migration-selection-drift balance, providing strong theoretical support for the creation of “genomic islands” of divergence when populations are still in the process of adapting to novel environments while exchanging migrants. Variation in migration rates can affect the genomic architecture of adaptation (Fig. 1) (Yeaman and Whitlock, 2011). With higher levels of migration (i.e., above the “phenotypic critical migration threshold”, Yeaman and Whitlock, 2011) adaptive divergence is unlikely because frequent migration impedes even a perfectly adapted mutation to overcome the homogenizing effects of gene flow. Below the “allelic critical migration threshold”, divergence
Fig. 1  Migration-selection balance and the interaction between adaptation and genetic architecture. Vertical bars represent the chromosomes of two diverging populations (following Wu and Ting, 2004)

One population is the source population and is found in "environment 1" (dark grey box), and the other population is adapting to a novel environment, "environment 2" (light gray box). Horizontal dashes on the chromosomes represent mutations at genetic loci conferring adaptation to environment 2. Thicker dashes denote larger effect mutations or the stacking of smaller effect mutations at a single locus. Blue shading shows chromosomal regions with reduced gene flow, and darker blue shading denotes greater neutral divergence in these regions. The right side of the figure shows a non-modular (dispersed) genetic architecture, and the left shows a modular (clustered) genetic architecture. Migration rate ($m$) increases from top to bottom in the figure.

Effect of adaptation on genetic architecture
("islands" of divergence can occur without modularity)

Effect of genetic architecture on adaptation
(modularity allows divergence even with high $m$)

occurs across the entire genome, even for mutations of small effect, and no genomic islands arise – as the distribution of mutation effect sizes follows the predictions of Fisher (1930) and Orr (1998). For the intermediate range of migration rates (between the two critical thresholds), adaptation is strongly influenced by the availability of large-effect mutations. Small-effect mutations are swamped out unless they are tightly linked to
other mutations, or if they “stack” at a single genetic locus, thereby creating a large-effect genetic change. Thus, the consequence of migration for adaptive population divergence may in part be dependent on the existing genomic architecture (Fig. 1). Given the short timeframe for adaptive divergence in postglacial fishes (e.g., 10,000 generations in stickleback, 6000 generations in dwarf lake whitefish) and large effective population sizes, variability in genomic architectures is likely insufficient to permit competition among architectures and the evolution of clustered (i.e., modular) genetic architectures de novo in such a short time frame. However, it is possible that such modules preexisted as standing genomic variation. These results appear consistent with other theoretical findings that suggest that the conditions that may generate divergence hitchhiking around loci may be limited because of the corresponding reductions in gene flow that implicate genome-wide divergence (Feder and Nosil, 2010). Hence, it is more likely that adaptation was facilitated for species with preexisting modularity or genomic architectures characterized by close linkage (i.e., clustering) of the relevant adaptive genetic loci, with the strength of selection acting directly on a new mutation as the most significant predictor for the likelihood of establishment (Feder et al., 2012b). For species with dispersed genetic architectures, adaptation will be constrained by migration and gene flow, and populations colonizing novel environments will remain maladapted (Fig. 1). In addition, higher than predicted gene flow between diverging populations exhibiting strong RI may still exhibit genomic mosaics of divergence in cases of large effective population sizes (Sambatti et al., 2012). Empirical tests of these predictions are increasing in frequency, and the identification of genomic islands of diversification have been identified in a handful of postglacial fish taxa with histories of colonization to novel environments (Table 1). Future empirical tests should examine genetic architecture and adaptation among species or populations that have colonized novel environments, potentially using genome scans (see Fig. 2) to compare cases with or without ongoing gene flow, populations of different ages since colonization of the novel environment, and species with different genetic architectures (e.g., with or without chromosomal rearrangements). Populations of postglacial fishes make excellent candidates for future work in this area.

6 Genomic Architecture, Peak Shifts, and Mutation Effect Sizes

Transition of an evolving population to a new phenotypic optimum can shape the genomic architecture of populations in a predictable manner during ecological speciation. For example, genetic changes of large effect are expected to contribute infrequently to adaptations in new or changing environments (Fisher, 1930; Orr, 2005;
This follows from Fisher’s geometric model of adaptation whereby all mutations are considered to be pleiotropic. Consequently, under this assumption large effect mutations are more likely than small effect mutations to direct the population away from, rather than towards, a local adaptive peak (Fisher, 1930). Following a peak shift, the disadvantage of large effect genetic changes should decrease when a population is “far” from the optimal phenotype (Orr, 1998; Schluter et al., 2010). Rogers et al. (2012) recently tested whether adaptation to a farther phenotypic optimum involved a higher frequency of large effect genetic changes than adaptation to a nearer optimum in wild populations of threespine stickleback adapting to postglacial lake environments. Their test compared populations of approximately equal age that have adapted independently to two types of freshwater environments, representing near (intraguild predators Cottus asper present) and far (intraguild predators Cottus asper absent) morphological optima, beginning from the same ancestral state at the end of the last ice age (Fig. 3). They showed that postglacial colonization to sculpin-absent lakes, representing far-optimum environments, consisted of larger steps in the adaptive walk than populations adapting to sculpin-present lakes, representing the near optimum (Fig. 3). Pleiotropy, not just optimum overshoot, was the most likely mechanism contributing to the genomic architecture of adaptive divergence. This was the first demonstration that the theoretical expectations of the geometric model of adaptation result in predictable changes to genomic architecture in nature.

Importantly, their finding of a higher frequency of large effect genetic changes in the population with a farther optimum may explain why genes of large effect are only found to contribute to the genetics of adaptation between populations occasionally (Rogers et al., 2012). A number of studies on the genomic architecture of speciation, including in postglacial fishes, have found contrasting support for both the oligogenic model (few genes of large effect) and the infinitesimal model (many genes of small effect) (reviewed in Orr, 2005). These results suggest that the probability of large effect genes underlying the genetic architecture of adaptive walks depend on the existing genomic architecture and, importantly, the nature of the peak shift (Rogers et al.,

---

**Fig. 3** The consequences of adaptation to freshwater environments in threespine stickleback in the context of Fisher’s geometric model of adaptation

Trait means of the same ancestral marine population are indicated by a blue dot in both panels. Arrows indicate mutations, with the length of the arrow representative of the effect size. In the left panel the optimum (red dot) is nearer to the same ancestral population, whereas in the right panel the ancestral population begins the process of adaptation to a more distant optimum (red dot). Black circles are contours of equal fitness, with fitness higher inside the circle than outside (modified from Rogers et al. 2012). Here, QTL maps of shape and metric traits on linkage group 4 (Lg4) reinforce that large effect genetic changes are more likely to increase fitness when the optimum is far than when it is near, and that this difference is enhanced with a greater numbers of traits. Colored boxes show QTL, with darker shading representative of larger effect sizes.
2012). The consequences of peak shift on the likelihood of ecological speciation may be significant if large effect mutations are more likely to be associated with RI between diverging populations than small effect mutations. In addition, if population divergence is captured at a transient stage of a continuing process of adaptation to an unreached distant optimum, populations that have experienced more large effect mutations will have attained a greater distance from the ancestral state by that time. If population divergence involves adaptation to an optimum phenotype that continues to shift away from the ancestor gradually, only small effect mutations will be advantageous and the distribution of genetic effect sizes should not differ between present-day environments (Kopp and Hermisson, 2007). These studies reinforce the importance of how the evolution of genomic architecture is closely associated with consequences of adaptive peak shift, and may have important implications for the likelihood of RI in the process.

Such adaptive peak shift may result in the reduction of gene flow at the genes underlying adaptive phenotypes, or those genes genetically correlated with them, producing a “signature of selection” on the genomes of diverging populations during ecological speciation. Several examples from numerous taxa, including postglacial fishes, have supported the hypothesis that the genomic architecture of ecological speciation is associated with these signatures of selection in natural populations (Table 1). In some postglacial species complexes, genome scans have revealed that the genomic architecture of ecological speciation is genic, with QTL more likely to be associated with regions of heterogeneous genomic divergence than non-QTL markers (Rogers and Bernatchez, 2007; Renaut et al., 2012). Of course, a limitation of QTL approaches is whether these regions represent a single fixation event or several. For example, large effect QTL may be built up of multiple smaller effect, linked mutations that were fixed sequentially during adaptation (McGregor et al., 2007; Bickel et al., 2011). It may be possible to test these alternatives by measuring effect sizes in populations of different ages at different stages in the process of adapting to a similar environment.

7 Summary

Butlin et al. (2012) recently identified the genetics of speciation as one of the central outstanding questions in relation to the speciation process, namely that distinguishing between different speciation processes requires much greater knowledge of the genetic architecture of reproductive isolation than is currently available. Studies of the genomic architecture of postglacial fishes in ecological speciation will therefore continue to be an avenue of research contributing to the quest for the origin of species. Our objective was to investigate evidence for the role of genomic architecture in ecological speciation in postglacial fishes. Research on the number, location, effects, and interactions among the genes underlying adaptation reinforce the prediction that there are indeed consequences of genomic architecture for ecological speciation. The genomic architectures of postglacial fish taxa implicate modularity. The degree of genomic modularity may be associated with the probability of population persistence, so future ecological speciation studies should investigate the consequences of hybridization and introgression for RI at such modules associated with adaptive divergence. The genomic architecture of adaptation is also associated with migration and gene flow, with variation in migration rates predicted to affect the likelihood of heterogeneous genomic divergence. The genomic architecture of some postglacial fishes seems to support this prediction, and, given the very recent colonization of postglacial environments, adaptive divergence may have been facilitated by preexisting modularity or genomic architectures with clustering of adaptive loci (Fig. 1), and with the strength of selection influencing the likelihood of establishment (Feder et al., 2012b). A growing number of studies in postglacial fishes have elucidated isolation by adaptation, with genomic islands of divergence appearing to implicate extensive linkage disequilibrium in line with recent predictions that strong selection can maintain such patterns of genomic architecture (Via, 2012; Hohenlohe et al., 2012). These genic reductions of gene flow at adaptive traits clearly indicate the development of reproductive barriers, but an understanding of which outliers and the corresponding phenotypic mechanism contributing to RI is less clear. In postglacial fishes, the formation of reproductive barriers at adaptive genes could be the result of trait associations (Smadja and Butlin, 2011), pleiotropy (Rogers et al., 2012), or the multifarious evolution of mating preferences. Reduced hybrid performance may favour the evolution of assortative mating, but recent studies finding disadvantages in limnetic/benthic hybrid stickleback are an important reminder of the need for experiments in the wild (Taylor et al., 2012). We emphasize that RI is not always expected to evolve from the process of adaptive divergence, as it depends on the existing genomic architecture and the nature of the adaptive peak shift.
For all of these reasons, there is much work to be done on all fronts. On the genomics front, the constraint will not be sequencing molecular polymorphisms but rather obtaining information about the architecture itself. This is particularly evident in the limited number of lab-reared crosses between diverging species, including crosses that would never occur in nature, to infer elements of genomic architecture. More controlled crosses between and within ecotypes and populations are needed to fully understand elements of genomic architecture. Admixture mapping may provide flexibility in cases where controlled crosses are not possible (Winkler et al., 2010; Lexer et al., 2010; Malek et al., 2012). On the ecology front, more empirical tests comparing genomic architectures in cases with or without ongoing gene flow, or different stages of adaptation since colonization of novel environments, are needed to test these questions. Overall, studies of the genomic architecture of postglacial fishes will therefore continue to be an avenue of research contributing to the quest for the origin of species.

Acknowledgements We would like to thank the thorough reviews by the associate editor and two anonymous reviewers that greatly improved the manuscript. SMR is funded by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada and a New Faculty Award from Alberta Innovates Technology Futures.

References


Mayr E. 1963. Animal Species and Evolution. Cambridge:


Whiteley AR, Derome N, Rogers SM, St-Cyr J, Larroche J et al., 2008. The phenomenics and expression quantitative trait locus mapping of brain transciptomes regulating adaptive divergence in lake whitefish species pairs (Coregonus sp.). Genetics 180: 147–164.


