

## IDEA AND PERSPECTIVE

### A guide to the genomics of ecological speciation in natural animal populations

Amber M. Rice,\* Andreas Rudh, Hans Ellegren and Anna Qvarnström

Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden

\*Correspondence: E-mail: amber.rice@ebc.uu.se

#### Abstract

Interest in ecological speciation is growing, as evidence accumulates showing that natural selection can lead to rapid divergence between subpopulations. However, whether and how ecological divergence can lead to the buildup of reproductive isolation remains under debate. What is the relative importance of natural selection vs. neutral processes? How does adaptation selection generate reproductive isolation? Can ecological speciation occur despite homogenizing gene flow? These questions can be addressed using genomic approaches, and with the rapid development of genomic technology, will become more answerable in studies of wild populations than ever before. In this article, we identify open questions in ecological speciation theory and suggest useful genomic methods for addressing these questions in natural animal populations. We aim to provide a practical guide for ecologists interested in incorporating genomic methods into their research programs. An increased integration between ecological research and genomics has the potential to shed novel light on the origin of species.

#### Keywords

Gene expression, gene flow, genetics, genome scan, genomics, mapping approaches, natural selection, next-generation sequencing, reproductive isolation, ecological speciation.

Ecology Letters (2011) 14: 9–18

#### INTRODUCTION

Research on ecological speciation, i.e. the evolution of reproductive isolation between populations as a result of adaptation to divergent environments (Schluter 2001; Funk *et al.* 2006), has been dominated by observational and experimental studies of reproductive isolation and fitness of alternative phenotypes in divergent ecological environments (e.g. Schluter 1993; Pfennig & Rice 2007; Hendry *et al.* 2009). These studies usually focus on natural populations that may still exchange genes and have diverged relatively recently. By focusing on the initial stages of speciation, research on ecological speciation can reveal genetic changes and selective forces associated with the actual evolution of reproductive isolation instead of targeting species differences that may have evolved after the speciation process was completed (Coyne & Orr 2004). However, not all diverging populations or ecotypes will ultimately reach the final stages of speciation (Coyne & Orr 2004; Hendry 2009; Nosil *et al.* 2009b). In contrast, the genetics of speciation has traditionally been studied using laboratory crosses of long-diverged species in which hybrids express some form of intrinsic reproductive isolation, such as hybrid sterility or inviability (reviewed in Presgraves 2010). By complementing ecological speciation of natural populations with emerging genomic methods, the gap between these approaches can be bridged. This will provide a deeper understanding of how ecological speciation proceeds, from initial adaptation to different environments to complete reproductive isolation and incompatible genomes.

Over the last 15 years, the technology behind genomic studies has undergone rapid development, including the introduction of microarrays for high-throughput DNA and gene expression analyses, and, most importantly, the development of next-generation sequencing (NGS) allowing large-scale genome or transcriptome sequencing even for non-model organisms. The improved technology means that genomic methods are becoming more useful for addressing questions in wild populations (Ellegren & Sheldon 2008) and more affordable. The aim of this article is to give a broad overview of how ecologists and evolutionary biologists, even those who have little previous experience with molecular techniques, might use these methods to address crucial questions related to ecological speciation. We discuss three major aspects of ecological speciation: population divergence through ecological selection, adaptation and reproductive isolation and divergence despite gene exchange. The examples focus on wild populations of sexually reproducing animals. While several of the scientific questions and genomic approaches discussed are relevant for studying ecological speciation in a wider range of organisms, there are several additional important questions, e.g. concerning genome duplication in plants, which we will not bring to attention here.

We have included several points of entry into the genomic methods for studying ecological speciation. First, we briefly review open theoretical questions in ecological speciation research and provide examples of how combinations of genomic methods can be used to address these questions. Second, Table 1 lists these questions, in the same order as in the text, and provides corresponding genomics questions and suggested methods. Table 2 follows up Table 1 and provides more detailed information on each method by outlining the types of data or materials that are needed, some of the advantages and disadvantages and references for further reading. By using open theoretical questions as a starting point, we aim to emphasize the importance of clearly formulating which hypotheses are being tested before embarking on a genomic analysis.

#### POPULATION DIVERGENCE THROUGH ECOLOGICAL SELECTION

For ecological speciation to occur, divergent natural selection should be the main cause of divergence. A combination of population genomics, quantitative genetics, gene expression studies and ecological data can be used to determine: (1) whether divergence between populations is mainly caused by natural selection vs. neutral processes and (2) whether exposure to similar environments leads to the evolution of similar adaptations. Given these goals, a necessary pre-requisite is to have information about environmental differences or divergent traits. From that starting point, two non-mutually exclusive paths may be taken to compare the populations of interest: a genome-wide approach (left path in Fig. 1) and a targeted functional approach (right path in Fig. 1).

The genome-wide approach (Fig. 1) starts with comparisons of DNA or RNA of individuals from different populations to identify differentiated loci or loci showing signatures of divergent selection. The initial step for scientists choosing to follow this pathway is to develop a set of marker loci, such as amplified fragment length polymorphisms (AFLPs) or single nucleotide polymorphisms (SNPs) (Box 1). Individuals should then be sampled from each of the divergent environments or populations. A genome scan, in which many individuals are genotyped at many loci across the genome, is then performed. When the genotyping is performed, outlier analyses can be used to identify loci putatively undergoing and/or closely linked to loci under divergent selection. Loci exhibiting higher levels of differentiation (commonly expressed as  $F_{ST}$ ) than expected under neutrality between populations sampled from different ecological environments are identified as linked to regions potentially under selection. Using genotypes of SNPs generated from sequenced restriction site associated DNA (RAD) tags (Box 1), Hohenlohe *et al.* (2010) performed an outlier analysis to identify regions under divergent selection in populations of freshwater and marine sticklebacks. They identified several regions with elevated differentiation between marine and freshwater stickleback populations, consistent with divergent selection. Moreover, because Hohenlohe *et al.* compared multiple populations within each environment, they addressed our second question as well (Table 1), identifying instances of parallel evolution in which multiple freshwater populations showed elevated divergence at the same sites.

Genome scans may also be used for cline analyses, which can test for signatures of selection in divergent, but adjacent environments. For this type of analysis, the most useful markers would be SNPs that are fixed among individuals from the same environment, but variable between individuals sampled from different environments. The DNA samples used for cline analyses must come from individuals sampled across the two environments and the transition zone between them. Once the genome scan is performed, allele frequencies at each marker locus (usually expressed as the proportion of alleles from one environment) are plotted against sampling location (usually expressed as distance from one end of the sampled transect), and a cline is fitted to the data. Cline analyses have been used frequently for investigating selection in hybrid zones (e.g. Teeter *et al.* 2008), but usually use a limited number of marker loci. Ideally, a large number of markers that are spread evenly throughout the genome should be genotyped to identify markers showing narrow cline width. Marker loci with the narrowest cline widths may be linked to loci with little genetic introgression between the two populations (i.e. foreign

**Table 1** Ecological speciation questions with corresponding questions from a genomics perspective

Ecological speciation question	Corresponding genomics question	Useful method(s)
<b>Population divergence through ecological selection</b>		
To what extent does divergence result from divergent selection and adaptation vs. neutral processes?	<ul style="list-style-type: none"> <li>• Are loci that are more divergent than expected under neutral evolution present?</li> <li>• Do these loci also show an association with phenotypic traits under divergent selection?</li> <li>• How many genes show signatures of divergent selection?</li> </ul>	<i>Genome scans</i> – outlier and cline analyses <i>Gene expression</i> – outlier and cline analyses <i>NGS</i> – marker development, outlier analyses and digital gene expression <i>QTL/association/admixture mapping</i> – linking divergent phenotypes with genetic loci <i>Targeted sequencing</i> – candidate gene approach
Does exposure to similar environments lead to the evolution of similar adaptations?	<ul style="list-style-type: none"> <li>• Do the same genes show signatures of selection in individuals sampled from <i>different</i> populations but exposed to <i>similar</i> ecological environments?</li> <li>• Are the same genes responsible for parallel phenotypic changes?</li> </ul>	Replicated across populations: <i>Genome scans</i> – outlier and cline analyses <i>Gene expression</i> – outlier and cline analyses <i>NGS</i> – marker development, outlier analyses and digital gene expression <i>QTL/association/admixture mapping</i> – linking divergent phenotypes with genetic loci <i>Targeted sequencing</i> – candidate gene approach
<b>Adaptation and reproductive isolation</b>		
What is the relative importance of different adaptations for reproductive isolation?	How does the amount of introgression between populations (e.g. in populations forming ring species) correlate with the amount of divergence in different traits?	<i>Association/admixture mapping</i> – linking genomic regions to divergent adaptations and isolation traits <i>Genome scans</i> – measure introgression (outlier and cline analyses)
Are the same adaptations important for reproductive isolation over the time course of speciation?	Are the same genes coding for traits underlying ecological adaptations and sexual isolation eventually also causing incompatibilities at later stages of divergence?	Using ring species or systems with clines: <i>QTL/association/admixture mapping</i> – linking genomic regions to divergent adaptations and isolation traits <i>Genome scans</i> – outlier, cline or admixture analyses
Do parallel adaptations achieved through different genetic pathways have a similar impact on the buildup of reproductive isolation?	What is the relationship between variation in the genetic architecture underlying ecological adaptation and the degree of reproductive isolation?	Using systems with parallel evolution: <i>QTL/association/admixture mapping</i> – linking genomic regions to divergent adaptations <i>Targeted sequencing/gene expression</i> – dissecting the genetic basis of adaptations
How does ecological divergence lead to genetic incompatibilities?	What are the identity and genomic locations of genes causing genetic incompatibility?	<i>QTL mapping</i> – linking genetic interactions with phenotypes showing ecologically dependent hybrid dysfunction
<b>Divergence despite gene exchange</b>		
What is the breadth and location of natural hybrid zones?	<ul style="list-style-type: none"> <li>• Where does the transition between species-specific alleles occur?</li> <li>• How steep is the cline for different marker loci?</li> <li>• Which populations show the greatest degree of admixture?</li> </ul>	<i>Genome scans</i> – cline or admixture analyses
How much realized gene flow occurs between populations across the genome?	<ul style="list-style-type: none"> <li>• What is the frequency of heterospecific genes in natural populations?</li> <li>• What is the cline breadth for multiple loci at secondary contact zones?</li> <li>• Do some loci show asymmetric clines or geographically displaced cline centres?</li> </ul>	<i>Genome scans</i> – cline or admixture analyses
How does the relationship between genes that have undergone adaptive divergence and genes associated with reproductive isolation allow the buildup of linkage disequilibrium?	<ul style="list-style-type: none"> <li>• Do outlier loci map to regions associated with reproductive isolation?</li> <li>• Where are genes for reproductive isolation located in the genome relative to genes for adaptive divergence?</li> <li>• Do such genes tend to be located in chromosomal inversions, on sex chromosomes, or in other regions of reduced recombination?</li> </ul>	<i>Association/admixture mapping/QTL mapping</i> – linking genomic regions to divergent adaptations and isolation traits <i>Linkage mapping</i> – presence of rearrangements

NGS, next-generation sequencing; QTL, quantitative trait locus.

alleles are strongly selected against at those loci, or they may be positioned at genomic sites with reduced intraspecific recombination rates for other reasons). After genotyping 39 SNP markers spread throughout the genome in mice collected across the European house mouse hybrid zone, Teeter *et al.* (2008) were able to identify three markers showing especially narrow cline widths; loci near these markers may be under selection. If populations in the two environments under study are not adjacent, making it difficult to obtain the samples necessary for a cline analysis, regions showing reduced introgression, which can indicate linkage to a locus under selection, may still be identified by quantifying the number of alleles from the opposite population that are present at a given locus (e.g. Sætre *et al.* 2003). Additionally, Gompert & Buerkle (2009) have recently proposed another method for identifying regions potentially under selection, which estimates introgression at single loci relative to the genome as a whole. This method may also be used when the divergent environments are not directly adjacent. Those loci with introgression patterns differing substantially from the genome-wide expectation are identified as potentially under selection.

Although a useful way to highlight genomic regions of interest, on their own, outlier and cline analyses cannot rule out population bottlenecks as causes for enhanced divergence, nor can they reveal the function of genes identified as targets of selection. Several other factors that affect the ability to conclusively detect signatures of divergent selection should also be considered, including population history, the strength of selection and the number of markers used (see Butlin 2010 for a discussion of relevant issues). Moreover, whether any individual outlier locus is actually under selection remains unclear until additional information is obtained, e.g. through quantitative trait locus (QTL) mapping (Via & West 2008). These tests are therefore most informative when they are combined with additional analyses (Fig. 2).

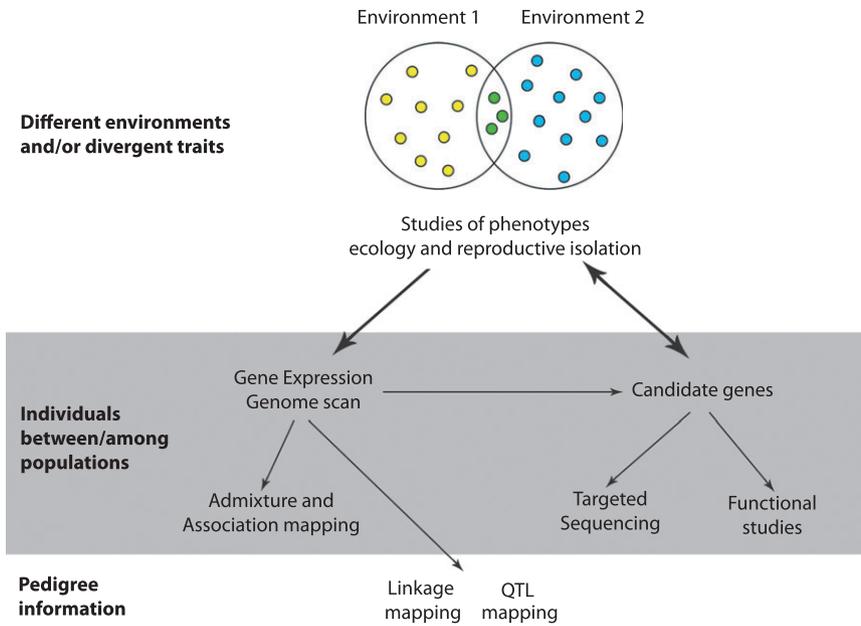
Adaptation to different environments may not always proceed through divergence in protein-coding genes; it may also occur through changes in levels of gene expression (Table 2; Fay & Wittkopp 2008). Gene expression may be measured using high-throughput microarrays (e.g. Derome *et al.* 2006; Hutter *et al.* 2008), by focusing on

**Table 2** Genomic methods for studying ecological speciation

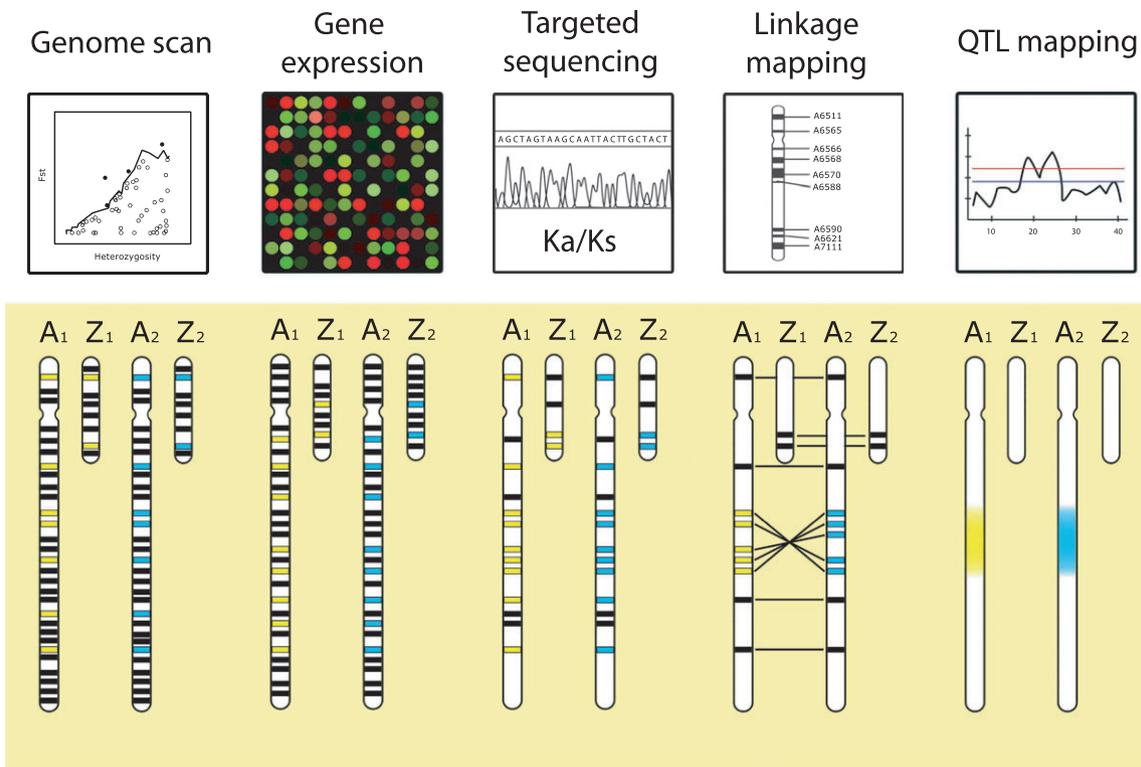
Genomic method	Required data/materials	Advantages	Disadvantages	References*
Next-generation sequencing (NGS)	<ul style="list-style-type: none"> <li>Prepared RNA or DNA libraries from individuals collected in different environments</li> <li>Optional: closely related reference genome for assembly</li> </ul>	<ul style="list-style-type: none"> <li>Massive amount of data produced which can have many uses (e.g. marker development, designing microarrays and sequence analysis)</li> <li>Possible to focus sequencing on particular parts of the genome</li> <li>Data can be used in many different analyses (e.g. outlier analyses, cline analyses, mapping and measuring introgression)</li> <li>Markers can be developed with little prior genetic information</li> <li>Microarray technology becoming more accessible in many species</li> <li>Allows identification of gene functional categories</li> <li>Digital gene expression approaches do not require <i>a priori</i> candidate genes (no microarray required). Produces information (variation in expression levels) that is not explicitly addressed by other methods</li> </ul>	<ul style="list-style-type: none"> <li>Still costly for large population studies</li> <li>Trade-off between length and number of DNA fragments sequenced</li> <li>Bioinformatics processing of data is very challenging</li> </ul>	<p>Hudson (2008), Wang <i>et al.</i> (2009), Wheat (2010), Baird <i>et al.</i> (2008), Torres <i>et al.</i> (2008), Schwarz <i>et al.</i> (2009) and Hohenlohe <i>et al.</i> (2010)</p>
Genome scan	<ul style="list-style-type: none"> <li>Genome-wide set of marker loci (e.g. AFLPs, SNPs)</li> <li>DNA samples of individuals from different environments</li> </ul>	<ul style="list-style-type: none"> <li>Data can be used in many different analyses (e.g. outlier analyses, cline analyses, mapping and measuring introgression)</li> <li>Markers can be developed with little prior genetic information</li> </ul>	<ul style="list-style-type: none"> <li>Without mapping information, assumption of even genome coverage cannot be evaluated</li> <li>Testing of multiple loci requires adjustment of significance threshold</li> </ul>	<p>Nosil <i>et al.</i> (2009a), Butlin (2010), Sætre <i>et al.</i> (2003), Teeter <i>et al.</i> (2008), Via &amp; West (2008), Gompert &amp; Buerkle (2009) and Hohenlohe <i>et al.</i> (2010)</p>
Gene expression	<ul style="list-style-type: none"> <li>RNA of individuals from different environments</li> <li>If using qPCR: primer sequences</li> </ul>	<ul style="list-style-type: none"> <li>Microarray technology becoming more accessible in many species</li> <li>Allows identification of gene functional categories</li> <li>Digital gene expression approaches do not require <i>a priori</i> candidate genes (no microarray required). Produces information (variation in expression levels) that is not explicitly addressed by other methods</li> </ul>	<ul style="list-style-type: none"> <li>Large amount of data must be corrected for multiple testing</li> <li>Microarrays can be expensive to develop in non-model organisms</li> <li>Possibility of nucleotide differences leading to a false signature of gene expression divergence</li> </ul>	<p>Bar-Or <i>et al.</i> (2007), Wang <i>et al.</i> (2009), Shapiro <i>et al.</i> (2004), Abzhanov <i>et al.</i> (2006), Derome <i>et al.</i> (2006), Steiner <i>et al.</i> (2007), Hutter <i>et al.</i> (2008), Torres <i>et al.</i> (2008), Baxter <i>et al.</i> (2010) and Matsumura <i>et al.</i> (2010)</p>
Targeted sequencing	<ul style="list-style-type: none"> <li>DNA or RNA of individuals from different environments</li> <li>Primer sequences</li> </ul>	<ul style="list-style-type: none"> <li>Many methods for detecting selection may be easily applied</li> </ul>	<p>The importance of unsequenced regions may be overlooked</p>	<p>Shapiro <i>et al.</i> (2004), Linnen <i>et al.</i> (2009) and Baxter <i>et al.</i> (2010)</p>
Admixture mapping	<ul style="list-style-type: none"> <li>Genome-wide marker set (species- or population-specific markers)</li> <li>DNA samples of individuals from admixed populations</li> <li>Phenotypic information</li> </ul>	<ul style="list-style-type: none"> <li>Useful in hybrid zones or strongly structured populations</li> <li>High power for detecting genotype-phenotype associations</li> </ul>	<p>Regions with significant associations may be large due to extensive linkage disequilibrium</p>	<p>Smith &amp; O'Brien (2005), Buerkle &amp; Lexer (2008), Stinchcombe &amp; Hoekstra (2008), Rieseberg <i>et al.</i> (1999) and Baird <i>et al.</i> (2008)</p>
Linkage mapping	<ul style="list-style-type: none"> <li>Genome-wide marker set</li> <li>DNA from pedigreed individuals</li> </ul>	<ul style="list-style-type: none"> <li>Provides physical location information</li> <li>Inversions (potential speciation hotspots) may be identified by comparing maps between divergent lineages</li> </ul>	<p>Linkage can be difficult to detect between distant loci or when few offspring are available</p>	<p>Stinchcombe &amp; Hoekstra (2008), Backström <i>et al.</i> (2008b), Backström <i>et al.</i> (2010) and Ball <i>et al.</i> (2010)</p>
Association mapping (also known as linkage disequilibrium mapping)	<ul style="list-style-type: none"> <li>Genome-wide marker set</li> <li>Phenotypic information</li> <li>DNA samples of individuals from different environments</li> </ul>	<ul style="list-style-type: none"> <li>Can also be used on candidate genes</li> </ul>	<p>If linkage disequilibrium is limited, a large number of markers may be needed to detect associations</p>	<p>Stinchcombe &amp; Hoekstra (2008) and Baxter <i>et al.</i> (2010)</p>
QTL mapping	<ul style="list-style-type: none"> <li>Genome-wide marker set</li> <li>DNA from pedigreed individuals</li> <li>Phenotypic information</li> </ul>	<ul style="list-style-type: none"> <li>Can also be used for comparative mapping</li> <li>Identification of epistasis is possible</li> </ul>	<ul style="list-style-type: none"> <li>Power to detect genes of small effect may be limited</li> <li>Requires fairly extensive pedigree when applied to natural populations, which can be difficult to obtain</li> </ul>	<p>Malmberg &amp; Mauricio (2005), Stinchcombe &amp; Hoekstra (2008), Oka <i>et al.</i> (2007) and Steiner <i>et al.</i> (2007)</p>

AFLPs, amplified fragment length polymorphisms; SNPs, single nucleotide polymorphisms; qPCR, quantitative PCR; QTL, quantitative trait locus.

\*General review articles listed in bold-faced type; empirical papers listed in regular type.



**Figure 1** Multiple paths to identify and locate the genes underlying divergently selected phenotypes that are involved in reproductive isolation. Bold text with accompanying white or grey fields describes necessary samples/information. At the top is the pre-requisite for such a study – information about either different environmental conditions (the large circles) or divergent traits (small yellow, blue and green circles within environments). Below are two, non-mutually exclusive paths for identifying genes involved in ecological speciation: one that takes a more genome-wide approach (on the left) and one that takes a more targeted functional approach (on the right). Arrows connecting the different method types indicate some possible follow-up analyses to provide an increasingly detailed picture of the genetic basis of ecological speciation.



**Figure 2** An illustration of one way to combine a variety of genomic methods to identify genomic regions important in ecological speciation. The figures at the top illustrate the genomic methods employed, whereas the chromosomes at the bottom are a simplified illustration of the genomes (A, autosomes; Z, sex chromosomes) of individuals from two different environments (1 and 2). The black bars indicate the loci examined by each approach, whereas the colours (yellow or blue) indicate the loci that show significant signatures of interest. Moving from the left to the right, an increasing amount of information is required for each method, which results in an increasing ability to tie specific genomic regions to specific traits. Both the genome scan and gene expression microarrays examine many loci spread throughout the genome and potentially find a large number of genes that show greater differentiation than expected under neutrality. Once a set of candidate loci are identified, targeted sequencing may be used on surrounding genes, allowing additional tests for signatures of selection. With the addition of a pedigree, linkage mapping becomes possible, and gene order (synteny) can be compared (shown here by lines connecting orthologous genes) between individuals from different environments. Regions where an inversion has occurred (shown here with crossing lines) are especially likely to harbour speciation genes. With the further addition of phenotypic data, quantitative trait locus (QTL) mapping can link genomic locations with specific phenotypes.

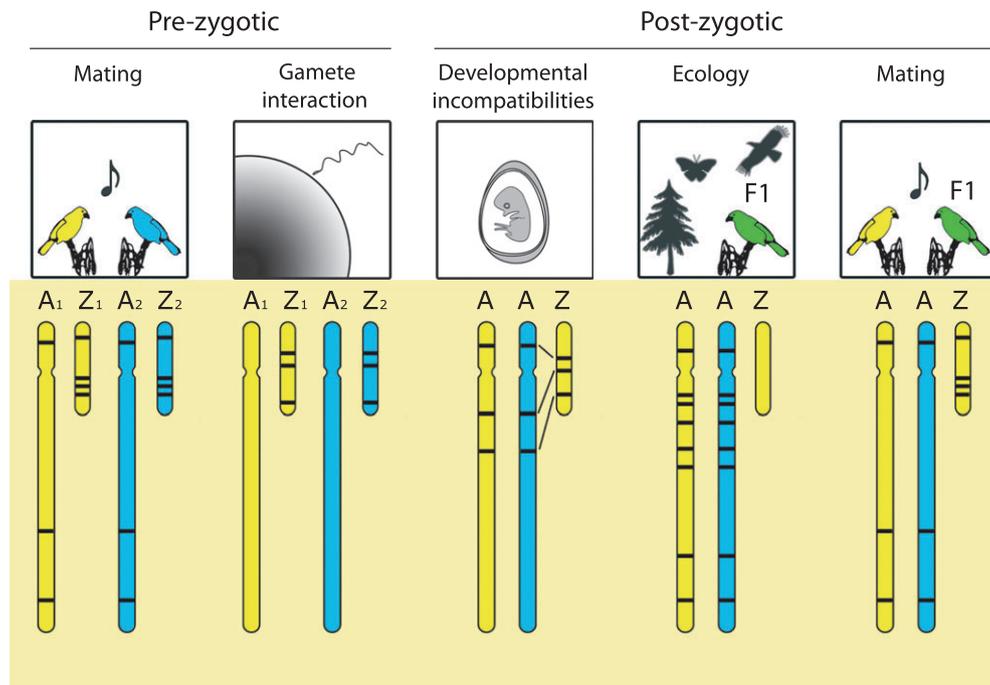
specific genes with quantitative PCR (qPCR; e.g. Baxter *et al.* 2010) or reverse transcription PCR (RT-PCR; e.g. Steiner *et al.* 2007), or, as is increasingly important, using NGS technology to sequence expressed sequence tags (ESTs) and determine the number of mRNA transcripts produced (digital gene expression; Box 2). Microarray-based studies are designed to compare gene expression levels at thousands of loci between two conditions (e.g. two treatments, two sexes and two species), and in theory, could be easily tailored to studies of ecological speciation and divergence between populations experiencing different environments (Abzhanov *et al.* 2006). A microarray developed for a closely related species could be used, but signals of low expression levels may also arise due to sequence divergence between the two species (Bar-Or *et al.* 2007). Alternatively, microarrays for non-model organisms can be constructed, but the costs for developing such an array may be prohibitive. By directly counting the number of transcripts produced, the newer sequence-based approaches (i.e. digital gene expression; Box 2) avoid some of the problems associated with hybridization-based comparisons of gene expression (i.e. microarrays) between divergent populations or species (Bar-Or *et al.* 2007). The digital gene expression approach has several other advantages over microarrays, including a greater ability to detect very low or very high levels of expression, and no limitation to a predefined set of genes (Wang *et al.* 2009; Wheat 2010). These features especially should be quite useful for comparisons between recently diverged populations, such as those often featured in ecological speciation research. Limitations and challenges do exist, however, not the least of which is difficulties linking expression divergence to gene function in systems without reference genomes (see Wang 2009; Wheat 2010 for discussion of additional considerations).

Expression profiles potentially under divergent selection may be identified using methods similar to the outlier analyses described before, with the prediction that loci showing significant differences in expression levels between environments may be under divergent selection (e.g. Derome *et al.* 2006). Indeed, when multiple populations are found in two different ecological environments, comparisons between environments can be replicated. In this way, loci can be identified that show repeated expression divergence between environments (Table 1). The existence of such parallel changes, such as those found in whitefish ecotypes (Derome *et al.* 2006), strongly

implicates ecologically based divergent selection as the cause of gene expression differences between environments.

To determine whether genetic loci or gene expression profiles showing signatures of divergent selection are actually associated with divergent phenotypes, a variety of mapping approaches can be used (Table 2). An important role for selection in driving divergence (Table 1) is upheld when loci identified as potentially under divergent selection by outlier or cline analyses are also linked to a divergent phenotype by mapping approaches. Because these mapping approaches may also be used to link traits contributing to reproductive isolation to particular loci, we describe these methods in more detail below. Alternatively, the targeted functional approach (Fig. 1) may be employed to establish links between genotypes and divergent phenotypes. This pathway requires that candidate loci potentially under divergent selection are already identified either through the genome-wide approach or through a combination of ecological knowledge about divergent phenotypes and detailed information on gene function from other systems. There are many well-studied examples of ecologically driven divergence in phenotypes. Beak sizes in various species and populations of Darwin's finches are specialized for feeding on different seed types (Hendry *et al.* 2006 and references therein) and two host races of *Rhagoletis pomonella* have phenology adapted either to a life on hawthorn or apple trees (Filchak *et al.* 2000). Colour patterns may also diverge, as seen in races of *Heliconius melpomene* butterflies which mimic different geographic races of *Heliconius erato* (Jiggins *et al.* 2004; Baxter *et al.* 2010) and in two subspecies of oldfield mice selected for camouflage in either dark or light habitats (Vignieri *et al.* 2010). These examples are by no means exhaustive, and many other types of traits may also diverge. By incorporating diverse fields such as molecular genetics and developmental biology, the targeted functional approach employs methods to establish functional links including (but not limited to) targeted sequencing of specific loci (Table 2; Shapiro *et al.* 2004), spatial patterns of gene expression (Shapiro *et al.* 2004; Abzhanov *et al.* 2006) or even phenotypic measurements of laboratory-bred animals with and without functional copies of specific alleles at the locus of interest (Linnen *et al.* 2009).

In some pioneering cases, the genomic basis of divergent ecological adaptations has already been revealed. In two subspecies of oldfield mice (*Peromyscus polionotus*), coat



**Figure 3** Expected genomic locations for genes involved in various forms of reproductive isolation. This example is from the point of view of two bird populations or species adapted to two different environments, indicated by the colours blue and yellow. The figures at the top illustrate two forms of pre-zygotic reproductive isolation and three forms of post-zygotic isolation. Below the figures are simplified representations of chromosomes in the genome (A, autosomal chromosomes; Z, sex chromosomes). The black bars on the chromosomes indicate hypothetical locations for genes involved in each form of reproductive isolation. For the two pre-zygotic forms of reproductive isolation, we depict chromosomes from two interacting individuals from environments 1 and 2 ( $A_1$  and  $Z_1$  vs.  $A_2$  and  $Z_2$ ). For the post-zygotic forms of reproductive isolation, a 'hybrid' individual (shown in green) has already been formed, hence we depict the chromosomes from such an individual (a female). Note that the autosomes come from both parental populations but the sex chromosome comes only from one parental population because we do not show the W chromosome. From the left: during mating interactions, theory predicts that most of the genes involved should be on the sex chromosomes, which is also the case with interactions between gametes. Developmental incompatibilities are likely to arise by interactions between genes from different parental populations, often between genes on the sex chromosome and the autosomes (interactions depicted with black lines). Performance in a given ecological niche is predicted to involve many genes, generally located on the autosomes. Finally, genes for mating in the 'hybrid' individual should be located in the same places as for the parental populations. However, the 'hybrid' may have an intermediate, and likely disfavoured, phenotype due to the contribution of genes from both parental populations.

colour has diverged in response to selection for camouflage on two different substrates: dark mainland habitat or light sand dune coastal habitat (Vignieri *et al.* 2010). Steiner *et al.* (2007) used QTL mapping to identify and locate three candidate genes explaining substantial parts of the pigmentation differences between mice adapted to alternate substrates. After adding in qPCR and RT-PCR analysis of the two candidate genes with the largest effect size (*Agouti* and *Mc1r*), Steiner *et al.* (2007) concluded that divergence in coat colour is caused by an interaction between a mutation in the coding region of *Mc1r* and changes in *Agouti* gene expression. Subsequent studies confirmed that variation at *Mc1r* and at a cis-regulatory region of *Agouti* contribute to pigmentation differentiation in *Peromyscus* mice (Mullen & Hoekstra 2008; Mullen *et al.* 2009). These studies thus illustrate the potential importance of changes in both genes and gene expression in adaptation to ecological selection.

By using genomic methods in natural populations, the general importance of divergent selection in driving speciation can be better understood. Outlier and cline analyses can be used to pinpoint loci showing signatures of divergent selection, at the levels of both gene coding regions and gene expression. When combined with ecological and phenotypic information, mapping approaches may be used to tie these loci to specific targets of selection, providing a picture of how selection at the phenotypic level is manifested at the genetic level. When such tests are performed in multiple populations across different environments, a greater understanding of how selection interacts with existing genetic backgrounds can be reached. Once such studies are performed in many natural populations, we can begin to draw general conclusions about the importance of selection in driving population divergence at both phenotypic and genetic levels.

#### ADAPTATION AND REPRODUCTIVE ISOLATION

Establishing a link between adaptation to divergent environments and the evolution of reproductive isolation is the central challenge for studies of ecological speciation. A comparative study across divergent taxa found a strong positive correlation between greater reproductive isolation and ecological divergence (Funk *et al.* 2006), which is consistent with the promotion of reproductive isolation by ecological adaptation. However, how reproductive isolation builds up over time as adaptation to divergent environments proceeds is not well understood. For example, whether or not ecological adaptations may themselves result in genetic incompatibilities remains largely unknown. Genomic approaches can be used to reveal links between adaptive divergence and reproductive isolation by addressing central questions (Table 1) such as: (1) What is the relative importance of different adaptations for reproductive isolation? (2) Are the same adaptations important for reproductive isolation over the time course of speciation? (3) Do parallel adaptations achieved through different genetic pathways have a similar impact on the buildup of reproductive isolation? and (4) How does ecological divergence lead to genetic incompatibilities?

Reproductive isolation typically arises through a combination of isolation traits acting at different stages of the reproductive cycle (Fig. 3). There are theoretical expectations about the genomic locations of genes coding for different forms of reproductive isolation (Fig. 3), e.g. whether they are expected to be sex-linked or not (Qvarnström & Bailey 2009). While the genomic regions associated with reproductive isolation have been mapped in several systems (reviewed in Qvarnström & Bailey 2009), it is still difficult to draw general conclusions, especially in natural populations. Additionally, reproductive isolation as a result of ecological adaptation may occur over very short time scales, as in host shifts in host-specific insects (Matsubayashi *et al.* 2010 and references therein), as well as in a more gradual fashion such as when directional selection on quantitative traits drives the evolution of reproductive isolation. The relative importance of different isolation traits (Fig. 3) at various points in the time course of speciation remains unclear.

Because mapping approaches allow links to be made between divergent phenotypes and genotypes, these methods will be particularly useful for investigating the relationship between ecological adaptation and reproductive isolation at the genetic level. By adding phenotypic data to genetic marker data, association mapping (Table 2; Stinchcombe & Hoekstra 2008 and references therein; Baxter *et al.* 2010) or admixture mapping (Table 2; Smith & O'Brien 2005; Baird *et al.* 2008; Buerkle & Lexer 2008) can link genomic regions to specific traits, such as traits under divergent selection and/or reproductive isolation traits. These methods are quite similar; both require large marker sets (Box 1) and seek to establish a statistical association between a genotype at a particular locus and a phenotype of interest (i.e. a trait involved in ecological divergence and/or reproductive isolation). The marker sets used have different characteristics, however. Admixture mapping requires markers that are fixed (or nearly so) within each environment (or population or species) but variable between, such as those used in cline analyses. Association mapping usually utilizes markers that are variable within a species (or population or environment). Both of these approaches rely on linkage disequilibrium between the marker locus and the gene underlying the phenotype, and thus, power to detect phenotype–genotype associations is largely dependent on marker density and coverage (Buerkle & Lexer 2008). Because admixture mapping takes advantage of the extensive linkage disequilibrium formed by gene flow between two species or structured populations, fewer markers are required to detect associations

between genotypes and phenotypes than for association mapping (Buerkle & Lexer 2008). For the same reason, admixture mapping may be an especially useful analysis for studies of ecological speciation.

We are not aware of any studies in animals that have used admixture mapping to identify genetic regions associated with reproductive isolation traits. In *Helianthus* sunflowers, however, Rieseberg *et al.* (1999) were able to map 16 genomic regions associated with pollen sterility in hybrids. Recently, admixture mapping was used to identify three loci associated with variation in lateral plate phenotypes between marine and freshwater threespine sticklebacks (Baird *et al.* 2008). The dense marker set Baird *et al.* (2008) developed from sequenced RAD tags (Box 1) helped to make such fine-scale mapping possible. Although prior mapping efforts and the availability of an annotated reference genome in this system enhanced the authors' ability to draw detailed conclusions, an analysis performed without using the reference genome successfully identified candidate genes for subsequent follow-up studies (Baird *et al.* 2008). This approach thus has great potential for studies of ecological speciation in natural populations (Buerkle & Lexer 2008), and as genomic technology makes developing dense marker sets easier and cheaper, we predict an increase in the use of admixture mapping for linking divergent phenotypes and isolation traits to genotypes.

Given enough time, genetic drift in isolated population pairs will cause genetic incompatibilities and intrinsic post-zygotic isolation. This occurs when isolated populations independently accumulate genetic changes that do not have negative effects on fitness within each population but cause inviability or infertility when they interact in hybrids formed between individuals from the two populations (Dobzhansky 1940; Muller 1940; Fig. 3). Ecological adaptation itself may also lead to the evolution of intrinsic isolation, through two non-exclusive pathways. First, ecological adaptation may promote neutral divergence by genetic drift, which may then result in genetic incompatibilities when hybrids are formed. The logic behind this argument is based on the assumption that if ecological adaptation leads to the evolution of isolating mechanisms, gene flow between populations is reduced and neutral loci will also diverge (reviewed in Gavrillets 2004; Nosil *et al.* 2009a). Second, the genes that are themselves undergoing adaptive divergence, or those hitchhiking along (Via & West 2008), may *per se* cause incompatibilities in hybrids. In both of these cases, interactions between divergent loci in hybrids formed between individuals from different populations may result in negative fitness effects. Adaptation may thus promote intrinsic incompatibilities, either indirectly or directly. How common it is for ecological adaptation to lead to the evolution of intrinsic isolation is unknown; very few examples have been described of loci undergoing adaptive divergence that are also involved in genetic incompatibilities, and we are not aware of any examples from animal systems. Whether or not most of the genes currently known to underlie cases of hybrid dysfunction originate as side effects of adaptive evolution is under debate (Orr *et al.* 2007; Presgraves 2010).

If pedigree information is available, QTL mapping (Fig. 1; Table 2; Steiner *et al.* 2007; Ellegren & Sheldon 2008; Stinchcombe & Hoekstra 2008) may be undertaken. QTL mapping may be used not only to identify the numbers and locations of loci associated with particular phenotypes, but may also reveal interacting loci (reviewed in Malmberg & Mauricio 2005). This approach is thus especially powerful for identifying loci involved in both genetic incompatibilities and ecological adaptation. If a set of interacting loci is revealed by QTL mapping to affect the expression of intrinsic reproductive isolation, and one or both of the interacting loci also affects a trait under divergent selection, then ecological adaptation *per se* may cause genetic incompatibilities. Once QTL have been identified, more detailed investigations of candidate loci may follow, perhaps using some of the methods from the targeted functional approach (Fig. 1). For instance, in house mice, Oka *et al.* (2007) used QTL mapping to identify two autosomal regions that interact with the X-chromosome to cause abnormally shaped sperm, a major contributor to sterility in backcrossed males. Using laboratory crosses, they subsequently replaced the chromosomes carrying those regions with chromosomes from the same genetic background as the X-chromosome, and were able to partially restore fertility (Oka *et al.* 2007).

In combination with these genomic mapping methods, certain types of study systems may become especially valuable for understanding the link between ecological divergence and reproductive isolation, and how these links may change over the time course of speciation. First, using geographic distance as a proxy for time, ring species (e.g. Irwin *et al.* 2001) provide a unique opportunity to study different stages in the speciation process. In ring species, there is a continuous distribution of nearby individuals interbreeding, but when individuals from the extremities of the distribution meet (where the ring closes), they express marked reproductive isolation. The relative importance of different adaptations for reproductive isolation could be investigated in such a system by correlating levels of introgression and phenotypic divergence in different traits between different pairs of populations while controlling for geographic distance. Furthermore, this type of system could also be used to investigate the relative importance of different adaptations and isolation traits at various times along the path to speciation. Traits involved in ecological adaptations and reproductive isolation between population pairs could be mapped, and by incorporating variation in geographic distance between the population pairs, the traits important for isolation at

different times could be identified (Table 1). Systems exhibiting a geographic cline in ecologically important traits may also be used in this context.

Cases of parallel evolution may be used to learn more about the relationship between adaptation to divergent ecological environments and the buildup of reproductive isolation (Table 1). Populations in the same ecological environment may have responded to similar selection pressures by parallel phenotypic changes; however, the underlying genetic changes may differ (e.g. Hoekstra *et al.* 2006). If this occurs, the relationship between adaptation to the environment and the buildup of reproductive isolation may vary depending on the genetic pathway taken to achieve the adaptation (Table 1). To examine this process, researchers could dissect the genetic changes underlying parallel adaptations in systems with several population pairs that have diverged as the result of the same selective pressures. This could be performed using mapping approaches, potentially combined with subsequent candidate gene analyses (Fig. 1; e.g. Steiner *et al.* 2007). Additionally, detailed tests of the degree of reproductive isolation present between populations in the same and in different environments should be performed (e.g. McKinnon *et al.* 2004). The basic assumption of ecological speciation is that reproductive isolation should be stronger between population experiencing different ecological environments than between pairs of populations experiencing similar ecological environments. If the degree of reproductive isolation between population pairs inhabiting similar environments varies in relation to differences in the genetic architecture underlying their phenotypically similar ecological adaptations, this would suggest that the pathway by which adaptation proceeds also could influence the link between adaptation and the buildup of reproductive isolation (i.e. mutation-order speciation; Schluter 2009). Although the feasibility of these kinds of analyses would be highly dependent on the number of available population pairs, which are naturally restricted, they could begin to shed new light on the links between ecological adaptation and the buildup of reproductive isolation.

Linking divergent ecological adaptations to the evolution of reproductive isolation is the central challenge for studies of ecological speciation, and genomic methods show great potential for revealing these links. Various mapping approaches will allow researchers to determine the relationships between the genetics underlying both divergent traits and reproductive isolation traits. QTL mapping, which can uncover genetic interactions affecting phenotypes, will be especially useful in understanding whether adaptation may lead directly to genetic incompatibilities. Also, when mapping approaches are used in combination with special types of systems, such as ring species or cases of parallel evolution, questions about the development of reproductive isolation over the time course of speciation and the influence of genetic architecture (Table 1) can be addressed. A deeper understanding of these central issues will become especially important for understanding whether and how ecological speciation may proceed when diverging populations continue to exchange genes.

#### DIVERGENCE DESPITE GENE EXCHANGE

Classical theoretical models on the balance between selection and immigration have found that strong natural selection is needed to achieve local adaptation even when there are relatively low levels of immigration. The homogenizing effect of gene flow is therefore traditionally seen as a major obstacle for ecological speciation, especially in sympatry and/or for the completion of the speciation process at secondary contact, because recombination can destroy the association between genes involved in adaptive divergence, pre-zygotic isolation and post-zygotic isolation (Felsenstein 1981). The relationship between these genes therefore has a large influence on whether speciation can proceed to completion in the face of gene flow. By allowing us to gain a deeper understanding of both the amount of realized gene flow occurring between populations and the physical relationships between genes involved in adaptive divergence and reproductive isolation, genomic methods can be used to investigate how ecological speciation may proceed in the face of gene flow. In this context, three crucial questions are: (1) What is the breadth and location of natural hybrid zones?, (2) How much realized gene flow occurs between populations across the genome? and (3) How does the relationship between genes that have undergone adaptive divergence and genes associated with reproductive isolation allow the buildup of linkage disequilibrium?

The 'problem' of gene flow can be avoided, or its homogenizing effects mitigated, if reproductive isolation follows directly as a side effect of ecological divergence (e.g. as a consequence of specific habitat preferences). Alternatively, if sufficient divergence occurred in allopatry, then complete reproductive isolation at secondary contact or strong enough selection against hybrids may exist to allow for the buildup of further isolation. The location and width of hybrid zones can reveal important information on the role of ecology in speciation. The main expectation is that the relative fitness of the two species will depend on the environment and that a narrow hybrid zone will be formed in association with a change in environmental conditions.

The location and breadth of hybrid zones is traditionally studied using geographic cline or admixture analyses (e.g. Teeter *et al.* 2008; Gompert *et al.* 2010). The same techniques can also be used to estimate variation in levels of gene flow across the genome by estimating introgression at multiple loci or by performing cline analyses at multiple loci. Again, for these types of analyses, a set of markers that are fixed within

one species or environment but variable between is required (Box 1). Regions of the genome showing a greater number of alleles from the opposite species or a cline with a shallow slope are likely experiencing relatively higher levels of gene flow than other regions. Indeed, if the cline for one marker is geographically displaced from the average location of the cline centre, one explanation may be that the marker is linked to a locus that is selectively favoured in the foreign genetic background. Such a pattern was found in a house mouse hybrid zone, in which alleles from one species appear to be spreading into the other (Teeter *et al.* 2008).

Another key prediction of ecological speciation is that the fitness of hybrids should depend on the environment (e.g. Hatfield & Schluter 1999; Pfennig & Rice 2007). It is therefore possible that hybrids have superior fitness as compared to the parental species in certain environments. The bounded hybrid-superiority model assumes that this is the case for the central part of a hybrid zone but there may also be temporal changes in environmental conditions and therefore in the relative fitness of hybrids. In Darwin's finches, selection on medium ground finches (*Geospiza fortis*), cactus finches (*Geospiza scandens*) and their hybrids changes depending on fluctuations in environmental conditions (e.g. Grant & Grant 2002). Given that hybrid fitness is mainly determined by the ability to utilize the environment to ensure survival and successful reproduction (i.e. there are few or no genetic incompatibilities), hybrids could form new species (Mavarez *et al.* 2006). Thus, natural hybrid zones make it possible not only to investigate the role of ecology in reducing gene flow between parental species but also in promoting hybrid speciation.

The completion of ecological speciation in spite of gene exchange may also be possible if mechanisms are present causing linkage disequilibrium between genes coding for traits involved in pre- and post-mating isolation such that species integrity can be maintained. This may be achieved through a dual function of traits (i.e. traits with a role in both pre- and post-zygotic isolation; Slatkin 1982; Gavrilets 2004; van Doorn *et al.* 2009). A dual function can arise through two different ways. First, the same genes (or complex of genes) may code for one phenotypic trait with a dual function (Slatkin 1982). Such traits are often referred to as 'magic traits' (Gavrilets 2004). Possible examples of magic traits can, for example, be found in sticklebacks and in Neotropical poison frogs. Body size in sticklebacks – which is associated with ecological performance (Schluter 1993), also acts as an assortative mating cue (e.g. McKinnon *et al.* 2004). Likewise, colour in poison frogs serves both as a warning signal to predators (Noonan & Comeault 2009) and as a mate choice cue (Reynolds & Fitzpatrick 2007). Second, the same gene (or complex of genes) could code for a trait used in sexual isolation and at the same time influence another trait causing selection against hybrids (Slatkin 1982). A possible example is the major histocompatibility complex (MHC). These genes play a role in immune function and are expected to be adapted to local parasite fauna. Moreover, the MHC may also be used as a cue in mate choice, potentially through an effect on odour (e.g. Milinski *et al.* 2005). Through choosing a mate based on MHC genotype, offspring may gain the best adapted complex of MHC alleles for the local parasite environment, while hybrids may harbour too much MHC diversity for an effective immune response (Eizaguirre *et al.* 2009). However, the most common pattern is probably that different traits are associated with pre- and post-zygotic isolation, placing mechanisms influencing the buildup of linkage disequilibrium of speciation genes at the heart of models on sympatric speciation (e.g. Felsenstein 1981; van Doorn *et al.* 2009) and of models on reinforcement (reviewed by Servedio & Noor 2003). The mapping approaches we discuss (i.e. admixture, association and QTL mapping) can be used to link genomic regions with pre- and post-zygotic isolation traits. If both pre- and post-zygotic isolation traits are linked to the same genomic regions, then dual function traits may explain why speciation can occur in the presence of gene exchange.

The homogenizing effects of gene flow may also be mitigated if the relevant genes are sheltered against interspecific recombination through their location in the genome. Possible sheltered genomic sites include physical linkage on the same chromosome (e.g. the sex chromosome; Sæther *et al.* 2007) or positions in the genome where there have recently been some chromosomal rearrangements (Fig. 2; Noor *et al.* 2001). The basic idea is that sites that are sheltered against interspecific recombination may function as 'speciation islands' where genes with species-specific fitness effects can become clustered (Via & West 2008) and where incompatibilities can build up (Navarro & Barton 2003). Recent research on ecological speciation indeed indicates that when gene flow between populations is present, most of the genome will not diverge between populations; instead, divergence will occur only at loci under selection and at closely linked loci (Via & West 2008; Thibert-Plante & Hendry 2009; but see Michel *et al.* 2010).

To determine whether the relationship between genes involved in adaptation and reproductive isolation allows the buildup of linkage disequilibrium (Table 1), linkage mapping can be used in combination with the previously discussed mapping approaches which can link genotypes to phenotypes, such as association, admixture and QTL mapping (which requires a linkage map). Linkage mapping traces the inheritance of specific alleles through a pedigree, and identifies alleles that tend to be inherited together, signifying reduced recombination and likely physical linkage. Linkage maps thus provide location information for genes in terms of recombination distance. Information about which genes underlie ecological adaptation and reproductive isolation, discovered using a mapping approach, could then be combined with the physical location information from the linkage map, shedding light on whether a mechanism allowing linkage disequilibrium,

such as multiple functions mapped to a single gene, may explain divergence in the face of gene flow. Moreover, linkage maps for populations of interest can be compared, and chromosomal inversions – potential hotspots for speciation genes – can be identified (Table 1; Fig. 2; Noor *et al.* 2001). If inversions are not present, close physical linkage can similarly reduce recombination risk. Recent research on *Ficedula* flycatcher hybrid zones provides an example of how a combination of methods (reviewed in Qvarnström *et al.* 2010; Sætre & Sæther 2010) can be used to understand divergence in the face of gene flow. Based on an analysis of SNP markers, the Z-chromosome (i.e. one of the two sex chromosomes in birds) appears to be subject to less genetic introgression between pied flycatchers (*Ficedula hypoleuca*) and collared flycatchers (*Ficedula albicollis*) as compared to the autosomes (Sætre *et al.* 2003), pointing to the potential presence of speciation genes on the Z-chromosome. This was confirmed when the same marker set was used in combination with detailed behavioural studies to establish that the Z-chromosome harbours genes involved in sexual isolation and genetic incompatibility (Sæther *et al.* 2007). However, a comparison of Z-chromosome linkage maps between the two species found no evidence of chromosomal rearrangements (Backström *et al.* 2010), suggesting that some other mechanism may be maintaining linkage disequilibrium between genes involved in sexual isolation and genetic incompatibilities.

To understand how speciation can occur in the face of gene flow between diverging populations, knowledge of the genetic architecture underlying traits involved in both ecological adaptation and reproductive isolation is required. Although this has been possible on a gross level, as in the *Ficedula* flycatcher example where such genes reside on the Z-chromosome (Sætre *et al.* 2003; Sæther *et al.* 2007), developing genomic methods are making fine-scale mapping a very real possibility in natural populations. Moreover, by making it possible to measure levels of gene flow across the genome, the dense marker sets that can be developed (Box 1) will allow insight into how gene flow and selection interact to cause genomic divergence (e.g. Via & West 2008; Nosil *et al.* 2009a; Michel *et al.* 2010) and eventual speciation.

#### CONCLUSIONS: THE MOST POWERFUL APPROACH IS TO COMBINE METHODS

We have provided a practical guide for how existing genomic technology can be used to study ecological speciation. As illustrated in Fig. 2, some methods require more prior information than others. Both the genome scan and gene expression microarrays examine many loci spread throughout the genome and potentially find a large number of genes that show greater differentiation than expected under neutrality. Once a set of candidate loci are identified, targeted sequencing may be used on surrounding genes, allowing additional tests for signatures of selection. With the addition of a pedigree, linkage mapping becomes possible, and gene order (synteny) can be compared between individuals from different environments. With the further addition of phenotypic data, QTL mapping can finally link genomic locations with specific phenotypes. Many exciting questions become answerable as a combination of genomic and ecological information is built up. There are emerging examples of studies on speciation using genomic methods in combination both with each other (e.g. Shapiro *et al.* 2004; Via & West 2008) and with ecological details (e.g. Chamberlain *et al.* 2009). For example, in two host races of pea aphids, outlier loci identified using a genome scan were found to cluster around mapped QTL loci associated with reproductive isolation (Via & West 2008). Not only did this analysis allow a link to be made between loci under divergent selection and reproductive isolation (Table 1), but it also provided a better understanding of the extent to which introgression was reduced around selected loci (Via & West 2008). By taking advantage of the fact that the butterfly subspecies *Heliconius cydno alithea* has a wing colour polymorphism that mirrors the divergence seen between two related species, *Heliconius cydno galanthus* and *Heliconius paebinus*, Chamberlain *et al.* (2009) were able to examine the genetics underlying adaptation and reproductive isolation at what may be the earliest stages of speciation (Table 1). These studies show that genomic technology is steadily becoming a more powerful tool in ecological studies.

Until recently, studying the genetic basis of ecological speciation has been unfeasible in most natural animal populations. While many quality observational and experimental studies have documented the importance of adaptation to divergent environments for the evolution of reproductive isolation, a lack of available genetic resources, not to mention the high cost of developing such resources, has made a deeper understanding of this process on a genetic level nearly impossible in most systems. Thus, many questions dealing with how gene flow can be overcome in ecological speciation and how divergence in ecologically important traits is related to the buildup of reproductive isolation have remained unanswered. Recent developments in genomic technology are now bringing such questions into reach for scientists working with natural animal populations. NGS technology has made it possible to develop dense marker sets in non-model organisms at a fraction of the time and cost required by traditional methods (Box 1). NGS technology even now continues to advance, as technology for genome sequencing from a single DNA molecule is under development (Rusk 2009), and costs continue to decrease. This advancement in technology will facilitate the use of a variety of mapping methods to locate genomic regions under selection and link genomic regions to divergent traits and reproductive isolation. Our comprehension of the importance of pleiotropy and genetic incompatibilities resulting from divergent

adaptation will be especially advanced. In order for ecological speciation research to maximize the possible benefits of this rapid development in genomic technology, however, a solid foundation of ecological studies and clear hypotheses must not be neglected. Combining detailed ecological and phenotypic information with these new genomic insights will bring a deeper understanding of the process of ecological speciation than has yet been possible.

#### Box 1 Marker development

Efficient development of molecular markers is important for most molecular studies of ecological speciation. AFLP markers can be developed in systems with no prior genetic information to search for signatures of selection in the genome (reviewed in Nosil *et al.* 2009a). However, marker loci developed using this technique cannot be easily linked to specific genomic locations or to gene functional information, rendering follow-up studies difficult. Furthermore, these markers are not suitable for comparing results from different studies. SNP markers are now gaining in use, not least because they are codominant and can be found at high density across the genome (Wheat 2010). Schwarz *et al.* (2009) sequenced ESTs to generate a large SNP marker set for future use in *Rhagoletis*, and recently, Baird *et al.* (2008) showed that large numbers of SNPs can also be developed by sequencing RAD tags. Indeed, with this approach, the SNP markers may be identified at the same time as the population samples of interest are genotyped (Hohenlohe *et al.* 2010). The density of SNP markers can be adjusted by using different restriction enzymes, and the method may prove to be very useful for generating large numbers of SNPs in a short amount of time, even in systems with little prior genetic information (Baird *et al.* 2008). If a reference genome from a closely related species is also available, however, marker loci may be more easily linked to genomic locations and possible functions (e.g. Backström *et al.* 2008a).

#### Box 2 Next-generation sequencing

Next-generation sequencing refers to high-throughput technologies that produce vast amounts of sequence data in a much shorter time and at much lower costs than the capillary sequencing methods previously in use. Three sequencing technologies are currently available, from three different companies: 454 pyrosequencing from Roche (Branford, CT, USA), SOLiD sequencing from Applied Biosystems (Carlsbad, CA, USA) and Solexa sequencing from Illumina (San Diego, CA, USA). Each of these technologies uses slightly different methods to produce sequence data, and they differ in the length and number of the sequence fragments (reads) produced (Hudson 2008). When designing a sequencing project, it is important to consider these differences, because they will impact the depth of coverage (the number of reads of a given DNA sequence) that can be achieved [see Wheat (2010) and references therein for a discussion of this and other important considerations in designing a NGS project]. Because ecological speciation research is concerned with variation within and between populations, whole genome sequencing will rarely, if ever, be the goal of such projects. NGS technology can be used for other more relevant applications, however. Nearly all of the methods we discuss in the text require developing a set of polymorphic markers (Box 1). NGS technology can be applied to marker development (Hudson 2008; Wheat 2010), through sequencing ESTs or RAD tags. Sequenced RAD tags have successfully been applied to ecological speciation in marine and freshwater sticklebacks (Box 1; Baird *et al.* 2008; Hohenlohe *et al.* 2010), while 454 sequencing of ESTs in *R. pomonella* identified several candidate speciation genes and provided over 40 000 markers for future studies (Box 1; Schwarz *et al.* 2009). NGS technology can also be used when studying gene expression, either in the design of microarrays (Wheat 2010) or in directly counting the number of mRNA transcripts produced by each gene (digital gene expression). The latter approach can be performed using several methods [e.g. RNA-Seq (Wang *et al.* 2009), massively parallel signature sequencing (MPSS; Torres *et al.* 2008), high-throughput super serial analysis of gene expression (HT-Super SAGE; Matsumura *et al.* 2010)]. All of these methods are sequencing approaches, meaning that gene expression can be measured in a species even if no relevant microarray exists. Each method has other potential sources of error or bias (Torres *et al.* 2008; Wang *et al.* 2009; Matsumura *et al.* 2010; Wheat 2010), but because they do not rely on the existence of a closely related model organism, these new NGS based approaches may be especially useful for studying divergence in natural populations.

## GLOSSARY

Admixture mapping and association mapping	Methods that identify statistical associations between a genotype at a particular locus and a phenotype of interest.
Amplified fragment length polymorphisms (AFLPs)	Molecular markers developed with restriction enzymes and PCR. Markers are dominant and biallelic.
Candidate gene	A gene suspected to be involved in the expression of a trait of interest.
cDNA	Complementary DNA, synthesized from messenger RNA (mRNA) transcripts. Contains only protein-coding regions.
Ecological speciation	The evolution of reproductive isolation between populations as a result of adaptation to divergent environments.
Expressed sequence tags (ESTs)	Short DNA sequences generated by sequencing cDNA.
High-throughput super serial analysis of gene expression (HT-Super SAGE)	Digital gene expression quantification method with barcoding of transcripts from different individuals, allowing transcriptome sequencing of many individuals in a single NGS run.
Linkage disequilibrium (LD)	Non-random association of alleles, sometimes, but not always, due to physical linkage.
Linkage mapping	Method that uses pedigree information to identify alleles that tend to be inherited together, producing a map of gene physical locations.
Microarray	Multiplex method using arrayed series of probes (oligonucleotide spots) which hybridize with a sample of targets (cDNA or RNA) later quantified by labelled targets (e.g. with fluorophores).
Massively parallel signature sequencing (MPSS)	Digital gene expression quantification method in which cDNA is fragmented, sequenced and the number of transcripts from each gene is quantified.
Next-generation sequencing (NGS)	High-throughput, massively parallel sequencing methods.
Quantitative PCR (qPCR)	PCR method that allows simultaneous amplification of a DNA template and quantification with the use of fluorescent dyes.
Quantitative trait locus (QTL)	Gene or genomic region that affects the expression of a trait showing continuous variation.
Quantitative trait locus (QTL) mapping	Approach for estimating location and effect sizes of QTLs. Requires genome-wide polymorphic marker loci, linkage map and variation in trait(s) of interest.
Restriction site associated DNA (RAD) markers	Polymorphic genetic markers obtained from DNA sequences flanking restriction enzyme sites (RAD tags).
Reverse transcription PCR (RT-PCR)	Reverse transcription of RNA to cDNA followed by regular or quantitative PCR amplification.
RNA-Seq	Digital gene expression quantification method in which cDNA is sequenced using NGS, aligned to a reference genome (or assembled de novo) and the number of transcripts from each gene is quantified.
Single nucleotide polymorphism (SNP)	DNA sequence variation by differences in a single nucleotide.
Speciation islands	Sites in the genome that are sheltered against interspecific recombination where genes with species-specific fitness effects can become clustered and where incompatibilities can build up.
Transcriptome	All RNA molecules produced at any one time, including mRNA, rRNA, tRNA and non-coding RNA.

## ACKNOWLEDGEMENTS

We thank Richard Bailey and Arild Husby for helpful discussion and comments; three anonymous referees, Sara Via, Jordi Bascompte and Arne Mooers for suggestions that greatly improved the manuscript; and the Swedish Research Council (AQ) and the European Science Foundation (AQ) for financial support.

## REFERENCES

- Abzhanov, A., Kuo, W.P., Hartmann, C., Grant, B.R., Grant, P.R. & Tabin, C.J. (2006). The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature*, 442, 563–567.
- Backström, N., Fagerberg, S. & Ellegren, H. (2008a). Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. *Mol. Ecol.*, 17, 964–980.
- Backström, N., Karaiskou, N., Leder, E.H., Gustafsson, L., Primmer, C.R., Qvarnström, A. *et al.* (2008b). A gene-based genetic linkage map of the collared flycatcher (*Ficedula albicollis*) reveals extensive synteny and gene-order conservation during 100 million years of avian evolution. *Genetics*, 179, 1479–1495.
- Backström, N., Palkopoulou, E., Qvarnström, A. & Ellegren, H. (2010). No evidence for Z-chromosome rearrangements between the pied and collared flycatcher as judged by gene-based comparative genetic maps. *Mol. Ecol.*, 19, 3394–3405.
- Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A. *et al.* (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, 3, e3376.
- Ball, A.D., Stapley, J., Dawson, D.A., Birkhead, T.R., Burke, T. & Slate, J. (2010). A comparison of SNPs and microsatellites as linkage mapping markers: lessons from the zebra finch (*Taeniopygia guttata*). *BMC Genomics*, 11, 218.
- Bar-Or, C., Czosnek, H. & Koltai, H. (2007). Cross-species microarray hybridizations: a developing tool for studying species diversity. *Trends Genet.*, 23, 200–207.
- Baxter, S.W., Nadeau, N.J., Maroja, L.S., Wilkinson, P., Counterman, B.A., Dawson, A. *et al.* (2010). Genomic hotspots for adaptation: the population genetics of Müllerian mimicry in the *Heliconius melponene* clade. *PLoS Genet.*, 6, e1000794.
- Buerkle, C.A. & Lexer, C. (2008). Admixture as the basis for genetic mapping. *Trends Ecol. Evol.*, 23, 686–694.
- Butlin, R.K. (2010). Population genomics and speciation. *Genetica*, 138, 409–418.
- Chamberlain, N.L., Hill, R.I., Kapan, D.D., Gilbert, L.E. & Kronforst, M.R. (2009). Polymorphic butterfly reveals the missing link in ecological speciation. *Science*, 326, 847–850.
- Coyne, J.A. & Orr, H.A. (2004). *Speciation*. Sinauer Associates, Sunderland, MA.
- Derome, N., Duchesne, P. & Bernatchez, L. (2006). Parallelism in gene transcription among sympatric lake whitefish (*Coregonus clupeaformis* Mitchell) ecotypes. *Mol. Ecol.*, 15, 1239–1249.
- Dobzhansky, T. (1940). Speciation as a stage in evolutionary divergence. *Am. Nat.*, 74, 312–321.
- van Doorn, G.S., Edelaar, P. & Weissing, F.J. (2009). On the origin of species by natural and sexual selection. *Science*, 326, 1704–1707.
- Eizaguirre, C., Lenz, T.L., Traulsen, A. & Milinski, M. (2009). Speciation accelerated and stabilized by pleiotropic major histocompatibility complex immunogenes. *Ecol. Lett.*, 12, 5–12.
- Ellegren, H. & Sheldon, B.C. (2008). Genetic basis of fitness differences in natural populations. *Nature*, 452, 169–175.
- Fay, J.C. & Wittkopp, P.J. (2008). Evaluating the role of natural selection in the evolution of gene regulation. *Heredity*, 100, 191–199.
- Felsenstein, J. (1981). Skepticism towards Santa Rosalia, or why are there so few kinds of animals. *Evolution*, 35, 124–138.
- Filchak, K.E., Roethele, J.B. & Feder, J.L. (2000). Natural selection and sympatric divergence in the apple maggot *Rhagoletis pomonella*. *Nature*, 407, 739–742.
- Funk, D.J., Nosil, P. & Etges, W.J. (2006). Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proc. Natl. Acad. Sci. USA*, 103, 3209–3213.
- Gavrilets, S. (2004). *Fitness Landscapes and the Origin of Species*. Princeton University Press, Princeton, NJ.
- Gompert, Z. & Buerkle, C.A. (2009). A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Mol. Ecol.*, 18, 1207–1224.
- Gompert, Z., Lucas, L.K., Fordyce, J.A., Forister, M.L. & Nice, C.C. (2010). Secondary contact between *Lycyaides idas* and *L. melissa* in the Rocky Mountains: extensive admixture and a patchy hybrid zone. *Mol. Ecol.*, 19, 3171–3192.
- Grant, P.R. & Grant, B.R. (2002). Unpredictable evolution in a 30-year study of Darwin's finches. *Science*, 296, 707–711.
- Hatfield, T. & Schluter, D. (1999). Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution*, 53, 866–873.
- Hendry, A.P. (2009). Ecological speciation! Or the lack thereof? *Can. J. Fish. Aqu. Sci.*, 66, 1383–1398.
- Hendry, A.P., Grant, P.R., Grant, B.R., Ford, H.A., Brewer, M.J. & Podos, J. (2006). Possible human impacts on adaptive radiation: beak size bimodality in Darwin's finches. *Proc. R. Soc. B-Biol. Sci.*, 273, 1887–1894.
- Hendry, A.P., Huber, S.K., De Leon, L.F., Herrel, A. & Podos, J. (2009). Disruptive selection in a bimodal population of Darwin's finches. *Proc. R. Soc. B-Biol. Sci.*, 276, 753–759.
- Hoekstra, H.E., Hirschmann, R.J., Bunde, R.A., Insel, P.A. & Crossland, J.P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science*, 313, 101–104.

- Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A. & Cresko, W.A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.*, 6, e1000862.
- Hudson, M.E. (2008). Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol. Ecol. Resour.*, 8, 3–17.
- Hutter, S., Saminadin-Peter, S.S., Stephan, W. & Parsch, J. (2008). Gene expression variation in African and European populations of *Drosophila melanogaster*. *Genome Biol.*, 9, R12.
- Irwin, D.E., Bensch, S. & Price, T.D. (2001). Speciation in a ring. *Nature*, 409, 333–337.
- Jiggins, C.D., Estrada, C. & Rodrigues, A. (2004). Mimicry and the evolution of premating isolation in *Heliconius melpomene* Linnaeus. *J. Evol. Biol.*, 17, 680–691.
- Linnen, C.R., Kingsley, E.P., Jensen, J.D. & Hoekstra, H.E. (2009). On the origin and spread of an adaptive allele in deer mice. *Science*, 325, 1095–1098.
- Malmberg, R.L. & Mauricio, R. (2005). QTL-based evidence for the role of epistasis in evolution. *Genet. Res.*, 86, 89–95.
- Matsubayashi, K.W., Ohshima, I. & Nosil, P. (2010). Ecological speciation in phytophagous insects. *Entomol. Exp. App.*, 134, 1–27.
- Matsumura, H., Yoshida, K., Luo, S., Kimura, E., Fujibe, T., Albertyn, Z. *et al.* (2010). High-throughput SuperSAGE for digital gene expression analysis of multiple samples using next generation sequencing. *PLoS ONE*, 5, e12010.
- Mavarez, J., Salazar, C.A., Bermingham, E., Salcedo, C., Jiggins, C.D. & Linares, M. (2006). Speciation by hybridization in *Heliconius* butterflies. *Nature*, 441, 868–871.
- McKinnon, J.S., Mori, S., Blackman, B.K., David, L., Kingsley, D.M., Jamieson, L. *et al.* (2004). Evidence for ecology's role in speciation. *Nature*, 429, 294–298.
- Michel, A.P., Sim, S., Powell, T.H.Q., Taylor, M.S., Nosil, P. & Feder, J.L. (2010). Widespread genomic divergence during sympatric speciation. *Proc. Natl. Acad. Sci. USA*, 107, 9724–9729.
- Milinski, M., Griffiths, S., Wegner, K.M., Reusch, T.B.H., Haas-Assenbaum, A. & Boehm, T. (2005). Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. *Proc. Natl. Acad. Sci. USA*, 102, 4414–4418.
- Mullen, L.M. & Hoekstra, H.E. (2008). Natural selection along an environmental gradient: a classic cline in mouse pigmentation. *Evolution*, 62, 1555–1570.
- Mullen, L.M., Vignieri, S.N., Gore, J.A. & Hoekstra, H.E. (2009). Adaptive basis of geographic variation: genetic, phenotypic and environmental differences among beach mouse populations. *Proc. R. Soc. B-Biol. Sci.*, 276, 3809–3818.
- Muller, H.J. (1940). Bearing of the *Drosophila* work on systematics. In: *The New Systematics* (ed. Huxley, J.S.). Clarendon Press, Oxford, UK, pp. 185–268.
- Navarro, A. & Barton, N.H. (2003). Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution*, 57, 447–459.
- Noonan, B.P. & Comeault, A.A. (2009). The role of predator selection on polymorphic aposematic poison frogs. *Biol. Lett.*, 5, 51–54.
- Noor, M.A.F., Grams, K.L., Bertucci, L.A. & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proc. Natl. Acad. Sci. USA*, 98, 12084–12088.
- Nosil, P., Funk, D.J. & Ortiz-Barrientos, D. (2009a). Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.*, 18, 375–402.
- Nosil, P., Harmon, L.J. & Seehausen, O. (2009b). Ecological explanations for (incomplete) speciation. *Trends Ecol. Evol.*, 24, 145–156.
- Oka, A., Aoto, T., Totsuka, Y., Takahashi, R., Ueda, M., Mita, A. *et al.* (2007). Disruption of genetic interaction between two autosomal regions and the X chromosome causes reproductive isolation between mouse strains derived from different subspecies. *Genetics*, 175, 185–197.
- Orr, H.A., Masly, J.P. & Phadnis, N. (2007). Speciation in *Drosophila*: from phenotypes to molecules. *J. Hered.*, 98, 103–110.
- Pfennig, D.W. & Rice, A.M. (2007). An experimental test of character displacement's role in promoting postmating isolation between conspecific populations in contrasting competitive environments. *Evolution*, 61, 2433–2443.
- Presgraves, D.C. (2010). The molecular evolutionary basis of species formation. *Nat. Rev. Gen.*, 11, 175–180.
- Qvarnström, A. & Bailey, R.I. (2009). Speciation through evolution of sex-linked genes. *Heredity*, 102, 4–15.
- Qvarnström, A., Rice, A.M. & Ellegren, H. (2010). Speciation in *Ficedula* flycatchers. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 365, 1841–1852.
- Reynolds, R.G. & Fitzpatrick, B.M. (2007). Assortative mating in poison-dart frogs based on an ecologically important trait. *Evolution*, 61, 2253–2259.
- Rieseberg, L.H., Whitton, J. & Gardner, K. (1999). Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, 152, 713–727.
- Rusk, N. (2009). Cheap third-generation sequencing. *Nat. Met.*, 6, 244–245.
- Sæther, S.A., Sætre, G.P., Borge, T., Wiley, C., Svedin, N., Andersson, G. *et al.* (2007). Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science*, 318, 95–97.
- Sætre, G.-P. & Sæther, S.A. (2010). Ecology and genetics of speciation in *Ficedula* flycatchers. *Mol. Ecol.*, 19, 1091–1106.
- Sætre, G.-P., Borge, T., Lindroos, K., Haavie, J., Sheldon, B.C., Primmer, C.R. *et al.* (2003). Sex chromosome evolution and speciation in *Ficedula* flycatchers. *Proc. R. Soc. B-Biol. Sci.*, 270, 53–59.
- Schluter, D. (1993). Adaptive radiation in sticklebacks: size, shape, and habitat use efficiency. *Ecology*, 74, 699–709.
- Schluter, D. (2001). Ecology and the origin of species. *Trends Ecol. Evol.*, 16, 372–380.
- Schluter, D. (2009). Evidence for ecological speciation and its alternative. *Science*, 323, 737–741.
- Schwarz, D., Robertson, H.M., Feder, J.L., Varala, K., Hudson, M.E., Ragland, G.J. *et al.* (2009). Sympatric ecological speciation meets pyrosequencing: sampling the transcriptome of the apple maggot *Rhagoletis pomonella*. *BMC Genomics*, 10, 633.
- Servedio, M.R. & Noor, M.A.F. (2003). The role of reinforcement in speciation: theory and data. *Ann. Rev. Ecol. Evol. Syst.*, 34, 339–364.
- Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jónsson, B. *et al.* (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, 428, 717–723.
- Slatkin, M. (1982). Pleiotropy and parapatric speciation. *Evolution*, 36, 263–270.
- Smith, M.W. & O'Brien, S.J. (2005). Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. *Nat. Rev. Gen.*, 6, 623–632.
- Steiner, C.C., Weber, J.N. & Hoekstra, H.E. (2007). Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biol.*, 5, e219.
- Stinchcombe, J.R. & Hoekstra, H.E. (2008). Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, 100, 158–170.
- Teeter, K.C., Payseur, B.A., Harris, L.W., Bakewell, M.A., Thibodeau, L.M., O'Brien, J.E. *et al.* (2008). Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Res.*, 18, 67–76.
- Thibert-Plante, X. & Hendry, A.P. (2009). Five questions on ecological speciation addressed with individual-based simulations. *J. Evol. Biol.*, 22, 109–123.
- Torres, T.T., Metta, M., Ottenwälder, B. & Schlötterer, C. (2008). Gene expression profiling by massively parallel sequencing. *Genome Res.*, 18, 172–177.
- Via, S. & West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol. Ecol.*, 17, 4334–4345.
- Vignieri, S.N., Larson, J.G. & Hoekstra, H.E. (2010). The selective advantage of crypsis in mice. *Evolution*, 64, 2153–2158.
- Wang, Z., Gerstein, M. & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Gen.*, 10, 57–63.
- Wheat, C.W. (2010). Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. *Genetica*, 138, 433–451.

Editor, Arne Moores

Manuscript received 7 April 2010

First decision made 19 May 2010

Second decision made 20 September 2010

Manuscript accepted 4 October 2010