Speciation genetics: evolving approaches

Mohamed A. F. Noor and Jeffrey L. Feder‡*

Abstract | Much progress has been made in the past two decades in understanding Darwin's mystery of the origins of species. Applying genomic techniques to the analysis of laboratory crosses and natural populations has helped to determine the genetic basis of barriers to gene flow which create new species. Although new methodologies have not changed the prevailing hypotheses about how species form, they have accelerated the pace of data collection. By facilitating the compilation of case studies, advances in genetic techniques will help to provide answers to the next generation of questions concerning the relative frequency and importance of different processes that cause speciation.

Gene flow
The movement of alleles between local populations that is due to the migration of individuals.

Hybrid zone
A location where the hybrid offspring of two divergent, partially geographically overlapping groups are prevalent. Hybrid zones are sometimes stable for many generations and there is often variation in the fitness of hybrids within the zone.

Genomic conflict
Competition within a genome for transmission to or success of gametes.

**DCMB Group/ Biology Department, Duke University, BOX 91000, Durham, North Carolina 27708, USA.
‡Department of Biological Sciences, University of Notre Dame, Galvin Life Science Center, PO BOX 369, Notre Dame, Indiana 46556-0369, USA.
e-mails: noor@duke.edu; jfeder@nd.edu*
doi:10.1038/nrg1968
Published online
3 October 2006

What types of genetic change bring about speciation is one of the most basic questions in biology. Speciation is a fundamental outcome of life. Given metabolism, reproduction, mutation, heredity, and the spatial–temporal subdivision of the environment and of individuals into populations, new species form through time, increasing biodiversity. The evolution of this biodiversity can only be fully explained if we identify the heritable underpinnings of species formation and the forces responsible for their origins.

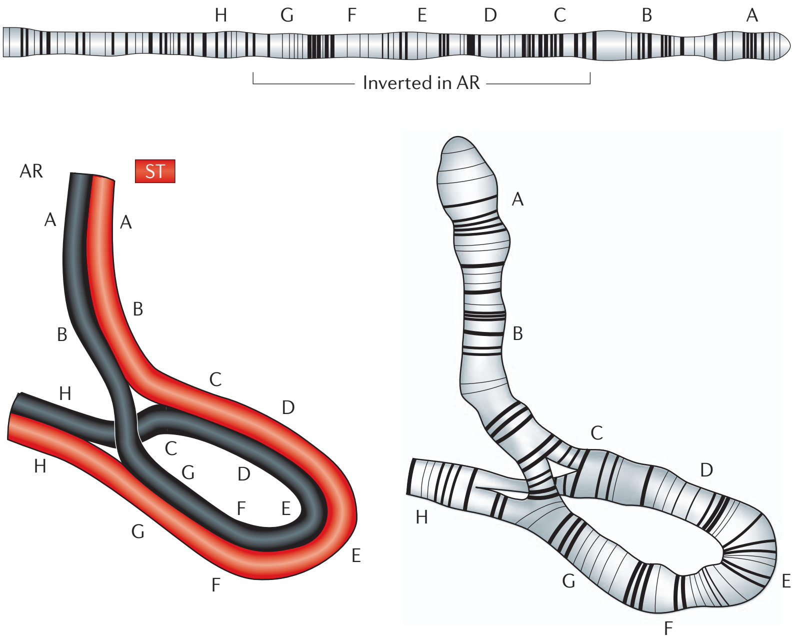
Here, we review how recent advances in molecular and genomic techniques are helping to achieve a greater understanding of the genetics of speciation. For the purpose of this review, we focus on technical advances rather than theoretical concepts, which are discussed extensively elsewhere[1]. We define speciation for sexually reproducing organisms as the transformation of within-population variation into taxonomic differences through the evolution of inherent barriers to gene flow. This definition is not universally accepted, but it remains the most commonly used by students of speciation and is of the greatest utility to dissecting the genetics of the process[1]. We discuss whether and how molecular techniques are helping to discern the genetic bases and evolutionary origins of barriers that contribute to population divergence. We present some recent discoveries from laboratory and field studies that apply molecular and genomics techniques to the speciation question. Our examples are chosen to illustrate some of the breadth of approaches that are used to tackle this exciting question.

We begin by framing the problem from a historical perspective, tracing how the development of ever-more-sophisticated methods has led to finer dissection of the genetic origins of species down to the level of the individual loci and eventually nucleotides. Technical and statistical advances are also extending laboratory-based discovery to natural populations, allowing researchers to investigate barriers to gene flow, genomic interactions and the genetic permeability of species boundaries in hybrid zones where differentiated taxa overlap and interbreed. Whole-genome surveys can provide representative snapshots of differentiation across the entire genomes of model genetic systems. Ultimately this progress is leading to large-scale comparative genomic analysis of entire taxonomic groups (model and emerging-model systems alike) from which general patterns and rules might emerge.

We conclude by discussing what the new methods have and potentially will reveal about the genetics of speciation. We argue that the new technology is not necessarily providing results that are inherently different from those of earlier studies. Rather, the new methodology is adding detail by accelerating the rate and ease of data collection on a genome-wide scale. We contend that this aspect of technology will have the greatest immediate impact — it will allow us to move from isolated case studies to compilations of results for representative groups to answer relative frequency questions about processes and factors that contribute to speciation. For example, how often do natural selection, sexual selection, genomic conflict and ecological interactions drive divergence? What proportion of genetic change is regulatory versus functional? How important is chromosomal change in speciation?

How often does speciation occur in the face of periodic or regular gene flow, generating mosaic patterns of genomic differentiation? It is not clear that the new methodology will lead to a major shift in our thinking about the ways in which speciation can occur. We propose that this limitation is not technical, but mostly rests in the imagination of students in the field.

Box 1 | **Classical approaches to the study of speciation genetics**



Genetic studies of speciation, which date back to the early twentieth century, attempt to identify the number, distribution and type of genes that contribute to phenotypes that prevent gene flow between species. They also examine patterns of differentiation between the genomes of diverging taxa.

Laboratory or garden hybridization studies

Laboratory or garden hybridization studies have traditionally had a central role in understanding the genetic basis of barrier (speciation-related) traits. Examining the distribution of phenotypes over one or two generations of controlled crosses has been used to infer the contributions of the sex chromosomes versus autosomes, identify patterns of dominance and provide estimates of the minimum number of genes that affect the speciation-related traits. An advance was made with the first true genetic mapping efforts, when associations were assessed between genetic (now, often molecular) markers and barrier traits in F_2 or other hybrids. For example, Dobzhansky[133] correlated hybrid testis size (a proxy for fertility) with genotype at seven mutant markers in a cross between races of *Drosophila pseudoobscura* (now, described as separate species). In principle, this is no different from modern QTL mapping studies.

Genetic diversity within and between species

Speciation genetics has also been explored using studies of genetic diversity within and between species. Hybridization or differentiation between two species can be inferred from the ratios of shared alleles to 'fixed' differences. Even before DNA was known to be the hereditary material, studies of genetic diversity within and between species were carried out using either phenotypes or polytene chromosome arrangements (see figure, which shows a chromosomal heterozygote for the standard (ST) and arrowhead (AR) arrangements in *D. pseudoobscura*). These were succeeded in the 1960s by protein electrophoretic approaches to the study of diversity and speciation, and today by the use of DNA-based genetic markers (BOX 2) and DNA sequences.

These approaches form much of the foundation for many modern studies of speciation genetics.

Classical approaches

Descriptive and conceptual. Although he entitled his famous book *On the Origin of Species*, Darwin[2] had only vague insights into the speciation process itself, viewing it as a later stage in a continuum from adaptive divergence among 'varieties' within species. The Russian geneticist Theodosius Dobzhansky and the German systematist Ernst Mayr instilled broader excitement in the field of speciation genetics with three conceptual advances in the 1930s and 1940s. First, Dobzhansky[3] and Mayr[4] compiled lists of traits that prevent gene flow between species, such as habitat divergence and hybrid sterility (here, we refer to them collectively as barriers). They noted that some of these barriers prevent the formation of hybrid offspring, whereas others prevent the success and propagation of hybrid offspring once formed. Second, Dobzhansky and Mayr argued that species can be defined in the context of these traits. This second advance identified a means for the study of speciation genetics: the genetics of barriers acting to reduce gene flow in nature indicates the genetics of speciation itself. Third, they studied these barriers directly: Mayr defined which ones operate in nature and Dobzhansky determined their genetic basis in the laboratory through controlled crosses (BOX 1). The implications of this last endeavour were profound, demonstrating that barriers are traits that can be genetically mapped and that species boundaries can be quantified genetically.

Despite the dramatic advances in molecular and genomic techniques, the 'old-school' approaches of Dobzhansky and Mayr for studying the genetics of speciation still apply today[5] (BOX 1). Since that time, similar laboratory and field studies have become plentiful enough for elegant meta-analyses of broader taxonomic data sets to be carried out. These meta-analyses have found, unsurprisingly, that barriers between long-diverged species are typically much stronger or more effective than barriers between very recently diverged species (for example, see REFS 6–11). Therefore, genetic divergence is associated with the accumulation of more (and/or stronger) barriers, irrespective of how genetic divergence is measured.

Echoing the differences in Dobzhansky's and Mayr's approaches, a slight divide has nonetheless emerged within the speciation community. At one extreme are the researchers who investigate barriers to gene flow using laboratory crosses of well-established model species, often focusing on easily scored phenotypes such as hybrid sterility or inviability. They have been successful in identifying genes that contribute to these traits (see below). However, the specific traits or genes identified might not directly reduce or have previously restricted gene flow between the focal species in nature (indeed, they often do not, as several species studied in this way do not occur together in nature). Moreover, these genes could contribute to reproductive isolation in the laboratory, but if they arose after gene flow was essentially complete between species in nature, they would not represent 'speciation' loci in the strict sense. Nonetheless, these studies have provided valuable genetic and evolutionary insights that can be applied to naturally hybridizing but less tractable species, and have identified the rate of accumulation of alleles that cause hybrid incompatibilities in isolated species. At the other extreme are naturalists who study very recently diverged, often hybridizing populations. Although their studies are directly applicable to natural divergence in its early stages, the populations might be ephemeral and never actually speciate: such taxa could eventually fuse (for example, see REF. 12). So, an unproven assumption of eventual speciation underlies these studies. The authors of this review were trained at opposite ends of this divide. We argue that complementary insights are yielded by these two approaches, and genetic and genomic techniques have facilitated progress in both arenas.

Laboratory-crossable species. Genetic crosses of laboratory-amenable organisms have been a touchstone for dissecting the genetics of speciation since the early 1930s, providing information on the minimum number of genes and the relative contributions of different chromosomes (especially the sex chromosomes) to barriers to gene flow between taxa. Although the basic protocol for gene mapping remains the same, several technical and methodological advances have increased the pace of discovery since Dobzhansky's[13] initial study of hybrid male sterility in *Drosophila pseudoobscura* 'races'. These advances include improvements in molecular genotyping (for example, see REFS 14,15; also see below), statistical advances in detecting association between markers and traits (for example, see REFS 16,17) and improved genetic cross methodologies (for example, see REF. 18). Repeated backcrosses have been used to introgress small and well-defined segments from one species into another to determine the effects of single loci. For example, in a tour-de-force, True *et al.*[19] inserted transposable *P*-elements into 87 positions in the genome of *Drosophila mauritiana*, and introgressed each of these segments into *Drosophila simulans* by backcrossing for 15 generations to determine the genetic positions of hybrid-sterility-conferring loci. Even now, a decade later, this represents an impressive accomplishment.

Data from nature. One can only learn so much by studying model organisms in the laboratory. One problem is that a reproductive barrier between geographically isolated populations that is identified on the basis of laboratory crosses might not be an effective barrier to gene flow in the field if and when these populations encounter each other. Surveys of natural hybrid zones and/or transplant studies in the field are therefore needed to complement laboratory-based studies to establish the significance and strength of specific barriers in nature. Another problem is the lack of taxonomic representation. Successful mapping studies have, to date, been mainly limited to model systems such as *Drosophila* species, where detailed linkage maps and tools such as deletions, inbred lines, genomic sequences and transformation systems are available. These model systems might not be representative of the types of trait or genetic architecture involved in speciation across taxonomic groups.

In this regard, analysis of natural populations can broaden our surveys by providing entries for identifying candidate speciation genes in non-model genetic organisms[20–22]. In hybrid zones, genomic regions that are relatively impermeable to introgression (admixture) and display enhanced differentiation probably contain genes that generate barriers between populations[23–28]. In addition, QTL analysis in natural hybrid zones can look for correlations of markers with phenotypic traits that distinguish taxa, several of which could be associated with barriers[29,30]. Therefore, hybrid zones not only provide valuable genetic information on the natural history of speciation and its causes, but can also be used to help to move from broad-scale characterization of genomic architecture to the specific genes responsible for barriers[20,31]. In the past decade, we have witnessed the transformation of many non-model genetic organisms, such as *Heliconius* butterflies and *Gasterosteus* sticklebacks, into 'emerging' model systems that are of particular interest because they integrate complementary information on speciation from the field and laboratory to provide a fuller understanding of the process.

As with genetic crosses, genetic studies of samples from natural populations have greatly benefited from technical advances in the sensitivity, speed, types and numbers of marker that can be scored. Sequence-based studies of multiple nuclear loci are now standard and combinations of different types of marker are being used to investigate different aspects of the problem of gene flow[32,33] (see also Luikart *et al.*[34] and BOX 2 for a description of the types of molecular genetic marker used in population genomics studies). Analyses of hybrid zones have yielded estimates of the number of genes that contribute to barriers, insights into how the balance between selection and migration shapes gene frequency clines, and assessments of the relative importance of inherent genomic incompatibilities, population demography and ecology in maintaining the genetic integrity of taxa[20–22,30,35–45]. In addition, an increasing (although still limited) number of case studies provide evidence for divergence-with-gene-flow speciation (for example, see REFS 33,46–49) and related phenomena, including sympatric speciation through genic mechanisms that do not involve changes in ploidy number[50–52], and hybrid speciation and the creative role of introgression in divergence[1,53–55]. Many of the case studies cited are not universally accepted[1]. Nonetheless, the modern approaches we discuss below can be and have been used to bolster the evidence for these controversial hypotheses in particular cases, such as sympatric speciation.

Modern approaches

Recreation or dissolution of speciation events. Genetic studies of speciation often suffer from the problem that they are investigating a process that is either complete or nearly so. One innovative yet relatively underused empirical solution to this problem involves the experimental recreation of new species and hybrid zones in nature. Applying molecular techniques to the traditional experimental hybrid approach has been fruitful in studies of hybrid speciation. Using *Helianthus*

Introgression
The movement of alleles from one species into the gene pool of another through repeated backcrossing of an interspecies hybrid with one of the parent species.

Backcross
The mating of an individual with its parent, or with an individual of the same genotype as its parent, to follow the inheritance of alleles and phenotypes.

Transplant studies
Studies in which organisms are moved from a native to an introduced setting to examine the effects of the environment or of related species.

Admixture
The mixing of genetically differentiated groups.

Cline
The gradual change of a genotype or phenotype in a species over a geographical area that is often associated with an environmental gradient.

Divergence-with-gene-flow speciation
Speciation that progresses without the complete absence of gene exchange between diverging taxa. This type of speciation includes but is not limited to sympatric speciation.

Sympatric speciation
Speciation through divergence in geographically overlapping taxa.

Hybrid speciation
Hybridization between two species gives rise to a new, pure-breeding taxon.

Box 2 | **Molecular genetic markers used in population genetic and genomic studies**

The types of molecular marker used in population genomics have been reviewed elsewhere[34], but this box presents a few examples of some of the most popular markers.

AFLPs

Amplified fragment length polymorphisms (AFLPs) are produced by selective PCRs that amplify many segments of the genome, and variability is scored by the presence or absence of particular amplicons. The advantage of this approach is that it can be high-throughput and fast; the disadvantages are that the markers are dominant (heterozygotes cannot be differentiated from homozygotes) and there can be some problems with reproducibility.

Microsatellites

These are arrays of 1–6 bases of repetitive sequence. They are often highly polymorphic, and are co-dominant markers. However, development of these markers is more expensive, and fewer can be scored simultaneously than with AFLPs. Additionally, their mutational properties are such that independent mutations often recreate the same allele, so identity does not indicate co-ancestry (a problem called homoplasy).

Sequence data

Sequence data are the ultimate source of all variation. Sequence variation could be in the form of SNPs or possibly gross features such as insertions or deletions. Direct examination is generally the most rigorous approach and can be done on a large scale when sequenced genomes are available, but, at present, this is often too laborious and expensive for non-model species.

SNPs

These are often misleadingly referred to as a class of genetic marker, but the means whereby SNPs are identified and/or scored are variable. They can be identified by techniques such as those using RFLPs (restriction fragment length polymorphisms), direct sequencing of alleles and comparing published DNA sequences. They can be subsequently scored by the same methods or by many others including pyrosequencing, real-time PCR and microspheres. Developing SNPs as markers can be quick if genome sequences or other sequence databases are available.

Transposable elements

Transposable elements are genomic mobile genetic elements. Some of these, such as SINEs (short interspersed elements), have been used for population genetics or phylogenetics because the probability of independent insertions into a particular spot is near zero, suggesting essentially no homoplasy.

sunflower species, Rieseberg and colleagues[56–58] were able to recreate the complex phenotypes of ancient hybrid species from early generation synthetic hybrids. Consistent with expectations, the same combinations of parental chromosomal segments required to generate extreme phenotypes in synthetic hybrids are those found in the natural hybrids.

Experimental hybridization has also proved useful for reconstituting possible sequential phenotypic steps from ancestral to derived states in the evolution of barriers between taxa, such as in the study of flower colour and shape in monkeyflowers[59,60] and wing patterns in *Heliconius* butterflies[61]. These experiments are natural extensions of some of the classical work that selected for reduced gene flow in maize[62] and *Drosophila* species[63,64]. In a recent example, Leu and Murray[65] identified the genetic basis for altered mate preference by selection for increased assortative mating between two yeast strains. Although this study did not yield information on particular examples from nature, it provided a hypothesis for one physiological means of assortative mating in yeasts.

Yeast have also been used in the experimental 'de-evolution' of a species. In an elegant study that examined the effect of genome rearrangements on hybrid sterility, Delneri *et al.*[66] experimentally reconfigured the *Saccharomyces cerevisiae* genome to make it collinear with that of its relative *Saccharomyces mikatae*. Hybrids of these two species normally produce only inviable spores, whereas hybrids of the experimentally manipulated *S. cerevisiae* with *S. mikatae* produced hybrids with some (but incomplete) spore viability. This experiment showed a direct contribution of the genome rearrangement to hybrid spore viability, but also demonstrated that genic effects must also contribute.

Despite past imperfections in design, Rieseberg and colleagues[47,67] concluded from a review of the collective experimental hybrid literature that most traits that differentiate species seem to be under selection in the wild, that hybrid fitness tends to be contingent on both hybrid genotype and the habitat into which they are placed, that intrinsic isolating factors are not necessarily more stable and irreversible than extrinsic, ecologically related barriers, and that hybrid incompatibilities could be quickly purged in both experimental and natural hybrid zones. These conclusions demonstrate that the coupling of experimental hybridization with genomics holds great promise for understanding the speciation process.

Empirical and statistical analyses for introgressive hybridization. The new molecular approaches are having a major impact on resolving the genetic architecture and permeability of species boundaries. In the past decade, we have witnessed a shift in surveys of natural populations from studies that are based on mitochondrial sequences to studies of multiple nuclear loci to high-throughput scans of entire genomes for detecting introgression and regions of higher differentiation ('islands of speciation') between diverging taxa. It is now possible to use microarray hybridization techniques on samples from natural populations to rapidly paint

Pyrosequencing
A method for DNA sequencing, in which the inorganic pyrophosphate that is released from a nucleoside triphosphate on DNA chain elongation is detected by a bioluminometric assay.

Microspheres
(Also known as microparticles or microbeads). Small 1–100 μm diameter particles that are used as solid supports in bioassays. They can carry a probe or primer, and can contain internal magnetic compounds to allow magnetic separation or internal fluorescent compounds for labelling.

a broad view of genomic differentiation. Follow-up sequence analysis is used to finely map the boundaries of introgressed versus diverged regions and to quantify the extent of differentiation for these genes.

Genetic data alone do not answer the question of whether taxa are or have undergone introgressive hybridization unless the results are viewed in the context of an appropriate evolutionary model for rigorous statistical testing. Concomitant with the advances in technology, more sophisticated analytical methods are also being developed to discern whether shared variation is due to introgression from a related species or represents the persistence of ancestral polymorphism. These methods can be categorized as being based on either summary population genetic statistics or phylogenetic gene-tree-building approaches. Gene-tree approaches can reveal alleles that show discordant phylogenetic patterns (paraphyly) that could indicate introgression[68–71], but could also be explained by incomplete lineage sorting. Gene trees have also been used to distinguish historical processes through nested clade analysis[72,73]. However, because it is difficult to devise powerful statistical tests of competing hypotheses for gene-tree-based methodologies such as nested clade analysis[74], one can incorrectly accept a best-fit model for a process that never occurred. So, gene trees could find their most practical application as a first, qualitative approach to identify loci that might have introgressed in the past or to identify taxa of possible hybrid origin.

Variation among loci in summary F_{ST} values of interpopulation differentiation[75,76] and their multi-allele G_{ST} extension[77] has also been used to test for introgression (for example, see REFS 26,78). The basis for the test is that divergent selection pressures on a trait can produce large between-population allele frequency differences for genetic markers that encode or are linked to the trait, and these differences are detectable by high, outlier F_{ST} values for the markers[79]. The test has been applied in an analogous manner to interspecific comparisons. In this case, outlier F_{ST} values point to genomic regions that might be relatively impermeable to introgression between taxa, presumably owing to association with a barrier. Although informative as a descriptive measure, F_{ST} comparisons have their shortcomings. First, they do not fully use the genealogical information inherent in DNA sequences. Second, F_{ST} values can be affected by historical and stochastic population processes that are unrelated to introgression, one of the most problematic being differences in mutation rate among genes or genomic regions. F_{ST} and G_{ST} values should therefore be adjusted by appropriate estimates of mutation rate when testing for introgression, which are based on either levels of intraspecific polymorphism or interspecific divergence derived from outgroup comparisons. Modified approaches such as the analysis of molecular variation[80] and the use of G_{ST} values[81] have been developed to ameliorate some of these problems. Accompanying biogeographical information can also strengthen the case for differential introgression; for example, through comparisons of cline width in hybrid zones among loci[24]. However, new generation methods for DNA sequence data that are based on the coalescent theory, such as the Wang–Wakeley–Hey test of shared polymorphism to fixed differences[48,82], the linkage disequilibrium test of gene flow[83], the isolation with migration model devised by Nielsen and Wakeley[84] and its multilocus extension[85], and the relative node depth approach[23,86,87], provide more sophisticated approaches to test for possible introgression (BOX 3).

Recent analyses suggest that many evolutionarily related, geographically overlapping taxa exchange genes through introgressive hybridization[1,55,88,89]. Moreover, these taxa commonly possess mosaic (composite) genome structures[88]. One particularly relevant finding is that genomic regions that are relatively impermeable to many instances of introgression have been associated with low recombination rates, such as is caused by chromosomal rearrangement. Studies in sunflowers[30], the *D. pseudoobscura* subgroup[83,90–92] and *Rhagoletis pomonella*[23,93] found evidence for greater introgression (lower divergence) in collinear segments of the genome than in inverted segments. Other regions of restricted recombination, such as pericentromeric regions or translocated/inverted regions, have also been implicated in divergence between races and species (for example, see REFS 27,28,94). The continued application of new molecular approaches coupled with powerful statistical analyses promises to lead to even greater insights into the relationship between genome structure and the persistence of species despite gene flow.

Use of whole-genome sequence assemblies. Whole-genome sequence assemblies have been completed for a number of eukaryotes, including several closely related taxa that can then be studied in a comparative framework. Coupling whole-genome sequences with functional studies can yield important insights into the genetic changes that underlie speciation. One recent example involves comparisons among three distantly related yeasts. Roughly 100 million years ago, a yeast ancestor experienced a whole-genome duplication[95]. Subsequently, alternative copies of the duplicated loci, including several essential genes, were reciprocally lost in different yeast lineages[96]. If two taxa were to hybridize, then for every reciprocally deleted gene, 25% of the resulting hybrid's spores would lack a functional copy of the gene[96], providing a simple mechanism for hybrid dysfunction[97]. Although this study fails to identify a particular speciation event that is facilitated by this process, it is possible that such reciprocal gene loss contributed to at least some barriers to gene flow between yeast species.

The completion and assembly of several eukaryotic genome sequences has also been a boon for speciation research in several indirect ways. With respect to genetic mapping studies, researchers can now select markers at any location in the genome to help pinpoint genes that contribute to barriers (for example, see REF. 98). Similarly, whole-genome sequences allow for the construction of microarrays and other genetic tools to assess the expression of all known and inferred transcripts for differences between species or races and strains (for example, see REF. 99). We discuss these advances in the next section.

Nested clade analysis
A coalescent approach to disentangling the effects of long-term population history from gene flow by reconstructing the sequence of events that have generated the current genetic pattern among populations within a species.

F_{ST}
A measure of population subdivision that is based on genetic polymorphism data derived from comparing the genetic variability within and between populations.

G_{ST}
An extension of the F_{ST} measure for multiple alleles, where G_{ST} is equal to the weighted average of F_{ST} for all alleles.

Box 3 | **Several methods used to infer introgression between taxa using genetic data**

Direct gene-tree comparisons
Gene trees are inspected to detect alleles or loci that have discordant genealogical patterns which could be indicative of introgression. Generally speaking, this is a qualitative approach, although tests can be devised that are based on the coalescent theory[134] or variations of the Shimodaira–Hasegawa incongruence test[135] to provide a statistical framework.

Assignment tests
Multilocus genotypes are used to assess the number of possible populations that are represented in a sample of individuals and to assign individuals to these populations, based on maximum likelihood or Bayesian approaches[136]. This method could be used to identify potential hybrid individuals with roughly equal probabilities of belonging to alternative taxa, implying ongoing gene flow.

Nested clade analysis
Haplotype gene trees are used to define a nested series of branches (clades), allowing an evolutionary analysis of the spatial distribution of genetic variation[75,76]. Nested clade analysis can be used to qualitatively detect possible gene flow. For example, when taxa overlap in a portion of their range, a pattern in which populations are genealogically more similar in sympatry than allopatry would imply introgression.

Outlier loci
Summary statistics such as F_{ST} values[75,76] or their multi-allele G_{ST} extensions[77] are calculated to identify outlier loci that have reduced inter-taxa differentiation compared with other genes. Analysis of molecular variation is more informative for DNA sequence data[80] and a new G_{ST}[81] value has been developed that ameliorates some of the problems affecting estimates that are due to differences in mutation rate (polymorphism levels) among loci.

Relative node depth
Node depths (coalescence times) between taxa relative to an outgroup species are compared among loci for evidence of significantly greater variation than predicted by the coalescent theory, suggesting introgression[23,86,87,137].

Wang–Wakeley–Hey
An isolation model of no gene flow between taxa since their time of separation is evaluated through comparisons of shared nucleotide polymorphisms with exclusive polymorphisms and fixed differences. Gene flow will increase the numbers of the shared polymorphisms and reduce the numbers of the fixed polymorphisms relative to the null neutral expectations of the coalescent theory. Furthermore, if gene flow occurs at some loci and not at others, it will increase the variance among loci in numbers of shared polymorphisms and fixed differences. The Wang–Wakeley–Hey test uses the difference between the highest and lowest counts of shared polymorphisms among a set of loci plus the difference between the highest and lowest counts of fixed differences observed over the same group of loci and compares the observed value with a simulated distribution. Alternatively, one can use a χ^2 statistic to measure the overall fit of the data to the isolation model[46].

Linkage disequilibrium
The difference (*X*) between the average linkage disequilibrium among all pairs of shared polymorphisms between taxa for a locus (DSS) and the average among all pairs of sites for which one member is a shared polymorphism and the other is an exclusive polymorphism (DSX) is calculated[90]. If polymorphisms are shared owing to gene flow, then DSS should tend to be positive, DSX negative, and *X* should be positive and large. However, in the absence of gene flow, there should be relatively little linkage disequilibrium for all cases and *X* should be small and close to zero.

Isolation with migration model
The isolation with migration model is a Markov chain Monte Carlo method that was developed by Nielsen and Wakeley[84] for estimating the relative effects of migration and isolation on genetic diversity in a pair of populations from DNA sequence data. The method allows for the joint estimation of multiple demographic parameters in either a Bayesian or a likelihood framework, including the migration rate for each population, the divergence time from a common ancestor, and the relative sizes of the ancestral and two current populations. The method was originally based on data from a single non-recombining locus and has been extended to multiple loci by Hey and Nielsen[85].

High-throughput approaches for genotyping or expression analysis. Methods that allowed for high-throughput genotyping have dramatically increased the speed and precision with which one can localize genes that confer barriers between species. Until recently, the bulk of molecular genotyping used in genetic mapping relied on electrophoretic separation of PCR products or proteins. Capillary-based approaches have also been available, but these are still rather limited in throughput capability. Several microarray-based marker methods for scoring SNPs have been developed that can allow one to genotype literally thousands of markers simultaneously from one individual with a small quantity of DNA (for example, see REF. 100). Although most of these array-based genotyping methods require *a priori* sequence information, a few, such as DArT[101], do not.

High-throughput approaches can also be used to look for divergence between species or races. Most studies have examined genetic divergence either across the genome using the single genome sequence assemblies for each species (for example, see REF. 102) or using multiple individuals of each species but typically only 30 or fewer loci (for example, see REFS 46,103–107). The former gives a genome-wide view but lacks the ability to distinguish divergence from polymorphism within species, whereas the opposite is true for the latter approach. With newly available genomic tools, one can get a glimpse of divergence between species using

DArT
Short for 'diversity arrays technology'; this is a technique for analysing DNA polymorphism, which is based on hybridization to microarrays, that does not require DNA sequence information.

multiple individuals in a high-throughput format. A recent example is the study by Turner *et al.*[28], in which samples of genomic DNA from seven strains of each of two hybridizing *Anopheles gambiae* mosquito races were hybridized to oligonucleotide microarrays. The researchers identified three regions of the genome that bore significant differentiation between the two races, suggesting that genes responsible for ecological and behavioural differentiation are likely to be located there. These results were generally (albeit not completely) consistent with other studies that use single marker scoring approaches[27].

Oligonucleotide and cDNA microarrays are also useful for rapidly assessing differences in gene expression for thousands of loci simultaneously; these tools might provide insights into divergent or disrupted genes that could be associated with barriers. Gene expression often evolves quickly between related species[108,109], and disruptions in transcriptional regulation could contribute to reduced fitness in sterile or inviable hybrids[110]. Disruptions of gene expression (defined as levels that are higher or lower than in both parental strains) have been documented in interspecies hybrids (for example, see REFS 111–113), and have at least sometimes been associated with sterility in hybrids (for example, see REFS 114,115). If disruptions of gene expression are the cause (and not the consequence) of some hybrid incompatibilities, then high-throughput 'reverse-genetics' approaches have the potential to quickly identify candidate genes or candidate pathways that contribute to speciation. That said, the challenge then is to determine whether the association is through causation. If hybrids have severely underdeveloped gonads, then sterility and underexpression of gonad-specific transcripts would both result, but the underexpression says nothing of the underlying causation.

Direct gene manipulations or assays. For all candidate genes, the final standard of proof for causality rests on direct genetic manipulations. One can insert the candidate gene of species A into species B to demonstrate that a barrier is formed or disrupted by this insertion. Recent advances in technologies for gene manipulations can be broadly categorized as transposon-based, reverse-genetics approaches (for example, replacement by homologous recombination) or transgenic[116]. These approaches have been applied to several genes that confer barriers between species. For example, hybrids of *D. simulans* and *D. mauritiana* are sterile, and one contributor might be the putative hybrid sterility gene *Odysseus* (*Ods*). This gene was initially identified by traditional mapping and introgression approaches[117–119]. When full-length *Ods* cDNA from each of the two species was injected into a fertile hybrid line of these two species, a strong and statistically significant effect was seen which depended on which of the two parent species the inserted alleles came from[120]. This result provides a molecular confirmation of the effect of alleles at this locus on hybrid fertility. However, this effect was only observed when the alleles were inserted into introgression lines and not when inserted into pure *D. simulans*. As such, while *Ods* probably contributes to hybrid fertility, insertion of a foreign allele alone is not sufficient to cause sterility.

The above example could have suffered an additional complication if the native copies of *Ods* in the hybrid genome altered the phenotypic effect of the inserted copy. In a similar study, Greenberg *et al.*[121] investigated the effect of the *desaturase2* (*desat2*) gene on various adaptive differences between two *D. melanogaster* populations. They used the elegant gene-replacement technique of Rong and Golic[122], which, unlike many transgenic methods, leaves only a single (transgenic) copy of the target gene of interest per genome. Using this technique on *desat2*, the authors found differences in cold tolerance and starvation susceptibility between geographical alleles. Although these elegant manipulations showed much potential, their result was questioned by a subsequent study using much larger sample sizes and greater numbers of replicates[123].

Another possibility is to knock out the function of a particular candidate gene, and then show by complementation that it conferred the barrier of interest. Presgraves *et al.*[124] mapped a hybrid inviability gene using overlapping chromosomal deficiencies from *D. melanogaster* to a particular cytological region. They then tested investigator-generated loss-of-function mutations at 12 loci that span the region, testing individual *D. melanogaster* mutations for their ability to uncover hybrid lethality when heterozygous with the *D. simulans* wild-type allele. They found that mutant alleles at only one locus failed to complement the *D. simulans* hybrid lethal factor, and thereby confirmed their candidate hybrid inviability locus[125].

For emerging model systems in which such tests are more difficult, *in vitro* tests of gene function can add support to a proposed candidate gene's involvement for a barrier that separates two species. Such studies have been done using interpopulation hybrids of *Tigriopus californicus* copepods, which showed reduced performance in several fitness-related traits relative to their parents'. Rawson and Burton[126] proposed that this hybrid breakdown was associated with co-adaptation between cytochrome *c* and cytochrome *c* oxidase. Using *in vitro* assays of enzyme activity, they observed that the cytochrome *c* variants isolated from two different populations each had significantly higher activity with the cytochrome *c* oxidase derived from their respective source populations, providing a mechanistic explanation for the observed hybrid fitness reduction. In a subsequent study, Harrison and Burton[127] used site-directed mutagenesis to construct cytochrome *c* variants and showed that interpopulation hybrid breakdown can be attributed to a single, naturally occurring amino-acid substitution. Alas, the story is more complicated, as F_2 hybrid offspring do not show consistently higher fitness when cytochrome *c* genotype matches maternal mtDNA-type in a constant 20°C environment[128,129]. Nonetheless, seeing how potential barriers can be mapped to a single amino acid in non-model systems is impressive and promising.

Box 4 | Examples of putative barrier genes that have been isolated to date

Several putative 'barrier genes' have been identified in various species, most of them in *Drosophila* species. Below we provide several examples and their purported barrier effects. These genes are associated with reproductive incompatibilities, but, in some cases, might not act as barrier genes in nature, as other aspects of their biology might already prevent gene flow between the focal species. For a more extensive discussion of these and other examples, see REFS 1,47.

bindin
Bindins are acrosomal proteins on the outside of sperm that promote adhesion to eggs, and their role in species-specific binding has been studied extensively in *Echinometra* sea urchins[138,139].

cytc
Co-adaptation between cytochrome *c*, cytochrome *c* oxidase, and other parts of the electron transport system could mediate differences in hybrid viability or development time among *Tigriopus californicus* copepod populations[126,140]. However, F_2 hybrid offspring do not have a consistent pattern of higher fitness when the cytochrome *c* genotype matches the maternal mitochondrial DNA-type in a constant 20°C environment[128,129].

desat2
The *desaturase2* (*desat2*) gene is associated with female cuticular hydrocarbon differences between some African and most non-African *Drosophila melanogaster*[141]. The African allele is also associated with behavioural discrimination by African strains[123,142], although Caribbean strains possess the African allele while being behaviourally indiscriminate. There are also mixed reports of its effects on ecological adaptation[121,123].

Hmr
When mutated, the *Hybrid male rescue* (*Hmr*) transcriptional regulator gene rescues male hybrids that die at the larval–pupal transition in crosses between *D. melanogaster* and *Drosophila simulans*, and restores fertility in hybrid females[143–145].

lysin
Within acrosomes of abalone (*Haliotis*) sperm lysin interacts with VERL (vitelline envelope receptor for lysin)[146] on eggs to mediate species-specific binding of sperm to eggs[147]. Lysins could also mediate fertilization specificity in other gastropods.

Nup96
A hemizygous (and presumably homozygous) copy of the *D. simulans Nup96* gene bears sequence variations that cause inviability when associated with the *D. melanogaster* X chromosome[125]. However, the effect of this gene is only noticeable in deficiency lines, as F_1 hybrids of these species are typically fully sterile.

Ods
Odysseus (*Ods*) is a homeobox-containing gene that is associated with hybrid male sterility between the geographically isolated species *D. simulans* and *D. mauritiana*[119]. Its effect has been shown to require several interactors[120,148].

per
The *period* (*per*) gene contributes to differences in courtship song between *D. melanogaster* and *D. simulans* that are associated with behavioural discrimination[149,150]. It also functions as a regulator of circadian rhythms. When transformed between species, the gene changes the important courtship song difference[151], but, curiously, no behavioural discrimination is apparent[152].

Xmrk2
The *Xmrk2* gene is overexpressed in some hybrid classes that are produced by *Xiphophorus maculatus* platyfish and *Xiphophorus helleri* swordtails, and causes invasive melanomas that can be fatal[153]. Functional tests have confirmed the role of this gene in tumour formation, but its effects are inconsistent when different strains of platyfish are used.

YUP
The *YELLOW UPPER* (*YUP*) region controls the presence or absence of yellow pigments in the petals of pink-flowered *Mimulus lewisii*, which is pollinated by bumblebees, and its red-flowered sister species *Mimulus cardinalis*, which is pollinated by hummingbirds[59]. This region could have a role in mediating a species barrier through differential pollinator attraction, but it has not yet been localized to a single gene.

Synthesis, advances and prospects

The application of molecular genetic and genomics techniques to the study of speciation has led to significant progress in two areas. First, technical advances have facilitated the mapping and characterization of specific genes that are responsible for barriers to gene flow. Second, more extensive and sensitive genetic and statistical surveys of natural populations have allowed inferences to be made about the speciation process and the nature of species boundaries. Most, although not all, of this progress involves improvements in scale and speed, although we have moved into a phase in which we now have several 'barrier genes' in hand and potentially many more to come soon (BOX 4). Collectively, these loci include both housekeeping and regulatory functions, and frequently seem to be targets of natural selection[1,130,131]. Nevertheless, we could be surprised in the future as more barriers are mapped and prove to involve mechanisms such as meiotic drive or genomic conflict to a greater extent than is currently appreciated.

We are also gaining a clearer picture of the genetic architecture of species barriers through the analysis of naturally occurring and experimentally generated hybrids. These studies have confirmed many traditional views about speciation, such as the role that geography often has in population divergence. They have also yielded evidence supporting contentious processes, especially in animals, related to divergence-with-gene-flow speciation. For example, many more potential cases of sympatric speciation have been proposed[50–52].

As a result of these discoveries, questions are now shifting from whether specific types of gene and types of process occur to what their relative frequency and importance for speciation are. Genomics is therefore having a major impact on speciation research, not through a fundamental paradigm shift in theory, but by providing rigorous confirmations of hypotheses and facilitating a methodological transition to meta-analysis. The new techniques are providing the means for compiling extensive catalogues of compelling case studies from representative groups to answer the next generation of frequency issues. Additionally, the techniques can be used in combination for a more complete understanding of specific cases. For example, genetic mapping initially identified the *Ods* gene that causes hybrid sterility[117,118], its effect was confirmed by gene manipulation[120], and possible downstream targets or consequences of its allelic replacement were described through microarrays and real-time PCR[115,132]. These types of combinatorial approach bring us closer to understanding the molecular mechanisms that underlie speciation.

As we continue to move into the era of 'comparative speciation genomics' it is important to recognize both the strengths and limitations of genomic technologies. Unquestionably, the pace and scale of data acquisition will accelerate, which is due in large part to advances in sequencing technology. It is therefore not unreasonable to think that in the next decade a single-investigator grant could propose to obtain and compare whole-genome sequences for multiple species or individuals. Annotations of genes will also improve such that the functional consequences of any observed differences could be determined computationally, and candidate genes for a particular trait could be identified almost instantaneously. In this regard, it is important not to neglect non-model organisms and to develop comparative genomics for both emerging and model systems to link field and laboratory-based studies of speciation genetics.

The 'genomics revolution' is not the be-all or end-all for the study of speciation, however, and the new purely sequence-based and/or expression-based approaches to studying speciation merely generate hypotheses to be tested. Another liability is that our current capacity to acquire data has exceeded our analytical ability to interpret the results, especially with regards to microarray experiments. Consequently, many barely interpreted data sets appear in the literature that are based on overly simplistic models. The assumption is often that with enough data, irrespective of how poorly analysed, someone will eventually 'divine' their true significance and meaning. For now, the new genomic tools are most reliably used to generate hypotheses, with careful and precise old-school reductionist bench work still needed to follow up and test these hypotheses.

A second problem is establishing the chronology of the genetic and phenotypic changes that lead to speciation. This problem pertains to determining both the sequence in which mutations in the same and different genes arose to generate a particular barrier and the order in which different barriers arose during a speciation event. Under certain instances of divergence-with-gene-flow speciation it might be possible to discern the order in which changes evolved between taxa, based on levels of neutral genetic divergence separating genes. Genes that restricted gene flow early in a speciation event will tend to show greater differentiation for linked sequences than those that arose later. For other speciation events, however, the best we might be able to do is to infer the possible order of change, based on information from experimental hybridization studies. In addition, large-scale comparative meta-analyses of taxa in different chronologically or temporally ordered stages of divergence could reveal trends in which certain types of barrier appear more often than not early in the speciation process relative to others (for example, see REFS 6–11).

In conclusion, the technical advances of the past decade have facilitated and accelerated speciation research. However, contrary to the hype, the new genomics approaches have not led to major conceptual insights into how we think new species form. Nor do we, the authors, foresee any theoretical breakthrough in the near future that would be driven by technology alone. The new methods are the tools that help to test and shape imaginative new hypotheses — the real 'engines' of progress. We now know much more about the genetic changes that make new species and the processes that are responsible for their evolution. One day we might even reach consensus on the nagging question of what exactly constitutes a species.

1. Coyne, J. A. & Orr, H. A. *Speciation* (Sinauer Associates, Sunderland, Massachusetts, 2004). **This book is a rigorous, authoritative and provocative review of the entire field of speciation.**
2. Darwin, C. R. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (John Murray, London, 1859).
3. Dobzhansky, T. *Genetics and the Origin of Species* (Columbia Univ. Press, New York, 1937).
4. Mayr, E. *Systematics and the Origin of Species* (Columbia Univ. Press, New York, 1942).
5. Howard, D. J. *et al.* The genetics of reproductive isolation: a retrospective and prospective look with comments on ground crickets. *Am. Nat.* **159**, S8–S21 (2002).
6. Archibald, J. K., Mort, M. E., Crawford, D. J. & Kelly, J. K. Life history affects the evolution of reproductive isolation among species of *Coreopsis* (Asteraceae). *Evolution* **59**, 2362–2369 (2005).
7. Coyne, J. A. & Orr, H. A. Patterns of speciation in *Drosophila*. *Evolution* **43**, 362–381 (1989).
8. Presgraves, D. C. Patterns of postzygotic isolation in Lepidoptera. *Evolution* **56**, 1168–1183 (2002).
9. Price, T. D. & Bouvier, M. M. The evolution of F_1 postzygotic incompatibilities in birds. *Evolution* **56**, 2083–2089 (2002).
10. Sasa, M. M., Chippindale, P. T. & Johnson, N. A. Patterns of postzygotic isolation in frogs. *Evolution* **52**, 1811–1820 (1998).
11. Moyle, L. C., Olson, M. S. & Tiffin, P. Patterns of reproductive isolation in three angiosperm genera. *Evolution* **58**, 1195–1208 (2004).
12. Taylor, E. B. *et al.* Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Mol. Ecol.* **15**, 343–355 (2006).
13. Dobzhansky, T. Studies on hybrid sterility. I. Spermatogenesis in pure and hybrid *Drosophila pseudoobscura*. *Z. Zellforch. Microsk. Anat.* **21**, 169–221 (1934).
14. Borevitz, J. O. *et al.* Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Res.* **13**, 513–523 (2003).
15. Winzeler, E. A. *et al.* Direct allelic variation scanning of the yeast genome. *Science* **281**, 1194–1197 (1998).
16. Kao, C.-H., Zeng, Z.-B. & Teasdale, R. D. Multiple interval mapping for quantitative trait loci. *Genetics* **152**, 1203–1216 (1999).
17. Zeng, Z. B. Precision mapping of quantitative trait loci. *Genetics* **136**, 1457–1468 (1994).
18. Cockerham, C. C. & Zeng, Z.-B. Design III with marker loci. *Genetics* **143**, 1437–1456 (1996).
19. True, J. R., Weir, B. S. & Laurie, C. C. A genome-wide survey of hybrid incompatibility factors by the introgression of marked segments of *Drosophila mauritiana* chromosomes into *Drosophila simulans*. *Genetics* **142**, 819–837 (1996). **The first genome-wide scan for recessive hybrid incompatibility factors, which was well ahead of its time in terms of its scale.**
20. Harrison, R. G. Hybrid zones: windows on evolutionary processes. *Oxford Sur. Evol. Biol.* **7**, 69–128 (1990).
21. Kocher, T. D. & Sage, R. D. Further genetic anayses of a hybrid zone between leaopard frogs (*Rana pipiens* complex) in central Texas. *Evolution* **40**, 21–33 (1986).
22. Szymura, J. M. & Barton, N. H. The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. *Evolution* **45**, 237–261 (1991).
23. Feder, J. L. *et al.* Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in *Rhagoletis*. *Proc. Natl Acad. Sci. USA* **102**, 6573–6580 (2005).
24. Payseur, B. A. & Nachman, M. W. The genomics of speciation: investigating the molecular correlates of X chromosome introgression across the hybrid zone between *Mus domesticus* and *Mus musculus*. *Biol. J. Linn. Soc.* **84**, 523–534 (2005).
25. Rogers, S. M. & Bernatchez, L. Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*). *Mol. Ecol.* **14**, 351–361 (2005).
26. Scotti-Saintagne, C. *et al.* Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics* **168**, 1615–1626 (2004).
27. Stump, A. D. *et al.* Centromere–proximal differentiation and speciation in *Anopheles gambiae*. *Proc. Natl Acad. Sci. USA* **102**, 15930–15935 (2005).
28. Turner, T. L., Hahn, M. W. & Nuzhdin, S. V. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* **3**, e285 (2005). **A novel microarray-based genome-wide scan for DNA sequence differentiation between two mosquito forms.**
29. Hilbish, T. J., Bayne, B. L. & Day, A. Genetics of physiological differentiation within the marine mussel genus *Mytilus*. *Evolution* **48**, 267–286 (1994).
30. Rieseberg, L. H., Whitton, J. & Gardner, K. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* **152**, 713–727 (1999).
31. Rieseberg, L. H. & Buerkle, C. A. Genetic mapping in hybrid zones. *Am. Nat.* **159**, S36-S50 (2002).
32. Avise, J. C. *Phylogeography: the History and Formation of Species* (Harvard Univ. Press, Cambridge, Massachusetts, 2000).
33. Hey, J., Won, Y.-J., Sivasundar, A., Nielsen, R. & Markert, J. A. Using nuclear haplotypes with microsatellites to study gene flow between recently separated cichlid species. *Mol. Ecol.* **13**, 909–919 (2004).
34. Luikart, G., England, P. R., Tallmon, D., Jordan, S. & Taberlet, P. The power and promise of population genomics: from genotyping to genome typing. *Nature Rev. Genet.* **4**, 981–994 (2003).
35. Barton, N. H. & Hewitt, G. M. Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* **16**, 113–148 (1985).
36. Burke, J. M., Carney, S. E. & Arnold, M. L. Hybrid fitness in Louisiana irises: analysis of parental and F1 performance. *Evolution* **52**, 684–691 (1998).
37. Campbell, D. R., Waser, N. M. & Wolf, P. G. Pollen transfer by natural hybrids and parental species in an *Ipomopsis* hybrid zone. *Evolution* **52**, 1602–1611 (1998).
38. Dod, B. *et al.* Counterselection on sex chormosomes in the *Mus musculus* European hybrid zones. *J. Evol. Biol.* **6**, 529–546 (1993).
39. Hatfield, T., Barton, N. H. & Searle, J. B. A model of a hybrid zone between two chromosomal races of shrew (*Sorex araneus*). *Evolution* **46**, 1129–1145 (1992).

40. Levin, D. A. & Schmidt, K. P. Dynamics of hybrid zone in Phlox: and experimental demographic investigation. *Am. J. Bot.* **72**, 1404–1409 (1985).
41. Mallet, J. *et al.* Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. *Genetics* **124**, 921–936 (1990).
42. Nürnberger, B., Barton, N., MacCallum, C., Gilchrist, J. & Appleby, M. Natural selection on quantitative traits in the *Bombina* hybrid zone. *Evolution* **49**, 1224–1238 (1995).
43. Porter, A. H., Wenger, R., Geiger, H., Scholl, A. & Shapiro, S. M. The *Pontia daplidice–edusa* hybrid zone in northwestern Italy. *Evolution* **51**, 1561–1573 (1997).
44. Rand, D. M. & Harrison, R. G. Ecological genetics of a mosaic hybrid zone: mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. *Evolution* **43**, 432–449 (1989).
45. Wang, H., McArthur, E. D., Sanderson, S. C., Graham, J. H. & Freeman, D. C. Narrow hybrid zone between two subspecies of big sagebrush (*Artemisia tridentata*: Asteraceae). IV. Reciprocal transplant experiments. *Evolution* **51**, 95–102 (1997).
46. Kliman, R. M. *et al.* The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* **156**, 1913–1931 (2000).
47. Rieseberg, L. H., Church, S. & Morjan, C. L. Integration of populations and differentiation of species. *New Phytol.* **161**, 59–69 (2004).
48. Wang, R. L., Wakeley, J. & Hey, J. Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics* **147**, 1091–1106 (1997).
49. Won, Y. J. & Hey, J. Divergence population genetics of chimpanzees. *Mol. Biol. Evol.* **22**, 297–307 (2005).
50. Berlocher, S. H. & Feder, J. L. Sympatric speciation in phytophagous insects: moving beyond controversy? *Annu. Rev. Entomol.* **47**, 773–815 (2002).
51. Dres, M. & Mallet, J. Host races in plant-feeding insects and their importance in sympatric speciation. *Philos. Trans. R. Soc. Lond. B* **357**, 471–492 (2002).
52. Via, S. Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol. Evol.* **16**, 381–390 (2001).
53. Dowling, T. E. & Secor, C. L. The role of hybridization and introgression in the diversification of animals. *Annu. Rev. Ecol. Syst.* **28**, 593–619 (1997).
54. Schwarz, D., Matta, B. M., Shakir-Botteri, N. L. & McPheron, B. A. Host shift to an invasive plant triggers rapid animal hybrid speciation. *Nature* **436**, 546–549 (2005).
55. Seehausen, O. Hybridization and adaptive radiation. *Trends Ecol. Evol.* **19**, 198–207 (2004).
56. Gross, B. L. *et al.* Reconstructing the origin of *Helianthus deserticola*: survival and selection on the desert floor. *Am. Nat.* **164**, 145–156 (2004).
57. Rieseberg, L. H. *et al.* Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**, 1211–1216 (2003).
58. Rosenthal, D. M., Rieseberg, L. H. & Donovan, L. A. Re-creating ancient hybrid species' complex phenotypes from early-generation synthetic hybrids: three examples using wild sunflowers. *Am. Nat.* **166**, 26–41 (2005).
59. Bradshaw, H. D. & Schemske, D. W. Allele substitution of a flower color locus produces a pollinator shift in monkeyflowers. *Nature* **426**, 176–178 (2003). **This study identifies a single major-effect locus that controls a pigmentation difference causing differential pollination by bees versus hummingbirds.**
60. Schemske, D. W. & Bradshaw, H. D. Jr. Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proc. Natl Acad. Sci. USA* **96**, 11910–11915 (1999).
61. Joron, M., Jiggins, C. D., Papanicolaou, A. & McMillan, W. O. *Heliconius* wing patterns: an evo–devo model for understanding phenotypic diversity. *Heredity* (in the press).
62. Paterniani, E. Selection for reproductive isolation between two populations of maize, *Zea mays* L. *Evolution* **23**, 534–547 (1969).
63. Kessler, S. Selection for and against ethological isolation between *Drosophila pseudoobscura* and *Drosophila persimilis*. *Evolution* **20**, 634–645 (1966).
64. Koopman, K. F. Natural selection for reproductive isolation between *Drosophila pseudoobscura* and *Drosophila persimilis*. *Evolution* **4**, 135–148 (1950).
65. Leu, J.-Y. & Murray, A. W. Experimental evolution of mating discrimination in budding yeast. *Curr. Biol.* **16**, 280–286 (2006).
66. Delneri, D. *et al.* Engineering evolution to study speciation in yeasts. *Nature* **422**, 68–72 (2003). **This elegant study provides the best evidence to date for both genic and chromosomal incompatibilities contributing to hybrid sterility in two yeast species that differ by a translocation.**
67. Lexer, C., Randell, R. A. & Rieseberg, L. H. Experimental hybridization as a tool for studying selection in the wild. *Ecology* **84**, 1688–1699 (2003).
68. Beltran, M. *et al.* Phylogenetic discordance at the species boundary: comparative gene genealogies among rapidly radiating *Heliconius* butterflies. *Mol. Biol. Evol.* **19**, 2176–2190 (2002).
69. Salzburger, W., Baric, S. & Sturmbauer, C. Speciation via introgressive hybridization in East African cichlids? *Mol. Ecol.* **11**, 619–625 (2002).
70. Schliewen, U. K. & Klee, B. Reticulate sympatric speciation in Cameroonian crater lake cichlids. *Front. Zool.* **1**, 5 (2004).
71. Taylor, D. J., Hebert, P. D. N. & Colbourne, J. K. Phylogenetics and evolution of the *Daphnia longispina* group (Crustacea) based on 12S rDNA sequence and allozyme variation. *Mol. Phyl. Evol.* **5**, 495–510 (1996).
72. Posada, D., Crandall, K. A. & Templeton, A. R. GenDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol. Ecol.* **9**, 487–488 (2000).
73. Templeton, A. R. Nested clade analysis of phylogeographic data: testing hypotheses about gene flow and population history. *Mol. Ecol.* **7**, 381–397 (1998).
74. Hey, J. & Machado, C. A. The study of structured populations — new hope for a difficult and divided science. *Nature Rev. Genet.* **4**, 535–543 (2003).
75. Wright, S. The genetical structure of populations. *Ann. Eugenics* **15**, 323–354 (1951).
76. Wright, S. The interpretation of population structure by *F*-statistics with special regards to systems of mating. *Evolution* **19**, 395–420 (1965).
77. Nei, M. Genetic distance between populations. *Am. Nat.* **106**, 283–292 (1972).
78. Campbell, D. & Bernatchez, L. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Mol. Biol. Evol.* **21**, 945–956 (2004).
79. Beaumont, M. A. & Balding, D. J. Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.* **13**, 969–980 (2004).
80. Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).
81. Hedrick, P. W. A standardized genetic differentiation measure. *Evolution* **59**, 1633–1638 (2005).
82. Wakeley, J. & Hey, J. Estimating ancestral population parameters. *Genetics* **145**, 847–855 (1997).
83. Machado, C. A., Kliman, R. M., Markert, J. A. & Hey, J. Inferring the history of speciation from multilocus sequence data: the case of *Drosophila pseudoobscura* and its close relatives. *Mol. Biol. Evol.* **19**, 472–488 (2002).
84. Nielsen, R. & Wakeley, J. W. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* **158**, 885–896 (2001).
85. Hey, J. & Nielsen, R. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**, 747–760 (2004).
86. Feder, J. L. *et al.* Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proc. Natl Acad. Sci. USA* **100**, 10314–10319 (2003). **This paper presents the unexpected finding that the genetic variation leading to sympatric differentiation between *Rhagoletis* host races originated in an isolated, allopatric population.**
87. Osada, N. & Wu, C.-I. Inferring the mode of speciation from genomic data: a study of the great apes. *Genetics* **169**, 259–264 (2005).
88. Mallet, J. Hybridization as an invasion of the genome. *Trends Ecol. Evol.* **20**, 229–237 (2005).
89. Rieseberg, L. H. & Wendel, J. in *Hybrid Zones and the Evolutionary Process* (ed. Harrison, R. G.) 70–109 (Oxford Univ. Press, New York, 1993).
90. Machado, C. A. & Hey, J. The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proc. R. Soc. Lond. B* **270**, 1193–1202 (2003).
91. Noor, M. A. F. *et al.* The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. *Evolution* **55**, 512–521 (2001).
92. Noor, M. A. F., Grams, K. L., Bertucci, L. A. & Reiland, J. Chromosomal inversions and the reproductive isolation of species. *Proc. Natl Acad. Sci. USA* **98**, 12084–12088 (2001).
93. Feder, J. L., Roethele, J. B., Filchak, K., Niedbalski, J. & Romero-Severson, J. Evidence of inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. *Genetics* **163**, 939–953 (2003).
94. Panithanarak, T. *et al.* Linkage-dependent gene flow in a house mouse chromosomal hybrid zone. *Evolution* **58**, 184–192 (2004).
95. Wolfe, K. H. & Shields, D. C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708–713 (1997).
96. Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S. & Wolfe, K. H. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* **440**, 341–345 (2006). **These authors used whole-genome sequences and functional data to suggest that differential gene loss following genome duplication could have contributed to speciation in yeast.**
97. Lynch, M. & Force, A. G. The origin of interspecific genomic incompatibility via gene duplication. *Am. Nat.* **156**, 590–605 (2000).
98. Ortiz-Barrientos, D., Counterman, B. A. & Noor, M. A. F. The genetics of speciation by reinforcement. *PLoS Biol.* **2**, e416 (2004).
99. Rifkin, S. A., Kim, J. & White, K. P. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nature Genet.* **33**, 138–144 (2003).
100. Gresham, D. *et al.* Genome-wide detection of polymorphisms at nucleotide resolution with a single DNA microarray. *Science* **311**, 1932–1936 (2006).
101. Jaccoud, D., Peng, K., Feinstein, D. & Kilian, A. Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res.* **29**, e25 (2001).
102. Richards, S. *et al.* Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and *cis*-element evolution. *Genome Res.* **15**, 1–18 (2005).
103. Llopart, A., Lachaise, D. & Coyne, J. A. Multilocus analysis of introgression between two sympatric sister species of *Drosophila*: *Drosophila yakuba* and *D. santomea*. *Genetics* **171**, 197–210 (2005).
104. Won, Y. J., Sivasundar, A., Wang, Y. & Hey, J. On the origin of Lake Malawi cichlid species: a population genetic analysis of divergence. *Proc. Natl Acad. Sci. USA* **102** (Suppl. 1), 6581–6586 (2005).
105. Borge, T., Webster, M. T., Andersson, G. & Saetre, G. P. Contrasting patterns of polymorphism and divergence on the Z chromosome and autosomes in two Ficedula flycatcher species. *Genetics* **171**, 1861–1873 (2005).
106. Ramos-Onsins, S. E., Stranger, B. E., Mitchell-Olds, T. & Aguade, M. Multilocus analysis of variation and speciation in the closely related species, *Arabidopsis halleri* and *A. lyrata*. *Genetics* **166**, 373–388 (2004).
107. Dolman, G. & Moritz, C. A multilocus perspective on refugial isolation and divergence in rainforest skinks (Carlia). *Evolution* **60**, 573–582 (2006).
108. Ranz, J. M. & Machado, C. A. Uncovering evolutionary patterns of gene expression using microarrays. *Trends Ecol. Evol.* **21**, 29–37 (2006).
109. Whitehead, A. & Crawford, D. L. Variation within and among species in gene expression: raw material for evolution. *Mol. Ecol.* **15**, 1197–1211 (2006).
110. Ortiz-Barrientos, D., Counterman, B. & Noor, M. A. F. Gene expression divergence and the origin of hybrid dysfunctions. *Genetica* (in the press).
111. Haerty, W. & Singh, R. S. Gene regulation divergence is a major contributor to the evolution of Dobzhansky–Muller incompatibilities between species of *Drosophila*. *Mol. Biol. Evol.* (in the press).
112. Nielsen, M. G., Wilson, K. A., Raff, E. C. & Raff, R. A. Novel gene expression patterns in hybrid embryos between species with different modes of development. *Evol. Dev.* **2**, 133–144 (2000).
113. Ranz, J. M., Namgyal, K., Gibson, G. & Hartl, D. L. Anomalies in the expression profile of interspecific hybrids of *Drosophila melanogaster* and *D. simulans*. *Genome Res.* **14**, 373–379 (2004).
114. Braidotti, G. & Barlow, D. P. Identification of a male meiosis-specific gene, *Tcte2*, which is differentially spliced in species that form sterile hybrids with laboratory mice and deleted in *t* chromosomes showing meiotic drive. *Dev. Biol.* **186**, 85–99 (1997).

115. Michalak, P. & Noor, M. A. F. Association of misexpression with sterility in hybrids of *Drosophila simulans* and *D. mauritiana*. *J. Mol. Evol.* **59**, 277–282 (2004).
The first large-scale test of gene misexpression in species hybrids.
116. Venken, K. J. T. & Bellen, H. J. Emerging technologies for gene manipulation in *Drosophila melanogaster*. *Nature Rev. Genet.* **6**, 167–178 (2005).
117. Coyne, J. A. & Charlesworth, B. Genetic analysis of X-linked sterility in hybrids between three sibling species of *Drosophila*. *Heredity* **62**, 97–106 (1989).
118. Perez, D. E., Wu, C.-I., Johnson, N. A. & Wu, M.-L. Genetics of reproductive isolation in the *Drosophila simulans* clade: DNA marker-assisted mapping and characterization of a hybrid-male sterility gene, *Odysseus* (*Ods*). *Genetics* **134**, 261–275 (1993).
119. Ting, C.-T., Tsaur, S.-C., Wu, M.-L. & Wu, C.-I. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282**, 1501–1504 (1998).
120. Sun, S., Ting, C.-T. & Wu, C.-I. The normal function of a speciation gene, *Odysseus*, and its hybrid sterility effect. *Science* **305**, 81–83 (2004).
121. Greenberg, A. J., Moran, J. R., Coyne, J. A. & Wu, C.-I. Ecological adaptation during incipient speciation revealed by precise gene replacement. *Science* **302**, 1754–1757 (2003).
122. Rong, Y. S. & Golic, K. G. Gene targeting by homologous recombination in *Drosophila*. *Science* **288**, 2013–2018 (2000).
123. Coyne, J. A. & Elwyn, S. Does the *desaturase-2* locus in *Drosophila melanogaster* cause adaptation and sexual isolation? *Evolution* **60**, 279–291 (2006).
124. Presgraves, D. C. A fine-scale genetic analysis of hybrid incompatibilities in *Drosophila*. *Genetics* **163**, 955–972 (2003).
This author thoroughly scanned the *D. simulans* genome for autosomal hybrid inviability factors when paired with a *D. melanogaster* X chromosome, identified the locations of at least 20 such factors, and estimated that recessive incompatibility factors were about eightfold more abundant than dominant ones.
125. Presgraves, D. C., Balagopalan, L., Abmayr, S. M. & Orr, H. A. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature* **423**, 715–719 (2003).
126. Rawson, P. D. & Burton, R. S. Functional coadaptation between cytochrome *c* and cytochrome *c* oxidase within allopatric populations of a marine copepod. *Proc. Natl Acad. Sci. USA* **99**, 12955–12958 (2002).
127. Harrison, J. S. & Burton, R. S. Tracing hybrid incompatibilities to single amino acid substitutions. *Mol. Biol. Evol.* **23**, 559–564 (2006).
The first study to identify a particular single amino-acid change that is sufficient to contribute to a likely barrier to gene flow between copepod populations.
128. Willett, C. S. & Burton, R. S. Viability of cytochrome *c* genotypes depends on cytoplasmic backgrounds in *Tigriopus californicus*. *Evolution* **55**, 1592–1599 (2001).
129. Willett, C. S. & Burton, R. S. Environmental influences on epistatic interactions: viabilities of cytochrome *c* genotypes in interpopulation crosses. *Evolution* **57**, 2286–2292 (2003).
130. Noor, M. A. F. Evolutionary biology: genes to make new species. *Nature* **423**, 699–700 (2003).
131. Wu, C.-I. & Ting, C.-T. Genes and speciation. *Nature Rev. Genet.* **5**, 114–122 (2004).
132. Michalak, P. & Noor, M. A. F. Genome-wide patterns of expression in *Drosophila* pure species and hybrid males. *Mol. Biol. Evol.* **20**, 1070–1076 (2003).
133. Dobzhansky, T. Studies of hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* **21**, 113–135 (1936).
134. Knowles, L. L. & Maddison, W. P. Statistical phylogeography. *Mol. Ecol.* **11**, 2623–2635 (2002).
135. Shimodaira, H. & Hasegawa, M. Multiple comparisons of Log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**, 1114–1116 (1999).
136. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
137. Patterson, N., Richter, D. J., Gnerre, S., Lander, E. S. & Reich, D. Genetic evidence for complex speciation of humans and chimpanzees. *Nature* **441**, 1103–1108 (2006).
138. Palumbi, S. R. All males are not created equal: fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proc. Natl Acad. Sci. USA* **96**, 12632–12637 (1999).
139. McCartney, M. A. & Lessios, H. A. Adaptive evolution of sperm bindin tracks egg incompatibility in neotropical sea urchins of the genus *Echinometra*. *Mol. Biol. Evol.* **21**, 732–745 (2004).
140. Willett, C. S. Deleterious epistatic interactions between electron transport system protein-coding loci in the copepod *Tigriopus californicus*. *Genetics* **173**, 1465–1477 (2006).
This study presents the unexpected (and unexplained) finding that hybrid incompatibility is more strongly associated with heterozygote/homozygote deleterious interactions than homozygote/homozygote (recessive) deleterious interactions.
141. Dallerac, R. *et al.* A Δ9 desaturase gene with a different substrate specificity is responsible for the cuticular diene hydrocarbon polymorphism in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **97**, 9449–9454 (2000).
142. Fang, S., Takahashi, A. & Wu, C.-I. A mutation in the promoter of desaturase 2 is correlated with sexual isolation between *Drosophila* behavioral races. *Genetics* **162**, 781–784 (2002).
143. Barbash, D. A., Roote, J. & Ashburner, M. The *Drosophila melanogaster Hybrid male rescue* gene causes inviability in male and female species hybrids. *Genetics* **154**, 1747–1771 (2000).
144. Barbash, D. A., Siino, D. F., Tarone, A. M. & Roote, J. A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. *Proc. Natl Acad. Sci. USA* **100**, 5302–5307 (2003).
145. Orr, H. A. & Irving, S. Genetic analysis of the *Hybrid male rescue* locus of *Drosophila*. *Genetics* **155**, 225–231 (2000).
146. Swanson, W. J. & Vacquier, D. The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule. *Proc. Natl Acad. Sci. USA* **94**, 6724–6729 (1997).
147. Shaw, A., McRee, D. E., Vacquier, V. D. & Stout, C. D. The crystal structure of lysin, a fertilization protein. *Science* **262**, 1864–1867 (1993).
148. Perez, D. E. & Wu, C.-I. Further characterization of the *Odysseus* locus of hybrid sterility in *Drosophila*: one gene is not enough. *Genetics* **140**, 201–206 (1995).
149. Ritchie, M. G., Halsey, E. J. & Gleason, J. M. *Drosophila* song as a species-specific mating signal and the behavioural importance of Kyriacou & Hall cycles in *D. melanogaster* song. *Anim. Behav.* **58**, 649–657 (1999).
150. Alt, S., Ringo, J., Talyn, B., Bray, W. & Dowse, H. The *period* gene controls cycles in courtship song of *Drosophila melanogaster*. *Anim. Behav.* **56**, 87–97 (1998).
151. Wheeler, D. A. *et al.* Molecular transfer of a species-specific behavior from *Drosophila simulans* to *Drosophila melanogaster*. *Science* **251**, 1082–1085 (1991).
152. Ritchie, M. G. & Kyriacou, C. P. Reproductive isolation and the *period* gene of *Drosophila*. *Mol. Ecol.* **3**, 595–599 (1994).
153. Walter, R. B. & Kazianis, S. *Xiphophorus* interspecies hybrids as genetic models of induced neoplasia. *ILAR J.* **42**, 299–321 (2001).

Acknowledgements
We thank N. Johnson, J. P. Masly, D. Presgraves, M. Taylor and anonymous reviewers for helpful comments on this manuscript. Our research programmes are funded by grants from the US National Science Foundation.

Competing interests statement
The authors declare no competing financial interests.

DATABASES
The following terms in this article are linked online to:
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
desat2 | *Hmr* | *per*

FURTHER INFORMATION
Noor laboratory homepage:
http://www.biology.duke.edu/noorlab
Access to this links box is available online.