A genomic view of introgression and hybrid speciation
Eric J Baack1 and Loren H Rieseberg1,2

Hybridization in plants and animals is more common and has more complex outcomes than previously realized. Genome-wide analyses of introgression in organisms ranging from oaks to sunflowers to fruit flies show that a substantial fraction of their genomes are permeable to alleles from related species. Hybridization can lead to rapid genomic changes, including chromosomal rearrangements, genome expansion, differential gene expression, and gene silencing, some of which are mediated by transposable elements. These genomic changes may lead to beneficial new phenotypes, and selection for fertility and ecological traits may in turn alter genome structure. Dramatic increases in the availability of genomic tools will produce a new understanding of the genetic nature of species and will resolve a century-old debate over the basis of hybrid vigor, while the natural recombinants found in hybrid zones will permit genetic mapping of species differences and reproductive barriers in nonmodel organisms.

Addresses
1 Department of Botany, University of British Columbia, 6270 W University Blvd, Vancouver, BC V6T 1Z4, Canada
2 Department of Biology, Indiana University, 1001 E 3rd Street, Bloomington, IN 47405, USA

Corresponding author: Rieseberg, Loren H (lriesebe@interchange.ubc.ca)

Current Opinion in Genetics & Development 2007, 17:513–518
This review comes from a themed issue on Genomes and evolution
Edited by Susan Wessler, Kelly Dawe and Jim Leebens-Mack
Available online 22nd October 2007
0959-437X/ $ – see front matter © 2007 Elsevier Ltd. All rights reserved.
DOI 10.1016/j.gde.2007.09.001

In 1742, seven years after writing ‘nullae dantur species novae’ (there are no new species) [1] Carolus Linnaeus was brought a fertile floral mutant of Linaria that he called ‘Peloria’. The unusual floral structure convinced Linnaeus that the plant was of hybrid origin, and the fertility of Peloria and other hybrids led Linnaeus to abandon his earlier certainty in the fixed nature of species. Instead, he proposed the radical evolutionary hypothesis that new species could arise via hybridization [2]. Despite this illustrious pedigree, hybrid speciation had little scientific support until early in 20th century when cytogenetic studies showed that hybridization may lead to speciation, especially if accompanied by chromosomal doubling (allopolyploidy). While these studies persuaded many 20th century botanists that hybridization was a common and significant force in evolution, this view was often disputed by zoologists. Now, three centuries after the birth of Linnaeus, hybridization is seen as an important phenomenon in many taxa, contributing to adaptation and speciation in plants, fish, and insects.

The resurgent interest in hybridization is closely linked to the shift from genetic to genomic approaches. In this review, we take a genomic perspective on introgression and hybrid speciation. We limit ourselves to hybridization between sexually reproducing organisms, and so do not consider horizontal gene transfer in prokaryotes, between organelles and the nucleus, or interspecific transfers between organelles (reviewed in [3]). We focus on detecting hybrids and on the genomic and evolutionary consequences of introgression and hybrid speciation, while ignoring the effects of genome duplication.

Introgression and hybrid speciation
Hybridization, the production of viable offspring from interspecific matings, occurs in roughly 10% of animal species and 25% of plant species [4]. It may sometimes lead to introgression, defined here as the stable integration of genetic material from one species into another through repeated back-crossing [5]; the transient presence of alien alleles is excluded as being of little evolutionary importance. Hybridization may also lead to speciation, in which the new hybrid lineages become reproductively isolated from parental populations. While allopolyploid species have complete genomes of each parental species, homoploid hybrid species, which have not undergone chromosomal duplication, typically have unequal parental genomic contributions because of back-crossing.

Challenges in detecting introgression and hybrid speciation, and recent progress
The extent of introgression and hybrid speciation is unclear: occasional hybridization may not lead to permanent genetic exchange, so the frequency of hybridizing species may overestimate the rate of introgression. However, young species have weaker isolating barriers than older species on average, so species that do not currently hybridize may have done so in the past [4]. This uncertainty highlights both the challenges of detecting past hybridization and introgression events and the recent development of many of the essential tools.

Introgression is frequently given as an explanation when phylogenetic trees inferred from different genes are in conflict. However, processes other than introgression can produce incongruent gene trees. For example, if
a large population that is polymorphic at many loci splits into several species, well-supported phylogenetic trees inferred for different genes may reflect the different histories of those genes in the parent population rather than introgression. This uncertainty can lead to conflicting interpretations of genetic data, exemplified by the recent debate over introgression in European oaks [6–8] and in the domestication of rice [9].

An alternative approach to detecting introgression uses comparisons of sequence data from multiple individuals per species. If species do not hybridize, then interspecific differences in sequence should accrue over time, while introgression will lead to decreased levels of interspecific divergence. The patterns of polymorphism and fixed differences depend on the migration rates among species, as well as the population size of ancestral and daughter populations — larger ancestral populations contain higher levels of sequence polymorphism and larger daughter populations retain polymorphisms longer. The statistical models required to estimate this set of parameters are complex and computationally intensive, and estimates of population sizes and migration rates often have wide confidence intervals. However, this approach has been successfully applied to detect introgression in chimps, fish, and butterflies [10–12]. Recent work has introduced an additional complication: patterns of introgression vary across the genome [13]. Alleles at some loci do not introgress while others do so freely, presumably because of variation in the fitnesses of foreign alleles in different genetic or ecological backgrounds. These results suggest that coalescence-based approaches have considerable power, but larger genome-scale analyses will be required to understand whether rampant genetic exchange is the rule for hybridizing species or whether much of the genome is resistant to introgression.

Genomic studies allow robust detection of interspecific genetic exchange, estimation of genome-wide patterns of introgression, and determination of the sizes of parental chromosomal blocks in introgressed populations and hybrid species [14,15]. From a diagnostic standpoint, associations among alleles of one species in the genetic background of a close relative provide compelling evidence of recent introgression. This interspecific linkage disequilibrium also makes it possible to distinguish between ancestral polymorphism and introgression as explanations for phylogenetic incongruence among genes. If a group of species has a history of introgression, two classes of gene trees should be generated, representing the two species involved in the hybridization event. Also, closely linked loci should share gene trees [16].

What have genomic studies of introgression revealed about species barriers?

Genome-wide analyses of introgression in many species, including cottonwood trees [17], mice [18,19], oaks [8], fruit flies [20], mosquitoes [21], and sunflowers [22] consistently find that different portions of the genome vary in their permeability to foreign alleles (Figure 1). There is also tremendous variation in the overall porosity of these genomes; most of the markers studied in cottonwoods did not introgress [17], while only a small fraction of the genome was differentiated in mosquitoes [21].

How do we account for these different introgression patterns? The ease of introgression depends on the

Figure 1

Genetic differentiation between three species of sunflower. *Helianthus annuus* and *H. argophyllus* are sister taxa that have nonoverlapping ranges, while *H. petiolaris* is a more distantly related species that overlaps with *H. annuus*. Genetic differentiation varies dramatically across the genome.
selection: recombination ratio [23], so reproductive barriers should not impede the introgression of neutral or advantageous alleles unless they are tightly linked to sites that contribute to reproductive isolation. Thus, for any given level of hybridization, less introgression is expected between species with complex, multilocus reproductive barriers. Likewise, introgression should be reduced in genomic regions with low levels of recombination. The latter prediction has been confirmed by recent work in Drosophila and Helianthus, in which introgression is diminished in low recombination regions adjacent to chromosomal breakpoints [20,22*]. In both cases, the scale of suppression was surprisingly limited: 5 cM in sunflowers, 2.5 Mb in fruit flies (ca. 12.5 cM).

Linking patterns of introgression to the fitness effects of different loci has been slower, but the reduced introgression of X-linked loci in mice compared to autosomal loci is consistent with the greater abundance of species incompatibilities on sex chromosomes [18,19]. These studies illustrate the potential power of genomic analyses of introgression for identifying the genetic basis of species differences, as these loci are least likely to introgress. Also, the natural recombinants found in hybrid zones or stabilized hybrid lineages permit admixture mapping of species differences and reproductive barriers [24,25]. In sunflowers, for example, chromosomal blocks associated with pollen sterility in natural hybrid zones also exhibit reduced introgression [26].

**Genomic alterations in hybrids**

Hybridization can result in genomic changes including alterations to gene expression, chromosomal structure, and genome size. A series of studies on synthetic allopolyploids (which combine hybridization with genome duplication) reveals that substantial changes occur immediately upon hybridization, including gene loss, gene silencing, changes in gene expression, and tissue-specific expression of copies of some loci from the two genomes [27*]. Genome duplication appears to have a weaker effect on genome change, in one case stabilizing expression patterns that were altered in the diploid hybrid [28]. Mechanistic explanations for these changes are preliminary, but are known to involve transposons in some cases [29,30].

Genomic alterations seen in homoploid hybrid species may be even more dramatic. Analysis of genome structure in three synthetic and three natural hybrid species of sunflower revealed massive karyotypic change over a handful of hybrid generations [31,32]. While some of the karyotypic differences arose through the sorting of chromosomal rearrangements that distinguish the parental species, most arose *de novo*. The natural hybrid species also exhibit increased genome sizes — nearly 50% larger than the parental species [33]. Despite multiple independent origins [34], all species show similar increases in genome size because of the proliferation of retrotransposons [35**]. Screens of F₁ to F₅ synthetic hybrids and of hybrid zone plants failed to detect an increase in genome size [33], so both the trigger and timing of this retrotransposon explosion remains unknown. Lastly, microarray analyses of one of the three hybrid species, Helianthus deserticola, revealed that that approximately 2% of genes had extreme or transgressive expression in the hybrid species [36]. Although these expression differences maybe adaptively important (transport-related proteins were over-represented), it is not clear whether they arose as a consequence of hybridization.

**Evolutionary consequences of introgression and hybrid speciation**

First generation hybrids often exhibit an increase in size, growth rate, and yield. This hybrid vigor or heterosis has been exploited by plant and animal breeders since Darwin's time, but its genetic basis remains controversial [37]. The dominance hypothesis, which has enjoyed long theoretical support, posits that deleterious recessive alleles are complemented in hybrids by fitter alleles from the alternate parent, generating an increase in vigor. In contrast, the overdominance hypothesis attributes hybrid vigor to the synergistic interactions of alleles at a heterozygous locus. However, it is difficult to distinguish true overdominance from the reciprocal complementation of deleterious alleles at linked loci (pseudo-overdominance), leading to scepticism about the former's importance [38]. More recent hypotheses include the possibility that heterosis is caused by synergistic interactions among alleles at different loci (epistasis) or by nonadditive changes in gene expression [39,40]. However, the latter hypothesis is best interpreted as a possible molecular mechanism underlying classical genetic models [41].

Recent gene mapping and expression studies support a pluralist explanation for heterosis. Dominance and overdominance are most frequently implicated by QTL studies, with epistasis reported less frequently [37]. Although QTL studies cannot rule out pseudo-overdominance, the preferential association of overdominant QTLs with heterotic phenotypes in tomato [42], and the demonstration of heterotic effects of the *erecta* mutant in *Arabidopsis thaliana* [43], are consistent with true overdominance. Less progress has been made toward determining the molecular mechanism(s) underlying heterosis. While nonadditive gene expression is common in hybrids, it has not yet been linked to heterotic phenotypes. Also, nonadditive gene expression may sometimes result from mis-expression in hybrids rather than synergistic interactions among alleles [44,45]. Lastly, studies in maize and rice suggest that small-scale duplications and deletions may contribute to heterosis via the complementation of deleted loci [46].

Although heterosis may play an important role in the establishment of asexual or allopolyploid hybrids [47],
its effects are diminished in later generation hybrid segregants such as introgressive lineages or homoploid hybrid species because of increasing homozygosity. Here, success depends on the fixation of favorable new gene combinations from the two species [48]. For example, there are suggestions that new hybrid gene combinations facilitated critical ecological changes in several recently proposed hybrid species, including sculpins [49], Rhagoletis flies [50], and butterflies [51,52]. In these cases, species of hybrid origin have colonized novel habitats, most probably through the expression of transgressive traits. Transgressive segregation, in which hybrids have more extreme trait values than either parent, can occur when parental species contain alleles with opposing effects. For example, a recent study in irises documented that although the upland Iris brevifolia cannot tolerate flooding, the introgression of its alleles into the flood-tolerant Iris fulva increases fulva's survival [53].

The most detailed analysis of the evolutionary forces responsible for fixing favored combinations of parental alleles or chromosomal segments in hybrids focused on recombination in the homoploid hybrid sunflowers [31]. Analysis revealed that selection on ecological traits had slightly higher power than selection for fertility when predicting hybrid genomic composition, though both forms of selection were significant. Also, simulation studies suggest that while reproductive isolation between the recombinant hybrid species and the parental species arose in a few generations, the complete elimination of interspecific chromosomal polymorphism required several hundred generations [14]. The genomes of animal species of hybrid origin have received little attention as yet, so the factors determining their composition are unknown.

Progress is also being made toward understanding the molecular mechanisms underlying phenotypic variation in hybrids [54]. For example, comparison of flowering time in synthetic tetraploids of Arabidopsis arenosa and A. thaliana revealed that while the autotetraploids flower early, the synthetic allotetraploid plants flower late, mimicking the late flowering of the natural allotetraploid between the same two parents. The late flowering phenotype results from complementation of reciprocally down-regulated loci in the two parental species, as well as epigenetic changes, which in turn correlate with histone acetylation and methylation. These changes brought about by hybridization probably have important fitness consequences. Early flowering plants may escape drought while late flowering plants may be more reliably pollinated and may have more resources to allocate to flowering. In addition, the shift in flowering time results in an immediate reproductive barrier between allotetraploid and its parental species.

**Future steps**

Genomic data have already reshaped our understanding of the nature of species barriers, revealing that key morphological and ecological differences can persist even as genes move between species at many loci. So far, most of these insights derive from species that are currently hybridizing, but inexpensive sequence data make it feasible to estimate the extent and timing of introgression at all stages of species divergence. Likewise, the ability to rapidly and inexpensively estimate the size and parentage of chromosomal segments in hybrids makes it possible to determine the importance of different kinds of reproductive barriers in resisting introgression, estimate the ages of hybrid zones, calculate the speed of hybrid speciation, and identify the factors responsible for hybrid genomic composition. Finally, advances in genomics and molecular biology should rapidly reveal the genetic and molecular mechanisms underlying heterosis and other phenotypic changes in hybrids.

**Acknowledgements**

The authors were supported in this work by grants from the National Institutes of Health (USA), the National Science Foundation (USA), and the National Sciences and Engineering Research Council (Canada).

**References and recommended reading**

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

In this study, four loci were sequenced from two species of butterfly, Heliconius cydno and H. melpomene, both of which occur in Panama. Two loci were reciprocally monophyletic but alleles of the other two were shared between the two species. Bayesian modeling suggests that this is because of introgression at one of the loci and ancestral polymorphism at the other.


This study looks at differentiation of coding and noncoding loci in three species of sunflower that originated from recent exposure to cultivated sorghum. Mol Ecol 2005, 14:2143-2154.


518 Genomes and evolution


Hybrid speciation in animals was thought to be much rarer than in plants, but recent reports are narrowing the gap. Here, genetic data document the hybrid origin of a butterfly in the genus *Lycaenides*. The hybrid species occupies a novel alpine habitat distinct from that of either parent.


One of the best studies connecting hybridization to changes in a phenotype with adaptive significance, including a detailed analysis of the genetic changes. The authors elucidate the factors leading an allotetraploid species to have a later flowering time than either of its diploid progenitors.