Genomics of hybridization and its evolutionary consequences

RICHARD J. ABBOTT,* NICHOLAS H. BARTON† AND JEFFREY M. GOOD‡

*School of Biology, University of St Andrews, Mitchell Building, St Andrews, Fife KY16 9TH, UK; †Institute of Science and Technology (IST Austria), Am Campus 1, A-3400 Klosterneuburg, Austria; ‡Division of Biological Sciences, University of Montana, 32 Campus Drive, Missoula, MT 59812, USA

Hybridization and its consequences have been of longstanding interest to evolutionary biologists. Darwin (1859) included a chapter on hybrids and the expression and causes of hybrid sterility in *The Origin of Species*, while the main proponents of the neo-Darwinian synthesis discussed the topic at varying length in the mid-1900s (Haldane 1932; Dobzhansky 1937; Stebbins 1950; Mayr 1963). Subsequently, Grant (1981) underlined the importance of hybridization to plant evolution, devoting six chapters of *Plant Speciation* to the topic, while Arnold (1997; 2006; 2015) has written three books on hybridization and its evolutionary consequences over the last 20 years.

Interest in hybridization stems from its importance in understanding the origin and maintenance of breeding barriers between diverging populations, the introduction of adaptive variation to species through introgression, and as a mechanism for the origin of new species through either homoploid hybrid speciation or allopolyploidy (Abbott *et al.* 2013). More broadly, study of natural hybridisation can help identify selected genes, and can yield estimates of selection and gene flow that would be impossible to measure directly.

Since the mid-1960s, molecular markers have been used to examine these diverse aspects of hybridization with the most recent approaches involving analyses of single nucleotide polymorphisms (SNPs) through comparison of genome sequences. By means of genomic analyses, signatures of hybridization

can be detected at a fine scale, shedding light on the reticulate evolutionary history of organisms, rates of current and historical gene flow between diverging lineages, and levels of selection that prevent hybridizing species from merging. Furthermore, in combination with functional gene analyses, genomic comparisons can identify the genes and gene combinations involved in maintaining hybridizing species as discrete entities, determine which genes and gene combinations have been introgressed into species, and which of these are of adaptive significance. In this special issue on the **Genomics of Hybridization** a collection of papers is presented which touches on these diverse but interrelated aspects of hybridization. The issue leads off with a review by Payseur & Rieseberg of the genomic methods used to detect and characterize hybridization between diverging lineages, and is followed by papers that address the detection and adaptive significance of introgression, hybrid zones, isolating barriers, hybrid speciation and the stabilization of hybrid genomes.

Payseur & Rieseberg (2016) make clear that a wide range of analytical methods have been used in the genomic analysis of hybridization in diverse organisms. They note the strengths, weaknesses and limitations of these methods pointing out that "few approaches reconstruct the magnitude and timing of gene flow, estimate the fitness of hybrids, or incorporate knowledge of recombination rate." Proposals are made for future research to remedy these deficiencies and, in turn, provide an improved understanding of the speciation process. They observe that consequences of hybridization appear to differ across taxa, although given the limited range of taxa analysed so far, no firm conclusions are drawn on this matter other than that there is clear variation of effect. They also note that loci resistant to gene flow and contributing to isolating barriers

between diverging lineages tend to be more common in sex chromosomes and in regions of reduced recombination.

Detection of introgression and adaptive significance.

The ability to detect introgression (the exchange of genes between diverging lineages that results from backcrossing of hybrids with one or both parents), and to determine how and why it varies across the genomes of hybridizing taxa is fundamental to most genomic studies of hybridization.

Incongruence between phylogenetic trees constructed from different gene sequences has traditionally provided a first indication that hybridization and introgression have occurred between the lineages analysed. However, phylogenomic analyses of taxa in which hybridization is known to occur need to distinguish between the effects of incomplete lineage sorting (ILS) and introgression (reticulation) as causes of such incongruence. It has recently been shown using the multispecies network coalescent (MSNC) that it is possible to construct a phylogenetic network accounting simultaneously for both ILS and reticulation effects. Using this approach, Wen *et al.* (2016) confirm a reticulate evolutionary history of the six genomes comprising the *Anopheles gambiae* (mosquito) complex, and further reveal the reticulate history of all chromosomes within these genomes.

Conflicts between taxonomic relatedness based on morphology and molecular phylogenetic relatedness can also result from introgression, or alternatively ILS or convergent evolution. A phylogenomic analysis using transcriptome SNP data by Ru *et al.* (2016) of related taxa of spruce (*Picea*) native to the Qinghai Tibet Plateau, China, shows that two varieties of one spruce

species are polyphyletic with one variety nesting in the clade of another species. Further analysis rules out ILS as a cause of this conflict and indicates that there has been extensive and bidirectional introgression between the two varieties. This, together with parallel evolution in arid habitats, is considered to have caused the morphological similarity between the two spruce varieties.

Particular problems with detecting introgression arise when it is rare and restricted to small parts of the genome. Rosenzweig *et al.* (2016) report the development of a new test for introgression, *RNDmin*, based on summary statistics between sister lineages relative to divergence to an outgroup, showing that it is robust to variation in mutation rate and inaccurate estimates of divergence time. They apply this test to two sister species of mosquito and identify three novel candidate regions for introgression one of which is located on the X chromosome, but outside an inversion that distinguishes the two species.

Further problems with detecting introgression are especially evident when one lineage involved is extinct. However, such problems may be overcome if ancient DNA is available for the extinct lineage. Research conducted on detecting admixture using archaic genomes is reviewed by Schaeffer *et al.* (2016). Notable successes have been obtained for a range of animals, although the majority of investigations have focused on admixture between humans and archaic hominins. Most recent estimates indicate that the proportion of Neanderthal ancestry in modern Eurasians is 1.5-2.1%, while the proportion of Denisovan ancestry in Melanesian and Australian aboriginal genomes is 3-6%. The latter is reduced to 0.2% in East Asian and Native American genomes.

Problems with detecting introgression also arise when hybridizing taxa differ in ploidy level (e.g. diploid versus tetraploid) due to difficulties in determining allele frequencies in polyploids because of gene replication. Zohren *et al.* (2016) develop a new method for genotyping tetraploid individuals across a large number of RAD loci and use this to show that unidirectional introgression has occurred from two diploid species into a widespread tetraploid species of birch (*Betula*). This is in keeping with Stebbins' rule (1971) that when introgression occurs across a ploidy barrier it normally proceeds from a diploid into a tetraploid, rather than in the reverse direction. Zohren *et al.*'s study also shows that the introgression pattern in birches indicates that the tetraploid picked up genes from one of the diploid species as it expanded its range northwards in the UK during the Holocene, replacing the diploid throughout much of its former range.

Although introgression is now regarded as common between hybridizing species, few studies have demonstrated that it is of adaptive significance. Suarez-Gonzalez *et al.* (2016) report a study of patterns of introgression and signals of local adaptation in two species of poplar (*Populus*) in North America. In admixed individuals of one species they locate a telomeric region of one chromosome that was introgressed from the other species and contains several candidate genes for local adaptation. Interestingly, this introgressed region is paralogous to a block of genes found in another chromosome that shows no signs of introgression or signatures of selection.

Hybridization and isolation across hybrid zones

Hybrid zones have been termed natural laboratories for evolutionary studies (Hewitt 1988) and are excellent systems for simultaneously examining introgression and isolating barriers between diverging lineages. Harrison & Larson (2016) review the importance of hybrid zones for studying speciation, placing particular emphasis on patterns of heterogeneous genome divergence and whether these may reflect the genetic architecture of adaptation and reproductive isolation between taxa. They further point out the problems of using such patterns to distinguish primary intergradation from secondary contact in the formation of hybrid zones. Distinguishing between these two modes of origin of hybrid zones is a major focus of a paper by Filatov *et al*. (2016), which examines a hybrid zone between two ragwort (*Senecio*) species that occurs across an elevational gradient on Mount Etna in Sicily. Using a model based analysis of nextRAD SNP data, rates of gene flow between the two species are estimated to be significantly higher at the time of divergence than currently. It is proposed that this rules out an origin of the hybrid zone by secondary contact and favours one of divergence with gene flow.

Additional studies of introgression and isolation across hybrid zones are presented in papers by Christe *et al.* (2016), Kenney & Sweigart (2016), Senerchia *et al*. (2016) and von Rönn *et al.* (2016). Christe *et al.* report the occurrence of replicate F_1 dominated hybrid zones between two species of poplar in Central Europe based on an analysis of SNPs generated by RAD sequencing. Although ancient introgression was detected between these two species, early-generation recombinants exhibit low relative survivorship under common garden conditions. Thus, despite F_1 fertility, selection against recombinant genotypes acts as an effective barrier to gene flow between these

hybridizing species. Kenney & Sweigart show that at one location in the USA introgression proceeds asymmetrically from a selfing to an outcrossing species of *Mimulus* (monkeyflowers). They report that average level of mixed ancestry varies greatly across the genome of the outcrossing species and hybrids, and that a QTL previously identified as affecting photoperiod overlaps a region of particularly low introgression. They further report that high levels of divergence between the two species are associated with a low recombination rate within genomic regions. It is concluded that isolating barriers between the two species involve differences in mating system and flowering time, and most likely selection against alleles of the selfing species in the genetic background of the outcrossing species.

Senerchia *et al*. focus on the introgression of long-terminal repeat retrotransposons (LTR-RTs) between two tetraploid, selfing species of wild wheat (*Aegilops*) under both wild and experimental conditions. They show that in contrast to AFLPs and quiescent LTR-RTs, active LTR-RTs are specifically (non-randomly) reorganised in the genomes of viable hybrids. Furthermore, it is apparent that significantly more active LTR-RT loci are absent from natural hybrids relative to AFLPs and quiescent LTR-RTs. Such preferential elimination of active LTR-RT sequences suggests that only hybrids that have repressed conflicting LTR-RT interactions survive in natural hybrid zones. It is concluded that active LTR-RTs, which distinguish the two species, are important contributors to intrinsic post-zygotic reproductive isolation between them.

The study by von Rönn *et al*. (2016) concerns morphological and genomic comparisons between two groups of European barn swallow (*Hirundo rustica rustica*), which differ in migratory phenotype. One group migrates from Northern

Europe to southern Africa, the other from Southern Europe to West-Central Africa. A contact (hybrid) zone between these two groups occurs in northern Germany and contains birds of both parental types as well as intermediates. Although there is clear divergence between the two parental types in migration-relevant morphological phenotype, an examination of ~24,000 RAD SNP loci revealed only one to be a high Fst outlier. Birds with intermediate morphology in the contact zone show reduced survival, but there is no indication of assortative mating of parental types. The very low genomic divergence found between the two migratory types of this species indicates that gene flow across the migratory divide is sufficient to prevent the build-up of genome-wide differentiation despite divergent selection.

More on isolating barriers

Harrison and Larson emphasize in their paper that divergent selection in the face of gene flow is only one of several possible causes of heterogeneous genome divergence between taxa. Genomic islands of relative divergence, as measured by F_{st}, may be due to reduced diversity, rather than increased divergence (Cruickshank and Hahn, 2014). Reduced diversity may in turn be due either to high background selection, associated with reduced recombination, or to selective sweeps. In the latter case, genomic islands indicate selected loci, and that selection may contribute to reproductive isolation. However, the islands themselves would not have been generated by a barrier to gene flow.

The problem of discounting alternative explanations for presence of genomic islands of divergence is analysed in depth by Yeaman *et al.* (2016). They explore conditions for the evolution of such islands, focusing primarily on the effect of linkage on the establishment probability of new mutations and how

migration and selection affect this process. Further development of this approach should lead to an ability to distinguish between alternative explanations for the occurrence of genomic islands of divergence between taxa and whether they contain genes contributing to local adaptation and reproductive isolation.

Several studies have identified associations amongst unlinked loci in admixed populations, and have suggested that these reveal epistatic interactions between incompatibility loci. Schumer and Brandvain (2016) simulate the distribution of such "ancestry disequilibrium", and find an excess of positive associations, even in the absence of selection against hybrids. Simulations that include incompatibilities show that these can in principle be detected through excess ancestry disequilibrium, but that the signal is hard to disentangle from the neutral background.

Genomic analyses of isolating barriers between particular pairs of hybridizing taxa are presented in turn by Pease *et al.* (2016) and Hu & Filatov (2016). Pease *et al.* report the results of a transcriptomic analysis of mechanisms of postmating prezygotic isolation between a self-incompatible (SI) and a self-compatible (SC) species of tomato (*Solanum*). Unilateral incompatibility (UI) exists between these species such that pollen from the SI species germinates and grows in the pistil of the SC species, whereas pollen of the SC species fails to fully grow in the pistil of the SI species. Such unilateral incompatibility is often, though not always, found between pairs of hybridizing SI and SC flowering plants. Pease et al. identify five strong candidate genes for involvement in UI between the two tomato species, two of which are firmly supported by prior developmental, functional and QTL mapping studies.

Hu & Filatov (2016) demonstrate that the disproportional effect of the X-chromosome on reproductive isolation between hybridizing species (the large X-effect), often reported in animals, is also apparent in two dioecious species in the plant genus *Silene* (campions). Sex chromosomes are rarely found in flowering plants and in the two species of *Silene* investigated most X-linked genes have Y-linked homologs, unlike the situation that is common for animal sex chromosomes (see Dufresnes *et al.* 2016 for exceptions). Thus, whereas in most animals the large X-effect is often explained by the fact that X-linked genes are hemizygous and therefore recessive genes are frequently expressed resulting in Bateson-Dobzhansky-Muller type incompatibility in interspecific hybrids, other causes must explain the large X-effect in *Silene*.

An under-appreciated postzygotic isolating barrier between plants is seed lethality, stemming from the failure of endosperm development in hybrid seed produced from crosses between species of either equivalent or different ploidy. This isolating barrier depends on the direction of the cross and thus of parent-of-origin effects. Epigenetic mechanisms regulating parental genome dosage in the endosperm, which is triploid (two maternal and one paternal genomes), are proposed as causes of hybrid endosperm defects. Lafon-Placette & Köhler (2016) review recent research that is clarifying the developmental mechanisms involved and the evolutionary forces that drive the establishment of this barrier between flowering plant species.

Hybrid speciation and genome stabilization

Hybrid speciation may occur in two ways, either with no change in chromosome number (homoploid hybrid speciation), or with a doubling of chromosome number in the hybrid relative to its parents (allopolyploidy). Homoploid hybrid speciation is

considered to be rare (Yakimowski & Rieseberg 2014), whereas allopolyploidy is regarded as common in plants (e.g., Barker *et al*. 2016), though not in animals. Detection of a homoploid hybrid species is difficult (Rieseberg 1997), and clearly a first step is to show that the species is truly of hybrid origin, a process greatly improved by genomic analysis. It was originally proposed by Grant (1981) that a Californian sunflower (*Helianthus bolanderi*) originated through homoploid hybridization, though this was not supported by results from a later allozyme analysis (Rieseberg *et al*. 1988). To overcome any doubts on this matter, Owens *et al*. (2016) conducted a genomic analysis on the putative hybrid taxon and its putative parent species, to show convincingly that the supposed hybrid taxon was clearly not hybrid. Their analysis revealed, however, that introgression has occurred from this taxon into what had previously been considered one of its putative parents, the invader *H. annuus*, as predicted by theory.

Gallagher *et al*. (2016) review and recommend network approaches for analyzing the effects of genome duplication on the ecology and evolution of polyploid plants, and particularly the novel interactions that occur among duplicated genomes that affect gene expression and patterns of divergence, recruitment, and loss of duplicated genes. They demonstrate the utility of this approach by focusing on the evolution of cotton fibres in the allopolyploid *Gossypium hirsutum,* the species that provides most of the world's cotton. They show how co-expression network analysis of the whole fibre transcriptome can identify the co-expression and interaction of genes involved in the development of this trait, and whether there is differential (biased) usage of duplicated genes (homeologs).

Early generation hybrids are genomically unstable and will generate large amounts of genetic variation through segregation, transposon activation and

epigenetic reprogramming etc. Stable hybrid genomes evolve when parts of each parental haplotype become fixed across individuals enabling stable admixed ancestry to be maintained over subsequent generations. Previous studies suggest that the rate of hybrid genome stabilization varies across taxa and is dependent on demographic differences, selection or strong genetic drift. Here, Schumer *et al.* (2016) examine a swordtail fish whose genome shows considerable ancient hybrid ancestry. They show that certain functionally important regions of the genome are devoid of hybrid ancestry, most likely due to the effects of selection. In contrast, hybridization-derived regions tend to show reduced functional constraint. They emphasise that "More research is needed to understand whether there are general rules for how quickly genomes stabilize after hybridization and which processes most commonly drive this stabilization."

Concluding remarks

Genomic analysis provides the means for detecting hybridization and introgression between diverging lineages, and determining the evolutionary consequences of such processes. Although most species do not hybridize in the wild (Schwenk *et al.* 2008; Whitney *et al.* 2010), genomic analyses increasingly show that many animal and plant species have experienced admixture during their evolutionary history. Whether such admixture is of adaptive significance remains largely untested, although in some instances it is (see Suarez-Gonzalez *et al.* 2016). The papers presented in this special issue emphasise the power of genomic analysis in detecting admixture, determining its adaptive significance, detecting the genetic basis of isolating barriers that prevent admixture, and

resolving evolutionary relationships between lineages that have experienced

admixture. Genomic analyses will undoubtedly become more sophisticated in the

future to provide more detailed information on the effects of hybridization (e.g.,

Gallagher *et al.* 2016). Molecular ecologists, however, should not lose sight of the

equal importance of experimental and field studies to evaluate the evolutionary

and ecological consequences of hybridization and introgression. Very few

studies, for example, combine genomic analyses, with measures of fitness of

admixed individuals under common garden conditions or in the wild. In studies

of hybrid zones, in particular, this should be attempted whenever possible (see

Christe *et al.* 2016; von Rönn *et al.* 2016), as a complement to genomic analyses,

so as to obtain a more complete understanding of the factors that may allow

species to originate and be maintained in the face of gene flow.

References

Abbott RJ, Albach D, Ansell S et al. (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, 26, 229–246.

Arnold ML (1997) *Natural Hybridization and Evolution*. Oxford University Press, Oxford, UK.

Arnold ML (2006) *Evolution Through Genetic Exchange*. Oxford University Press, Oxford, UK.

Arnold ML (2015) *Divergence with Genetic Exchange*. Oxford University Press, Oxford, UK.

Barker MS, Arigo N, Baniaga AE, Li Z, Levin DA (2016) On the relative abundance of autopolyploids and allopolyploids. *New Phytologist*, 210, 391-398.

Christe C, Stölting KN, Bresadola L, Fussi B, Heinze B, Wegmann D, Lexer C (2016) Selection against recombinant hybrids maintains reproductive isolation in hybridizing *Populus* species despite F1 fertility and recurrent gene flow. *Molecular Ecology, ….*

Cruickshank T. E., Hahn M. W. (2014)   Reanalysis suggests that genomic

islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* 23, 3133–3157.

Dobzhansky T (1937) *Genetics and the Origin of Species*. Columbia University Press, New York.

Dufresnes C, Majtyka T, Baird SJE, Gerchen JF, Borzée A, Savary R, Ogielska M, Perrin N, Stöck M (2016) Empirical evidence for large X-effects in animals with undifferentiated sex chromosomes. *Scientific Reports*, 6, 21029.

Filatov DA, Osborne OG, Papadopoulos AS (2016) Demographic history of speciation in a *Senecio* altitudinal hybrid zone on Mt. Etna. *Molecular Ecology*, …

Gallagher JP, Grovera CE, Hua G-J, Wendel JF (2016) Insights into the ecology and evolution of polyploid plants through network analysis. *Molecular Ecology*, …

Grant V (1981) *Plant Speciation*, 2nd edition. Columbia University Press, New York.

Haldane JBS (1932) *The Causes of Evolution*. Longmans Green & Co., London.

Harrison RG, Larson EL (2016) Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. Molecular Ecology, ….

Hewitt GM (1988) Hybrid zones: natural laboratories for evolutionary studies. *Trends in Ecology & Evolution*, 3, 158–167.

Hu X-S, Filatov DA (2016) The large-X effect in plants: Increased species divergence and reduced gene flow on the *Silene* X-chromosome. Molecular Ecology, …

Kenney AM, Sweigart AL (2016) Reproductive isolation and introgression between sympatric *Mimulus* species. *Molecular Ecology*, ..

Lafon-Placette C, Köhler C (2016) Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers. *Molecular Ecology*, ..

Mayr E (1963) *Animal Species and Evolution*. Belknap Press, Cambridge, MA.

Owens GL, Baute GJ, Rieseberg LH (2016) Revisiting a classic case of introgression: hybridization and gene flow in Californian sunflowers. *Molecular Ecology*, ….

Payseur BA, Rieseberg LH (2016) A genomic perspective on hybridization and speciation. *Molecular Ecology*, ….

Pease JB, Guerrero RF, Sherman NA, Hahn MW Moyle LC (2016) Molecular mechanisms of post-mating prezygotic reproductive isolation revealed by transcriptome analysis. *Molecular Ecology*, ….

Rieseberg LH (1997) Hybrid origins of plant species. *Annual Review of Ecology and Systematics*, 28, 359 – 389.

Rieseberg LH, Soltis DE, Palmer JD (1988) A molecular reexamination of introgression between *Helianthus annuus* and *H. bolanderi* (Compositae). *Evolution*, 42, 227-238.

Ru D-F, Mao K-S, Zhang L, Wang X-J, Lu Z-Q, Sun Y-S (2016) Genomic evidence for polyphyletic origin and inter-lineage gene flow in *Picea brachytyla* in the Qinghai-Tibet Plateau. *Molecular Ecology*, ….

Schaefer NK, Beth Shapiro B, Green RE (2016) Detecting hybridization using ancient DNA. *Molecular Ecology*,…

Schumer M., Brandvain Y. (2016) Determining epistatic selection in admixed populations. *Molecular Ecology* …

Schumer M, Rongfeng Cui R-F, Powell DL, Rosenthal GG, Andolfatto P (2016) Ancient hybridization and genomic stabilization in a swordtail fish. *Molecular Ecology*,…

Schwenk K, Brede N, Streit B. 2008 . Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences,* 363, 2805 – 2811.

Senerchia N, Felber F, North B, Sarr A, Guadagnuolo R, Parisod C (2016) Differential introgression and reorganization of retrotransposons in hybrid zones between wild wheats. *Molecular Ecology*,…

Stebbins GL (1950) *Variation and Evolution in Plants*. Columbia University Press, New York.

Stebbins GL (1971) *Chromosomal Evolution in Higher Plants*. London, UK: Edward Arnold.

Suarez-Gonzalez A, Hefer CA, Christe C, Corea O, Lexer C, Cronk QCB, Douglas CJ (2016) Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). *Molecular Ecology*, …..

von Rönn JAC, Shafer ABA, Wolf JBW (2016) Disruptive selection without evolution across a migratory divide. *Molecular Ecology*, …..

Wen, D-Q, Yu Y, Hahn MW, Nakhleh L (2016) Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. *Molecular Ecology*, …..

Whitney KD, Ahern JR, Campbell LG, Albert LP, King MS (2010) Patterns of hybridization in plants. *Perspectives in Plant Ecology, Evolution and Systematics*, 12, 175-182..

Yeaman S, Aeschbacher S, Bürger R (2016) The evolution of genomic islands by

increased establishment probability of linked alleles. *Molecular Ecology*, …..

Yakimowski SB, Rieseberg LH (2014) The role of homoploid hybridization in evolution: a century of studies synthesizing genetics and ecology. *American Journal of Botany*, 101, 1247-1259.

Zohren J, Wang N, Kardailsky I, Borrell JS, Joecker A, Nichols RA, Buggs RJA (2016) Unidirectional diploid-tetraploid introgression among British birch trees with shifting ranges shown by RAD markers. *Molecular Ecology*, …..