

REVIEW

Next-generation hybridization and introgression

AD Twyford^{1,2} and RA Ennos³

Hybridization has a major role in evolution—from the introgression of important phenotypic traits between species, to the creation of new species through hybrid speciation. Molecular studies of hybridization aim to understand the class of hybrids and the frequency of introgression, detect the signature of ancient hybridization, and understand the behaviour of introgressed loci in their new genomic background. This often involves a large investment in the design and application of molecular markers, leading to a compromise between the depth and breadth of genomic data. New techniques designed to assay a large sub-section of the genome, in association with next-generation sequencing (NGS) technologies, will allow genome-wide hybridization and introgression studies in organisms with no prior sequence data. These detailed genotypic data will unite the breadth of sampling of loci characteristic of population genetics with the depth of sequence information associated with molecular phylogenetics. In this review, we assess the theoretical and methodological constraints that limit our understanding of natural hybridization, and promote the use of NGS for detecting hybridization and introgression between non-model organisms. We also make recommendations for the ways in which emerging techniques, such as pooled barcoded amplicon sequencing and restriction site-associated DNA tags, should be used to overcome current limitations, and enhance our understanding of this evolutionary significant process.

Heredity (2012) **108**, 179–189; doi:10.1038/hdy.2011.68; published online 7 September 2011

Keywords: next-generation sequencing; hybridization; introgression; reticulate evolution; single-nucleotide polymorphisms (SNPs); restriction site-associated DNA (RAD) tags

INTRODUCTION

Hybridization, the crossbreeding between individuals of different species, and introgression, the transfer of genes between species mediated primarily by backcrossing, have been the focus of evolutionary studies over many decades (see Anderson, 1949; Arnold, 1992; Rieseberg and Carney, 1998). Hybridization is potentially a creative evolutionary process, allowing genetic novelties to accumulate faster than through mutation alone (Anderson and Hubricht, 1938; Martinsen *et al.*, 2001). This may increase allelic variation at selectively neutral loci, and transfer adaptively important genetic variation, which may increase the fitness of the introgressed lineage (Choler *et al.*, 2004; Martin *et al.*, 2006; Castric *et al.*, 2008; Kim *et al.*, 2008). Moreover, hybridization can have a role in speciation. Hybridization in association with whole-genome duplication (polyploidy) is considered a likely route to speciation, particularly in plants (Hegarty and Hiscock, 2008). The difference in ploidy levels between the polyploid hybrid and diploid progenitors acts as a strong reproductive barrier (Soltis *et al.*, 2004), although there are examples of introgression across ploidy levels (for example, *Senecio*; Chapman and Abbott, 2010). Hybrid speciation can also occur without a change in chromosome number (homoploid hybrid speciation), where the hybrid lineage is ecologically or spatially divergent from the parental progenitors (Gross and Rieseberg, 2005; Abbott *et al.*, 2010).

The degree of hybridization and introgression in natural systems is limited by reproductive isolating barriers, and increasing evidence support these as permeable filters to gene flow, which may not prevent it entirely (Mallet, 2005; Slotte *et al.*, 2008). Therefore, rampant gene

flow may occur between species where their distribution patterns overlap and they interact, so much so that introgression has been described as an ‘invasion of the genome’ (Mallet, 2005). This is consistent with the increasing frequency with which hybridization is reported, with between 1 and 10% of animals and 25% of plant species known to hybridize with at least one other species (Mallet, 2005; Schwenk *et al.*, 2008). This ubiquity of hybridization and introgression confirms that it is a widespread evolutionary phenomenon.

Studies increasingly use detailed molecular tools to understand the dynamic nature of hybridization and introgression. Ideally, these studies aim to have a good coverage of markers distributed across the genome, with a high marker density, in order to accurately detect introgressed linkage blocks. However, this idealized situation is far from reality in all but the most well-developed model systems (Rieseberg *et al.*, 2000; Dempewolf *et al.*, 2010). A major limitation when assessing introgression is the availability of genetic resources to accurately estimate interspecific gene flow; where insufficient molecular markers or gene sequences are studied, cryptic introgression is likely to go undetected (Currat *et al.*, 2008). Moreover, new tools are required to assess the type of genes that may be passing across species boundaries, and how these interact with the recipient genome. Next-generation sequencing (NGS) technologies, which generate a large quantity of nucleotide sequence data from complex nucleic acid populations (Metzker, 2010), promise to improve vastly our ability to study hybridization and introgression, by allowing new genomic tools to be generated for organisms with no prior sequence data (Hohenlohe *et al.*, 2011). Recent reviews have described the technical

¹Royal Botanic Garden, 20A Inverleith Row, Edinburgh, UK; ²Institute of Molecular Plant Sciences, School of Biological Sciences, University of Edinburgh, Edinburgh, UK and

³Institute of Evolutionary Biology, School of Biological Sciences, Ashworth Laboratories, University of Edinburgh, Edinburgh, UK

Correspondence: AD Twyford, Royal Botanic Garden, 20A Inverleith Row, Edinburgh EH3 5LR, UK.

E-mail: a.twyford@rbge.org.uk

Received 26 February 2011; revised 17 June 2011; accepted 27 June 2011; published online 7 September 2011

background of these technologies (for example, Metzker, 2010) and a number of their diverse applications (for example, for understanding the genetic basis of adaptation, Stapley *et al.*, 2010; the use of transcriptomics, Bräutigam and Gowik, 2010).

In this review, we describe how NGS technologies can be used to study hybridization and introgression, and the theoretical issues that must be assessed before embarking on such studies. Our main aim is to highlight how NGS technologies can be used to bridge the traditional divide between population genetic studies, where many markers are surveyed for a large number of individuals from a few species, and molecular systematic studies, where in-depth sequence data are generated for a few loci in a limited number of individuals from each of a large number of species. The generation of in-depth genomic data for many individuals will significantly aid our understanding of the genetics of introgression, and we relate this to three major questions: What is the frequency of introgression between hybridizing species in the wild? How significantly has ancient hybridization contributed to the evolutionary process? What is the behaviour of introgressed loci in their new genomic backgrounds? We largely draw our examples from the plant literature, where hybridization has long been considered an important evolutionary force (Arnold, 1992; Rieseberg and Carney, 1998), but also include examples from studies of animal hybridization, where it is increasingly being appreciated as an evolutionary stimulus (Mallet, 2005; Jiggins *et al.*, 2008; Schwenk *et al.*, 2008). We start by comparing population genetic and phylogenetic approaches to studying hybridization. We then highlight the methodological difficulties with these current approaches, and suggest how NGS technologies will best be used to resolve these issues. Finally, we assess the potential implications of genomic introgression studies for understanding the significance of natural hybridization in evolution.

APPROACHES USED TO DETECT HYBRIDIZATION AND INTROGRESSION

Many methods have been used to detect hybridization, including critical examination of patterns of morphology, cytology, secondary chemistry and molecular markers (Rieseberg and Ellstrand, 1993). The increase in the use of molecular markers for studying patterns of hybridization is similar to that seen in other areas of evolutionary biology, and this is largely due to the ability to apply molecular markers in a wide range of situations and analyse the data by using a robust statistical framework based on our current knowledge of evolutionary theory (Rieseberg *et al.*, 2000). Two widely adopted approaches can be used to detect hybridization at different temporal and spatial scales. Molecular phylogenetic approaches can be used to identify hybridization events by surveying many species, whereas population genetic studies allow a more detailed assay to confirm the class of hybrids and the number of genes being introgressed.

Phylogenetic approach

Gene tree reconstructions can be used to infer incongruence and identify potential hybridization and introgression events (see Linder and Rieseberg, 2004). This is because phylogenetic reconstructions of hybridizing taxa using multiple independent sources of genetic information, such as low-copy nuclear markers often have polyphyletic signatures (Mao *et al.*, 2010). These approaches can not only be used to identify the parents of recent hybrids of unknown parentage, but also to infer ancient hybridization events. Where alternate genealogies are supported for tightly linked genes, this can be used to infer the introgression of chromosome blocks (Hobolth *et al.*, 2007). Interpreting phylogenetic reconstructions of formerly hybridizing taxa is

challenging (Willyard *et al.*, 2009) as reticulation may obscure the pattern of bifurcation, and the sequence evolution of these species will more closely fit a net-like rather than a tree-like pattern over time. Therefore, the aim of these studies is to identify non-recombinant sections of DNA for phylogenetic comparisons between species, and homoploid hybrids have successfully been detected using phylogenetic reconstructions incorporating intragenic recombination (for example, in tobacco plants, Kelly *et al.*, 2010; soft corals, Mcfadden and Hutchinson, 2004).

Phylogenetic reconstructions are usually based on specifically sequenced loci, rather than other types of molecular markers (such as microsatellites), because fast evolving markers do not contain enough information for resolving deeper level relationships (Schlötterer, 2004). In order to accurately identify introgressed loci, a large number of nuclear regions rather than high-copy organelle and nuclear ribosomal markers are required (Hobolth *et al.*, 2007; Hohenlohe *et al.*, 2011), and increasing effort is being made to identify informative nuclear markers in a range of different organisms (discussed below).

Population genetic approach

The genomic composition of putative hybrids, and the frequency of introgression, can be estimated by genotyping natural populations with molecular markers. The typical population genetic approach is to analyse the patterns of markers in hybrid zones, which are dynamic sites where species interact and cross-hybridize (Barton and Hewitt, 1985; Arnold, 1992; Buggs, 2007), and compare them to reference populations of individuals away from these zones (Rieseberg and Carney, 1998; Pinheiro *et al.*, 2010). The genomic contribution of the parental lineages in each hybrid individual can then be estimated (the 'hybrid index' or 'admixture proportion'; maximum likelihood implementation in HINDEX; Buerkle, 2005) as well as the hybrid class (for example, F1, F1 backcross, model-based Bayesian implementation in NEWHYBRIDS; Anderson and Thompson, 2002). Moreover, where detailed genomic data are available, genomic clines of introgressed alleles can be identified (using INTROGRESS; Gompert and Buerkle, 2010).

The criteria for markers used in hybridization studies are common to other population genetic studies, namely they should be inherited in a simple Mendelian manner, have reproducible results when repeated, be scorable across all individuals, have a low number of null alleles and have the maximum amount of information for the minimum cost and effort (Schlötterer, 2004). Markers used for hybridization studies need to amplify reliably in the divergent parental taxa, and should have diagnostic alleles distinguishing the putative parental species, or at least have a significant difference in allele frequency (Arnold, 1992; Moccia *et al.*, 2007). Preference should be given to mapped markers (see Rieseberg *et al.*, 2000 for more detail) as they can be selected to have a good coverage across each chromosome, and allow the size of introgressed linkage blocks and the rate of linkage block erosion to be estimated.

The basic properties of suitable markers are detailed in Schlötterer (2004) and summarized in Table 1. The choice of marker type depends on the genomic resources available for the organism of interest, and whether functional information about introgressed genes is required. For example, amplified fragment length polymorphisms (AFLPs) generate a multi-locus genotype from the whole genome even when no prior sequence data are available. However, the anonymous banding profile gives no information about the types of loci that are introgressed. By contrast, single-nucleotide polymorphism (SNP) assays designed in known genes are an increasingly popular high-throughput marker, which can be used to deduce functional information if comparative genomic resources are available.

Table 1 Types of commonly used molecular markers and their relative benefits (adapted from Schlötterer, 2004)

Marker	Marker type	Advantages	Disadvantages
Amplified fragment length polymorphisms (AFLPs)	Whole-genome fingerprinting	Quick to optimize, many loci, cheap, no ascertainment bias	Anonymous, dominant, reproducibility issues
Microsatellites (simple sequence repeat markers, SSRs)	Short tandem repeats	Polymorphic, co-dominant	Hard to automate, genomic SSRs often species-specific, homoplasy
DNA sequencing	Locus-specific sequence	High information content, universal primers available	Expensive to apply, currently no automation
Single-nucleotide polymorphisms (SNPs)	Nucleotide-specific amplification	Can be automated, highly abundant	Low information content, ascertainment bias

Organelle markers provide additional information to complement anonymous or nuclear markers in introgression studies, with mitochondrial sequencing being popular with animal geneticists and chloroplast sequencing widely employed in plant genetics. Many more examples of organelle introgression have been detected than nuclear introgression (Martinsen *et al.*, 2001; Gompert *et al.*, 2008), especially under scenarios of demographic expansion (Currat *et al.*, 2008). This has been explained largely by the maternal inheritance of the organelle genome, which means intraspecific gene flow at organelle loci occurs at a much lower rate than at nuclear loci. Therefore, local patterns of interspecific organelle capture will not be swamped and obscured by high levels of intraspecific gene flow (Petit and Excoffier, 2009). This explanation is supported by evidence from organisms that have an atypical mode of chloroplast inheritance (for example, paternal chloroplast inheritance in gymnosperms). Here there is a higher level of intraspecific gene flow for paternally inherited chloroplast markers than for maternally inherited mitochondrial markers. As predicted, paternally inherited chloroplasts have lower observed rates of introgression than maternally inherited mitochondria (Du *et al.*, 2009).

Additional advantages of using organelle genomes to study hybridization include their non-recombinant nature, which makes organelle introgression easier to detect, and their predominantly uniparental inheritance, which allows the initial direction of hybridization to be ascertained (see Galtier *et al.*, 2009 for a review of these assumptions). Organelle capture can easily be detected by surveying organelle haplotypes, and this can highlight potential introgression events that would otherwise not be identified. To do this, sufficient resolution to distinguish different haplotypes is all that is required, and the challenge is sampling enough individuals to detect rare organelle capture across a species' range. However, interpreting patterns of organelle sharing between species requires caution to distinguish between introgression and incomplete lineage sorting (see Zhou *et al.*, 2010 and discussion below).

Genetic maps are a valuable tool to support studies of hybrid swarms, and can be used to understand the behaviour of introgressed loci in different genomic backgrounds (Rieseberg *et al.*, 2000). Association studies, which compare phenotypic scores in a large number of mapping progeny to multi-locus genotypic data, can be used to search for chromosome sections associated with traits of interest (quantitative trait loci (QTLs)). This is a powerful framework for understanding the effects of introgressed loci, and markers associated with a particular QTL can be used to genotype natural populations to infer the introgression of functional genes. Fitness changes associated with introgressed genes can be assessed using reciprocal transplant or common garden experiments (Arnold, 1992; Rieseberg

et al., 2003; Martin *et al.*, 2006). Alternatively, admixture mapping in natural populations containing a mixture of early and later generation recombinant individuals can be used to understand the genetic architecture of introgressed candidate traits (reviewed by Buerkle and Lexer, 2008).

PROBLEMS OF THE CURRENT METHODOLOGIES

Insufficient availability of markers and low marker resolution

Population genetic studies and phylogenetic studies require highly informative markers to estimate the degree of interspecific gene flow, otherwise later generation backcrosses and recalcitrant introgression may go unnoticed (Barton, 2001). As few as four to five fixed markers between species can be sufficient to indicate early generation hybridization (Boecklen and Howard, 1997). However, 24–48 unlinked co-dominant markers may be required for the correct assignment of individuals to hybrid categories, depending on the population structure, the frequency of hybridization and the degree of genome divergence (Vähä and Primmer, 2006). This many markers are seldom produced in a traditional marker design protocol, nor are they easily transferred from related model species (Zane *et al.*, 2002). In addition to the number of markers or genes amplified, the resolution of genetic data is also a constraint. Diagnostic species-specific markers, which are highly differentiated between the putative parent species, are the most powerful type of markers for assigning later generation hybrids and detecting introgressed alleles in population genetic studies (Hohenlohe *et al.*, 2011). However, only a small proportion of loci sampled will fall into this category. These are particularly hard to identify between closely related taxa, which are the most likely to hybridize.

Similarly, molecular phylogenetic studies aiming to compare incongruent gene trees for hybrid identification are largely constrained by the phylogenetic resolution of each locus, and the number of loci that can be sampled. Comparisons of poorly resolved gene trees, using markers with limited sequence divergence between species, are likely to be uninformative in tracing the reticulate history of species (Linder and Rieseberg, 2004). Moreover, sequencing many loci is a major investment in time and money, and universal primers, which amplify conserved nuclear ribosomal and organellar markers, are often all that are available. Primers designed with broad phylogenetic scope typically do not vary greatly between species, unless they have conserved areas flanking more variable regions. To estimate the size of introgressed linkage blocks, low-copy nuclear markers are required. The primer design and cloning required to confirm a single amplified product is time-consuming (McFadden and Hutchinson, 2004; Steele *et al.*, 2008). Typically, conserved orthologous markers for amplifying nuclear genes are limited to well-studied families (for example, Asteraceae; Chapman *et al.*, 2007).

Low-throughput genotyping

Even with methodological improvements for detecting genetic variation between species and developing molecular markers, traditional genotyping is still relatively low throughput. For example, simple sequence repeats (SSRs) derived from available genomic resources, such as expressed sequence tags (ESTs), are much quicker to design than traditional microsatellites. M13-tailed fluorescent primers (Schuelke, 2000) decrease the number of expensive labelled primers required, and PCR multiplexing reduces the number of reactions that need to be completed. However, microsatellite studies still require a considerable amount of time and money to amplify a modest number of loci (10–15), representing a small fraction of the genome. In sequence-based approaches, each locus needs to be sequenced in a separate PCR. Therefore, sequencing effort (and the associated bioinformatic work) is a significant cost, limiting the accessions per species and the number of loci that can be sampled. This narrow sampling within species and poor depth of genetic data will overlook much of the interspecific gene flow.

Proving introgression as opposed to incomplete lineage sorting

The greatest theoretical challenge posed by genetic introgression studies is proving that the observed pattern of shared alleles between species is the product of recent introgression, as opposed to non-contemporary process such as ancient introgression or the incomplete lineage sorting of genes after speciation (deep coalescence; Pollard *et al.*, 2006; Willyard *et al.*, 2009), as shown in Figure 1. The genetic signature of incomplete lineage sorting is the same as ancient introgression soon after speciation, so we treat these together, and contrast this with recent introgression (introgression here). Empirical evidence is therefore required to infer introgression based on shared alleles between species (Slotte *et al.*, 2008; Du *et al.*, 2009). This is particularly a problem in species with a large effective population size (N_e), where high intraspecific allelic diversity makes inferences of hybridization difficult, or in recently speciating groups where levels of divergence are low (Pollard *et al.*, 2006).

This problem is illustrated by population studies in oaks (*Quercus*). Oaks exist in large, open-pollinated populations, where chloroplast

haplotypes between species are frequently shared, and nuclear microsatellite markers indicate common alleles between species (Muir and Schlötterer, 2005; Lepais *et al.*, 2009). The primary argument for ancestral polymorphism is the absence of a cline of introgressed genes between two hybridizing *Quercus* species (Muir and Schlötterer, 2005), whereas Lexer *et al.* (2006) argue that heterogeneous F_{ST} values between markers indicate different patterns of selection and homogenization, which obscure ongoing introgression. Further evidence from controlled pollinations, the occurrence of natural hybrids in other oak species and the number of chloroplast DNA substitutions between species is consistent with introgression rather than ancestral polymorphism in explaining limited interspecific divergence across different loci (Lexer *et al.*, 2006). This debate emphasizes the importance of obtaining data from multiple independent genetic sources and highlights that snapshot data are insufficient for inferring processes, when a number of different processes could have led to the same pattern, as shown in Table 2.

Having data from loci in which the phylogenetic relationship among allelic variants are known would allow us to distinguish among models of ancestral polymorphism and introgression. For example, Donnelly *et al.* (2004) showed the sharing of ancestral mitochondrial haplotypes (inferred from being internal in an organelle network) between two *Drosophila* species away from areas of sympatry, consistent with the retention of ancestral polymorphisms. Zhou *et al.* (2010) showed similar patterns of mitochondrial haplotype sharing owing to incomplete lineage sorting in two hybridizing pine species. Building on these experimental frameworks, studies that incorporate multiple loci with known phylogenetic relationship among their alleles, would allow us to reject the hypothesis that shared alleles arose from ancestral polymorphism and so unequivocally recognize the process of hybridization in natural populations.

Polyloid evolution and hybridization

Whole-genome duplication leading to polyploidy is often associated with hybridization and reproductive isolation (Rieseberg and Carney, 1998; Slotte *et al.*, 2008). It is now understood that polyploidy has been common throughout the history of flowering plants, effectively

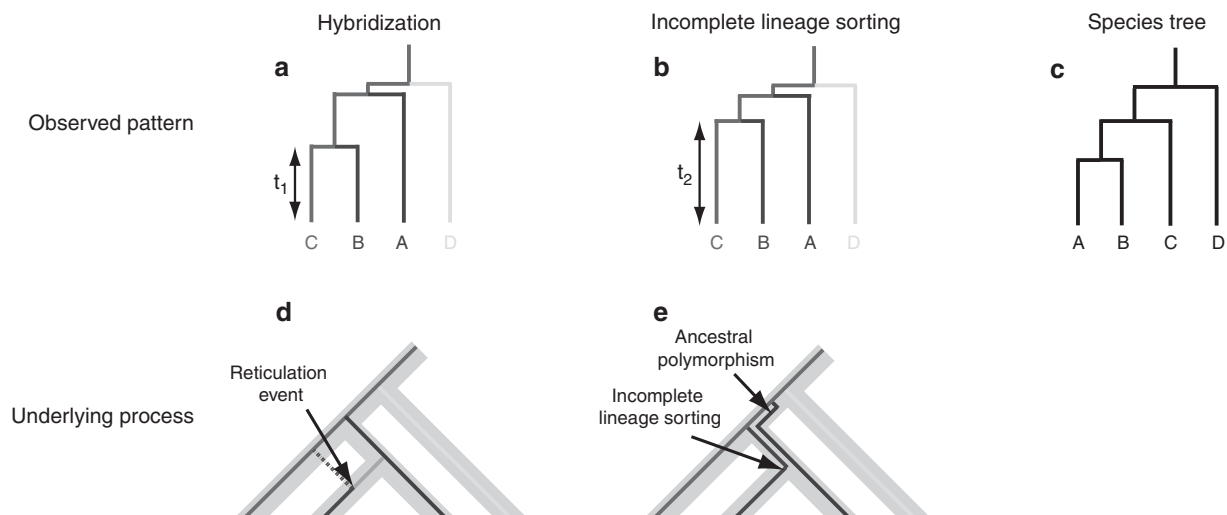


Figure 1 Hybridization and incomplete lineage sorting revealed by molecular phylogenetics. The phylogenetic relationship of alleles (coloured lines) are shown in the context of the species tree (grey bars, and the tree in panel c). The pattern of alleles when species hybridize (a) or when incomplete lineage sorting occurs (b) are the same, even though they are due to different processes (d and e, respectively). However, lineage sorting always results in coalescence with the other species prior to the speciation event (t_2). Coalescence of alleles is not expected where hybridization events are significantly later than the speciation event (t_1). Adapted from Pollard *et al.* (2006). A full colour version of this figure is available at the *Heredity* journal online.

Table 2 Evidence for introgression or shared ancestral polymorphism

<i>Support for introgression</i>	<i>Support for ancestral polymorphism</i>	<i>Biological reason</i>
Species are known to hybridize under greenhouse conditions and in the wild	Species cannot hybridize under greenhouse conditions, and/or little evidence of hybridization in the wild	Strength of reproductive barrier dictates potential for hybridization and introgression
Tightly linked genes show similar patterns of gene flow between species	Lack of correlation between allelic states at linked genes	Recombination less likely between linked genes, and introgressed linkage blocks are likely to be maintained
Different patterns of genetic distances between species at different loci, with the greatest degree of divergence between loci contributing to species-level differences	Equal divergence across different loci	Selection pressures different for different loci, with strong diversifying selection at loci underlying species-specific traits
Populations in sympatry or parapatry show less genetic divergence than those that occur in allopatry	Equivalent levels of divergence across species range	Where introgression occurs a cline of shared genes may be expected between species
Phylogenetic incongruency between uniparentally and biparentally inherited markers	Congruency between the phylogeny derived from uniparentally inherited markers and the maximum clade credibility tree from biparentally inherited markers	Uniparentally inherited markers (thus with a low effective population size) are particularly susceptible to introgression

Data from Willyard *et al.* (2009); Yatabe *et al.* (2007) and Donnelly *et al.* (2004).

making all angiosperms ancient polyploids (paleopolyploids), and recent polyploidy has been detected in many plant and some animal species (Soltis *et al.*, 2004; Hohenlohe *et al.*, 2011). The main difficulty for genomic studies of hybridization in polyploid taxa is distinguishing between homologs, similar gene copies that pair in meiosis, and homeologs, the duplicated gene copies from polyploidy (Buggs *et al.*, 2010). After polyploidization, duplicate gene copies undergo complex fates, including gene loss and gene silencing (Hegarty and Hiscock, 2008). Studies of polyploid taxa require homeolog-specific markers to distinguish duplicate gene copies (Buggs *et al.*, 2010; Hohenlohe *et al.*, 2011). However, most polyploid taxa have limited genetic resources available to design markers for distinguishing these duplicated gene copies.

HOW CAN NGS TECHNOLOGIES HELP US TO GET AROUND THESE LIMITATIONS?

Generating more markers with greater resolution

Advances in sequencing technologies allow genomic resources to be generated for non-model groups. These include cDNA sequences from transcriptomes, complete organelle genomes and even complete nuclear genomes (Dempewolf *et al.*, 2010; Stapley *et al.*, 2010). These large-scale genomic resources typically allow us to identify many hundreds of SSRs and tens of thousands of SNPs within species. Markers derived from expressed gene sequences have a number of benefits that make them ideal for hybridization studies. Firstly, searching genomic resources, such as transcriptomes, for variable markers (for example, with the program QDD for SSRs, Megléc *et al.*, 2010; SNPdetector for SNPs, Zhang *et al.*, 2005) is a much simpler process than traditional methods for anonymous marker design (Zane *et al.*, 2002; Lepais and Bacles, 2011). Secondly, the function of the locus can be inferred from BLAST searches to annotated sequences. This bridges the gap with the widely used candidate gene approach to functional genetics. Thirdly, the regions in which primers are designed are likely to be conserved between species, reducing the probability of null alleles, and making cross-amplification and direct comparisons between species a viable option (Woodhead *et al.*, 2005).

The main concern about coding sequence markers is that they may be acting in a non-neutral manner, and the subject of selection, biasing

calculations of population genetic parameters (Ellis and Burke, 2007). The assumption that they are not subject to selection is rarely tested, and an in-depth comparison between estimates of genetic diversity with coding sequence markers and anonymous markers in natural systems would validate this assumption. In addition to concerns about selection, we anticipate that EST markers are less likely to be polymorphic than anonymous markers owing to functional constraints in transcribed regions, and may contain less informative differences between the hybridizing taxa (Ellis and Burke, 2007). Woodhead *et al.* (2005) compared genetic diversity in the fern *Athyrium distentifolium* using EST-SSRs, genomic SSRs and AFLPs, and all marker types showed similar rank orders of population diversity and comparable F_{ST} values, suggesting polymorphism in EST-SSRs can often be considered effectively neutral. In a comparison of EST-SSRs and genomic SSRs in *Castanea*, Martin *et al.* (2010) found no significant differences in the F_{ST} values calculated with the two marker types, suggesting no deviation from selective neutrality. However, genomic SSRs have higher relative diversities these systems, and a number of others (see references in Martin *et al.*, 2010). It should be remembered, however, that decreased variation at EST-SSR loci may be considered a benefit for interspecific studies, where homoplasy may be a problem.

NGS resources are also promising for detecting variable gene sequences for comparative phylogenetic studies. This can be done by mining transcriptomes, nuclear genomes, or whole-organelle genomes, which is a much quicker way than traditional methods (Dunn *et al.*, 2008). Software to identify the most variable gene regions have been developed (for example, BMGE; Criscuolo and Gribaldo, 2010), and once located cloning or bioinformatic analyses can be used to ensure single copies are present, and PCR-optimized to ensure consistent amplification of a range of species. The completion of a suite of genetic and genomic resources for the asterid *Guizotia abyssinica* (Dempewolf *et al.*, 2010) illustrates the possible outputs from NGS data, which can be used for marker design.

High-throughput genotyping

Population genomic and comparative genomic studies aim to produce broad-scale genetic data, scoring many thousands of variable polymorphisms across the genome (Stapley *et al.*, 2010). As whole-genome

re-sequencing for a large number of individuals remains beyond the means of most researchers, genomic partitioning methods, where individual sequence libraries are enriched with subsections of the genome, will become increasingly popular (Ng *et al.*, 2009; Turner *et al.*, 2009). Suitable subsets of the genome for hybridization studies include SNP markers scattered through the genome, candidate loci that may be introgressed between species, and sequence markers at known genomic locations.

For such approaches to be used, a high level of automation is required, and the most widely used high-throughput genotyping methods are SNP marker panels (Chan, 2009). To design SNP markers, prior genomic resources are required to locate informative genetic variation, which is a major constraint of many projects. Moreover, SNP panels are most effective for scoring allelic variation that has been detected in the limited number of individuals in which genomic resources are generated. Therefore, the development of cost-effective genomic resources through NGS (for example, Buggs *et al.*, 2010) should be expanded to ensure good sampling of allelic variation.

For population genetic studies, SNP markers allow different alleles at each locus to be identified, and allelic diversity over many loci scored. Despite the low information content of each individual SNP, the large number of markers that can be scored yields a high-density coverage of markers across the genome. Moreover, SNP panels can be used directly on whole genomic DNA, removing the requirement of target enrichment and streamlining the experimental process. One example of an automated SNP panel is KASPar from KBioScience (Hertfordshire, UK), which relies on competitive allele-specific PCR, with fluorescent resonance energy transfer detection. This has been used for a range of genetic studies (for example, Tian *et al.*, 2011). An alternative for SNP typing is the Sequenom MassARRAY platform (Hamburg, Germany), which uses single termination mix multiplexed PCR and identifies different SNPs based on their different masses (applied by Thompson *et al.*, 2009). For phylogenetic analyses, 9000 informative SNPs were used to produce a well-supported phylogeny for the bacterial genus *Brucella*, with markers designed from whole genomes, with rigorous quality-control checks to ensure orthology (Foster *et al.*, 2009). The detection of informative SNPs for phylogenetic analyses of more complex genomes will require significant work and bioinformatic advances (discussed later).

Whereas the prospects for widespread high-throughput SNP detection and application seem good, the use of high-throughput genotyping with other commonly used markers, such as SSRs, is not viable. The difficulty here is that, in many cases, individual loci would have to be amplified and tagged prior to sequencing. Moreover, many NGS technologies have difficulties with sequencing repetitive sequences (particularly mononucleotide repeats; Chan, 2009). Even if high-throughput SSR amplification was achieved, stutter bands and PCR artefacts make accurate automated scoring of SSRs difficult (Schlötterer, 2004); therefore, these techniques may largely be discarded in preference for automatable SNP assays.

An alternative to SNP panels for large-scale genotyping is targeted re-sequencing, through sequencing of EST libraries, pooled amplicon sequencing of specific loci or genome-wide resequencing microarrays (Turner *et al.*, 2005, 2009; Griffin *et al.*, 2011). Comparative sequencing of EST libraries for multiple individuals is an effective method for reducing the complexity of the genome, and still allows the sequencing of a genome-wide sample of loci (Kane *et al.*, 2009). The main drawback with sequencing EST collections, apart from its cost, is that RNA is required rather than DNA, and the high rate of RNA degradation makes this technique impractical for sampling a large

number of wild-collected individuals for many organisms (Bräutigam and Gowik, 2010). In pooled amplicon sequencing, PCR products of target regions are mixed and sequenced using an NGS platform. The most basic application of this method is for mixed environmental samples where only a single PCR primer pair and DNA sample are used, such as sequencing the P6 loop (chloroplast *trnL* intron) to identify plant families in an ice core using 454 pyrosequencing (Sønstebo *et al.*, 2010). More complex applications, required in many population genetic and phylogenetic studies, require the ligation of a specific barcode to each sample so that the reads from pooled genetic data can be traced to the individual of interest. Alternatively, genome-wide array-based resequencing can be used, where cDNA or whole genomic DNA are hybridized onto a chip containing target oligonucleotide probes (reviewed by Turner *et al.*, 2009), which can then subsequently be sequenced using NGS. This approach allows individuals to be sampled at many loci of known genomic location; however, it is dependent on prior genomic data being available for the organism of interest (Turner *et al.*, 2005).

Combined approaches for marker detection and application

Increasing read lengths of NGS platforms, and significant bioinformatic improvements, are leading to promising developments integrating marker design and application. In these approaches, a large number of sequence reads are generated for all the individuals, and then a bioinformatic pipeline used to select informative characters. For example, the researcher may be targeting species-specific variation, screening out within- and between-population variation, and retaining the battery of markers that are fixed between species. Restriction site-associated DNA (RAD) tags is an example of this approach. RAD markers are short sequences of DNA adjacent to a restriction enzyme recognition site, in which SNPs are compared between individuals (full method in Miller *et al.*, 2007b, and summarized in Figure 2c). This technique has been used in various organisms such as *Drosophila* (Miller *et al.*, 2007b), *Neurospora* (Baird *et al.*, 2008), the pitcher plant mosquito (Emerson *et al.*, 2010), zebrafish (Miller *et al.*, 2007a), the three spine stickleback (Miller *et al.*, 2007b), trout (Hohenlohe *et al.*, 2011) and barley (Chutimanitsakun *et al.*, 2011). This approach sequences a subset of the genome to reduce costs and uses barcodes for individual samples to allow bulk sequencing. The number of SNP markers can be determined by the combination of restriction enzymes applied, to allow fewer markers for more individuals or higher resolution for fewer individuals. The use of a reference genome is recommended for introgression studies as this allows the assessment of synteny of introgressed linkage blocks, and their putative function. This approach meets the demands of population geneticists, providing many loci for potentially large numbers of individuals, while securing the cost-benefit of NGS by obtaining large amounts of data per lane of sequencing. This is exemplified in a study by Emerson *et al.* (2010) where 3741 SNPs fixed within and variable among populations were identified, for a total of 126 individuals, in two lanes of an ILLUMINA GAIIIX sequencer.

Proving introgression as opposed to incomplete lineage sorting

The greater resolution and depth of genetic data generated by NGS can be used to support hypotheses of introgression as opposed to incomplete lineage sorting. One signature of recent introgression is higher allelic diversity near hybrid swarms, and a cline of introgressed alleles as one samples away from them (Arnold, 1992). Increased sampling breadth and depth will increase the ability to detect such local phylogeographic structure. Further support comes from inferences made using high-resolution sequence data. Similar DNA sequences

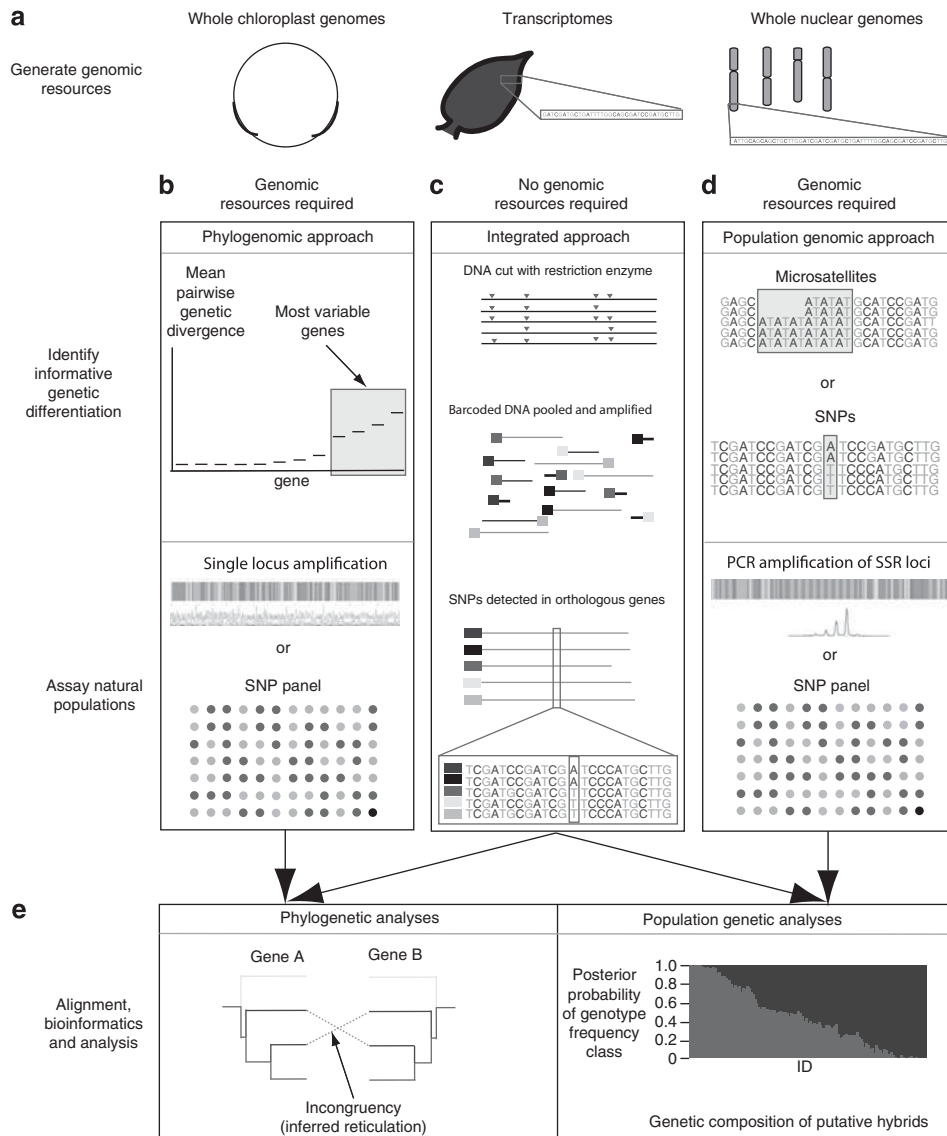


Figure 2 Workflow for using NGS in molecular phylogenomics and population genomics. Different genomic resources can be produced (a) and used in phylogenomic (b) and population genomic studies (d), where informative genetic differentiation is identified and markers designed to assay natural populations. These separate stages, including production of genomic resources, are not required for integrated studies (c) where marker detection and application are performed simultaneously, as shown here with RAD tags. Instead, in the integrated approach, DNA is cut with restriction enzymes (red arrows), an individual/population-specific adaptor (coloured box) is ligated and the product amplified on an NGS platform (not shown). The genomic data generated from these studies can then be analysed by using standard phylogenetic and population genetic programmes (e). A full colour version of this figure is available at the *Heredity* journal online.

at each locus from individuals of hybridizing species would support a recent common ancestry for this allele, as expected by the transfer of alleles mediated by hybridization (Slotte *et al.*, 2008). This is particularly valuable when set in the context of a large number of loci, where different patterns of introgression can be identified at each locus. To assess ancient introgression, the phylogenetic relationship between allelic variants can be calculated, and introgression is supported when the coalescent time for the alleles at a locus is after the point of speciation (Mao *et al.*, 2010). Finally, hyper-variable organelle markers can be identified from NGS data and high-throughput genotyping methods used to score many individuals at these loci. The relative frequency of ancient and derived organelle haplotypes, which are shared between species; the haplotype frequencies in hybrid swarms relative to a reference population and comparisons of nuclear and organelle DNA

diversity can also be used to support hypotheses of introgression or ancestral polymorphism (Donnelly *et al.*, 2004; Gompert *et al.*, 2008).

Polyloid evolution and hybridization

NGS is also promising for the study of polyploid evolution. Buggs *et al.* (2010) developed high-throughput SNP assays to distinguish differential expression of homeologs in the allopolyploid plant *Tragopogon miscellus*. They used a hybrid NGS sequencing approach combining 454 and Illumina sequencing on cDNA, followed by SNP validation with the Sequenom MassARRAY iPLEX. This experimental through-flow allows homologs and homeologs to be identified in non-model organisms with relative ease. Griffin *et al.* (2011) developed a mixed amplicon sequencing protocol for genotyping polyploid species. They amplified multiple low-copy nuclear markers and a chloroplast

Table 3 Comparison between a complex genome (angiosperm) and a more simple genome (mammal), and the subsequent problems faced by RAD tag sequencing in complex genomes, adapted from Kejnovsky *et al.* (2009)

	<i>Complex angiosperm genome</i>	<i>Simple mammal genome</i>	<i>Difficulty for RAD tag sequencing</i>
Occurrence of multi-gene families	Large multi-gene families common (37.4% of genes occur in families with > 5 members)	Fewer, smaller multi-gene families (in humans 1.4% of genes occur in families with > 5 members)	Paralog determination for studies without a reference genome. Paralog discrimination for pairwise SNP comparisons
Genome size variation	Very large: ~2000-fold	Relatively small: 5-fold	Larger genomes require more markers for high coverage—making studies more expensive
Retrotransposable elements	Diverse range, varying in copy number with frequent insertions. Retroelements contribute up to 80% of the genome	Low diversity of retroelement families, with a low frequency of retroelement insertion	Restriction digest sites in different retrotransposons may be hard to analyse, and many reads will be uninformative
Polyploidy	Frequent occurrence; all species have successive rounds of ancient polyploidization, with many having more recent events too	Absent in mammals, only one controversial report of a polyploid rodent	Discrimination of nearly identical paralogs difficult
Recombination and translocation	Frequent recombination and chromosome translocations	Lower rates of recombination combined with small number of large translocations	Pairwise comparison between recombined genes difficult

Abbreviation: RAD, restriction site-associated DNA.

marker in separate PCRs, and then used a pooled barcoded protocol prior to 454 sequencing. The PCR products were error-checked for PCR recombination, prior to distinguishing different alleles at each nuclear locus using the criteria of two or more basepair changes in > 80% of the reads. This technique could be extended to identifying different alleles in diploid heterozygotes as well as polyploids. This method is easier than bacterial cloning, and is useful for reconstructing polyploid evolution with a moderate number of species; however, the large number of PCR amplifications limits the number of samples that can be sequenced with this method. Moreover, the study by Griffin *et al.* (2011) used primers that are known to amplify a single locus; a different approach is required if divergent paralogs may be present (discussed later).

WHAT ARE THE PROBLEM AREAS THAT ARE STILL OUTSTANDING IN THE APPLICATION OF NGS?

Many of the emerging NGS techniques have yet to be rigorously tested on complex genomes, and the techniques must overcome the difficulties associated with repetitive genomes, heavily laden with transposable elements, in addition to the recurrent rounds of polyploidy some genomes have been through (Soltis *et al.*, 2004). Here, the challenge to design markers that reliably amplify homologous subsets of the genome, is difficult, as nearly identical paralogs can be difficult to distinguish from heterozygosity (Hohenlohe *et al.*, 2011). This leads to a demanding bioinformatic challenge, where a reference genome may be required to identify nearly identical paralogs. These issues are illustrated in methods that reduce the complexity of the genome through restriction fragment digestion, such as RAD tags, as summarized in Table 3. If the restriction digest site is present in a transposon, large numbers of reads will not be informative, thus stringent data filters are required. Moreover, restriction digest sites may not be shared between alleles (null alleles), making pairwise comparisons between alleles difficult.

Owing to the high cost of NGS, the sample size used for the design of markers is often low, frequently just a single or a few individuals. This is particularly the case for organisms with complex genomes,

where a large sequencing effort is required to ensure good coverage. Low sample size is particularly problematic for species-level comparisons, as it will be impossible to know whether sequence variation is fixed or variable within species if only a few individuals are sequenced (Excoffier *et al.*, 2009). Therefore, targeted re-sequencing of genomic subsets of interest in a broader sample of individuals may be a better use of resources than whole-genome sequencing of a few individuals for hybridization studies.

HOW CAN NGS BEST BE USED TO STUDY HYBRIDIZATION AND INTROGRESSION?

We believe NGS technologies should be used to increase our understanding of three important components of hybridization in natural systems: the spatio-temporal dynamics of hybrid zones, the significance of reticulate evolution in species formation, and the behaviour of introgressed loci in their new genomic background. A particular emphasis should be on studying these areas in ecologically well-characterized groups where no current genomic data are available.

In situations where hybrid zones and introgression between species in the wild is being investigated, NGS technologies should be employed to generate an array of informative molecular markers, even between closely related species. Large-scale SNP typing is already used for evolutionary model systems, such as *Populus*, where 35 diagnostic SNPs have been assayed for 635 individuals (Thompson *et al.*, 2009); however, NGS will aid marker design and application for systems with no current genomic resources. Focused studies of hybrid swarms should be expanded to include samples from across a species ranges, in order to accurately assess the degree of admixture and introgression. By expanding the study range, replicate hybrid swarms can be included, as patterns of hybridization may not be the same under different environmental and demographic scenarios (Excoffier *et al.*, 2009). Moreover, recent theoretical and empirical data (summarized in Buggs, 2007) highlight that hybrid zones may not be static in space and time (Barton and Hewitt, 1985). Increased sample ranges will allow a better understanding of the dynamic nature of past

hybridization events, and also influence future conservation policies of taxa that are known to hybridize. For example, high-resolution genomic data will allow progenitors of recent hybrid species in taxonomically complex groups to be distinguished, so that conservation work can focus on conserving the evolutionary process underlying the generation of genetic novelty in the group (Ennos *et al.*, 2005).

In phylogenetic research, genomic resources will provide a wealth of data from which loci evolving at the required speed for good resolution and widespread amplification can be selected. Moreover, integrated approaches combining the power of NGS with targeted capture of informative loci will allow many loci to be amplified in a more cost-efficient manner (Ng *et al.*, 2009; Turner *et al.*, 2009). This will have wide-ranging implications for evolutionary studies, allowing an increased sampling breadth within and between species, and the amplification of many more informative markers. Targeted NGS sequence data will be important in understanding hybridization in complex polyploid groups where distinguishing homologs and homeologs has previously hindered research (Slotte *et al.*, 2008; Buggs *et al.*, 2010; Hohenlohe *et al.*, 2011). Its use in reconstructing reticulate evolution in ancient homoploid hybrid taxa will also be significant, as the ability to study a large number of nuclear markers with high coverage to sample all alleles in a population will allow complex historical scenarios to be better understood.

To understand how introgressed alleles interact with the rest of the recipient genome, and the subsequent selection introgressed alleles may be subject to, genotyping natural populations at many loci will allow the identification of outlier loci. These include loci with a high rate of introgression, which are under positive selection (or tightly linked to a gene undergoing positive selection, 'hitchhiking' markers), as well as loci that are likely to have a decreased frequency of introgression. These include genes involved in co-adapted gene complexes (epistatic interactions) or structurally divergent areas of the chromosome where recombination is suppressed (Rieseberg *et al.*, 1995; Kane *et al.*, 2009). The identification of regions that show no introgression is of particular importance, as these areas may contain genes responsible for reproductive isolation, which have a role in maintaining the distinct identities of species which co-occur in allopatry (Turner *et al.*, 2005). Once outlier loci that show divergent patterns of introgression from the rest of the genome are identified, these can be removed from analysis, weighted accordingly, or tested to see if they are under selection (Luikart *et al.*, 2003).

In each of the above cases, an understanding of the frequency of introgression at each locus and the subsequent selection which occurs, may be the first step towards understanding the adaptive significance of introgressed genes. This may be done by identifying introgressed chromosome blocks and searching in these regions for candidate genes that may underlie introgressed functional traits. Alternatively, genome-wide comparisons can be made using transcriptional profiling or the use of markers in transcribed regions (Bräutigam and Gowik, 2010).

WHAT ARE THE MEDIUM-TERM PROSPECTS AND THE LONGER TERM VISION FOR APPLYING NGS TECHNOLOGIES TO STUDIES OF HYBRIDIZATION?

The main aim of obtaining an accurate estimate of interspecific gene flow certainly appears achievable if current methodological limitations, such as developing a large number of markers that distinguish between paralogous genes, can be achieved. Therefore, in the medium term, implementation of novel methods may vastly increase our understanding of how porous genomes are, the types of loci that

introgress, and the rate at which linkage blocks are eroded over time. Of the emerging group of new methods, those that integrate the design and application of markers, are perhaps the most exciting for hybridization studies. These methods harvest NGS technologies in the most effective manner, using them for both polymorphic marker design and automated high-throughput genotyping (Miller *et al.*, 2007b). RAD tag sequencing is one example that satisfies the target of sampling many individuals for a large number of loci, and the recent application of RAD tags in the complex genome of barley indicates this method can be successfully applied to repetitive genomes (Chutimanitsakun *et al.*, 2011). Alternative approaches to genome-wide sampling allow important sequence variation to be amplified in a more targeted manner. This includes genome-wide array based resequencing, or targeted capture of candidate genes for a specific introgressed phenotype, and these approaches will allow the introgression of functional traits to be examined in a range of ecologically important scenarios.

One exciting prospect of genomic introgression studies is the bridging of population genetic and molecular phylogenetic frameworks to understand the contribution of introgression over different temporal and spatial levels (Figure 2). The incorporation of in-depth within-species sampling of population genetics, and the breadth of genetic data from phylogenetic studies, will allow both ancient and more recent introgression events to be identified. To enable this quantity of genetic data to be handled, the emphasis should be placed on developing bioinformatic tools to match the power of new NGS technologies (Turner *et al.*, 2009).

CONCLUSION

Computer simulations and models of gene flow predict widespread introgression under different demographic scenarios (for example, Currat *et al.*, 2008), but until now genome-wide studies of introgression have been limited to a handful of model organisms (for example, sunflowers; Kane *et al.*, 2009). In this review, we promote the use of NGS technologies to design molecular markers spread throughout the genome, and encourage the use of high-throughput assays to genotype large numbers of individuals, especially in non-model organisms. Genomic introgression studies should use an increasing depth of genetic data to integrate population genomic and phylogenomic frameworks, as well as databases including gene function, to infer adaptive introgression. Such approaches will allow us not only to shed light on recent introgression events between taxa, but also to focus on this evolutionary process in a historical perspective, and deduce the adaptive function of introgressed genes.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Richard Abbott, Richard Buggs, Peter Hollingsworth, Catherine Kidner and Gill Twyford for useful comments on earlier drafts of the paper. This work forms part of the postgraduate thesis by ADT, funded by the Biotechnology and Biological Sciences Research council (UK).

-
- Abbott RJ, Hegarty MJ, Hiscock SJ, Brennan AC (2010). Homoploid hybrid speciation in action. *Taxon* **59**: 1375–1386.
- Anderson E (1949). *Introgressive Hybridization*. Wiley: New York.
- Anderson E, Hubricht L (1938). Hybridisation in *Tradescantia*. III The evidence for introgressive hybridisation. *Am J Bot* **25**: 396–402.
- Anderson EC, Thompson EA (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**: 1217–1229.

- Arnold ML (1992). Natural hybridization as an evolutionary process. *Annu Rev Ecol Syst* **23**: 237–261.
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA *et al.* (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **3**: e3376.
- Barton NH (2001). The role of hybridization in evolution. *Mol Ecol* **10**: 551–568.
- Barton NH, Hewitt GM (1985). Analysis of hybrid zones. *Annu Rev Ecol Syst* **16**: 113–148.
- Boecklen WJ, Howard DJ (1997). Genetic analysis of hybrid zones: number of markers and power of resolution. *Ecology* **78**: 2611–2616.
- Bräutigam A, Gowik U (2010). What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. *Plant Biol* **12**: 831–841.
- Buerkle CA (2005). Maximum-likelihood estimation of a hybrid index based on molecular markers. *Mol Ecol Notes* **5**: 684–687.
- Buerkle CA, Lexer C (2008). Admixture as the basis for genetic mapping. *Trends Ecol Evol* **23**: 686–694.
- Buggs RJA (2007). Empirical study of hybrid zone movement. *Heredity* **99**: 301–312.
- Buggs RJA, Chamala S, Wu WEI, Gao LU, May GD, Schnable PS *et al.* (2010). Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. *Mol Ecol* **19**: 132–146.
- Castric V, Bechsgaard J, Schierup MH, Vekemans X (2008). Repeated adaptive introgression at a gene under multi-allelic balancing selection. *PLoS Genet* **4**: e1000168.
- Chan EY (2009). Next-generation sequencing methods: impact of sequencing accuracy on SNP discovery. *Methods Mol Biol* **578**: 95–111.
- Chapman M, Chang J, Weisman D, Kesseli R, Burke J (2007). Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae). *Theor Appl Genet* **115**: 747–755.
- Chapman MA, Abbott RJ (2010). Introgression of fitness genes across a ploidy barrier. *New Phytol* **186**: 63–71.
- Choler P, Erschbamer B, Tribsch A, Gielly L, Taberlet P (2004). Genetic introgression as a potential to widen a species' niche: insights from alpine *Carex curvula*. *Proc Natl Acad Sci USA* **101**: 171–176.
- Chutimanitsakun Y, Nipper R, Cuesta-Marcos A, Cistué L, Corey A, Filichkina T *et al.* (2011). Construction and application for QTL analysis of a restriction site associated DNA (RAD) linkage map in barley. *BMC Genomics* **12**: 4.
- Crisuolo A, Gribaldo S (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* **10**: 210.
- Currat M, Ruedi M, Petit RJ, Excoffier L (2008). The hidden side of invasions: massive introgression of local genes. *Evolution* **62**: 1908–1920.
- Dempewolf H, Kane NC, Ostevik KL, Geleta M, Barker MS, Lai Z *et al.* (2010). Establishing genomic tools and resources for *Guizotia abyssinica* (L.f.) Cass.—the development of a library of expressed sequence tags, microsatellite loci, and the sequencing of its chloroplast genome. *Mol Ecol Resources* **10**: 1048–1058.
- Donnelly MJ, Pinto J, Girod R, Besansky NJ, Lehmann T (2004). Revisiting the role of introgression vs shared ancestral polymorphisms as key processes shaping genetic diversity in the recently separated sibling species of the *Anopheles gambiae* complex. *Heredity* **92**: 61–68.
- Du FK, Petit RJ, Liu JQ (2009). More introgression with less gene flow: chloroplast vs mitochondrial DNA in the *Picea asperata* complex in China, and comparison with other conifers. *Mol Ecol* **18**: 1396–1407.
- Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA *et al.* (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**: 745–749.
- Ellis JR, Burke JM (2007). EST-SSRs as a resource for population genetic analyses. *Heredity* **99**: 125–132.
- Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE *et al.* (2010). Resolving postglacial phylogeography using high-throughput sequencing. *Proc Natl Acad Sci USA* **107**: 16196–16200.
- Ennos RA, French GC, Hollingsworth PM (2005). Conserving taxonomic complexity. *Trends Ecol Evol* **20**: 164–168.
- Excoffier L, Foll M, Petit RJ (2009). Genetic consequences of range expansions. *Annu Rev Ecol Syst* **40**: 481–501.
- Foster JT, Beckstrom-Sternberg SM, Pearson T, Beckstrom-Sternberg JS, Chain PSG, Roberto FF *et al.* (2009). Whole-genome-based phylogeny and divergence of the genus *Brucella*. *J Bacteriol* **191**: 2864–2870.
- Galtier N, Nabholz B, Glémin S, Hurst GDD (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol Ecol* **18**: 4541–4550.
- Gompert Z, Buerkle CA (2010). Introgess: a software package for mapping components of isolation in hybrids. *Mol Ecol Resources* **10**: 378–384.
- Gompert Z, Forister ML, Fordyce JA, Nice CC (2008). Widespread mito-nuclear discordance with evidence for introgressive hybridization and selective sweeps in *Lycaeides*. *Mol Ecol* **17**: 5231–5244.
- Griffin P, Robin C, Hoffmann A (2011). A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to *Poa* grasses. *BMC Biol* **9**: 19.
- Gross BL, Rieseberg LH (2005). The ecological genetics of homoploid hybrid speciation. *J Hered* **96**: 241–252.
- Hegarty MJ, Hiscock SJ (2008). Genomic clues to the evolutionary success of polyploidy plants. *Curr Biol* **18**: 435–444.
- Hobolth A, Christensen OF, Mailund T, Schierup MH (2007). Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS Genet* **3**: e7.
- Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Mol Ecol Resources* **11**(Suppl 1): 117–122.
- Jiggins CD, Salazar C, Linares M, Mavarez J (2008). Hybrid trait speciation and *Heliconius* butterflies. *Philos Trans R Soc Lond B Biol Sci* **363**: 3047–3054.
- Kane NC, King MG, Barker MS, Raduski A, Karrenberg S, Yatabe Y *et al.* (2009). Comparative genomic and population genetic analysis indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. *Evolution* **63**: 2061–2075.
- Kejnovsky E, Leitch IJ, Leitch AR (2009). Contrasting evolutionary dynamics between angiosperm and mammalian genomes. *Trends Ecol Evol* **24**: 572–582.
- Kelly LJ, Leitch AR, Clarkson JJ, Hunter RB, Knapp S, Chase MW (2010). Intragenic recombination events and evidence for hybrid speciation in *Nicotiana* (Solanaceae). *Mol Biol Evol* **27**: 781–799.
- Kim M, Cui M-L, Cubas P, Gillies A, Lee K, Chapman MA *et al.* (2008). Regulatory genes control a key morphological and ecological trait transferred between species. *Science* **322**: 1116–1119.
- Lepais O, Bacles CFE (2011). Comparison of random and SSR-enriched shotgun pyrosequencing for microsatellite discovery and single multiplex PCR optimization in *Acacia harpophylla* F. Muell. Ex Benth. *Mol Ecol* **11**: 711–724.
- Lepais O, Petit RJ, Guichoux E, Lavabre JE, Alberto F, Kremer A *et al.* (2009). Species relative abundance and direction of introgression in oaks. *Mol Ecol* **18**: 2228–2242.
- Lexer C, Kremer A, Petit RJ (2006). COMMENT: shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Mol Ecol* **15**: 2007–2012.
- Linder CR, Rieseberg LH (2004). Reconstructing patterns of reticulate evolution in plants. *Am J Bot* **91**: 1700–1708.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003). The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet* **4**: 981–994.
- Mallet J (2005). Hybridization as an invasion of the genome. *Trends Ecol Evol* **20**: 229–237.
- Mao X, Zhang J, Zhang S, Rossiter SJ (2010). Historical male-mediated introgression in horseshoe bats revealed by multilocus DNA sequence data. *Mol Ecol* **19**: 1352–1366.
- Martin M, Mattioni C, Cherubini M, Turchini D, Villani F (2010). Genetic diversity in European chestnut populations by means of genomic and genic microsatellite markers. *Tree Genet Genomes* **6**: 735–744.
- Martin NH, Bouck AC, Arnold ML (2006). Detecting adaptive trait introgression between *Iris fulva* and *I. brevicaulis* in highly selective field conditions. *Genetics* **172**: 2481–2489.
- Martinsen GD, Whitham TG, Turek RJ, Keim P (2001). Hybrid populations selectively filter gene introgression between species. *Evolution* **55**: 1325–1335.
- McFadden CS, Hutchinson MB (2004). Molecular evidence for the hybrid origin of species in the soft coral genus *Alcyonium* (Cnidaria: Anthozoa: Octocorallia). *Mol Ecol* **13**: 1495–1505.
- Megléc E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N *et al.* (2010). QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* **26**: 403–404.
- Metzker ML (2010). Sequencing technologies—the next generation. *Nat Rev Genet* **11**: 31–46.
- Miller MR, Atwood TS, Eames BF, Eberhart JK, Yan YL, Postlethwait JH *et al.* (2007a). RAD marker microarrays enable rapid mapping of zebrafish mutations. *Genome Biol* **8**: R105.
- Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007b). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res* **17**: 240–248.
- Moccia MD, Widmer A, Cozzolino S (2007). The strength of reproductive isolation in two hybridizing food-deceptive orchid species. *Mol Ecol* **16**: 2855–2866.
- Muir G, Schlotterer C (2005). Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Mol Ecol* **14**: 549–561.
- Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C *et al.* (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**: 272–276.
- Petit RJ, Excoffier L (2009). Gene flow and species delimitation. *Trends Ecol Evol* **24**: 386–393.
- Pinheiro F, De Barros F, Palma-Silva C, Meyer D, Fay MF, Suzuki RM *et al.* (2010). Hybridization and introgression across different ploidy levels in the Neotropical orchids *Epidendrum fulgens* and *E. puniceoluteum* (Orchidaceae). *Mol Ecol* **19**: 3981–3994.
- Pollard DA, Iyer VN, Moses AM, Eisen MB (2006). Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. *PLoS Genet* **2**: 1634–1647.
- Rieseberg LH, Baird SJE, Gardner KA (2000). Hybridization, introgression, and linkage evolution. *Plant Mol Biol* **42**: 205–224.
- Rieseberg LH, Carney SE (1998). Tansley review no. 102 Plant hybridisation. *New Phytol* **140**: 599–624.
- Rieseberg LH, Ellstrand NC (1993). What can molecular and morphological markers tell us about Plant hybridization? *Crit Rev Plant Sci* **12**: 213–241.
- Rieseberg LH, Linder CR, Seiler GJ (1995). Chromosomal and genic barriers to introgression in *Helianthus*. *Genetics* **141**: 1163–1171.

- Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T *et al.* (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**: 1211–1216.
- Schlötterer C (2004). The evolution of molecular markers—just a matter of fashion? *Nat Rev Genet* **5**: 63–69.
- Schuelke M (2000). An economic method for the fluorescent labelling of PCR fragments. *Nat Biotechnol* **18**: 233–234.
- Schwenk K, Brede N, Streit B (2008). Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philos Trans R Soc Lond B Biol Sci* **363**: 2805–2811.
- Slotte T, Huang H, Lascoux M, Ceplitis A (2008). Polyploid speciation did not confer instant reproductive isolation in *Capsella* (Brassicaceae). *Mol Biol Evol* **25**: 1472–1481.
- Soltis DE, Soltis PS, Tate JA (2004). Advances in the study of polyploidy since *Plant speciation*. *New Phytol* **161**: 173–191.
- Sønsteby JH, Gielly L, Brysting AK, Elven R, Edwards M, Haile J *et al.* (2010). Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Mol Ecol Resources* **10**: 1009–1018.
- Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, Ekblom R *et al.* (2010). Adaptation genomics: the next generation. *Trends Ecol Evol* **25**: 705–712.
- Steele PR, Guisinger-Bellian M, Linder CR, Jansen RK (2008). Phylogenetic utility of 141 low-copy nuclear regions in taxa at different taxonomic levels in two distantly related families of rosids. *Mol Phylogenet Evol* **48**: 1013–1026.
- Thompson SL, Lamothe M, Meirmans PG, Périnet P, Isabel N (2009). Repeated unidirectional introgression towards *Populus balsamifera* in contact zones of exotic and native poplars. *Mol Ecol* **19**: 132–145.
- Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S *et al.* (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet* **43**: 159–162.
- Turner EH, Ng SB, Nickerson DA, Shendure J (2009). Methods for genomic partitioning. *Annu Rev Genomics Hum Genet* **10**: 263–284.
- Turner TL, Hahn MW, Nuzhdin SV (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol* **3**: e285.
- Vähä JP, Primmer CR (2006). Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Mol Ecol* **15**: 63–72.
- Willyard A, Cronn R, Liston A (2009). Reticulate evolution and incomplete lineage sorting among the ponderosa pines. *Mol Phylogenet Evol* **52**: 498–511.
- Woodhead M, Russell J, Squirrell J, Hollingsworth PM, Mackenzie K, Gibby M *et al.* (2005). Comparative analysis of population genetic structure in *Athyrium distentifolium* (Pteridophyta) using AFLPs and SSRs from anonymous and transcribed gene regions. *Mol Ecol* **14**: 1681–1695.
- Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH (2007). Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics* **175**: 1883–1893.
- Zane L, Bargelloni L, Patarnello T (2002). Strategies for microsatellite isolation: a review. *Mol Ecol* **11**: 1–16.
- Zhang J, Wheeler DA, Yakub I, Wei S, Sood R, Rowe W *et al.* (2005). SNPdetector: a software tool for sensitive and accurate SNP detection. *PLoS Comput Biol* **1**: e53.
- Zhou YF, Abbott RJ, Jiang ZY, Du FK, Milne RI, Liu JQ (2010). Gene flow and species delimitation: a case study of two pine species with overlapping distributions in south-east China. *Evolution* **64**: 2342–2352.