Review

Hybrid speciation in plants: new insights from molecular studies

Matthew J. Hegarty and Simon J. Hiscock
School of Biological Sciences, University of Bristol, Woodland Road, Bristol, BS8 1UG, UK

Summary

Abrupt speciation through interspecific hybridisation is an important mechanism in angiosperm evolution. Flowering plants therefore offer excellent opportunities for studying genetic processes associated with hybrid speciation. Novel molecular approaches are now available to examine these processes at the level of both genome organization and gene expression – transcriptomics. Here, we present an overview of the molecular technologies currently used to study hybrid speciation and how they are providing new insights into this mode of speciation in flowering plants. We begin with an introduction to hybrid speciation in plants, followed by a review of techniques, such as isozymes and other markers, which have been used to study hybrid species in the past. We then review advances in molecular techniques that have the potential to be applied to studies of hybrid species, followed by an overview of the main genomic and transcriptomic changes suspected, or known, to occur in newly formed hybrids, together with commentary on the application of advanced molecular tools to studying these changes.


Introduction

Interspecific hybridization, resulting in hybrid offspring reproductively or otherwise isolated from their parental taxa, has long been viewed as an important mechanism in plant speciation (Grant, 1981; Abbott, 1992; Arnold, 1997; Rieseberg, 1997; Ellstrand & Schierenbeck, 2000). Recently an elegant series of molecular-based studies has demonstrated that hybridization can promote adaptive evolution and speciation (Arnold et al., 2003; Rieseberg et al., 2003). It has highlighted the value of studying hybrid speciation at the genomic level, utilizing the many molecular tools currently available through technological advances in molecular biology.

Occasional natural hybridisation has always been regarded as the rule rather than the exception in plants (Stebbins, 1959, as cited in Grant, 1981; Ellstrand et al., 1996), but the frequency of spontaneous natural hybridisation varies considerably between different plant genera and families, being most common among outcrossing species with reproductive strategies that can stabilise hybridity, such as vegetative reproduction, permanent odd polyploidy or agamospermy (Ellstrand et al., 1996). Estimates of the extent of natural hybridisation within flowering plants are extremely variable and unreliable.
because floras rarely contain extensive lists of confirmed hybrids (Abbott, 1992). Nevertheless, a survey of five floras from regions that have been the focus of extensive botanical surveys – United Kingdom, Scandinavia, US Great Plains, US Intermountain and Hawaii (Ellstrand et al., 1996) – suggests that spontaneous hybridisation is common (c. 11% of the species listed in these floras are hybrids), although by no means ubiquitous – hybrids being concentrated in a small percentage of families and even fewer genera. Ellstrand et al. (1996) speculate that this apparently restricted occurrence of hybrid taxa does not necessarily impinge on the contribution made by hybridisation to plant evolution, as even a single hybrid may serve as the progenitor of a new species, provided it is at least partially fertile.

Hybrid speciation can occur by one of two routes. Homoploid speciation involves hybridisation between two taxa without a change in ploidy, whereas allopolyploid speciation is characterized by the offspring of a hybridisation having a different ploidy level than that of either parent. Hybrids often possess odd ploidy numbers, which generally cause low fertility or complete sterility, but these initial hybrids frequently undergo spontaneous chromosome doubling (allopolyploidy), which stabilises the genome, or, if partially fertile, may backcross to one of the parental species (introgression). Polyploidy (both allopolyploid and autopolyploid) is very common in plants, with estimates in angiosperms ranging from 30 to 80% of species predicted to have polyploid genomes, while in ferns estimates of the number of polyploid species are even higher, at > 90% (Leitch & Bennett, 1997; Soltis & Soltis, 2000). However, the evolutionary contribution of polyploidy to existing plant lineages may be greater than even these estimates suggest; recent studies have shown that species long thought to be diploid may actually be ancient polyploids – paleopolyploids (Simillion et al., 2002). By comparison, known cases of homoploid speciation are rare, possibly as a consequence of hybrid sterility and hybrid breakdown but, perhaps more significantly, because of the difficulty in identifying homoploid hybrids (Ferguson & Sang, 2001).

One obvious feature of hybrid speciation is that it has the potential to occur repeatedly at different times and in different geographical locations, which may also result in morphological differences leading to offspring of the same hybridising taxa being given different names. Despite earlier reservations about the possibility of recurrent hybrid formation (Grant, 1958), a growing body of evidence indicates that hybrid speciation can occur on multiple occasions (Abbott, 1992; Rieseberg et al., 1996). Indeed, Soltis & Soltis (1993) have highlighted over 30 examples of recurrent polyploid species formation in plants, mostly allopolyploids associated with an initial hybridization event. In the genus Senecio (Asteraceae), for example, the native UK species S. vulgaris (tetraploid) has hybridised on at least two separate occasions with alien S. squilidus (diploid) to produce two distinct, but morphologically similar, populations of the allohexaploid S. cambrensis in Wales and Edinburgh, respectively (Abbott & Lowe, 1996). Two examples from the USA are the allotetraploid hybrids Tragopogon mirus and T. miscellus, which may have formed on as many as 12 and 20 separate occasions within the last 70 yr (Soltis et al., 1995). Studies of the homoploid hybrid sunflower Helianthus anomalus, suggest that this hybrid, which is found across the western USA, has arisen on at least three separate occasions (Schwarzbach & Rieseberg, 2002) indicating that homoploid speciation, like allopolyploid speciation, can occur on a repeated basis.

As mentioned previously, newly formed hybrids are not true species unless they maintain their taxonomic identity and are reproductively isolated from both of their parental taxa, either genetically or ecologically (Ungerer et al., 1998). Allopolyploid hybrids are usually resistant to introgression with their parental taxa because of differences in ploidy, although this is not always the case (Petit et al., 1999). In the case of homoploid hybrids, other factors must influence their postzygotic isolation from parental taxa (Rieseberg, 1997). Grant (1981) proposed the ‘recombinational speciation’ model for this process, whereby the two parental species differ by two or more chromosomal rearrangements. In such a scenario, the hybrid will be heterozygous for these rearrangements and thus partially sterile because 75% of its gametes will be unbalanced and inviable due to deletions/insertion events. Half of the remaining viable gametes will recover parental chromosome structures, whilst the other half will possess recombinant karyotypes (Rieseberg, 1997). Should inbreeding occur in the hybrid, a small number of F₁ individuals will be produced that possess novel karyotypes (Fig. 1a). These offspring will be fertile among themselves but at least partially resistant to introgression with the parental species. However, the impact of chromosomal rearrangement on novel hybrid taxa does not end with this process. The possibility for recombination between the two parental chromosomes remains over successive generations and it is expected that this will lead to a progressive reduction in the size of parental linkage blocks (Ungerer et al., 1998). This process will eventually be countered by stabilisation of the hybrid genome such that further recombination is only possible between linkage blocks derived from the same parental species (Fig. 1b). A similar process apparently occurs in allopolyploid hybrids for different reasons (Fig. 1c, but also discussed in the section Chromosomal rearrangements).

A second possible mechanism for isolation of hybrids from their progenitors is ‘transgressive segregation’ (Grant, 1975), whereby new combinations of parental alleles in the hybrids may serve to ‘preadapt’ particular hybrids to survive in novel ecological niches unavailable to either parent (deVicente & Tanksley, 1993; Rieseberg et al., 1999). This process is predicted to arise from the additive and epistatic action of adaptively important alleles inherited from each parent, producing extreme, or transgressive phenotypes (Rieseberg et al., 1999). The reality of this prediction has now been demonstrated by a comparison of adaptive quantitative trait loci (QTL) in
natural and synthetic sunflower hybrids (Lexer et al., 2003a; Rieseberg et al., 2003). In this study, a hybrid sunflower population was resynthesised by crossing the diploid species *Helianthus annuus* and *H. petiolaris*, known to be the parents of three diploid hybrid species, *H. anomalus*, *H. deserticola* and *H. paradoxus*, adapted to sand dunes, desert basins and saline marshes, respectively (Rieseberg, 1997). It was then shown that certain extreme synthetic hybrid phenotypes could survive in the same extreme natural habitats as the hybrid species that they most closely resembled. Furthermore, chromosomal segments containing QTLs considered to be adaptive in each particular natural hybrid were also present in the corresponding synthetic hybrids (Lexer et al., 2003a, b; Rieseberg et al., 2003). This clearly demonstrates that extreme phenotypes, created by hybridization and chromosomal rearrangements, have suites of QTLs that preadapt them for survival in extreme habitats within which neither parental nor other hybrid phenotypes can survive. The new hybrid species are thus perfectly isolated ecologically from their parents and each other (Lexer et al., 2003a; Rieseberg et al., 2003). Given the difficulties in detecting adaptive trait introgression and hybrid speciation, this example of ecological adaptation linked...
to hybridisation is unique and provides a valuable model for future studies of plant speciation. Such studies will require the use of new molecular technologies available, as well as new approaches to utilising older techniques.

**Development of Molecular Approaches to the Study of Hybrid Speciation**

Historically, methods of studying hybrid species in animals or plants (often restricted to confirming a suspected hybridisation event) were restricted to observations of gross morphology (reviewed in Grant, 1981) or light microscopy of chromosome pairing at meiosis (Rieseberg et al., 2000). These approaches were often limited in plants because hybrids of closely related species do not necessarily show distinct morphological differences (Maki & Murata, 2001) and analysis of meiosis offers only minor insight into the interaction of different parental genomes in hybrid species (Rieseberg et al., 2000). As molecular techniques developed, it was possible to analyse hybrids in more detail, but these approaches tended to focus more on identifying ancient hybridisation events and changes to single genes or gene families rather than alterations to the entire genome (as is vital to understanding the full impact of hybridisation).

Initial molecular studies of hybridisation employed isozyme marker technology, which has been used extensively to determine the hybrid origins of plant species. Marker-based assays of hybrid origin rely on the assumption that species-specific markers identified in each parent can subsequently also be identified in the hybrid. This approach has proved useful in many studies of hybrid speciation: Soltis et al. (1995) used isozyme analysis in conjunction with restriction fragment length polymorphism (RFLP) analysis of chloroplast DNA to show the hybrid origins of *Tragopogon mirus* and *T. miscellus* (both tetraploid) via the inter-specific hybridization of *T. dubius* and *T. porrifolius* and *T. dubius* and *T. pratensis*, respectively. Similar studies of polymorphic isozyme loci in *Arisaema ehmense* supported the hypothesis that *A. ehmense* is a homoplody hybrid of the diploid species *A. serratum* and *A. touaense* (Maki & Murata, 2001). Isozyme marker assays were also used to confirm the origin of *Senecio squallidus* (Oxford ragwort) as a homoplody hybrid of two diploid Mediterranean taxa, *S. aethiopsis* and *S. chrysanthemifolius* (Abbott et al., 2000). This analysis showed that UK *Senecio squallidus* is almost certainly derived from material collected from a hybrid zone between these two species on Mount Etna (Sicily). Similar assays of the allohexaploid hybrid *Senecio cambrensis* (Welsh ragwort) showed that this species had arisen on at least two separate occasions in Wales and Scotland, respectively (Ashton & Abbott, 1992).

Analysis of the internal transcribed spacer (ITS) and intergenic spacer (IGS) regions of ribosomal RNA gene clusters have also been widely employed in studies of hybrid speciation (Appels & Dvorak, 1982; Bhatia et al., 1996; Lowe & Abbott, 1996; Rieseberg et al., 1996; Baumel et al., 2002). The rDNA genes of plants are organized within the genome in several clusters of highly repetitive sequences (Ritland et al., 1993). Although the genes themselves show little sequence divergence between closely related species, the transcribed regions between the genes (the ITS) display rates of divergence high enough to make the ITS a useful resource for phylogenetic studies (Baldwin, 1992; Baldwin et al., 1995). Because these sequences are transcribed but not translated, they tend to evolve rapidly and show high rates of mutation (Bhatia et al., 1996). RFLP analyses of the IGS region have proved very useful for determining the phylogenies of closely related plant species and can thus be used to determine the parental origins of existing hybrid species and even different populations of the same species (Appels & Dvorak, 1982). More recently, sequence level analysis has helped identify the divergent origins of diploid *Brassica* species, *B. nigra*, *B. campestris* and *B. oleracea*. ITS sequences predict that *B. nigra* diverged from *B. campestris* and *B. oleracea* relatively early in their evolutionary history because ITS sequences of *B. nigra* are very different from those of *B. campestris* and *B. oleracea*, which are relatively similar (Bhatia et al., 1996).

Despite the advantages of rDNA genes for phylogenetic studies, recent examination of drawbacks to the technique have suggested that relying solely on ITS and IGS data is problematic. rDNA genes are present in high copy numbers and are subject to concerted evolution (Wendel et al., 1995). Combined with the likely presence of pseudogenes, which can interfere with divergence studies, the potential for homoplasy in purely ITS/IGS-based studies is high (Alvarez & Wendel, 2003; Bailey et al., 2003). This does not invalidate the technique, as rDNA genes are the only currently available nuclear-encoded region that may be universally amplified between species (Bailey et al., 2003). However, comparative sequencing of single copy nuclear-encoded genes is also required and indeed recommended (Raymond et al., 2002; Alvarez & Wendel, 2003). The greater utility of studying single copy genes for phylogenetic analysis is demonstrated in the study of Baumel et al. (2002), who employed comparative sequencing of a region within the Waxy granule-bound starch synthase gene of several *Spartina* taxa as the basis for a phylogenetic analysis. The study compared the results of sequencing the *Waxy* gene with data from ITS and chloroplast DNA sequencing, and showed that comparative sequencing of single copy nuclear genes produced the most informative phylogenetic data. The work of Cronn et al. (2003) in determining the ancestry of the allotetraploid hybrid *Gossypium gossypoides* takes this one stage further: by differentiating between homoeologous gene copies; it was demonstrated that previous cpDNA and ITS sequencing had produced misleading results due to widespread introgression after hybrid formation.

RFLP analysis of chloroplast DNA has also proved useful in the detection of hybridisation events. Studies of chloroplast
DNA haplotypes using RFLP can aid construction of phylogenies because chloroplast genomes are maternally inherited and have low frequencies of structural change and sequence evolution (Olmstead & Palmer, 1994). Schwarzbach & Rieseberg (2002) used this approach to demonstrate the potential multiple origins of the diploid hybrid sunflower Helianthus anomalus and H. petiolaris, but these results conflict with those from simple sequence repeat (SSR) marker assays conducted as part of the same analysis, which suggested a single origin. Schwarzbach & Rieseberg (2002) suggest that the highly mutable nature of SSRs could account for this discrepancy, illustrating the importance of not relying on a single marker system when studying complex genetic events, such as speciation. This is an important caveat for such studies, which will require the accumulation of evidence from multiple sources before definitive answers can be given.

It is clear therefore that marker assays become more powerful when used in combination with each other. Studies of Senecio cambrensis using RFLP analysis of intergenic spacer (IGS) and chloroplast DNA, together with isozyme assays, showed that this species is closely related to a species from the Canary Islands, Senecio teneriffae (Lowe & Abbott, 1996). Lowe & Abbott (1996) hypothesised that, whilst S. cambrensis is a hybrid of S. squalidus and S. vulgaris, the related S. teneriffae was the result of hybridisation between S. vulgaris and S. glaucus, a diploid species closely related to S. squalidus. As these examples illustrate, marker assays have been a powerful tool for studies of hybrid speciation in the past, but efforts have focused primarily on detecting the parentage and origin of hybrids, rather than on resolving issues such as the genetic and transcriptomic impact of hybrid formation. Studies aimed at examining this impact require the use of new technologies, along with novel approaches to those already mentioned.

**Current Molecular Approaches to Studying Hybrid Speciation**

New techniques for high-throughput and rapid analysis have revolutionised the use of molecular markers in studies of hybrid speciation. These techniques provide the opportunity to gain new insights into long-standing questions about the consequences of hybridisation at the level of genome and transcriptome.

**Recent advances in marker technologies**

Marker systems, such as amplification fragment length polymorphisms (AFLPs) and SSRs lend themselves to high-throughput methods of data generation (if not analysis) and are proving to be the markers of choice for construction of high-density chromosome maps (Gupta et al., 1996; Powell et al., 1996), which are vital for studying genomic rearrangement in hybrid taxa (Rieseberg, 1998). Recent technological advances have permitted the use of automated sequencing systems to perform AFLP/SSR experiments employing fluorescently tagged products, either using polyacrylamide gel systems (Koelmer et al., 1998; Arnold et al., 1999) or capillary electrophoresis (Wenz et al., 1998; Myburg et al., 2001). These systems represent a quantum leap forward in AFLP and SSR analysis, as they are now easier to use and faster to perform (Koelmer et al., 1998), but have the drawback of being expensive. Despite the advantages of these systems, there are still problems when attempting to identify true alleles in the case of highly polymorphic loci, which may require extensive sampling of material to identify parental alleles. The true power of these techniques becomes apparent when the ability to multiplex is factored in; using high-throughput capillary sequencers, it is now possible to perform up to 70,000 genotyping assays in a single week (Myburg et al., 2001), which has revolutionised the way that marker assays can be employed. Advances in high-throughput molecular marker technologies, together with improvements in bioinformatics software to simplify and speed up data analysis, are now making molecular marker assay systems accessible to more users, particularly with the advent of commercial genotyping services.

**cDNA-SSCP and cDNA-AFLP**

Use of marker assays on first-strand cDNA rather than genomic DNA has provided two related techniques for assessing differences in gene expression and distinguishing between separate gene copies. Single-stranded conformation polymorphism (SSCP) relies upon slight differences in the behaviour of single-stranded DNA fragments on nondenaturing electrophoresis gels (Orita et al., 1989). Single nucleotide differences in single-stranded DNA (ssDNA) alter the secondary and tertiary structures of the DNA in nondenaturing conditions and thus allow separation of gene fragments containing even a single point mutation (Cotton, 1997). By performing SSCP analysis of first-strand cDNA, it is possible to determine differences in gene expression between homeologues of the same gene (Adams et al., 2003), although the possibility of intraspecific variation and presence of gene family members must be taken into account. One drawback to the technique is that it requires the design of gene-specific oligonucleotide primers and consequently is limited to relatively small-scale analysis of gene expression.

cDNA-AFLP is a related technique that relies upon similar polymorphic differences between homeologous gene copies. In this case, the standard AFLP procedure is performed using a pool of cDNA generated from a given tissue (Bachem et al., 1996). The system represents an improvement on previous techniques such as differential display (Liang & Pardee, 1992) because the possibility of competition between abundant and rare transcripts during the PCR amplification is eliminated (Bachem et al., 1996). As with cDNA-SSCP, the cDNA-AFLP
technique can differentiate between closely related gene copies, an advantage over other rapid whole-genome systems such as microarrays, which rely upon a hybridisation-based approach. However, one drawback to cDNA-AFLP not shared with cDNA-SSCP is that the identity of differentially expressed genes is not known immediately and must be confirmed by isolation and sequencing of the restriction fragment (Osborn et al., 2003). Despite their drawbacks, these two methods represent powerful molecular tools for studying genome-wide gene expression profiles.

Microarray analysis

The development of large-scale expression assays by hybridisation has enabled the determination of gene expression at the level of the whole transcriptome (Seki et al., 2001). Essentially, microarrays function like Northern blots, except that the ‘probe’ DNA is bound to the solid support (a treated glass slide) and the ‘target’ transcript is labelled with a fluorescent tag and hybridised to the array under a coverslip. By labelling two mRNA populations from different sources with two distinct fluorescent tags, the relative expression levels of transcripts from each source can be compared in a single hybridisation.

DNA-based microarrays currently use one of two formats. In the first format, the ‘probe’ DNA spots consist of PCR-amplified cDNAs (Schena et al., 1995). In the second, the ‘probe’ DNA molecules are oligonucleotides designed to complement the 3′ untranslated regions of genes (Aharoni & Vorst, 2001). The latter array type is more powerful, as it allows the user to distinguish different homeologues of the same gene (the 3′ UTR is variable between different gene copies), but has the disadvantage that extensive sequence data must be available. Consequently, such arrays are typically only viable for species that have been studied in whole genome sequencing programs such as Arabidopsis thaliana and Oryza sativa. Although the majority of microarray studies use fully sequenced unigene sets, cDNA microarrays can be created without extensive sequencing of the clones to be used, but these may be prone to cross-hybridisation between homologous gene copies.

Microarrays have the potential to provide rapid comparisons of gene expression across a large number of genes, but there are drawbacks. The technology is relatively expensive, and the cost increases if custom arrays must be produced. However, the recent advent of ‘anonymous’ cDNA arrays, whereby clones are sequenced only after they display interesting expression patterns, is helping to offset the problem of cost. Experimental problems, such as differences in sample preparation, labelling efficiency, slide quality and biological variation between samples means that experimental design must be rigorous in order to ensure statistically viable results (Churchill, 2002). The most unusual drawback, however, is that of data overload: Microarray experiments have the potential to generate vast amounts of data and so it is vital that sensible cut-offs be applied (Hess et al., 2001). Careful choice of bioinformatics software and statistical analyses is vital. Despite these drawbacks, microarray technology is perhaps the most powerful genomic tool currently available to molecular biologists.

Chromosome painting

Chromosome painting refers to the use of fluorescent in situ hybridisation (FISH) of specific DNA probes to specific chromosomes or chromosomal segments (Lysak et al., 2001). Until recently, chromosome painting of specific chromosome segments had been impracticable as a tool for studying euploid plants (Lysak et al., 2001), although FISH, using total genomic DNA had been successfully employed to distinguish different parental chromosome sets (Schwarzacher et al., 1989) and even specific chromosome segments introgressed into hybrid lines (Schwarzacher, 1997). Nevertheless, painting of specific chromosomes proved problematic. Difficulties with reproducibility were common, possibly due to the high levels of repetitive DNA sequences within many plant genomes (Fuchs et al., 1996), as well as active interchromosomal homogenisation of this repetitive DNA (Schwarzacher et al., 1997). Attempts to solve these problems by probing with individual large genomic clones (bacterial artificial chromosomes [BACs] and yeast artificial chromosomes [YACs]) in species with small genomes and low repetitive DNA content proved successful, for example in rice (Jiang et al., 1995) and cotton (Hanson et al., 1995), amongst others. Recently, Lysak et al. (2001) succeeded in applying these techniques to Arabidopsis thaliana, successfully painting an entire chromosome, and later two others (Lysak et al., 2003). This has opened the door for using the technique to study chromosomal rearrangements and homologue associations (Lysak et al., 2001) and also allows comparative chromosome painting in plants (Lysak et al., 2003). Such techniques provide an important additional tool to marker-generated chromosome maps for the study of chromosome evolution during hybrid speciation.

New Insights into Hybrid Speciation from New Molecular Technologies

The new molecular technologies discussed above are rapidly providing (or have the potential to provide) crucial insights into the predicted genetic consequences of hybrid speciation, namely: chromosomal rearrangements, transposon activation, rapid sequence elimination, and gene silencing.

Chromosomal rearrangements

There are two ways in which chromosomal rearrangements can influence hybrid taxa. The first of these, recombinational
Evidence that chromosomal rearrangements take place further isolation of the hybrid from its progenitors (Fig. 1b). Rearrangement is predicted to occur, whereby recombinants are established, the second form of chromosomal rearrangement in the diploid hybrid sunflower *Helianthus annuus* marker types (RAPDs and AFLPs) in a study of chromosomal markers. Rieseberg and coworkers employed a variety of marker assays. Rieseberg and coworkers employed a variety of marker types (RAPDs and AFLPs) in a study of chromosomal rearrangements in the diploid hybrid sunflower *Helianthus anomalus*. Since the hybrid contains markers from the parental taxa, *H. annuus* and *H. petiolaris*, it was possible to use species-specific markers to assess rearrangements of parental linkage groups in the hybrid over successive generations following hybridisation. Using a combination of RAPD (Rieseberg et al., 1996) and AFLP markers (Ungerer et al., 1998), Rieseberg and coworkers were able to demonstrate the rapid rearrangement of parental linkage blocks in newly synthesised hybrids to a form approximating those of the wild hybrid plants within 10–60 generations. Indeed, the RAPD study, which also employed ITS sequencing and isozyme assays, showed that the reconstruction of basic linkage blocks within synthetic hybrids could occur within as few as 3–5 generations after hybrid formation (Rieseberg et al., 1996).

Perhaps surprisingly, chromosomal rearrangement also occurs during allopolyploid speciation, potentially at even faster rates than in diploid hybrids (Gale & Devos, 1998a). This is surprising because theoretically there should be no problems with chromosome pairing because allopolyploid hybrids usually possess a complete copy of each parental genome. However, because closely related species often display colinearity of gene order (Gale & Devos, 1998b), it is possible that mispairing between highly homeologous chromosomes might occur during meiosis (Moore, 2002). Recombination of one or both parental genomes would reduce this possibility by making the two genomes non-homologous (Fig. 1c). Studies in cereals using a variety of markers for chromosome mapping have confirmed the rapid rate of genomic change in newly synthesised allopolyploids (Liu et al., 1998), but further studies are clearly needed to confirm the widespread nature of this phenomenon.

Chromosomal rearrangements following hybridisation have also been demonstrated visually using chromosome painting techniques. Comai et al. (2003) used chromosome-specific tags, and centromeric tags specific to Arabidopsis *thaliana* and *A. arenosa*, to study the genome composition of synthetic allotetraploids and the natural hybrid, *A. suecica*, derived from these two *Arabidopsis* species. Hybrids contained 16 chromosomes from *A. arenosa* and 10 from *A. thaliana*, with two chromosomes of *A. arenosa* being homeologous to chromosome 4 of *A. thaliana* based on hybridisation to *A. thaliana* probe DNA. Observations of centromere behaviour during meiosis revealed that, although chromosomes of different parental origins coalesced at early prophase I, they had resolved themselves into proper pairings by metaphase, so confirming the hypothesis that chromosomal rearrangements rapidly enable homeologous chromosomes to pair correctly.

Chromosome painting has also revealed two types of chromosomal rearrangement in new polyploids: random translocations, occurring in different chromosomes in different populations of the same hybrid, and species-specific translocations, which involve specific chromosomes and are found in all populations of a particular hybrid (Jiang & Gill, 1994). It has been proposed that these species-specific rearrangements may be a response to changes in nuclear–cytoplasmic interactions (Leitch & Bennett, 1997; Wendel, 2000). Because the cytoplasmic genome of a newly arisen allopolyploid is derived solely from the female parent, genome rearrangements may be necessary to restore nuclear-cytoplasmic compatibility (Soltis & Soltis, 1999). Indeed, studies by Song et al. (1995) using RFLP marker assays in synthetic *Brassica* polyploids showed that rearrangement events tended to occur primarily in the paternally contributed genome. This finding is reinforced by earlier observations that the genome of the natural allotetraploid *B. juncea* is more similar to that of its maternal diploid progenitor, *B. rapa*, than to its paternal parent, *B. nigra* (Song et al., 1988). Whilst it is a commonly held view that chromosomal rearrangement is prevalent in newly formed hybrids, there are exceptions to this apparent rule; findings from AFLP and methylation-sensitive AFLP (MSAP) analysis of the allopolyploid hybrid *Spartina anglica* indicate that there has been no significant genome reorganisation since the initial hybridisation event, suggesting that all phenotypic variation between the hybrid and the parental taxa may be caused by epigenetic factors rather than as a result of chromosome duplication (Ainouche et al., 2003). Such factors may include organ-specific silencing of genes from one parental genome (Adams et al., 2003) or genomic imprinting leading to different developmental regulation via parent-of-origin effects on gene expression (Alleman & Doctor, 2000). Similarly, work by Liu et al. (2001) showed no change in genome structure associated with hybridisation and polyploidisation in synthetic cotton hybrids. Most strikingly, RFLP analysis by Axelson et al. (2000) in resynthesised and natural *Brassica juncea* hybrids suggests no significant genome reorganisation has occurred in either, contrary to the findings of Song et al. (1995). These exceptions to the apparent rule of chromosomal rearrangement in allopolyploids suggest that certain hybrid species may well be able to adapt to a novel environment at the phenotypic level, without a corresponding
reorganisation of the genome. These adaptations may involve gene silencing, although Liu et al. (2001) employed a variety of methylation-sensitive and insensitive AFLP markers in their analysis, showing that there is no change in methylation state in cotton hybrids. Hence, other epigenetic factors, such as alterations in chromatin folding (Liu & Wendel, 2003) or novel combinations of homeologous regulatory systems (Riddle & Birchler, 2003), may play a role in this process.

Although the prevalent view of chromosomal rearrangement is that it is a consequence of speciation (Sites & Moritz, 1987), there is a growing school of thought that suggests chromosomal rearrangement may play a role in reinforcing speciation by limiting the potential for introgressive backcrossing to the progenitor species (Rieseberg, 2001). Using RAPD and RFLP markers it was shown that chromosomal differences restrict recombination across large linkage blocks (Rieseberg et al., 1995) and so can potentially limit gene flow due to introgression. Whilst restricted recombination alone is probably not sufficient to cause speciation, it may act to facilitate speciation in combination with other factors. For example, Buerkle et al. (2000) found that, in theory, genetic isolation between a newly arisen hybrid and its parent species could be maintained if chromosomal barriers to fertility were strong and there was at least a partial geographical/ecological divide between the hybrid and the parents. This may explain how hybrids can become established alongside their parental taxa, leading to the examples of parapatric and sympatric speciation we see today. For example, the hybrid sunflowers studied by Rieseberg et al. (2003) show dramatically reduced fertility with the parents in the F1 generation, typically 10% of normal (Ungeer et al., 1998), whilst subsequent generations display rapid increases in fertility amongst themselves. Indeed, decrease in hybrid-parent fertility is often a direct consequence of genetic events leading to increased fertility among the hybrids (Grant, 1981). Combined with the transgressive adaptation of these hybrids to favour nearby habitat zones that are unsuitable for either parent species (Rieseberg et al., 2003), such a reduction in fertility between the hybrids and their progenitors may reinforce their isolation and emergence as true species.

Another possible mechanism for causative effects of chromosomal restructuring on speciation is given by the r theory (Noor et al., 2001; Rieseberg, 2001), which suggests that genomic blocks protected from gene flow by chromosomal rearrangements may serve as sites for accumulation of genic factors preventing interfertility with parental taxa. As these factors accumulate, complete reproductive isolation between the hybrid and the parental species slowly develops. Evidence for this comes from studies of the Solanaceae, where a paracentric inversion identified on chromosome 10 of tomato (Lycopersicon esculentum) is believed to have arisen after divergence from the common ancestor with potato (Solanum tuberosum). Studies showed that the divergence of genes mapped to chromosome 10 is greater between the two species than that observed in genes on collinear chromosomes. Such a finding is consistent with the predictions of the r theory. Despite this, there are still many questions to be answered about the role of chromosomal rearrangements in speciation. These include whether rates of speciation are correlated more strongly with rearrangements that inhibit recombination directly rather than indirectly, and whether rearrangements contribute to selection for reproducitively isolated hybrids (reviewed in Rieseberg, 2001).

Transposon activation

It has been proposed that mobile genetic elements (transposons) may become highly active in newly arisen hybrid species as a consequence of genomic instability or ‘genomic shock’ (McClintock, 1984). Another theory suggests that transposons may facilitate rapid genomic reorganisation in new polyploid species (Matzke & Matzke, 1998). This is possible because polyploids contain duplicate copies of every gene, so that transposon insertions into ‘single’ copy genes are less deleterious to the plant. This means that transposons can multiply and persist for far longer in polyploids than in diploids. It has been argued that transposons have played a major role in the evolution of gene silencing mechanisms via methylation (McDonald, 1998) because many transposable element systems are inhibited by methylation. If this is the case, polyploid species – which possess a higher number of transposons – should also display higher levels of DNA methylation. Indeed, a rough correlation between transposon copy number and levels of DNA methylation has been shown (Matzke & Matzke, 1998), but more data is required before a clear generalisation can be made.

Transposon insertions into genic regions can serve as a source of genome rearrangement because the activity of transposons within the host genome can increase the likelihood of chromosome breakage (Weil & Wessler, 1993), sequence amplification/gene duplication (Jin & Bennetzen, 1994) and may even lead to altered patterns of gene expression (Martiensoenn et al., 1989). Osborn et al. (2003) discuss the possibility of using transposon display (Hanley et al., 2000; Melayah et al., 2001) to investigate whether epigenetic changes in polyploids target transposons. Transposon display relies on the generation of transposon-tagged fragments of DNA that provide a banding profile which shows transposable element insertions within a particular plant (Hanley et al., 2000). Differences in the banding patterns of polyploids compared with their progenitors would identify candidate transposon insertion sites for targeted studies of the DNA methylation status of those elements inserted in genes encoding regulatory functions, such as transcription factors. Such an approach would certainly provide clearer insights into how transposon activity might facilitate genome reorganisation in newly formed polyploids.
Rapid sequence elimination

Another consequence of hybrid species formation, possibly associated with chromosomal rearrangements, is the rapid and reproducible loss of low copy DNA sequences from hybrid genomes (Feldman et al., 1997). Feldman et al. (1997) studied several low copy sequences that were either chromosome-specific or genome-specific in existing and newly synthesised polyploids of bread wheat (*Triticum aestivum* L.). By hybridising these sequences to genomic DNA from the hybrids, it was found that differential and nonrandom elimination of sequences occurred on two of the three homeologous chromosome pairs in hexaploid wheat, and that these changes were reproduced in all plants tested, whether wild, cultivated or newly synthesised. Later studies using AFLP analysis showed that this sequence loss may occur on a large scale – c. 14% of sequence fragments tested (Shaked et al., 2001). Eight of nine AFLP fragments sequenced were found to correspond to low copy DNA (Shaked et al., 2001).

The fact that these sequences were eliminated in both the synthetic and natural hybrids suggests that this may be a mechanism to facilitate differentiation of homeologous chromosomes (Eckhardt, 2001). Whilst the mechanism by which this sequence removal occurs is unknown (although Shaked et al. (2001) provide some intriguing suggestions), it is clear that the process represents a further mechanism for differentiation of homeologous chromosomes and insurance of correct meiotic pairing in the hybrid (Fig. 1c). Hopefully, the recent advances in large-scale marker technologies will enable rapid experimental determination of deletion/recombination breakpoints and allow determination of how this mechanism operates.

Gene silencing

An important consequence of hybrid speciation in polyploids is gene silencing. The union of two divergent genomes can lead initially to ‘genome shock’ when new allopolyploid species are suddenly confronted with a situation where they possess redundant and divergent homeologues of many genes (McClintock, 1984). Recently it has been possible to demonstrate that genome shock leads to widespread gene silencing, perhaps as an attempt by the hybrid plant to stabilise its genome (Comai et al., 2000). The experimental technique used by Comai et al. (2000) employed cDNA-AFLP to screen c. 700 cDNAs from leaf and flower tissue of polyploid hybrids between *Arabidopsis thaliana* and *A. arenosa*. Their analysis identified several cDNAs present in both parents that were missing in the hybrids and two cases of novel cDNA-AFLP fragments appearing in the hybrids. Pursuing likely candidates for gene silencing using RT-PCR analysis, Comai et al. (2000) confirmed silencing in three genes for all hybrids tested, and two other genes silenced only in a single hybrid. This suggests that c. 0.4% of genes may be silenced in allopolyploids, but this is likely to be an underestimate because in this particular RT-PCR assay, partial gene silencing is scored as un-silenced (Comai et al., 2000).

The classical model of genome evolution set out by Ohno (1970) predicts that duplicate genes will be subject to silencing and eventually lost due to mutational events over time. However, if all duplicated genes in polyploids were silenced and eventually lost, the effect of polyploidy on the evolution of new species would be minimal (Otto & Whitton, 2000). Thus, many plant species show high numbers of duplicated genes retained over long periods of evolutionary time; maize, for example, possesses duplicate copies of c. 72% of its genes (Whitkus et al., 1992). Recent theories (Kellogg, 2003; Otto, 2003) have suggested that many of these duplicated genes are maintained in the polyploid genome due to subfunctionalisation, whereby the duplicate copies of a gene suffer deleterious but complementary mutations such that both copies are required for phenotypic normality (Lynch & Force, 2000).

Another mechanism of subfunctionalisation is discussed by Adams et al. (2003) who showed that duplicated genes display organ-specific reciprocal silencing. Adams et al. (2003) analysed the expression of homeologous copies of 40 gene pairs in polyploid cotton (*Gossypium hirsutum*). Their technique allowed approximate quantification of transcript levels between the two homeologs, and showed that 10 of the gene pairs displayed biased expression from one parental genome or the other. Further analysis of 16 gene pairs in 10 different tissues showed organ-specific silencing effects in 11 genes in at least one tissue type. This showed that a relatively high percentage of genes display silencing effects or biased expression in a developmentally regulated manner, although there did not appear to be preferential transcription of genes from one particular genome.

Developmental regulation of genes through silencing can be seen in the phenomenon of genomic imprinting, where parent-of-origin-specific control of gene expression is observed (Alleman & Doctor, 2000). Imprinting effects can have severe consequences for hybrid species. In studies of *Arabidopsis* polyploids, it was observed that developing seeds showed unusual phenotypes depending on whether the paternal or maternal genome contributions were increased (Scott et al., 1998). If the maternal genome contribution was larger, endosperm development in the seeds was inhibited, whilst if the paternal contribution was increased the seeds showed increased growth of endosperm and embryo. Scott et al. (1998) hypothesised that this was the result of imprinting for endosperm developmental genes.

As with most examples of DNA-based gene silencing, genomic imprinting involves the methylation of cytosine residues (Adams et al., 2000). Changes in DNA methylation frequently occur in the genes of newly formed allopolyploids, as shown by AFLP and cDNA-AFLP analysis (Comai et al., 2000; Shaked et al., 2001), but there are relatively few documented cases of early silencing of redundant gene copies in...
polyploid formation (Pikaard, 2001). As Pikaard discusses, one exception is nucleolar dominance, in which rDNA genes from one parental genome are silenced independently of paternal or maternal origin (Pikaard, 2000). Other recent findings indicate that some redundant protein-coding genes and putative transcription factors can also be silenced in this manner (Comai et al., 2000; Shaked et al., 2001). Such studies raise the possibility that levels of gene silencing in polyploids may be quite low, with estimates ranging from c. 0.4% in synthetic allotetraploid hybrids (Comai et al., 2000) to c. 2.5% in natural allotetraploids (Lee & Chen, 2001).

Interestingly, however, several of the silenced genes identified in these studies were transcription factors, so a knock-on effect of reduced transcription for genes under the control of these transcription factors might be predicted, suggesting an indirect route for gene silencing in allopolyploids. This could explain in part the developmentally regulated manner in which silencing was observed in cotton (Adams et al., 2003): different homeologues of a transcription factor might be silenced in different tissues, resulting in a further silencing of downstream genes common to the same parental genome.

Another reason why gene silencing may be necessary in newly formed polyploids is that there are many classes of protein that are dimeric or polymeric in their native states. Studies in Drosophila have shown that different gene homeologues can encode different monomeric subunits of these proteins, and that proteins, formed from subunits of different origins (heterodimeric proteins), may be functionally compromised (Phillips et al., 1995). If this is the case in allopolyploids, silencing of one genomic copy of the subunit-encoding genes may occur in order to prevent heterodimer formation. However, an alternative view (Gottlieb, 2003) proposes a scenario where heterodimeric proteins might actually contribute to hybrid vigour. Other mechanisms of gene silencing might also play a role in speciation. For instance, cosuppression where ‘silencing’ occurs at the post-transcriptional level, might be involved in reducing levels of protein translated from redundant gene copies in a dosage-dependent fashion (Wolffe & Matzke, 1999).

With the advent of new technologies that enable genome-wide expression assays, patterns of gene expression in new polyploids can be readily investigated. Microarray analysis has already been used to study the effects of polyploidisation in yeast (Galitski et al., 1999), highlighting the feasibility of such an approach in multicellular eukaryotes. Indeed, Chen and coworkers are currently using oligonucleotide arrays to perform genome-wide transcriptome analysis of autoploid and allotetraploid in Arabidopsis (Chen et al., 2004), a task facilitated by the commercial availability of these arrays. Comparison between pooled RNA from the parental taxa and the hybrids has identified over 1000 genes significantly up or down-regulated out of the 26 090 genes present on the array. Significantly, more of these changes are found in allotetraploids than in autoploid. Early indications in polyploids show differential expression of transposons, transcription factors and DNA repair enzymes, as well as factors involved in programmed cell death, signal transduction, light regulation, protein synthesis and some noncoding RNAs. Analysis of concurrent generations following polyploidisation has shown that some genes may be silenced as early as the second generation, whilst others change at a slower rate (Wang et al., 2004).

As an alternative to oligonucleotide arrays, cDNA microarrays enable investigation of hybrid speciation in species for which extensive genome sequence is not available. This is important because many of the species classically used to study polyploidy, Tragopogon for example, are unlikely to ever be sequenced en masse. Recently, this problem has been addressed by the development of so-called ‘anonymous’ cDNA microarrays, where the majority of clones spotted are of unknown sequence. Sequencing is then performed ex post facto only on those genes displaying interesting expression patterns. Our own work focuses on the use of such arrays to study both homoploid and allotetraploid speciation in the genus Senecio, and has already produced an extensive list of clones displaying differential expression between hybrids and their parental taxa (M.J. Hegarty & S.J. Hiscock, unpublished data).

Conclusion

Traditional models of speciation, such as the allopatric and sympatric models (Levin, 2002), are difficult to study experimentally at the genomic level. However, the formation of new species by interspecific hybridisation in sympatry offers a real opportunity to study speciation directly at the level of the genome and transcriptome, particularly because many hybrids can be re-synthesised experimentally. Recent experiments have combined traditional molecular tools with newer technologies to provide new and exciting insights into the genetic and evolutionary processes associated with hybrid speciation. Molecular approaches have, for the first time, demonstrated directly that interspecific hybridization is a mechanism for adaptive evolution (Rieseberg et al., 2003). An exciting new era in plant evolutionary biology has begun and researchers must embrace the new molecular technologies available to address more fundamental questions about the genetic processes of homoploid and allotetraploid speciation in plants and probe even more deeply into what happens when divergent genomes collide.

Acknowledgements

We thank Richard Abbott, Stephen Harris and Keith Edwards for valuable comments on earlier drafts of this manuscript. We would also like to thank two anonymous referees for useful advice on its improvement. Work in SJH’s laboratory is funded by the Natural Environment Research Council and the Biotechnology and Biological Sciences Research Council.
References


