Annals of Botany **106**: 439–455, 2010
doi:10.1093/aob/mcq126, available online at www.aob.oxfordjournals.org

ANNALS OF BOTANY
Founded 1887

GENES IN EVOLUTION: THE CONTROL OF DIVERSITY AND SPECIATION

Speciation genes in plants

Loren H. Rieseberg[1,2,*] and Benjamin K. Blackman[3]

[1]*Botany Department, University of British Columbia, 3529-6270 University Blvd, Vancouver, B.C. V6T 1Z4, Canada,* [2]*Biology Department, Indiana University, 1001 E Third St., Bloomington, IN 47405, USA and* [3]*Biology Department, Duke University, Box 90338, Duke University, Durham, NC 27708, USA*
**For correspondence. E-mail: lriesebe@interchange.ubc.ca*

Received: 12 February 2010 Returned for revision: 11 May 2010 Accepted: 13 May 2010 Published electronically: 24 June 2010

• *Background* Analyses of speciation genes – genes that contribute to the cessation of gene flow between populations – can offer clues regarding the ecological settings, evolutionary forces and molecular mechanisms that drive the divergence of populations and species. This review discusses the identities and attributes of genes that contribute to reproductive isolation (RI) in plants, compares them with animal speciation genes and investigates what these genes can tell us about speciation.
• *Scope* Forty-one candidate speciation genes were identified in the plant literature. Of these, seven contributed to pre-pollination RI, one to post-pollination, prezygotic RI, eight to hybrid inviability, and 25 to hybrid sterility. Genes, gene families and genetic pathways that were frequently found to underlie the evolution of RI in different plant groups include the anthocyanin pathway and its regulators (pollinator isolation), *S* RNase-SI genes (unilateral incompatibility), disease resistance genes (hybrid necrosis), chimeric mitochondrial genes (cytoplasmic male sterility), and pentatricopeptide repeat family genes (cytoplasmic male sterility).
• *Conclusions* The most surprising conclusion from this review is that identities of genes underlying both prezygotic and postzygotic RI are often predictable in a broad sense from the phenotype of the reproductive barrier. Regulatory changes (both *cis* and *trans*) dominate the evolution of pre-pollination RI in plants, whereas a mix of regulatory mutations and changes in protein-coding genes underlie intrinsic postzygotic barriers. Also, loss-of-function mutations and copy number variation frequently contribute to RI. Although direct evidence of positive selection on speciation genes is surprisingly scarce in plants, analyses of gene family evolution, along with theoretical considerations, imply an important role for diversifying selection and genetic conflict in the evolution of RI. Unlike in animals, however, most candidate speciation genes in plants exhibit intraspecific polymorphism, consistent with an important role for stochastic forces and/or balancing selection in development of RI in plants.

Key words: Speciation, reproductive isolation, mating system isolation, pollinator isolation, ecological isolation, unilateral incompatibility, hybrid necrosis, hybrid sterility, hybrid inviability, hybrid breakdown, cytoplasmic male sterility, restoration.

INTRODUCTION

No other recent advance in speciation studies has received as much attention as the cloning and characterization of genes that contribute to the cessation of gene flow or reproductive isolation (RI) between populations (Orr and Presgraves, 2000; Butlin and Ritchie, 2001; Noor, 2003; Orr *et al.*, 2004; Rieseberg *et al.*, 2004; Noor and Feder, 2006; Mallet, 2006; Bomblies and Weigel, 2007*b*; Rieseberg and Willis, 2007; Bomblies, 2010; Presgraves, 2010). These so-called 'speciation genes' are of interest because knowledge of their identities and attributes offers clues to the ecological settings, evolutionary forces and molecular mechanisms that drive the divergence of populations and species (Orr *et al.*, 2004, 2007). For example, information about gene identity and normal function may suggest particular traits or functions that are prone to disruption by deleterious interactions among divergent alleles. Patterns of sequence evolution can be used to detect the action of positive selection during speciation, and functional information may provide insight into the roles of different kinds of mutations (coding, regulatory, copy number, microchromosomal rearrangements, etc.) in diversification.

Barriers to gene flow caused by speciation genes can arise at multiple prezygotic and postzygotic life-history stages (Fig. 1). Despite widespread interest in speciation genes, most discussions have focused on genes that contribute to intrinsic postzygotic reproductive barriers such as hybrid sterility or inviability (Ting *et al.*, 1998; Coyne and Orr, 2004; Wu and Ting, 2004; Orr, 2005; Haerty and Singh, 2006; Mallet, 2006; Bomblies and Weigel, 2007*b*; Orr *et al.*, 2007; Presgraves, 2010). This focus has some justification because until very recently there were few *a priori* expectations regarding the identities and normal functions of genes that cause hybrid incompatibilities. In contrast, genes underlying prezygotic reproductive barriers were expected to be associated with the genetic pathways or networks that underlie the barrier phenotype of interest. However, as discussed below, candidate genes (or gene families) for both prezygotic and postzygotic isolation are increasingly predictable in a broad sense from barrier phenotype. Moreover, given the principal role of prezygotic barriers in the origins of many plant and animal species, a more inclusive discussion of the genes and mutations that underlie species formation is warranted.

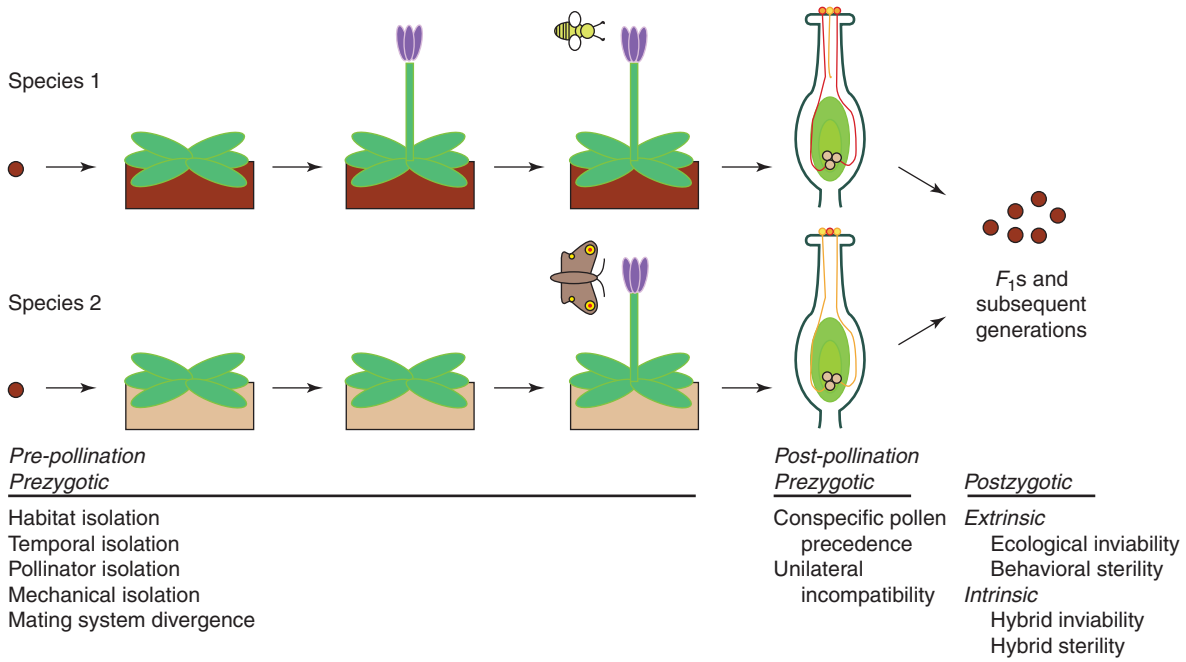

FIG. 1. Reproductive isolating barriers occur at multiple prezygotic and postzygotic life-history stages.

Another characteristic of the literature on speciation genes has been an emphasis on animal systems, particularly *Drosophila*. Botanists are partially to blame for this bias because relevant plant genes have rarely been discussed in the context of RI and speciation, at least until recently (although see Rieseberg *et al.*, 2004; Bomblies and Weigel, 2007*a*; Moyle, 2008; Widmer *et al.*, 2008; Bomblies, 2010). Also, many potential speciation genes in plants have been identified in crop species or exhibit intraspecific polymorphism, so their role in speciation is unclear (Rieseberg and Willis, 2007).

In this review, we partially rectify these earlier omissions by compiling a list of genes that underlie both prezygotic and postzygotic reproductive barriers in plants (Table 1). We also discuss the normal functions of these genes and describe the mutations that are the cause of RI (Appendix 1). We use this information to address a number of fundamental questions about the genetics of speciation, including the following. (1) What kinds of genes and mutations underlie reproductive barriers in plants and are there differences between plants and animals, prezygotic and postzygotic barriers, and hybrid sterility and inviability? (2) What is the relative importance of regulatory versus coding mutations in the evolution of RI? (3) What role has genetic redundancy played in the evolution of hybrid incompatibilities? (4) What evolutionary forces (i.e. natural selection, sexual selection, drift) are responsible for the divergence of the genes underlying RI? And (5) what are the taxonomic and geographical distributions of RI inducing mutations and do these patterns differ between plants and animals and between prezygotic and postzygotic barriers?

WHAT IS A SPECIATION GENE?

A speciation gene can be strictly defined as a gene that contributes to the splitting of two lineages by reducing the amount of gene flow between them. One concern with this definition is that if the cessation of gene flow is solely a consequence of geographical isolation, then one could argue that no genes contributed to speciation in these instances. Yet, we would not consider speciation to be complete until genetically based barriers to gene flow had evolved between the geographically isolated populations. Another concern is that speciation is an ongoing process and additional changes that result in further RI can accumulate after speciation is already complete. This creates two uncertainties. First, genes contributing to incipient speciation may be segregating within species. Second, for genes contributing to RI between species, it can be difficult to distinguish between those genetic changes that played causal roles in speciation and those that arose after speciation was complete. In considering these problems, we have chosen to treat genes contributing to prezygotic and postzygotic isolating RI differently. For postzygotic barriers, we have broadened the definition of speciation genes in this review to include any gene that contributes to reproductive isolation between populations, even if the contribution is likely to be small. This definition therefore accommodates genes that exhibit intraspecific polymorphism, as well as genes that diverged late in the speciation process and had no causative role. For prezygotic barriers, we retain the criteria that the genes contribute to barriers associated with a cladogenic event because prezygotic barriers may be more readily reversible. In an earlier discussion, Rieseberg *et al.* (2004) argued that the term 'isolation genes' would be a more accurate descriptor of genes that contribute to RI than 'speciation genes'. Likewise, Noor and Feder (2006) employ the term 'barrier genes'. However, these terms can refer to processes other than reproductive isolation (e.g. barriers to the absorption of nutrients in the small intestine); hence our use of 'speciation genes' herein, while recognizing that many of these genes are unlikely to be the cause of

TABLE 1. *Genes underlying reproductive barriers in plants, with complementary genes indicated by superscript numbers*

Gene	Normal function	Organism	Level	Barrier phenotype	Likely genetic cause	Reference
Genes that underlie pre-pollination barriers						
ANTHOCYANIN-2 (AN2) – a *myb*-type transcription factor	Regulation of anthocyanin production (floral colour)	*Petunia axillaris/P. integrifolia*	Inter-specific	Flower colour differences leading to pollinator shift	Loss of function mutations in coding sequence	(Quattrocchio *et al.*, 1999; Hoballah *et al.*, 2007)
ROSEA1, *ROSEA2* and *VENOSA* – all are *myb*-type transcription factors	Regulation of anthocyanin production (floral colour)	*Antirrhinum* species	Inter-specific	Flower colour differences leading to shifts in pollinator fauna	?	(Schwinn *et al.*, 2006)
FLAVONOID-3′-HYDROXYLASE (F3′H)	Key enzyme in anthocyanin pathway (floral colour)	*Ipomea* species	Inter-specific	Flower colour differences leading to pollinator shift	*cis*-regulatory mutations that downregulate *F3′H*	(Des Marais and Rausher, 2010)
FLOWERING LOCUS C (FLC) – a MADS-box transcription factor	Represses flowering	Allopolyploid *A. suecica* isolated from parental species, *A. thaliana* and *A. arenosa*	Inter-specific	Delayed flowering of allopolyploid, *A. suecica*	*cis*-regulatory mutations in both *A. thaliana* and *A. arenosa* FLC	(Wang *et al.*, 2006*a*)
STYLE2·1 – encodes a helix-loop-helix (HLH)	Transcription factor	*Solanum lycopersicon/ S. pennellii*	Inter-specific	Recessed style leading to mating system shift outcrossing to selfing	*cis*-regulatory mutation(s)	(Chen *et al.*, 2007)
Genes that underlie post-pollination prezygotic barriers						
S-RNase-SI	Rejection of self pollen	*Nicotiana* species	Inter-specific	Rejection of pollen from other species	?	(Murfett *et al.*, 1996)
Genes that underlie intrinsic postzygotic barriers: hybrid inviability						
[1]*Cf2* – an extracellular leucine-rich repeat receptor-like gene	Resistance to fungal pathogen	*Solanum lycopersicon/ S. pimpinellifolium*	Inter-specific	Hybrid necrosis	Gene copy number variation	(Dixon *et al.*, 1996; Kruger *et al.*, 2002)
[1]*RC3* – encodes an extracellular cysteine protease	Perception of fungal *Avr* proteins	*Solanum lycopersicon/ S. pimpinellifolium*	Inter-specific	Hybrid necrosis	Changes in protein sequence	(Kruger *et al.*, 2002; Rooney, 2005)
DANGEROUS MIX 1 (*DM1*) – an NBS-LRR gene	Disease resistance	*Arabidopsis thaliana*	Intra-specific	Hybrid necrosis	Gene copy number variation	(Bomblies *et al.*, 2007)
HISTIDINOL-PHOSPHATE AMINO-TRANSFERASE (*HPA1* and *HPA2*)	Synthesis of essential amino acid, histidine	*Arabidopsis thaliana*	Intra-specific	Arrest of hybrid seed development	Reciprocal silencing of duplicate genes	(Bikard *et al.*, 2009)
TRANSPARENT TESTA GLABRA2 (*TTG2*) – a WRKY transcription factor	Regulates epidermal cell fate	*Arabidopsis thaliana*	Intra-specific	Lethality of interploidal hybrids	*cis*-regulatory mutations	(Dilkes *et al.*, 2008)
HBD2 – encodes a casein kinase	Root development and hormone sensitivity	*Oryza sativa*	Inter-sub-specific	Hybrid necrosis	Change in protein sequence	(Yamamoto *et al.*, 2010)
HWH1 – encodes a GMC oxidoreductase	Catalyses oxidation–reduction reactions	*Oryza sativa*	Inter-sub-specific	Hybrid necrosis	?	(Jiang *et al.*, 2008)
Genes that underlie intrinsic postzygotic barriers: hybrid sterility						
S5 – encodes an aspartate protease	Disease resistance signalling and cell death	*Oryza sativa*	Inter-sub-specific	Hybrid female sterility (embryo sac sterility)	Changes in protein sequence	(Chen *et al.*, 2008)
[2]*SaM* – a SUMO E3 ligase-like gene	Post-translational modification	*Oryza sativa*	Inter-sub-specific	Hybrid male sterility (pollen abortion)	Substitution in intron-splicing site, leading to truncated protein	(Long *et al.*, 2008)
[2]*SaF* – encodes an F-box protein	Mediation of protein-protein interactions	*Oryza sativa*	Inter-sub-specific	Hybrid male sterility (pollen abortion)	Amino acid substitution	(Long *et al.*, 2008)
mtRPL27 – Nuclear-encoded mitochondrial ribosomal protein L27	Translation of mitochondrial genes	*Oryza sativa/O. glumaepatula*	Inter-specific	Hybrid male sterility (pollen abortion)	Reciprocal silencing of duplicate genes	(Yamagata *et al.*, 2010)

speciation. Also, although we applied strict selection criteria, many genes we have chosen to list have not been examined with a comprehensive battery of functional, organismal and population-level studies and thus should still be considered as candidate speciation genes.

CANDIDATE SPECIATION GENES AND THEIR ATTRIBUTES

For inclusion in Table 1, we required convincing genetic and/or functional validation of candidate speciation genes. As a consequence, many excellent candidates (e.g. Wang *et al.*, 1997; Whittall *et al.*, 2006; Escobar-Restrepo *et al.*, 2007; Case and Willis, 2008; Scalliet *et al.*, 2008) were not included on the list, although several of these are discussed briefly in the text. In some instances, changes in the expression of candidate genes were shown to be the cause of RI, but it was not clear whether the expression shift was controlled by *cis*- versus *trans*-acting factors (e.g. Wang *et al.*, 1997; Whittall *et al.*, 2006). In other cases, barrier phenotypes in interspecific crosses were similar to those of *Arabidopsis* mutants (Escobar-Restrepo *et al.*, 2007) and/or rapid adaptive diversification was observed in proteins that seem likely to contribute to RI (Mayfield *et al.*, 2001).

We also required evidence that the candidate speciation genes potentially contribute to RI between populations. For genes responsible for intra- or interspecific incompatibilities we considered this to be self-evident, although we recognize that incompatibility alleles at some of these genes might be too rare to have a significant effect on gene flow between populations. For pre-pollination barriers, we considered that floral changes associated with observed shifts in the pollinator community were likely to contribute to RI, even if this had not been proven by fieldwork in each instance. On the other hand, we found that genetic changes underlying adaptation to environmental differences were less clearly associated with RI. An example comes from *Arabidopsis halleri*, which is able to colonize heavy-metal-polluted soils due to enhanced expression of *HEAVY METAL ATPASE 4* (*HMA4*) (Hanikenne *et al.*, 2008). However, there is no evidence that *Ah-HMA4* contributes to RI because, although zinc tolerance in *A. halleri* is constitutive, the species is found in both metalliferous and non-metalliferous sites and there is little evidence of population structure associated with soil type (Pauwels *et al.*, 2006).

Genes and mutations that contribute to pre-pollination prezygotic isolation

Our review of the literature identified seven genes that have been shown to underlie pre-pollination postzygotic barriers (Table 1). Five of these genes were responsible for variation in floral pigmentation, one for differences in flowering time and one for a shift in mating system from outcrossing to selfing. Surprisingly, we were unable to find convincing examples of genes contributing to habitat isolation, even though ecogeographical isolation is viewed by many students of plant speciation as the most common and important type of reproductive barrier (Stebbins, 1950; Schemske, 2000; Rieseberg and Willis, 2007; Sobel *et al.*, 2010). However, there are some promising candidates, such as *Ha-CDPK3*, which co-segregates with quantitative trait loci (QTLs) for the uptake of toxic mineral ions and immigrant inviability in a hybrid sunflower and its parental species (Lexer *et al.*, 2004). Likewise, reciprocal transplant experiments have shown that the cytoplasm frequently contributes to habitat isolation (e.g. Wu and Campbell, 2007; Kimball *et al.*, 2008; Sambatti *et al.*, 2008), but the genes involved have not been isolated.

We were also surprised to find only a single case in which interspecific genetic differences in flowering times had been characterized (Chen *et al.*, 2007), and even in this case the main reproductive barrier is a ploidy shift. However, genes have been connected with flowering time differences among accessions of *Arabidopsis thaliana* and among cultivars of many domesticated plant species (e.g. Yano *et al.*, 2000; Caicedo *et al.*, 2004; Yan *et al.*, 2004, 2006; Turner *et al.*, 2005; Balasubramanian *et al.*, 2006; Takahashi *et al.*, 2009; Schwartz *et al.*, 2009; Blackman *et al.*, 2010). Although we did not include these genes on our list, a reasonable argument could be made that they contribute more to reproductive isolation between populations than many of the genes underlying hybrid incompatibilities that are included in Table 1.

Despite the small number of pre-pollination RI genes on our list, several trends have emerged. First, it appears that the same genes and pathways may frequently contribute to pre-pollination RI. Currently, the main evidence for this trend comes from studies of variation in flower colour (Table 1; Appendix 1). Differences in the intensity and patterning of anthocyanin production in flowers appear to result mostly from mutations in the same small subfamily of *MYB*-related transcription factors that regulate the anthocyanin biosynthetic pathway (Quattrocchio *et al.*, 1999; Schwinn *et al.*, 2006). Another frequent change – the transition from bee-pollinated, blue/purple flowers to red, hummingbird-pollinated flowers – appears to have repeatedly involved mutations in a structural gene, *FLAVONOID-3′-HYDROXYLASE* (*F3′H*), that redirect flow down different branches of the anthocyanin pathway (Des Marais and Rausher, 2010). Although these inferences are based on a small number of studies, they are corroborated by broader surveys of spontaneous mutations, natural intraspecific polymorphisms and inferences from partially characterized candidate genes (Rausher, 2008; Streisfeld and Rausher, 2009*a*, *b*). In addition to the anthocyanin biosynthetic pathway, we suspect that genes in the photoperiod, vernalization and scent-biosynthetic pathways will often be associated with speciation as genes from these pathways frequently appear in studies of cultivated and model plants (Dudareva *et al.*, 1996; Nam *et al.*, 1999; Shindo *et al.*, 2005; Jones *et al.*, 2008; Scalliet *et al.*, 2008; Takahashi *et al.*, 2009; Koeduka *et al.*, 2009; Schwartz *et al.*, 2009).

Another interesting trend is that six of the seven pre-pollination speciation genes listed have regulatory functions, whereas only one is a structural gene (Table 1). This ratio may be slightly biased in favour of regulatory genes because repeated involvement in speciation is not accounted for and the structural gene, *F3′H*, may have been repeatedly involved in floral colour changes (Des Marais and Rausher, 2010). Nonetheless, it does imply that genes involved in transcriptional regulation play a disproportionately significant role in the evolution of pre-pollination RI.

Mutations contributing to pre-pollination RI appear to be mainly loss-of-function (LOF) mutations in coding sequence or *cis*-regulatory mutations (Table 1). The phenotypes associated with pre-pollination RI are often fixed within species (or nearly so), possibly suggesting that the causative mutations are widespread as well. This conjecture may not be warranted, however, as LOF mutations in *Pa-ANTHOCYANIN2* (*AN2*), the only pre-pollination RI gene for which intraspecific variation has been examined, arose independently at least five times (Hoballah *et al.*, 2007). These observations may be reconciled if flower colour is a late evolving character during pollinator shifts (Quattrocchio *et al.*, 1999). Likewise, based on studies of these loci, little can be said about the role of selection versus drift in the evolution of pre-pollination speciation genes. Although the traits associated with pre-pollination RI seem likely to have experienced divergent natural selection because of their effects on pollinator behaviour, timing of reproduction or mating system, only for *ROSEA1* is there direct evidence of selection on the underlying genes (Whibley *et al.*, 2006).

Genes that contribute to post-pollination prezygotic barriers

As far as we are aware, the only genes contributing to post-pollination prezygotic barriers that have been functionally validated are *S* RNases, the functional products of the *S*-locus in the Solanaceae, which is responsible for self-incompatibility (SI) within species. *S*-RNase is secreted by stylar tissue and degrades the RNA of incompatible pollen tubes. The probable role of the *S*-locus in the evolution of interspecific or 'unilateral' incompatibility has long been suspected because of asymmetric patterns of pollen rejection in crosses between SI and self-compatible (SC) species in the Solanaceae, Plantaginaceae, Rosaceae and Brassicaceae (Lewis and Crowe, 1958; Hiscock and Dickinson, 1993; Hancock *et al.*, 2003). Unilateral incompatibility has been mapped to the *S*-locus in tomato (Chetelat and De Verna, 1991; Bernacchi and Tanksley, 1997), but only in *Nicotiana* have functional studies been performed to show that that *S* RNase genes contribute to interspecific pollen rejection (Murfett *et al.*, 1996). *S* RNase-SI genes are known to be under negative frequency-dependent selection, which results in extreme allelic polymorphism (Richman and Kohn, 2000).

In crucifers, the *S*-locus contains two tightly linked, highly polymorphic genes, *S*-receptor kinase (SRK) and its ligand *S-locus* cysteine rich protein (SCR), which determine SI in the stigma and pollen, respectively (Stein *et al.*, 1991; Schopfer *et al.*, 1999). Allele-specific binding of these proteins initiates a signalling cascade in the stigma that prevents self pollen from germinating. Transformation of SC *A. thaliana* with SRK and SCR alleles from its SI relative *A. lyrata* results in restoration of SI but has no demonstrated effect on unilateral incompatibility between these species (Nasrallah, 2002).

A promising candidate gene for post-pollination prezygotic RI in plants is *FERONIA* (*FER*), a receptor-like kinase involved in pollen tube reception. *FER* mutants in *A. thaliana* are characterized by a failure to arrest pollen tube growth in ovules. A similar phenotype is observed in interspecific crosses with pollen from *A. lyrata* or *Cardamine flexuosa*, which exhibit accelerated amino acid diversification in the extracellular domain of *FER*. Thus, coding changes in *FER* may contribute to the crossability barriers between these species, but this possibility has not yet been functionally verified (Escobar-Restrepo *et al.*, 2007).

Genes and mutations that contribute to hybrid inviability

Empirical studies in plants and animals have shown that intrinsic postzygotic barriers to reproduction – hybrid inviability and hybrid sterility – frequently evolve through mechanisms consistent with the classic Dobzhansky–Muller (DM) model (Dobzhansky, 1937; Muller, 1942). As adaptive or nearly neutral substitutions accumulate in diverging lineages, substitutions may be fixed in one lineage that are incompatible with substitutions in the other lineage. Therefore, hybrid dysfunction results when these incompatible alleles are brought together in hybrid progeny. Although most often considered as a two-locus model with one substitution at each locus occurring in each diverging lineage, analyses of genes underlying DM incompatibilities in plants (Tables 1 and 2; Appendix 1) have shown that diverse evolutionary paths also yield similar outcomes (Fig. 2).

In a recent review, Bomblies and Weigel (2007*b*) suggested that disease resistance genes might play an important role in the evolution of hybrid inviability. Their arguments were based on the following observations: (1) inter- and intra-specific hybrids often exhibit tumours, as well as hybrid necrosis or weakness; (2) symptoms of hybrid necrosis are similar to necrotic symptoms typically associated with environmental stress and pathogen attack; (3) hybrid necrosis usually results from the interactions of complementary genes, similar to classic DM incompatibilities; and (4) at least one example is known where hybrid necrosis impedes interspecific gene flow in the wild (McNaughton and Harper, 1960). Genetic characterization of hybrid necrosis in crosses between tomato species and between *Arabidopsis* ecotypes (Appendix 1) has revealed that incompatibilities among complementary disease resistance genes are indeed the cause of the necrosis (Kruger *et al.*, 2002; Rooney, 2005; Bomblies *et al.*, 2007). Disease resistance genes have also been implicated in mediating hybrid necrosis in crosses between cultivated rice varieties (Yamamoto *et al.*, 2010) as well as between cultivated lettuce and an evolutionarily distant wild relative (Jeuken *et al.*, 2009). Interestingly, in the tomato and *Arabidopsis* examples, one of the parental lines lacks a member of the interacting gene pair, a circumstance that might have facilitated the evolution of the incompatibility in the first place.

Copy number variation played a more obvious role in the evolution of a hybrid incompatibility involving *HISTIDINOL-PHOSPHATE AMINO-TRANSFERASE* (*HPA*) in *Arabidopsis* (Table 1; Appendix 1). Botanists have long speculated that the reciprocal silencing of duplicate genes might be a frequent cause of intrinsic postyzgotic barriers (Werth and Windham, 1991), but there has been little empirical evidence to support such a mechanism. The reciprocal silencing of HPA duplicates in natural populations of *A. thaliana* and mtRPL27 duplicates in AA genome species of rice represent the first clear examples of this process (Bikard *et al.*, 2009; Yamagata *et al.*, 2010). The transposition of *JYALPH*A, a hybrid sterility gene in *Drosophila*, was

TABLE 2. *Genes involved in cytoplasmic male sterility, with complementary genes indicated by superscript numbers*

Gene	Normal function	Organism	Level	Likely genetic cause	Reference
[1]Radish Ogura *and Kosena cytoplasms* (*ORF138* and *ORS125*)	Disruption of pollen development	*Raphanus sativus*	Intra-specific; Inter-generic	Recombination created chimeric gene	(Bonhomme *et al.*, 1991; Iwabuchi *et al.*, 1999)
[1]*Rfo/RFK1* – a mitochondria-targeting *PPR* gene	Regulation of organelle gene expression	*Raphanus sativus*	Intra-specific; Inter-generic	Amino-acid substitutions in PPR domains	(Koizuka *et al.*, 2000; Brown *et al.*, 2003; Desloire *et al.*, 2003)
Brassica *pol* cytoplasm (*ORF224*)	Disruption of pollen development	*Brassica napus*	Intra-specific	Recombination created chimeric gene	(Singh and Brown, 1991)
Brassica *nap* cytoplasm (*ORF222*)	Disruption of pollen development	*Brassica napus*	Intra-specific	Recombination created chimeric gene	(L'homme *et al.*, 1997)
Brassica *tour* cytoplasm (*ORF263*)	Disruption of pollen development	*Brassica juncea/ B. tournefortii*	Inter-specific	Recombination created chimeric gene	(Landgren *et al.*, 1996)
Brassica *tour* cytoplasm (*ORF193*)	Disruption of pollen development	*Brassica napus/ B. tournefortii*	Inter-specific	Recombination created chimeric gene	(Dieterich *et al.*, 2003)
Moricandia arvensis cytoplasm (*ORF108*)	Disruption of pollen development	*Moricandia arvensis/Brassica napus*	Inter-generic	Recombination created chimeric gene	(Ashutosh *et al.*, 2008)
Sunflower PET1 cytoplasm (*ORF522*)	Disruption of pollen development	*Helianthus annuus/ H. petiolaris*	Inter-specific	Recombination created chimeric gene	(Horn *et al.*, 1991)
[2]Petunia pcf cytoplasm (*ORF402*)	Disruption of pollen development	*Petunia hybrida*	Inter-specific	Recombination created chimeric gene	(Young and Hanson, 1987)
[2]*Rf-PPR592* – a mitochondria-targeting *PPR* gene	Regulation of organelle gene expression	*Petunia hybrida*	Inter-specific	Promoter deletion resulting in lack of expression	(Bentolila *et al.*, 2002)
Maize cytoplasm S (*ORF355/ORF77*)	Disruption of pollen development	*Zea mays*	Intra-specific	Recombination created chimeric gene	(Zabala *et al.*, 1997)
[3]Maize cytoplasm T (*URF13*)	Disruption of pollen development	*Zea mays*	Intra-specific	Recombination created chimeric gene	(Dewey *et al.*, 1987)
[3]*RF2* – encodes an aldehyde dehydrogenase	Oxidation of aldehydes	*Zea mays*	Intra-specific	Amino acid substitution leading to loss of *Aldh* activity	(Cui *et al.*, 1996)
Sorghum A3 cytoplasm (*ORF107*)	Disruption of pollen development	*Sorghum bicolor*	Intra-specific	Recombination created chimeric gene	(Tang *et al.*, 1996)
Sorghum *RF1* – a mitochondria-targeting PPR gene	Regulation of organelle gene expression	*Sorghum bicolor*	Intra-specific	?	(Tang *et al.*, 1996)
Wheat As or Tt cytoplasm (*ORF256*)	Disruption of pollen development	*Triticum aestivum/ T. timopheevi*	Inter-specific	Recombination created chimeric gene	(Hedgcoth *et al.*, 2002)
Common bean CMS (*ORF239*)	Disruption of pollen development	*Phaseolus vulgaris*	Intra-specific	Recombination created chimeric gene	(Abad *et al.*, 1995)
[4]Rice Boro II cytoplasm (*ORF79*)	Chimeric orf	*Oryza sativa*	Inter-sub-specific	Recombination created chimeric gene	(Wang *et al.*, 2006*b*)
[4]Rice *RF1A* – a mitochondria-targeting *PPR* gene	Regulation of organelle gene expression	*Oryza sativa*	Inter-sub-specific	Frameshift mutation in protein	(Wang *et al.*, 2006*b*)
[4]Rice *RF1B* – a mitochondria-targeting *PPR* gene	Regulation of organelle gene expression	*Oryza sativa*	Inter-sub-specific	Nine amino acid substitutions, one of which likely causes loss of restoration function	(Wang *et al.*, 2006*b*)
RETROGRADE-REGULATED MALE STERILITY (*RMS*) – encodes a 178-amino-acid protein of unknown function	unknown	*Oryza rufipogon*	Intra-specific	Upregulation of RMS due to *cis*-regulatory mutation(s)	(Fujii and Toriyama, 2009)

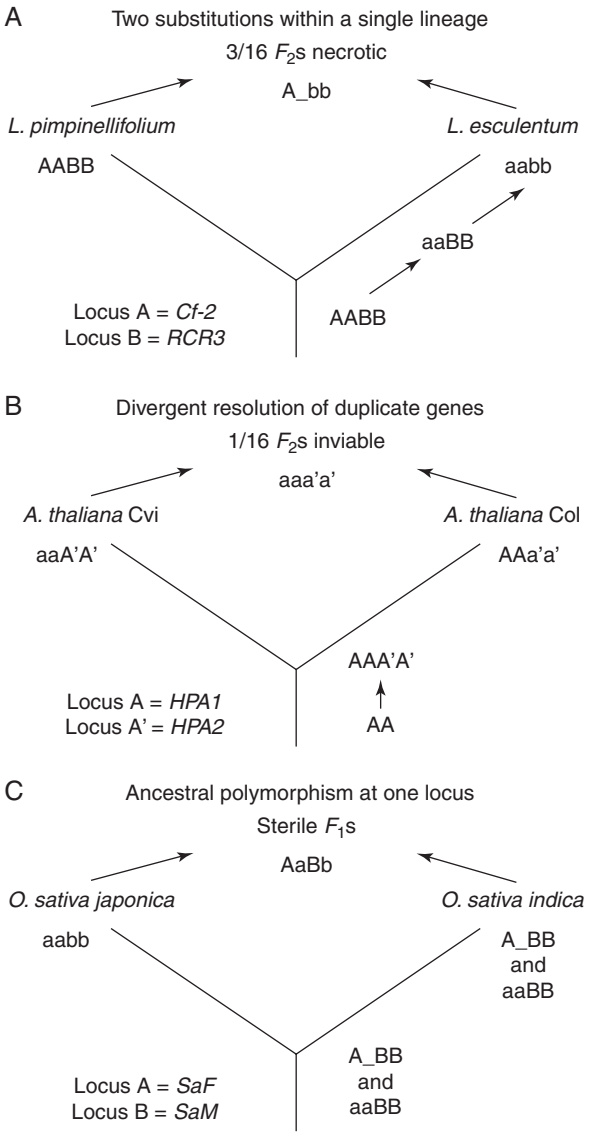

FIG. 2. Diverse evolutionary paths lead to the origin of Dobzhansky–Muller incompatibilities in plants. (A) Two substitutions within a single lineage illustrated by *Cf-2* and *RCR3* loci in *Solanum* species. (B) Divergent resolution of an ancestral gene duplication as illustrated by the *HPA1* and *HPA2* loci in *Arabidopsis thaliana* populations. (C) Divergent resolution of an ancestral polymorphism at one locus and lineage-specific substitution at a second tightly linked locus as illustrated by the *SaF* and *SaM* loci in domesticated *Oryza sativa* subspecies.

probably mediated by gene duplication as well, but this has not been proven (Masly *et al.*, 2006).

Another potential trend is a role for maternally expressed regulatory genes in the development of dosage-sensitive incompatibilities. Several dosage-sensitive incompatibilities involving *cis*-regulatory changes (Dilkes *et al.*, 2008) or epigenetic causes (loss of maternal imprinting) have been identified in interploidal crosses of *Arabidopsis* (Josefsson *et al.*, 2006; Walia *et al.*, 2009; Erilova *et al.*, 2009).

Unlike the genetics of pre-pollination barriers, transcriptional regulators are not over-represented among genes that cause hybrid inviability. Nor is there an excess of LOF or *cis*-regulatory mutations. Alleles contributing to hybrid inviability vary widely in their geographical distributions. For example, *DM1* is restricted to a small fraction of the range of *A. thaliana*, whereas reciprocal divergence at *HWH1* and *HWH2* between the *indica* and *japonica* subspecies of rice is substantially more complete (Jiang *et al.*, 2008). The role of positive selection has not been fully resolved. Disease resistance loci in plants typically contain multiple, related, tightly linked genes that have arisen through evolutionary recent gene duplication events and often are the targets of diversifying selection (Mondragon-Palomino *et al.*, 2002; Kuang *et al.*, 2004). On the other hand, the haphazard distribution of LOF mutations in duplicate *At-HPA* genes is consistent with near neutral processes (Bikard *et al.*, 2009).

Genes and mutations that contribute to hybrid sterility

Cytoplasmic male sterility. By far the largest numbers of sterility loci that have been characterized in plants (or animals) are found in the plant mitochondrial genome and cause cytoplasmic male sterility or CMS (Table 2; Appendix 1). Molecular genetic studies indicate that CMS typically results from rearrangements in the mitochondrial genome, most frequently via the formation and expression of chimeric open reading frames (ORFs; Hanson, 1991; Hanson and Bentolila, 2004; Chase, 2007; Carlsson *et al.*, 2008). In most species, CMS is characterized by the absence of pollen. However, in a few instances (e.g. sunflower, petunia and maize), anthers are missing as well. The chimeric ORFs that cause CMS often include ATP synthase subunit gene promoter regions and/or coding regions or occur near ATP synthase genes (Appendix 1). Less frequently, the chimeric genes include subunits of cytochrome oxidase or NADH dehydrogenase. Essentially all CMS-associated loci also include unique unidentified sequences that show no similarity to chloroplast or nuclear sequences in plants and whose origins remain unknown (Hanson and Bentolila, 2004). Although these genes are included in our overall tally, we have catalogued CMS and restorer loci in a separate table (Table 2) to acknowledge that many of these alleles are specific to crop lines or are rare and thus may have had little impact on speciation.

Mitochondria are usually maternally inherited, so CMS is typically transmitted through ovules (but see McCauley *et al.*, 2007). In contrast, nuclear genes are transmitted through both ovules and pollen. This difference in inheritance pattern creates a genetic conflict between nuclear and cytoplasmic genes. Theoretical studies indicate that CMS will spread in outcrossing populations if the CMS mutation provides even a slight fitness advantage in female function (Fig. 3; Lewis, 1941; Frank, 1989; Hodgins *et al.*, 2009). Because CMS plants are freed from the cost of producing pollen, some CMS mutants have been shown to increase the fitness of females relative to hermaphrodites (Delph *et al.*, 2007). That 'selfish' CMS elements can spread in spite of their effects on sterility casts doubt on the effectiveness of CMS as a reproductive barrier. In addition to genetic conflict, the successful origin of CMS loci represents a kind of diversifying selection, as it appears that novel proteins are required to evade regulation by restorer genes (see below).

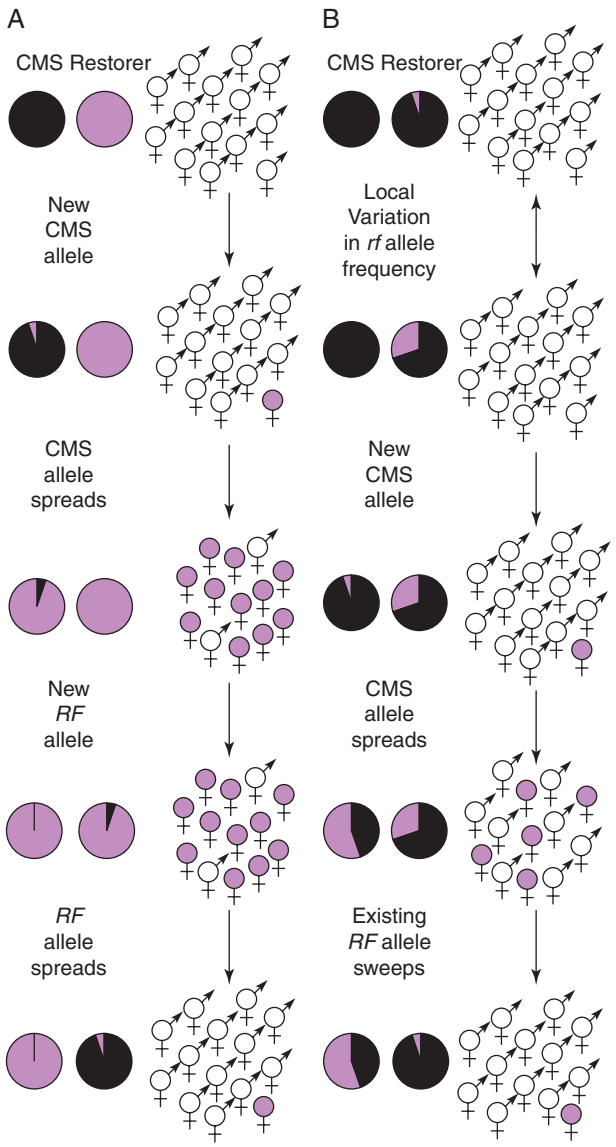

FIG. 3. Two evolutionary scenarios for the spread of CMS and *RF* alleles in hermaphroditic populations. (A) Mitochondrial CMS-causing allele arises first, followed by evolution of a new nuclear restorer allele. (B) Invasion of a CMS allele in a population segregating for *RF* and *rf* alleles. Females and hermaphrodites are represented by standard symbols. Frequencies of the female permissive alleles at the mitochondrial CMS locus and nuclear restorer locus are shown in pink.

Although all of the fully characterized CMS-associated loci are associated with crop plants, the CMS cytoplasms often derive from natural populations of a wild relative. For example, the *PET* CMS cytoplasm in cultivated sunflower (*Helianthus annuus*) was discovered in a naturally occurring population of the prairie sunflower, *Helianthus petiolaris*, from St Louis, USA (Leclercq, 1969). The only well-characterized CMS cytoplasm not associated with a crop is in *Mimulus guttatus*, where mtDNA transcripts containing *NAD6* were perfectly correlated with CMS (Case and Willis, 2008). However, functional studies to confirm which of the ORFs present in these *Mg-NAD6*-containing transcripts cause sterility have not yet been performed.

We are aware of only three studies that have examined the distribution of the molecular variants underlying CMS in natural populations. In hermaphroditic species of *Helianthus* and *Mimulus*, the CMS loci are exceedingly rare, known only from the wild population in which they were originally found (Rieseberg *et al.*, 2004; Case and Willis, 2008). In contrast, in gynodioecious populations of radish (*Raphanus sativus*) the frequency of the CMS locus ranged from 0 to 1 (Murayama *et al.*, 2004). Case and Willis (2008) hypothesize that in the absence of gynodioecy, CMS loci will be rare and unlikely to contribute significantly to reproductive isolation and speciation.

*Nuclear restorer of fertility (*RF*) genes.* CMS-associated loci represent only one member of the pair of DM genes that cause cytoplasmic male sterility. The other member of these DM pairs is the non-restoring allele (*rf*) or homologue of nuclear restorer of fertility genes. *RF* genes restore fertility in CMS plants in several different ways (reviewed in Hanson and Bentolila, 2004). Most frequently, they regulate the transcript profile or protein accumulation of the CMS locus. Less frequently, they act to ameliorate the negative consequences of metabolic effects (Liu *et al.*, 2001) or to reduce the abundance of mitochondrial segments carrying the CMS locus (Mackenzie and Chase, 1990).

We are aware of seven *RF* genes that have been cloned to date (Table 2; Appendix 1). Five of these belong to the pentatricopeptide repeat (*PPR*) family. The *PPR* family is one of the largest gene families in plants, with 441 and 655 *PPR* genes recognized in *Arabidopsis* and rice, respectively. The evolution of *PPR* genes resembles that of disease resistance genes in that they encode proteins with repeat domains, occur in small clusters of closely related genes that have arisen through evolutionarily recent gene duplication and transposition, and frequently exhibit an excess of non-synonymous substitutions indicative of diversifying selection (Geddy and Brown, 2007; Foxe and Wright, 2009). Notably, the nuclear restorer loci involved in CMS in *Mimulus* map to a region containing recent tandem expansions of PPR genes (Barr and Fishman, 2010).

The geographical distribution of *RF* genes and their non-restoring *rf* allele is only known from radish, where the Rf gene frequency varies from 0·41 to 1 (and the non-restoring allele from 0 to 0·59). There is indirect evidence that *rf* alleles are infrequent in hermaphrodites, as fertility is restored in most progeny from wild × CMS crosses. For example, all of the crosses that our group has made between natural populations of sunflower and cultivated lines carrying the *PET* CMS locus have generated fertile progeny (Rieseberg *et al.*, 1995, 1996; Burke *et al.*, 2002, 2004; Lai *et al.*, 2005). These patterns imply either that *rf* alleles occur at low frequency or that there are multiple RF genes capable of restoring a given cytoplasm (e.g. Wang *et al.*, 2006*b*).

The molecular changes responsible for the evolution of the *rf* sterility alleles have been characterized for five of the seven *RF* loci (Table 2; Appendix 1). Three of the mutations result in changes in protein sequence or structure, whereas two occur in *cis*-regulatory regions. Interestingly, most of these changes appear to represent LOF mutations, perhaps implying that the sequence of events responsible for the evolution of CMS

may differ from that typically envisaged (Fig. 3). Both theoretical and empirical studies of CMS typically assume that a mitochondrial rearrangement that causes CMS arises first. The spread of the CMS locus generates strong selection for the evolution of a nuclear restorer allele. This sequence of events may be correct for gynodioecious species and for some cytotypes in hermaphrodites. However, in other instances (perhaps the majority), a newly arisen CMS locus may simply exploit a defect in the regulation of organellar genes that was caused by an LOF mutation in an RF gene. This might also account for the apparently limited geographical distribution of CMS-associated loci and *rf* alleles in hermaphroditic species.

Other hybrid sterility genes. A mystery associated with the genetics of plant speciation has been the frequent discovery of loci that reduce fitness when heterozygous (i.e. underdominant loci). Such loci should be rare because the establishment of alleles with negative fitness effects is unlikely except in small, inbred populations (Hedrick, 1981; Walsh, 1982; Lande, 1984). Some of the earliest examples of underdominant loci were found in crosses between the *indica* and *japonica* subspecies of rice (Oka, 1953). These examples were interpreted by Stebbins (1958) to most likely result from small structural changes involving as few as one to five genes. However, subsequent genetic analyses have largely failed to find evidence of structural differentiation in the vicinity of these loci.

The recent characterization of two underdominant loci in rice has provided two different solutions to this mystery (Chen *et al.*, 2008; Long *et al.*, 2008). At the *S5* locus, a single underdominant gene was found to underlie hybrid female sterility; individuals heterozygous for alternative alleles from the *indica* and *japonica* are sterile. However, a third non-functional allele restores cross-compatibility between the subspecies. The existence of this compatible allele in both subspecies would make it feasible for the sterility-inducing mutations to arise without a loss of fitness. In contrast to *S5*, the *Sa* locus, which underlies male sterility, resolves into two adjacent DM genes. We suspect that other apparent examples of underdominance in plants will have similar explanations.

Although the evolutionary forces responsible for the establishment of the sterility-causing alleles in rice are unclear, the alleles are geographically widespread (Appendix 1). In contrast, of the incompatible alleles at two loci involved in a DM incompatibility between two *Mimulus* species, one allele is geographically widespread in one species whereas the other is extremely geographically restricted in the other species (Sweigart *et al.*, 2007).

DISCUSSION

The identities of candidate speciation genes

An important conclusion from this review is that for many reproductive barriers, the identities of many candidate speciation genes in plants can be predicted in a broad sense from the barrier phenotype. This is not surprising for prezygotic barriers, such as changes in floral colour or flowering time. As expected, the genes underlying these barriers do indeed derive from the genetic pathways or networks known to be associated with these phenotypes. A similar generalization can be made in animals: genes causing prezygotic RI generally belong to the expected genetic pathway (Wheeler *et al.*, 1991; Swanson and Vacquier, 1998; Palumbi, 1999; Fang *et al.*, 2002; Barrett *et al.*, 2009; Wittkopp *et al.*, 2009). A caveat is that many studies of prezygotic RI employed a candidate gene approach, which undoubtedly biases findings toward well-circumscribed genes and pathways.

What is more surprising is that this conclusion also holds for multiple, common intrinsic postzygotic barriers in plants. CMS, for example, is typically caused by interactions between chimeric mitochondrial genes (particularly ATP synthase subunits) and members of the *PPR* gene family. Likewise, hybrid necrosis often results from changes in disease resistance genes. In contrast, few if any predictions can be made in animals, and the identity of each new speciation gene is a surprise (Ting *et al.*, 1998; Froschauer *et al.*, 2001; Barbash *et al.*, 2003; Presgraves *et al.*, 2003; Brideau *et al.*, 2006; Harrison and Burton, 2006; Masly *et al.*, 2006; Lee *et al.*, 2008; Mihola *et al.*, 2009; Phadnis and Orr, 2009; Tang and Presgraves, 2009). Unlike plants, there is little evidence that disease resistance genes contribute to RI in animals. Mitochondrial genes have been implicated in reduced hybrid viability in *Tigriopus californicus* copepods (Harrison and Burton, 2006) and in hybrid sterility in yeast (Lee *et al.*, 2008), but mitochondrial-associated RI appears to be much rarer in animals than in plants. These trends aside, various other types of genes are represented among the few genes with known involvement in other hybrid sterility and inviability phenotypes. Only once additional genes involved in these forms of reproductive isolation are cloned will we know whether there is a broadly predictable relationship between these other barrier phenotypes and the normal functions of the underlying genes as well.

The nature of the genetic changes underlying the evolution of RI

A few general trends about the genetics of RI have emerged from this review. First, regulatory changes (both *cis* and *trans*) dominate the evolution of pre-pollination RI in plants. This contrasts with the genetics of intrinsic postzygotic barriers, where a mix of regulatory changes and changes in protein-coding genes are found. Second, LOF mutations frequently contribute to both prezygotic and postzygotic RI in plants. This trend is most apparent in genes underlying changes in floral colour and in the evolution of CMS, where LOF mutations in *PPR* genes appear to provide an opportunity for the invasion of a CMS-causing mitochondrial locus. A third emerging trend is the importance of copy number variation in the evolution of hybrid sterility and inviability. Both disease resistance genes and *PPR* genes belong to large gene families and high rates of gene turnover in these families appear to contribute to the evolution of RI. Even for low-copy-number genes, there is now evidence that gene duplication followed by reciprocal silencing of the duplicate copies can lead to RI (Bikard *et al.*, 2009; Yamagata *et al.*, 2010).

Few if any of these trends hold in animals based on the candidate speciation genes identified so far. Differences between

plants and animals include the rarity of LOF mutations in animal speciation genes, as well as frequent reports of rapid protein evolution underlying animal RI (e.g. Ting *et al.*, 1998; Presgraves *et al.*, 2003; Tang and Presgraves, 2009). Also, copy number variation appears to be less important in the evolution of animal than plant RI (although see Masly *et al.*, 2006).

The nature of the evolutionary forces underlying the evolution of RI

In contrast to animal speciation genes, which often exhibit the signature of positive selection, there is surprisingly little direct evidence that selection has been responsible for the divergence of candidate speciation genes in plants. The much higher levels of intraspecific polymorphism for RI alleles in plants versus animals (see below) is also consistent with a larger role for stochastic evolutionary forces in the evolution of plant RI. Nonetheless, there is indirect support for a key role for selection in the evolution of many of these genes, including: (1) disease resistance genes and *PPR* genes belong to large gene families and members of these families have been shown to be the targets of diversifying selection; (2) *S* RNase-SI genes are known to be under negative frequency-dependent selection; (3) CMS loci are selfish genetic elements that invade populations because they increase fitness through female function; and (4) studies in natural populations indicate that changes in floral pigmentation, floral scent and flowering time often are under divergent natural selection (Harder and Barrett, 2006).

Taxonomic and geographical distributions of RI-inducing mutations

Another apparent difference between plants and animals is the higher level of intraspecific polymorphism for RI alleles in the former (Rieseberg and Willis, 2007). This is especially true for CMS loci, *rf* alleles and *R* genes, which often have extremely restricted geographical distributions (Table 1). The seeming difference between plants and animals might be a consequence of ascertainment bias, as most of the plant genes in Table 1 were cloned for reasons other than their effects on RI. In contrast, the majority of speciation genes in animals were cloned because of their probable role in speciation, and intraspecific polymorphism may have been viewed as a liability. Alternatively, this is a real difference that stems from the greater importance of balancing selection and/or drift in the evolution of plant RI.

Future directions

We have reviewed what is known about candidate speciation genes in plants and identified possible patterns and trends that should be the subject of more rigorous testing in the future. However, there are notable gaps in our knowledge. In particular, with the possible exception of *Pa-AN2* (Quattrocchio *et al.*, 1999; Hoballah *et al.*, 2007), none of the studies reviewed here provides a truly comprehensive analysis of a speciation gene. Such an analysis should ideally include functional characterization of the gene and mutation(s) that underlie RI, analyses of the phylogenetic and geographical distribution of the RI-inducing alleles, evolutionary analyses to test for selection and field studies that examine the effects of the allelic variants on RI in natural populations. In addition, more studies are needed of speciation genes in non-model and non-crop systems. This would reduce possible biases or artefacts due to artificial selection and/or a selfing mating system. Studies that identify genes underlying plant reproductive barriers for which no genes are currently known (e.g. ecogeographical isolation, mechanical isolation, conspecific pollen precedence, extrinsic postzygotic isolation) are also sorely needed. Lastly, it would be useful to conduct more rigorous evolutionary analyses of the genes listed in Table 1. Although we found few reports of selection on these genes, in most cases the appropriate tests were not conducted. If such data were collected for multiple genes underlying multiple barriers within a single system, then insights into speciation dynamics – the timing and order in which different types of reproductive isolating barriers arise – may be gained as well.

ACKNOWLEDGEMENTS

We thank Kay Hodgins, Nolan Kane, Sally Otto, Dolph Schluter, David Lowry and John Willis for helpful discussions. This work was supported by the Natural Sciences and Engineering Research Council of Canada (DG 327475 to L.H.R.) and a US National Science Foundation Postdoctoral Research Fellowship in Biology (DBI-0905958 to B.K.B.).

LITERATURE CITED

Abad AR, Mehrtens BJ, Mackenzie SA. 1995. Specific expression in reproductive tissues and fate of a mitochondrial sterility-associated protein in cytoplasmic male-sterile bean. *Plant Cell* **7**: 271–285.

Ashutosh, Kumar P, Dinesh Kumar V, Sharma PC, Prakash S, Bhat SR. 2008. A novel *orf108* co-transcribed with the *atpA* gene is associated with cytoplasmic male sterility in *Brassica juncea* carrying *Moricandia arvensis* cytoplasm. *Plant & Cell Physiology* **49**: 284–289.

Balasubramanian S, Sureshkumar S, Agrawal M, *et al.* 2006. The PHYTOCHROME C photoreceptor gene mediates natural variation in flowering and growth responses of *Arabidopsis thaliana*. *Nature Genetics* **38**: 711–715.

Barbash DA, Siino DF, Tarone AM, Roote J. 2003. A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 5302–5307.

Barr CM, Fishman L. 2010. The nuclear component of a cytonuclear hybrid incompatibility in *Mimulus* maps to a cluster of pentatricopeptide repeat genes. *Genetics* **184**: 455–465.

Barrett RDH, Rogers SM, Schluter D. 2009. Environment specific pleiotropy facilitates divergence at the ectodysplasin locus in threespine stickleback *Evolution* **63**: 2831–2837.

Bentolila S, Alfonso AA, Hanson MR. 2002. A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 10887–10892.

Bernacchi D, Tanksley SD. 1997. An interspecific backcross of *Lycopersicon esculentum* × *L. hirsutum*: linkage analysis and a QTL study of sexual compatibility factors and floral traits. *Genetics* **147**: 861–877.

Bikard D, Patel D, Le Mette C, *et al.* 2009. Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science* **323**: 623–626.

Blackman BK, Strasburg JL, Raduski AR, Michaels SD, Rieseberg LH. 2010. The role of recently derived *FT* paralogs in sunflower domestication. *Current Biology* **20**: 629–635.

Bomblies K. 2010. Doomed lovers: mechanisms of isolation and incompatibility in plants. *Annual Review of Plant Biology* **61**: 109–124.

Bomblies K, Weigel D. 2007*a*. *Arabidopsis* – a model genus for speciation. *Current Opinion in Genetics & Development* **17**: 500–504.

Bomblies K, Weigel D. 2007*b*. Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nature Reviews Genetics* **8**: 382–393.

Bomblies K, Lempe J, Epple P, *et al*. 2007. Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *Plos Biology*, **5**: e236. doi:10·1371/journal.pbio.0050236

Bonhomme S, Budar F, Ferault M, Pelletier G. 1991. A 2·5 kb *NcoI* fragment of Ogura radish mitochondrial-DNA is correlated with cytoplasmic male-sterility in *Brassica* cybrids. *Current Genetics* **19**: 121–127.

Brideau NJ, Flores HA, Wang J, Maheshwari S, Wang X, Barbash DA. 2006. Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science* **314**: 1292–1295.

Brown GG, Formanova N, Jin H, *et al*. 2003. The radish *Rfo* restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *Plant Journal* **35**: 262–272.

Burke JM, Tang S, Knapp SJ, Rieseberg LH. 2002. Genetic analysis of sunflower domestication. *Genetics* **161**: 1257–1267.

Burke JM, Lai Z, Salmaso M, Nakazato T, *et al*. 2004. Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*. *Genetics* **167**: 449–457.

Butlin R, Ritchie MG. 2001. Evolutionary biology – Searching for speciation genes. *Nature* **412**: 31–33.

Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD. 2004. Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. *Proceedings of the National Academy of Sciences of the USA* **101**: 15670–15675.

Carlsson J, Leino M, Sohlberg J, Sundstroem JF, Glimelius K. 2008. Mitochondrial regulation of flower development. *Mitochondrion* **8**: 74–86.

Case AL, Willis JH. 2008. Hybrid male sterility in *Mimulus* (Phrymaceae) is associated with a geographically restricted mitochondrial rearrangement. *Evolution* **62**: 1026–1039.

Chase CD. 2007. Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. *Trends in Genetics* **23**: 81–90.

Chen JJ, Ding JH, Ouyang YD, *et al*. 2008. A triallelic system of *S5* is a major regulator of the reproductive barrier and compatibility of *indica-japonica* hybrids in rice. *Proceedings of the National Academy of Sciences of the USA* **105**: 11436–11441.

Chen KY, Cong B, Wing R, Vrebalov J, Tanksley SD. 2007. Changes in regulation of a transcription factor lead to autogamy in cultivated tomatoes. *Science* **318**: 643–645.

Chetelat RT, DeVerna JW. 1991. Expression of unilateral incompatibility in pollen of *Lycopersicon pennellii* is determined by major loci on chromosomes 1, 6 and 10. *Theoretical and Applied Genetics* **82**: 704–712.

Coyne J, Orr H. 2004. *Speciation.* Sunderland, MA: Sinauer Associates.

Cui XQ, Wise RP, Schnable PS. 1996. The *rf2* nuclear restorer gene of male-sterile T-cytoplasm maize. *Science* **272**: 1334–1336.

Delph LF, Touzet P, Bailey MF. 2007. Merging theory and mechanism in studies of gynodioecy. *Trends in Ecology & Evolution* **22**: 17–24.

Desloire S, Gherbi H, Laloui W, *et al*. 2003. Identification of the fertility restoration locus, *Rfo*, in radish, as a member of the pentatricopeptide-repeat protein family. *Embo Reports* **4**: 588–594.

Des Marais DL, Rausher MD. 2010. Parallel evolution at multiple levels in the origin of hummingbird pollinated flowers in *Ipomea*. *Evolution*, in press. doi: 10·1111/j.1558-5646·2010·00972.x

Dewey RE, Timothy DH, Levings CS. 1987. A mitochondrial protein associated with cytoplasmic male-sterility in the T-cytoplasm of maize. *Proceedings of the National Academy of Sciences of the USA* **84**: 5374–5378.

Dieterich JH, Braun HP, Schmitz UK. 2003. Alloplasmic male sterility in *Brassica napus* (CMS 'Tournefortii-Stiewe') is associated with a special gene arrangement around a novel *atp9* gene. *Molecular Genetics and Genomics* **269**: 723–731.

Dilkes BP, Spielman M, Weizbauer R, *et al*. 2008. The maternally expressed WRKY transcription factor *TTG2* controls lethality in interploidy crosses of Arabidopsis. *Plos Biology* **6**: e308. doi:10·1371/journal.pbio.0060308

Dixon MS, Jones DA, Keddie JS, Thomas CM, Harrison K, Jones JDG. 1996. The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell* **84**: 451–459.

Dobzhansky T. 1937. *Genetics and the origin of species.* New York: Columbia University Press.

Dudareva N, Cseke L, Blanc VM, Pichersky E. 1996. Evolution of floral scent in *Clarkia*: novel patterns of S-linalool synthase gene expression in the *C. breweri* flower. *Plant Cell* **8**: 1137–1148.

Erilova A, Brownfield L, Exner V, *et al*. 2009. Imprinting of the polycomb group gene *MEDEA* serves as a ploidy sensor in *Arabidopsis*. *PLoS Genetics* **5**: e1000663. doi:10·1371/journal.pgen.1000663

Escobar-Restrepo JM, Huck N, Kessler S, *et al*. 2007. The *FERONIA* receptor-like kinase mediates male-female interactions during pollen tube reception. *Science* **317**: 656–660.

Fang S, Takahashi A, Wu CI. 2002. A mutation in the promoter of desaturase 2 is correlated with sexual isolation between *Drosophila* behavioral races. *Genetics* **162**: 781–784.

Foxe JP, Wright SI. 2009. Signature of diversifying selection on members of the pentatricopeptide repeat protein family in *Arabidopsis lyrata*. *Genetics* **183**: 663–672.

Frank SA. 1989. The evolutionary dynamics of cytoplasmic male-sterility. *American Naturalist* **133**: 345–376.

Froschauer A, Korting C, Bernhardt W, *et al*. 2001. Genomic plasticity and melanoma formation in the fish *Xiphophorus*. *Marine Biotechnology* **3**: S72–S80.

Fujii S, Toriyama K. 2009. Suppressed expression of *RETROGRADE-REGULATED MALE STERILITY* restores pollen fertility in cytoplasmic male sterile rice plants. *Proceedings of the National Academy of Sciences of the USA* **106**: 9513–9518.

Geddy R, Brown GG. 2007. Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. *BMC Genomics* **8**: 130. doi:10·1186/1471-2164-8-130

Haerty W, Singh RS. 2006. Gene regulation divergence is a major contributor to the evolution of Dobzhansky-Muller incompatibilities between species of *Drosophila*. *Molecular Biology and Evolution* **23**: 1707–1714.

Hancock CN, Kondo K, Beecher B, McClure B. 2003. The *S*-locus and unilateral incompatibility. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **358**: 1133–1140.

Hancock CN, Kent L, McClure BA. 2005. The stylar 120 kDa glycoprotein is required for *S*-specific pollen rejection in *Nicotiana*. *Plant Journal* **43**: 716–723.

Hanikenne M, Talke IN, Haydon MJ, *et al*. 2008. Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of *HMA4*. *Nature* **453**: 391–395.

Hanson MR. 1991. Plant mitochondrial mutations and male-sterility *Annual Review of Genetics* **25**: 461–486.

Hanson MR, Bentolila S. 2004. Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell* **16**: S154–S169.

Harder LD, Barrett SCH. 2006. *Ecology and evolution of flowers.* New York: Oxford University Press.

Harrison JS, Burton RS. 2006. Tracing hybrid incompatibilities to single amino acid substitutions. *Molecular Biology and Evolution* **23**: 559–564.

Hedgcoth C, El-Shehawi AM, Wei P, Clarkson M, Tamalis D. 2002. A chimeric open reading frame associated with cytoplasmic male sterility in alloplasmic wheat with *Triticum timopheevi* mitochondria is present in several *Triticum* and *Aegilops* species, barley, and rye. *Current Genetics* **41**: 357–365.

Hedrick PW. 1981. The establishment of chromosomal variants. *Evolution* **35**: 322–332.

Hiscock SJ, Dickinson HG. 1993. Unilateral incompatibility within the Brassicaceae – further evidence for the involvement of the self-incompatibility (*S*)-locus. *Theoretical and Applied Genetics* **86**: 744–753.

Hoballah ME, Gubitz T, Stuurman J, *et al*. 2007. Single gene-mediated shift in pollinator attraction in *Petunia*. *Plant Cell* **19**: 779–790.

Hodgins KA, Rieseberg L, Otto SP. 2009. Genetic control of invasive plant species using selfish genetic elements. *Evolutionary Applications* **2**: 555–569.

Horn R, Kohler RH, Zetsche K. 1991. A mitochondrial 16-kDa protein is associated with cytoplasmic male-sterility in sunflower. *Plant Molecular Biology* **17**: 29–36.

Iwabuchi M, Koizuka N, Fujimoto H, Sakai T, Imamura J. 1999. Identification and expression of the kosena radish (*Raphanus sativus* cv.

Kosena) homologue of the ogura radish CMS-associated gene, *orf138*. *Plant Molecular Biology* **39**: 183–188.

Jeuken MJW, Zhang NW, McHale LK, *et al*. 2009. *Rin4* Causes hybrid necrosis and race-specific resistance in an interspecific lettuce hybrid. *Plant Cell* **21**: 3368–3378.

Jiang W, Chu S-H, Piao R, *et al*. 2008. Fine mapping and candidate gene analysis of *hwh1* and *hwh2*, a set of complementary genes controlling hybrid breakdown in rice. *TAG Theoretical and Applied Genetics* **116**: 1117–1127.

Jones H, Leigh FJ, Mackay I, *et al*. 2008. Population-based resequencing reveals that the flowering time adaptation of cultivated barley originated east of the fertile crescent. *Molecular Biology and Evolution* **25**: 2211–2219.

Jones KN, Reithel JS. 2001. Pollinator-mediated selection on a flower color polymorphism in experimental populations of *Antirrhinum* (Scrophulariaceae). *American Journal of Botany* **88**: 447–454.

Josefsson C, Dilkes B, Comai L. 2006. Parent-dependent loss of gene silencing during interspecies hybridization. *Current Biology* **16**: 1322–1328.

Kimball S, Campbell DR, Lessin C. 2008. Differential performance of reciprocal hybrids in multiple environments. *Journal of Ecology* **96**: 1306–1318.

Klein RR, Klein PE, Mullet JE, Minx P, Rooney WL, Schertz KF. 2005. Fertility restorer locus *Rf1* of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. *Theoretical and Applied Genetics* **111**: 994–1012.

Koeduka T, Orlova I, Baiga TJ, Noel JP, Dudareva N, Pichersky E. 2009. The lack of floral synthesis and emission of isoeugenol in *Petunia axillaris* subsp *parodii* is due to a mutation in the isoeugenol synthase gene. *Plant Journal* **58**: 961–969.

Koizuka N, Imai R, Iwabuchi M, Sakai T, Imamura J. 2000. Genetic analysis of fertility restoration and accumulation of *ORF125* mitochondrial protein in the kosena radish (*Raphanus sativus* cv. Kosena) and a *Brassica napus* restorer line. *Theoretical and Applied Genetics* **100**: 949–955.

Kruger J, Thomas CM, Golstein C, *et al*. 2002. A tomato cysteine protease required for *Cf-2*-dependent disease resistance and suppression of autonecrosis. *Science* **296**: 744–747.

Kuang H, Woo SS, Meyers BC, Nevo E, Michelmore RW. 2004. Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell* **16**: 2870–2894.

Lai Z, Nakazato T, Salmaso M, *et al*. 2005. Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. *Genetics* **171**: 291–303.

Lande R. 1984. The expected fixation rate of chromosomal inversions. *Evolution* **38**: 743–752.

Landgren M, Zetterstrand M, Sundberg E, Glimelius K. 1996. Alloplasmic male-sterile Brassica lines containing *B. tournefortii* mitochondria express an ORF 3′ of the *atp6* gene and a 32 kDa protein. *Plant Molecular Biology* **32**: 879–890.

Leclercq P. 1969. A case of cytoplasmic male-sterility in sunflower. *Annales De L'Amelioration Des Plantes* **19**: 99–106.

Lee HY, Chou JY, Cheong L, Chang NH, Yang SY, Leu JY. 2008. Incompatibility of Nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* **135**: 1065–1073.

Lewis D. 1941. Male sterility in natural populations of hermaphrodite plants: the equilibrium between females and hermaphrodites to be expected with different types of inheritance. *New Phytologist* **40**: 56–63.

Lewis D, Crowe LK. 1958. Unilateral interspecific incompatibility in flowering plants. *Heredity* **12**: 233–256.

Lexer C, Lai Z, Rieseberg LH. 2004. Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *New Phytologist* **161**: 225–233.

L'homme Y, Stahl RJ, Li XQ, Hameed A, Brown GG. 1997. *Brassica nap* cytoplasmic male sterility is associated with expression of a mtDNA region containing a chimeric gene similar to the *pol* CMS-associated *orf224* gene. *Current Genetics* **31**: 325–335.

Liu F, Cui XQ, Horner HT, Weiner H, Schnable PS. 2001. Mitochondrial aldehyde dehydrogenase activity is required for male fertility in maize. *Plant Cell* **13**: 1063–1078.

Long YM, Zhao LF, Niu BX, *et al*. 2008. Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 18871–18876.

Mackenzie SA, Chase CD. 1990. Fertility restoration is associated with loss of a portion of the mitochondrial genome in cytoplasmic male-sterile common bean. *Plant Cell* **2**: 905–912.

Mallet J. 2006. What does *Drosophila* genetics tell us about speciation? *Trends in Ecology & Evolution* **21**: 386–393.

Masly JP, Jones CD, Noor MAF, Locke J, Orr HA. 2006. Gene transposition as a cause of hybrid sterility in *Drosophila*. *Science* **313**: 1448–1450.

Mayfield JA, Fiebig A, Johnstone SE, Preuss D. 2001. Gene families from the *Arabidopsis thaliana* pollen coat proteome. *Science* **292**: 2482–2485.

McCauley DE, Sundby AK, Bailey MF, Welch ME. 2007. Inheritance of chloroplast DNA is not strictly maternal in *Silene vulgaris* (Caryophyllaceae): evidence from experimental crosses and natural populations. *American Journal of Botany* **94**: 1333–1337.

McClure BA, Haring V, Ebert PR, *et al*. 1989. Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases. *Nature* **342**: 955–957.

McNaughton IH, Harper JL. 1960. The comparative biology of closely related species living in the same area. *New Phytologist* **59**: 27–41.

Michaels SD, Amasino RM. 1999. *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**: 949–956.

Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J. 2009. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* **323**: 373–375.

Mondragon-Palomino M, Meyers BC, Michelmore RW, Gaut BS. 2002. Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Research* **12**: 1305–1315.

Moyle LC. 2008. Genetic underpinnings of postzygotic reproductive barriers among plants. *New Phytologist* **179**: 572–574.

Muller HJ. 1942. Isolating mechanisms, evolution, and temperature. *Biological Symposia* **6**: 71–125.

Murayama K, Yahara T, Terachi T. 2004. Variation of female frequency and cytoplasmic male-sterility gene frequency among natural gynodioecious populations of wild radish (*Raphanus sativus* L.). *Molecular Ecology* **13**: 2459–2464.

Murfett J, Strabala TJ, Zurek DM, Mou BQ, Beecher B, McClure BA. 1996. S RNase and interspecific pollen rejection in the genus *Nicotiana*: Multiple pollen-rejection pathways contribute to unilateral incompatibility between self-incompatible and self-compatible species. *Plant Cell* **8**: 943–958.

Nam KH, Dudareva N, Pichersky E. 1999. Characterization of benzylalcohol acetyltransferases in scented and non-scented *Clarkia* species. *Plant Cell Physiology* **40**: 916–923.

Nasrallah JB. 2002. Recognition and rejection of self in plant reproduction. *Science* **296**: 305–308.

Nivison HT, Hanson MR. 1989. Identification of a mitochondrial protein associated with cytoplasmic male-sterility in *Petunia*. *Plant Cell* **1**: 1121–1130.

Noor MAF. 2003. Evolutionary biology – Genes to make new species. *Nature* **423**: 699–700.

Noor MAF, Feder JL. 2006. Speciation genetics: evolving approaches. *Nature Reviews Genetics* **7**: 851–861.

Oka H. 1953. The mechanism of sterility in the intervarietal hybrid: phylogenetic differentiation of the cultivated rice plant. VI. *Japanese Journal of Breeding* **2**: 217–224.

Orr HA. 2005. The genetic basis of reproductive isolation: insights from *Drosophila*. *Proceedings of the National Academy of Sciences of the USA* **102**: 6522–6526.

Orr HA, Presgraves DC. 2000. Speciation by postzygotic isolation: forces, genes and molecules. *Bioessays* **22**: 1085–1094.

Orr HA, Masly JP, Presgraves DC. 2004. Speciation genes. *Current Opinion in Genetics & Development* **14**: 675–679.

Orr HA, Masly JP, Phadnis N. 2007. Speciation in *Drosophila*: from phenotypes to molecules. *Journal of Heredity* **98**: 103–110.

Palumbi SR. 1999. All males are not created equal: fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 12632–12637.

Pauwels M, Frerot H, Bonnin I, Saumitou-Laprade P. 2006. A broad-scale analysis of population differentiation for Zn tolerance in an emerging

model species for tolerance study: *Arabidopsis halleri* (Brassicaceae). *Journal of Evolutionary Biology* **19**: 1838–1850.

Phadnis N, Orr HA. 2009. A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science* **323**: 376–379.

Presgraves DC. 2010. The molecular evolutionary basis of species formation. *Nature Reviews Genetics* **11**: 175–180.

Presgraves DC, Balagopalan L, Abmayr SM, Orr HA. 2003. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature* **423**: 715–719.

Quattrocchio F, Wing J, van der Woude K, *et al*. 1999. Molecular analysis of the *Anthocyanin2* gene of *Petunia* and its role in the evolution of flower color. *Plant Cell* **11**: 1433–1444.

Rausher MD. 2008. Evolutionary transitions in floral color. *International Journal of Plant Sciences* **169**: 7–21.

Richman AD, Kohn JR. 2000. Evolutionary genetics of self-incompatibility in the Solanaceae. *Plant Molecular Biology* **42**: 169–179.

Rieseberg LH, Willis JH. 2007. Plant speciation. *Science* **317**: 910–914.

Rieseberg LH, Vanfossen C, Desrochers AM. 1995. Hybrid speciation accompanied by genomic reorganization in wild sunflowers *Nature* **375**: 313–316.

Rieseberg LH, Sinervo B, Linder CR, Ungerer MC, Arias DM. 1996. Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science* **272**: 741–745.

Rieseberg LH, Church SA, Morjan CL. 2004. Integration of populations and differentiation of species. *New Phytologist* **161**: 59–69.

Rooney HCE. 2005. Cladosporium *Avr2* inhibits tomato *Rcr3* protease required for *Cf-2*-dependent disease resistance (vol 308, pg 1783, 2005). *Science* **310**: 54–54.

Sambatti JBM, Ortiz-Barrientos D, Baack EJ, Rieseberg LH. 2008. Ecological selection maintains cytonuclear incompatibilities in hybridizing sunflowers. *Ecology Letters* **11**: 1082–1091.

Scalliet G, Piola F, Douady CJ, *et al*. 2008. Scent evolution in Chinese roses. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 5927–5932.

Schemske DW. 2000. Understanding the origin of species. *Evolution* **54**: 1069–1073.

Schopfer CR, Nasrallah ME, Nasrallah JB. 1999. The male determinant of self-incompatibility in *Brassica*. *Science* **286**: 1697–1700.

Schwartz C, Balasubramanian S, Warthmann N, *et al*. 2009. *Cis*-regulatory changes at *FLOWERING LOCUS T* mediate natural variation in flowering responses of *Arabidopsis thaliana*. *Genetics* **183**: 723–732.

Schwinn K, Venail J, Shang YJ, *et al*. 2006. A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *Plant Cell* **18**: 831–851.

Shindo C, Aranzana MJ, Lister C, *et al*. 2005. Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of *Arabidopsis*. *Plant Physiology* **138**: 1163–1173.

Singh M, Brown GG. 1991. Suppression of cytoplasmic male-sterility by nuclear genes alters expression of a novel mitochondrial gene region. *Plant Cell* **3**: 1349–1362.

Sobel JM, Chen GF, Watt LR, Schemske DW. 2010. The biology of speciation. *Evolution* **64**: 295–315.

Stebbins GL. 1950. *Variation and evolution in plants.* New York: Columbia University Press.

Stebbins GL. 1958. The inviability, weakness, and sterility of insterspecific hybrids. *Advances in Genetics Incorporating Molecular Genetic Medicine* **9**: 147–215.

Stein JC, Howlett B, Boyes DC, Nasrallah ME, Nasrallah JB. 1991. Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. *Proceedings of the National Academy of Sciences of the USA* **88**: 8816–8820.

Streisfeld MA, Rausher MD. 2009*a*. Altered *trans*-regulatory control of gene expression in multiple anthocyanin genes contributes to adaptive flower color evolution in *Mimulus aurantiacus*. *Molecular Biology and Evolution* **26**: 433–444.

Streisfeld MA, Rausher MD. 2009*b*. Genetic changes contributing to the parallel evolution of red floral pigmentation among *Ipomoea* species. *New Phytologist* **183**: 751–763.

Swanson WJ, Vacquier VD. 1998. Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. *Science* **281**: 710–712.

Sweigart AL, Mason AR, Willis JH. 2007. Natural variation for a hybrid incompatibility between two species of *Mimulus*. *Evolution* **61**: 141–151.

Takahashi Y, Teshima KM, Yokoi S, Innan H, Shimamoto K. 2009. Variations in *Hd1* proteins, *Hd3a* promoters, and *Ehd1* expression levels contribute to diversity of flowering time in cultivated rice. *Proceedings of the National Academy of Sciences of the USA* **106**: 4555–4560.

Tang HV, Pring DR, Shaw LC, *et al*. 1996. Transcript processing internal to a mitochondrial open reading frame is correlated with fertility restoration in male-sterile sorghum. *Plant Journal* **10**: 123–133.

Tang SW, Presgraves DC. 2009. Evolution of the *Drosophila* nuclear pore complex results in multiple hybrid incompatibilities. *Science* **323**: 779–782.

Ting CT, Tsaur SC, Wu ML, Wu CI. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282**: 1501–1504.

Turner A, Beales J, Faure S, Dunford RP, Laurie DA. 2005. The pseudo-response regulator *Ppd-H1* provides adaptation to photoperiod in barley. *Science* **310**: 1031–1034.

Uyttewaal M, Arnal N, Quadrado M, *et al*. 2008. Characterization of *Raphanus sativus* pentatricopeptide repeat proteins encoded by the fertility restorer locus for Ogura cytoplasmic male sterility. *Plant Cell* **20**: 3331–3345.

Walia H, Josefsson C, Dilkes B, Kirkbride R, Harada J, Comai L. 2009. Dosage-dependent deregulation of an *AGAMOUS-LIKE* gene cluster contributes to interspecific incompatibility. *Current Biology* **19**: 1128–1132.

Walsh JB. 1982. Rate of accumulation of reproductive isolation by chromosome rearrangements. *American Naturalist* **120**: 510–532.

Wang JH, Dudareva N, Bhakta S, Raguso RA, Pichersky E. 1997. Floral scent production in *Clarkia breweri* (Onagraceae). 2. Localization and developmental modulation of the enzyme S-adenosyl-L-methionine: (iso)eugenol O-methyltransferase and phenylpropanoid emission. *Plant Physiology* **114**: 213–221.

Wang J, Tian L, Lee H-S, Chen ZJ. 2006*a*. Nonadditive regulation of *FRI* and *FLC* loci mediates flowering-time variation in *Arabidopsis* allopolyploids. *Genetics* **173**: 965–974.

Wang ZH, Zou YJ, Li XY, *et al*. 2006*b*. Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. *Plant Cell* **18**: 676–687.

Werth CR, Windham MD. 1991. A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression *American Naturalist* **137**: 515–526.

Wheeler DA, Kyriacou CP, Greenacre ML, *et al*. 1991. Molecular transfer of a species-specific behaviour from *Drosophila simulans* to *Drosophila melanogaster*. *Science* **251**: 1082–1085.

Whibley AC, Langlade NB, Andalo C, *et al*. 2006. Evolutionary paths underlying flower color variation in *Antirrhinum*. *Science* **313**: 963–966.

Whittall JB, Voelckel C, Kliebenstein DJ, Hodges SA. 2006. Convergence, constraint and the role of gene expression during adaptive radiation: floral anthocyanins in *Aquilegia*. *Molecular Ecology* **15**: 4645–4657.

Widmer A, Lexer C, Cozzolino S. 2008. Evolution of reproductive isolation in plants. *Heredity* **102**: 31–38.

Wittkopp PJ, Stewart EE, Arnold LL, *et al*. 2009. Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in *Drosophila*. *Science* **326**: 540–544.

Wolfe LM, Sowell DR. 2006. Do pollination syndromes partition the pollinator community? A test using four sympatric morning glory species. *International Journal of Plant Sciences* **167**: 1169–1175.

Wu CA, Campbell DR. 2007. Leaf physiology reflects environmental differences and cytoplasmic background in *Ipomopsis* (Polemoniaceae) hybrids. *American Journal of Botany* **94**: 1804–1812.

Wu CI, Ting CT. 2004. Genes and speciation. *Nature Reviews Genetics* **5**: 114–122.

Yamagata Y, Yamamoto E, Aya K, *et al*. 2010. Mitochondrial gene in the nuclear genome induces reproductive barrier in rice. *Proceedings of the National Academy of Sciences of the USA* **107**: 1494–1499.

Yamamoto E, Takahashi T, Morinaka Y, *et al*. 2010. Gain of deleterious function causes an autoimmune response and Bateson-Dobzhansky-Muller incompatibility in rice. *Molecular Genetics and Genomics* **283**: 305–315.

Yan L, Loukoianov A, Blechl A, *et al*. 2004. The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science* **303**: 1640–1644.

Yan L, Fu D, Li C, *et al*. 2006. The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. *Proceedings of the National Academy of Sciences of the USA* **103**: 19581–19586.

Yano M, Katayose Y, Ashikari M, *et al*. 2000. *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the arabidopsis flowering time gene *CONSTANS*. *Plant Cell* **12**: 2473–2483.

Young EG, Hanson MR. 1987. A fused mitochondrial gene associated with cytoplasmic male-sterility is developmentally regulated. *Cell* **50**: 41–49.

Zabala G, GabayLaughnan S, Laughnan JR. 1997. The nuclear gene *Rf3* affects the expression of the mitochondrial chimeric sequence R implicated in S-type male sterility in maize. *Genetics* **147**: 847–860.

APPENDIX 1

Descriptions of genes that appear to contribute to reproductive isolation in plants

Genes that contribute to pre-pollination barriers.

A *ANTHOCYANIN2 (AN2).* Loss of function (LOF) mutations in *Pa-AN2*, a *myb*-type transcription factor, are responsible for flower colour differences between the white, hawk-moth-pollinated flowers of *Petunia axillaris*, and the violet-reddish flowers of *P. integrifolia* (Quattrocchio *et al*., 1999). Although all accessions of *P. axillaris* tested carry LOF mutations in *Pa-AN2*, the LOF mutations arose independently at least five times and show little evidence of having been favoured by positive selection (Hoballah *et al*., 2007). This observation is puzzling because pollinator preference experiments using genetic introgressions and transgenics show that the *Pa-AN2* mutations have a major effect on pollinator attraction.

B *ROSEA1*, *ROSEA2* and *VENOSA*. Wild snapdragon (*Antirrhinum*) species differ in flower colour and intensity – traits that appear to be associated with pollinator attraction and possibly speciation in the group (Jones and Reithel, 2001; Whibley *et al*., 2006). Three genes underlying differences between at least six different wild species have been shown to be allelic to loci identified from mutations within the garden snapdragon, *A. majus* (Schwinn *et al*., 2006). All are *Myb* transcription factors. *ROS1* and *ROS2* are tightly linked and are responsible for differences in pigment intensity among *Antirrhinum* species, whereas *VENOSA* underlies variation in pigmentation patterning. *ROS1* was recently shown to be under divergent natural selection in a hybrid zone between *Antirrhinum striatum* and *A. pseudomajus* (Whibley *et al*., 2006).

C *FLAVONOID-3′-HYDROXYLASE* (*F3′H*). Like snapdragon, transitions in flower colour among morning glory (*Ipomea*) species often appear to be associated with shifts in the pollinator fauna and speciation (Wolfe and Sowell, 2006). Independent transitions from primarily bee-pollinated, blue/purple flowers to red, hummingbird-pollinated flowers in the *Mina* lineage (which includes *I. quamoclit*) and *I. horsfalliae* result from down-regulation in the same gene, *Ipo-F3′H* (Streisfeld and Rausher, 2009b; Des Marais and Rausher, 2010). In the former case, *cis*-regulatory mutations in *Ipo-F3′H* have been demonstrated to be at least partly responsible, but involvement of a *trans*-acting factor has not been ruled out in *I. horsfalliae*.

D *FLOWERING LOCUS C (FLC).* Natural and synthetic allotetraploids of *Arabidopsis thaliana* × *A. arenosa* flower later than either parental species due to up-regulation of *At-FLC*, which is trans-activated by *Aa-FRI* (Wang *et al*., 2006*a*). *FLC* is a MADS-box transcription factor that represses flowering, whereas FRI is a coiled-coil nuclear protein that positively regulates *FLC* expression (Michaels and Amasino, 1999). The non-additive effects on flowering in synthetic allopolyploid hybrids are due to divergence of both *FRI* and *FLC* during the 60Myr divergence of the parental species. While cis-regulatory mutations in *At-FLC* and *Aa-FLC* derived FLC copies appear to be involved in the flowering time divergence of the natural allopolyploid *A. suecica*, sequence evolution of *As-FRI* copies is unexamined.

E *STYLE2·1.* The cultivated tomato, *Solanum lycopersicon*, is partially reproductively isolated from related species due to a recessed style that results in predominantly self-pollination. Style length is controlled by a short 99-amino-acid protein containing a helix-loop-helix (*HLH*) motif (Chen *et al*., 2007). HLH proteins are thought to act as transcription factors. Functional studies indicate that reduced style length is caused by a mutation in the *Sl-STYLE2·1* promoter, which downregulates expression during floral development.

Genes that contribute to post-pollination prezygotic barriers.

A S-RNase-SI. In plant groups that exhibit self-incompatibility (SI), the *S* locus sometimes also causes interspecific or 'unilateral' incompatibility. The situation in *Nicotiana* is typical: species with SI reject pollen from other species, whereas species that are self-compatible (SC) will accept pollen from both SI and SC species. Genetic studies have shown that unilateral incompatibility maps to the *S* locus and that different *S*-alleles can sometimes have different impacts on interspecific compatibility (Hancock *et al*., 2003). As in other plants from the Solanaceae, the functional products of the *S* locus are S-RNAses (McClure *et al*., 1989). To verify that *S*-RNAse genes were responsible for interspecific incompatibility, Murfett *et al*. (1996) transformed three SC species (*N. glutinosa*, *N. tabacum* and *N. plumbaginifolia*) with *S-RNAse-SI* genes from an SI species, *N. alata*. *Na-S-RNase-SI* contributed to rejection of pollen from all three species, but via different mechanisms. *Na-S-RNase-SI* alone was sufficient to reject *N. glutinosa* and *N. tabacum* pollen, but *N. alata* also rejected pollen from these species via an *S-RNase*-independent mechanism. Rejection of *N. plumbaginifolia* pollen required both *S-RNase-SI* and other genetic factors from *N. alata*, including *HT-B* [a pistil-expressed, extracellular, small asparagine/aspartate (N/D)-rich protein] from *N. alata* (Hancock *et al*., 2005). *Np-HT-B* (a non-S factor) does not appear to be expressed in styles of *N. plumbaginifolia*, but the genetic factor responsible for reduced expression is unknown.

Genes that contribute to intrinsic postzygotic barriers: hybrid inviability.

A *Cf2* and *RCR3*. Domesticated lines of tomato (*Solanum lycopersicon*) that lack the fungal resistance gene, *Cf2*, exhibit weakness or necrosis when two tightly linked and essentially identical copies of the gene are introgressed from a wild species, *S. pimpinellifolium* (Dixon *et al*., 1996;

Kruger *et al.*, 2002). However, necrosis is not observed when a second gene, *RCR3* (an extracellular cysteine protease), from *S. pimpinellifolium* is also introduced. *Sl-Cf2* encodes an extracellular leucine-rich repeat receptor-like protein that activates a hypersensitive response when in the presence of AVR2 protein from the fungal pathogen, *Cladosporium fulvum*. Cf-AVR2 binds and inhibits Sl-RCR3, apparently leading to a conformation change in the latter that triggers the Sl-Cf2-mediated hypersensitive response (Rooney, 2005). The RCR3 protein from *S. lycopersicon* differs from Sp-RCR3 by one amino acid deletion and six amino acid substitutions. These changes are speculated to mimic the conformation imposed on Sl-RCR3 by Cf-AVR2 binding, thereby stimulating a hypersensitive response by Sl-Cf2 protein even in the absence of infection (Rooney, 2005).

B *DANGEROUS MIX* (*DM1* and *DM2*?). Hybrids between different wild accessions of *Arabidopsis thaliana* sometimes exhibit necrosis due to deleterious epistatic interactions that induce autoimmune responses. Several of the genes underlying these interactions have been cloned and characterized (D. Weigel, Max Planck Institute for Developmental Biology, Germany and K. Bomblies, Harvard University, USA, pers. comm.). In the one published study (Bomblies *et al.*, 2007), mapping and functional analyses show that an allele of an NB-LRR gene family (the largest family of resistance genes in plants) member triggers necrosis when combined with an allele from a second locus (*DM2*) from another accession. As in the *Sl-CF2* example (above), a functional copy of *DM1* is missing in the innocuous accession. *DM2* has been mapped to an approx. 148-kb region that includes two NB-LRR genes, which can be viewed as strong candidates. Interestingly, the expression of this incompatibility is temperature sensitive, with necrosis observed in F_1 plants at 16 °C but not at 23 °C. However, F_2 plants that are partially or doubly homozygous at *DM1* and *DM2* suffer from necrosis, suggesting that increased dosage of incompatible alleles can overcome the ameliorating effects of temperature. A species-wide geographical survey revealed that *DM1* was locally restricted to a region near Umkirch, Germany.

C *HISTIDINOL-PHOSPHATE AMINO-TRANSFERASE* (*HPA1* and *HPA2*). *HPA* encodes a key enzyme in the synthesis of histidine, an essential amino acid. Columbia-0 (Col) and Cape Verde Island (Cvi) accessions of *Arabidopsis thaliana* contain different functional copies of the essential *HPA* gene (Bikard *et al.*, 2009). Seed development is arrested in progeny that are homozygous for silenced copies at both genes. A survey of the 30 other accessions indicates that silencing of one or the other gene copy occurred in at least six different ways, including deletions, early stop codons and/or loss of expression. Silencing is widespread so that approx. 25 % of crosses in *A. thaliana* are likely to exhibit *At-HPA* incompatibility.

D *TRANSPARENT TESTA GLABRA2* (*TTG2*). Polyploid species are frequently reproductively isolated from their diploid progenitors due to dosage-sensitive incompatibilities in progeny from inter-ploidal crosses. Landsberg erecta (Ler) and Columbia-0 (Col) accessions of *Arabidopsis* vary in their tolerance to inter-ploidal matings. Genetic analyses of the progeny from Ler × Col recombinant inbred lines crossed with Col indentified a major QTL for interploidal lethality (Dilkes *et al.*, 2008). Fine-mapping and mutant analyses revealed that a maternally expressed *WRKY* transcription factor, *At-TTG2*, was mainly responsible for the QTL effects. Sequence and expression comparisons of the Ler and Col alleles at *At-TTG2* further imply that *cis*-regulated differences in expression levels rather than changes in coding sequence are the cause of variability in inter-ploidal hybrid inviability.

E *HBD2* and *HBD3*? Fine mapping studies in near isogenic lines derived from a cross between the *indica* variety, 'Habataki', and the *japonica* variety, 'Koshihikari', showed that hybrid necrosis was caused by an interaction between two unlinked loci (Yamamoto, 2010). The first, *hbd2*, is a 17-kb region on Lg2 that contains a single predicted gene encoding a casein kinase 1 (*CKI1*) homologue. Although the two varieties did not differ in HDB2 expression level, the coding sequences differed by a single charge-changing amino acid substitution, and overexpression of the 'Habataki' allele in the 'Koshihikari' background causes necrosis. The second locus, *hbd3*, maps to a region in Lg11 that contains an NBS-LRR gene cluster and has a great deal of sequence and structural variation. Although the causal mutations have not been identified, immune response genes are upregulated in the *hbd2/hbd3* double mutant and HBD2 overexpression lines, consistent with a role for disease response genes. The causal HBD2 mutation is limited to only a few varieties of *O. sativa indica* derived from the Indonesian landrace 'Peta'.

F *HWH1* and *HWH2*? Segregation analyses of recombinant inbred lines of the *indica* and *japonica* subspecies of rice (*Oryza sativa*) showed that hybrid inviability was caused by recessive alleles at two unlinked loci, *HWH1* and *HWH2* (Jiang *et al.*, 2008). *HWH 1* mapped to an interval of 11·8 kb, which contains a single predicted gene, a putative glucose-methanol-choline (*GMC*) oxidoreductase family protein. Unfortunately, no functional studies have been performed, so it is not clear whether changes in coding versus regulatory regions are responsible for the hybrid incompatibility phenotype. The identity of *HWH2* is less clear. Mapping studies place it in a 117-kb region on chromosome 11, which contains 12 predicted genes. One of these, a putative hexose transporter, was put forward by the authors as the best candidate, but no functional data are provided to support this claim. Cultivar surveys indicate that the *HWH1* haplotype is fairly common in *indica*, whereas the *HWH2* haplotype is almost fixed in *japonica*, so this incompatibility should be frequently observed in inter-subspecific hybrids.

Genes that contribute to intrinsic postzygotic barriers: hybrid sterility.

A *S5*. Hybrids between *indica* and *japonica* exhibit a reduction in female (embryo sac) fertility. Map-based cloning and transgenic complementation indicated that sterility is caused by interactions between alleles of an aspartate protease gene (Chen *et al.*, 2008). Thus, it appears to be a rare example of true underdominance. Aspartate proteases are a large family of proteolytic enzymes that contribute to disease resistance signalling and cell death in reproductive tissues in *Arabidopsis*. Other than its effect on embryo sac fertility, the function of the *S5* locus in rice is unknown. Interestingly, a third non-functional allele at this locus restores compatibility

between the two subspecies, which might provide a means by which an underdominant mutation could become established. The functional alleles of *S5* in *indica* and *japonica* differ by two amino acid substitutions in the central domain of the protein, but it is unclear how these might contribute to hybrid sterility.

B *SaM* and *SaF*. The *Sa* locus causes pollen to abort in hybrids between the *indica* and *japonica* varieties of rice. A combination of fine-mapping studies and functional analyses by transformation indicate that the *Sa* locus comprises two adjacent genes, *SaM*, which encodes a small ubiquitin-like modifier (*SUMO*) E3 ligase-like protein, and *SaF*, which encodes an F-box protein (Long *et al.*, 2008). Pollen abortion requires the presence of three alleles: both the *indica* and *japonica* alleles at *SaM* (*SaM* + and *SaM*−, respectively), as well as the *indica* allele at *SaF* (*SaF* +). Abortion occurs in pollen grains carrying *SaM*− alleles, but not those carrying *SaM* + . The *SaM*− allele encodes a truncated protein due to a substitution in an intron-splicing site, resulting in negative interactions with *SaF* + . In contrast, the SaM+ protein contains a self-inhibitory domain that blocks interactions with both *SaF* alleles, thereby preventing sterility of *SaM* + pollen. The *SaF*− and *SaF* + alleles differ by a single amino acid substitution that does not alter physical interactions with *SaM*−, but apparently affects male sterility in another way. SUMO proteins are involved in the post-translational modification of other proteins, whereas F-box proteins contain a motif that mediates protein−protein interactions and are frequently involved in signal transduction and cell cycle regulation.

C Nuclear-encoded mitochondrial ribosomal protein L27 (*mtRPL27*). A chromosomal segment containing *mtRPL27* appears to have been duplicated prior to the evolution of the AA genome species of rice (Yamagata *et al.*, 2010). Subsequent to duplication, one copy (on chromosome 8) has been lost in a wild rice species, *Oryza glumaepatula*. The other copy (on chromosome 4) has lost functionality in domesticated rice, *O. sativa*, apparently due to a loss of promoter activity. Functional mtRPL27 protein appears to be required for pollen development. As a consequence, first-generation hybrids between *O. glumaepatula* and *O. sativa* exhibit complete pollen sterility. The loss of the chromosome 8 copy of *mtRPL27* has occurred in all populations of *O. glumaepatula* and in some populations of *O. barthii* and *O. longistaminata* (Yamagata *et al.*, 2010). The taxonomic and geographical distribution of LOF mutations in the chromosome 4 copy of *mtRPL27* was not reported.

D Chimeric ORFs in mitochondrial DNA. Rearrangements in plant mitochondrial genomes frequently lead to the formation of chimeric genes, which are the cause of CMS. CMS loci typically include unique, unidentified sequences, as well as portions of one or several mitochondrial genes (see below). Most commonly, plants carrying the CMS locus fail to produce pollen, but sometimes they fail to produce anthers as well. We list 13 well-characterized CMS genes and the mitochondrial gene fragments they are associated with, but our list is by no means exhaustive. References can be found in Table 2.

(1) Radish Ogura cytoplasm (*ORF138*) and Kosena cytoplasm (*ORF125*), two alleles of the same ORF, associated with ATP synthase subunits.
(2) Brassica *pol* cytoplasm (*ORF224*), associated with ATP synthase subunits.
(3) Brassica *nap* cytoplasm (*ORF222*), associated with ATP synthase and NADH dehydrogenase subunits.
(4) Brassica *tour* cytoplasm (*ORF263*), associated with ATP synthase and NADH dehydrogenase subunits.
(5) *Moricandia arvensis* cytoplasm (*ORF108*), associated with ATP synthase subunits.
(6) Sunflower PET1 cytoplasm (*ORF522*), associated with ATP synthase subunits.
(7) Petunia pcf cytoplasm (*ORF402*), associated with ATP synthase, NADH dehydrogenase, and cytochrome oxidase subunits, as well as ribosomal protein genes.
(8) Maize cytoplasm S (*ORF355/ORF77*), associated with ATP synthase subunits.
(9) Maize cytoplasm T (*URF13*), associated with ATP synthase subunits.
(10) Sorghum A3 cytoplasm (*ORF107*), associated with ATP synthase subunits.
(11) Wheat As or Tt cytoplasm (*ORF256*), associated cytochrome oxidase subunits.
(12) Common bean CMS (*ORF239*), composed entirely of unique sequence.
(13) Rice Boro II cytoplasm (*ORF79*), associated with ATP synthase and cytochrome oxidase subunits.

Although knowledge of the distribution of the CMS and their (restorers) within and among species is incomplete, Brassica *pol*, Brassica *nap*, Maize cytoplasm S and Maize cytoplasm T were identified in intraspecific crosses; Rice Boro II cytoplasm in crosses between subspecies; Brassica *tour*, Sunflower PET1, Petunia pcf, and Wheat As or Tt cytoplasms in interspecific crosses; and Radish Ogura cytoplasm in an intergeneric cross.

A Nuclear Restorers of CMS. Nuclear loci that restore male fertility in CMS plants are referred to as restorer of fertility (*RF*) genes. Restorers typically act by regulating the transcript profile and/or protein accumulation of the CMS locus (Hanson and Bentolila, 2004). Thus far, seven *RF* genes have been cloned and characterized:

(1) Maize *RF2*. The first restorer gene to be cloned was an aldehyde dehydrogenase that restores the fertility of the *URF13* CMS locus in maize (Cui *et al.*, 1996). However, restoration only occurs in the presence of a second restorer gene, *Zm-RF1*. *Zm-RF1* is known to downregulate *URF13* expression, whereas *Zm-RF2* appears to biochemically compensate for detrimental consequences of residual *URF13* expression, possibly by oxidizing toxic aldehyde. A non-restoring *Zm-RF2* homologue is characterized by an amino acid substitution in the substrate binding pocket that appears to result in a loss of aldehyde dehydrogenase activity (Liu *et al.*, 2001).

(2) *RF-PPR592*. The *RF* gene in *Petunia* restores fertility by reducing the expression of the *PCF* CMS locus, leading to a substantial reduction in the amount of PCF protein (Nivison and Hanson, 1989). Map-based cloning indicates that the *Ph-RF* locus comprises at least two duplicated genes containing pentatricopeptide repeats (Bentolila *et al.*, 2002). The pentatricopeptide repeat (PPR) family is one of the largest gene families in plants and is known to be involved in regulating

organelle expression (Hanson and Bentolila, 2004). Transgenic experiments indicate that one of the duplicated *PPR* genes, *RF-PPR592*, can restore fertility of plants carrying the *PCF* CMS locus. A non-restoring homologue of *RF-PPR592* has a 530-bp deletion in the promotor region, which appears to account for its lack of expression in floral buds of CMS plants. Sequence analyses indicate that the non-restoring allele, *rf-PPR592*, probably arose via recombination between the duplicate PPR genes, similar to *RF-PPR591* and *RF-PPR592*.

(3) *RFK1/RFO*. Two nuclear loci are known to be capable of restoring fertility of the *ORF125* CMS locus (the so-called Kosena cytoplasm) in radish. One of these loci, *RFK1*, has been cloned and shown to be a member of the *PPR* gene family (Koizuka *et al.*, 2000). There was no change in the expression of *ORF125* in restored plants, but the amount of ORF125 protein was reduced. *RFK1* differs from a non-restoring homologue by four amino acid substitutions in the region of PPR repeats. *RFK1* is allelic to *RFO*, which restores the Ogura CMS, which is caused by the chimeric mitochondrial gene, *ORF138* (Brown *et al.*, 2003; Desloire *et al.*, 2003). Like *RFTK1*, *RFO* appears to downregulate *ORF138* at the translational or post-translational level (Uyttewaal *et al.*, 2008). *RFK1/RFO* is flanked by two additional *PPR* genes, one of which appears to be a pseudogene, implying a role for gene duplication in the evolution of the *RFO* locus.

(4) Rice *RF1A* (see below).

(5) Rice *RF1B*. In rice, co-transcription of mitochondrial genes yields a cytotoxic protein (ORF79) that is responsible for the Boro II (BT) CMS. BT CMS is restored by the *Os-Rf1* locus, which is composed of a cluster of *PPR* genes, two of which can independently restore male fertility: *Os-RF1A* and *Os-RF1B* (Wang *et al.*, 2006*b*). *Os*-RF1A reduces levels of the ORF79 protein by endonucleolytic cleavage, whereas *Os*-RF1B degrades *ORF79* transcripts. Sequence analyses of non-restorer *Os-RF1B* alleles revealed nine amino acid substitutions, one of which seems likely to cause a loss of restoration function. Likewise, *Os-RF1A* alleles were found to encode a truncated protein due to a frameshift mutation.

(6) *RETROGRADE-REGULATED MALE STERILITY* (*RMS*). The *RMS* gene was identified by positional cloning of *Os-RF17*, which restores the Chinese wild rice (CW)-type cytoplasmic male sterility (Fujii and Toriyama, 2009). The *RMS* gene encodes a 178-amino-acid protein of unknown function. Sterility is caused by upregulation of *RMS*, apparently due to *cis*-regulatory changes.

(7) Sorghum *RF1*. The *Sh-RF1* gene restores fertility of the sorghum A1 cytoplasm, which is the primary cytoplasm used for hybrid seed production. Fine-mapping localized the *RF1* gene to a 19-kb region containing three genes (Klein *et al.*, 2005). Two of these were completely conserved between restored and non-restored plants, whereas 19 differences were observed in the region spanning the third gene, which encoded a PPR protein.