Divergent Evolution of Duplicate Genes Leads to Genetic Incompatibilities Within *A. thaliana*

David Bikard,[1] Dhaval Patel,[2] Claire Le Metté,[1] Veronica Giorgi,[1] Christine Camilleri,[1] Malcolm J. Bennett,[2] Olivier Loudet[1]*

Genetic incompatibilities resulting from interactions between two loci represent a potential source of postzygotic barriers and may be an important factor in evolution when they impair the outcome of interspecific crosses. We show that, in crosses between strains of the plant *Arabidopsis thaliana*, loci interact epistatically, controlling a recessive embryo lethality. This interaction is explained by divergent evolution occurring among paralogs of an essential duplicate gene, for which the functional copy is not located at the same locus in different accessions. These paralogs demonstrate genetic heterogeneity in their respective evolutionary trajectories, which results in widespread incompatibility among strains. Our data suggest that these passive mechanisms, gene duplication and extinction, could represent an important source of genetic incompatibilities across all taxa.

W hen crossing individuals from differ-ent species is feasible, offspring often have reduced viability or fertility (*1*, *2*). The Bateson-Dobzhansky-Muller model explains such incompatibilities on the basis of the syner-gistic interaction of genes that have functionally diverged among the respective parents (*3–5*). Elucidating the molecular basis of such genetic incompatibilities is of great importance to the

science of evolution as well as to plant breeding. Whether these incompatibilities mostly appear concurrently with speciation (arising, for exam-ple, in geographically isolated populations) or after speciation has occurred as a consequence of their divergence remains a considerable question (*4*); it is now known that such incompatibilities can seg-regate within species first (*5*, *6*). A limited number of genes interacting to cause hybrid incompati-bility have been identified at the molecular level, such as the *Lhr*/*Hmr* system responsible for lethality of male F1s from a cross between two *Drosophila* species (*7*). Recently, an interaction between *zeel-1* and *peel-1* loci was discovered to cause widespread genetic incompatibility among *Caenorhabditis elegans* strains (*8*). Also at the

intraspecific scale, a typical dominant case of in-compatibility in *Arabidopsis thaliana* has been identified that may establish a link between hy-brid necrosis and the plant immune system (*9*).

While generating homozygous progeny from crosses between *A. thaliana* wild strains, it is fre-quently witnessed that physically unlinked loci do not always segregate independently and that, often, one homozygous allelic combination at two independent loci is rare or totally absent in the descendants of a specific cross (*10*, *11*). This phe-nomenon explains part of the segregation distor-tion inherited in such material and is viewed as the recessive version of Bateson-Dobzhansky-Muller–type incompatibilities, although other models than functional divergence could apply (*12*, *13*). Such epistasis-based recessive incompatibilities could result in reduced fitness in the progeny and limit the extent of rearrangements among parental ge-nomes. Furthermore, if several incompatibilities were to segregate within a cross they should lead to conflicts between the genomes of diverged strains, which could result in isolating barriers and, ultimately, speciation (*14*, *15*).

A cross between the *Arabidopsis* reference accession Columbia-0 (Col) and the Cape Verde Island accession Cvi-0 (Cvi) was generated, and 367 F6 recombinant inbred lines (RILs) were genotyped, revealing that two pairs of unlinked loci did not segregate independently from each other (*10*). For both of these pairs, a specific com-bination of Col and Cvi alleles was not found in the RIL set, resulting in a transchromosomal link-age disequilibrium pattern (pseudo-LD) and ex-hibiting segregation distortion (fig. S1). By focusing on one of these two-locus interactions (labeled LD1 in fig. S1), we realized that a homo-zygous combination of the Col allele at the LD1.1

locus (bottom of chromosome 1) and the Cvi allele at the LD1.5 locus (top of chromosome 5) caused arrested embryo development, resulting in seed abortion (Fig. 1 arrows). A large F2 population generated from the same cross recapitulated this recessive incompatibility (Fig. 1) and showed complete penetrance. We also noted that, in one intermediate heterozygous combination, the embryo developed normally but the primary root of the resulting seedling was shortened to ~one-third of its regular size, adding a quantitative phenotype we named "weak root" to the complexity of this epistasis. Taking advantage of heterogeneous inbred families [HIFs; (16)] derived from F6 RILs segregating for each locus while being fixed for the appropriate (incompatible) genotype at the other locus, we fine-mapped the LD1 epistatic interaction (17), reducing the candidate intervals of LD1.1 and LD1.5 to 65 and 15 kb, containing 11 and 4 annotated genes, respectively.

Gene annotation within these intervals revealed a candidate gene pair; the histidinol-phosphate amino-transferase gene codes for a protein (HPA) that catalyzes an important step in the biosynthetic pathway leading to histidine (His), an essential amino acid incorporated into proteins and hence required in many aspects of plant growth and development (18). Two paralogous copies of this gene are found in the Col genome; one (HPA1/HISN6A/At5g10330) lies within the LD1.5 candidate interval, and the other (HPA2/HISN6B/At1g71920) lies within the LD1.1 candidate interval. These paralogs appear to have

arisen from a recent single gene duplication event resulting in a dispersed duplicate pair, a mode of gene duplication that typically affects fewer genes than tandem or segmental duplication events in A. thaliana (19). In Col, At1g71920 and At5g10330 coding sequences differ by two synonymous single nucleotide polymorphisms (SNPs). Intraspecific sequence analyses and comparison to A. lyrata show that the ancestral locus is represented by At1g71920 (19) and that At5g10330 arose from a 3.3-kb duplication centered on At1g71920. In the Col background, transferred DNA (T-DNA)–insertion mutants in At5g10330 are embryo-lethal at the homozygous state [emb2196; (20)], and we have confirmed this result in other mutant lines, including SALK_089516. Recently, a weak point mutation allele of At5g10330 (hpa1) was shown to affect the maintenance of the root meristem and primary root elongation (21), an effect typically associated with the lack of free His in the plant. Both arrested embryo development and root growth impairment have been complemented by supplying emb2196 and hpa1 plants with exogenous histidine (18, 21).

The SNPs in the coding sequence were used to distinguish between the two paralogs' mRNA, confirming that there is no detectable transcriptional activity from At1g71920 in Col, whereas At5g10330 is expressed in both shoot and root [fig. S2; (19)]. Sequence and reverse transcription polymerase chain reaction (RT-PCR) analyses in Cvi showed that the gene at LD1.1 (At1g71920) was expressed, whereas there were no traces of an HPA coding sequence at LD1.5 in

this background (fig. S2). Instead, according to sequence results, a 6.4-kb region appears to be deleted in Cvi (compared with the Col sequence) that encompasses the entire Col duplication stretching to 3 kb beyond the Col duplicated region on one side. Furthermore, the borders of this additional 3-kb deletion contained traces of transposable elements: ATGP9-LTR and VANDAL-18NA. The region homologous to LD1.5 in A. lyrata lacks a HPA gene but is also slightly different from Cvi, with deletions and insertions of hundreds of base pairs at the locus. Then, the possibility that the LD1.5 locus might have undergone multiple rearrangements makes it difficult to determine whether the HPA gene was deleted in Cvi or whether it was ever present in this background.

The situation at each paralog is summarized in Fig. 1, with Col and Cvi retaining alternate functional copies of the essential HPA gene. The combination of the two silenced copies in a progeny homozygous for the Col allele at LD1.1 and the Cvi allele at LD1.5 leads to arrested seed development, presumably because the embryo is unable to synthesize His. Similarly, the limited primary root growth in weak-root plants could be explained by a reduced His quantity in these plants because they have a single functional HPA copy originating from the Col LD1.5 locus (Fig. 1). Given that the Col LD1.5 allele is less expressed than the Cvi LD1.1 allele (fig. S2), this allelic combination may be specifically limiting for the root, an organ particularly sensitive to a shortage in His (21). All other genotypes should result in greater HPA activity, enough to sustain normal

Fig. 1. Genetic mechanism underlying LD1 interaction and incompatibility. All combinations of alleles at LD1.5 and its interactor LD1.1 result in phenotypically normal plants, except for the combination Col at LD1.1/Heterozygous (Het) at LD1.5, which shows a reduced primary root length (weak-root phenotype), and another combination (Col at LD1.1/Cvi at LD1.5), which shows embryo lethality at an early stage in the silique (seeds under development containing lethal embryos are indicated by arrows). Fine-mapping identified a duplicate gene for which Col and Cvi show reciprocal gene loss explaining the interaction and incompatibility.
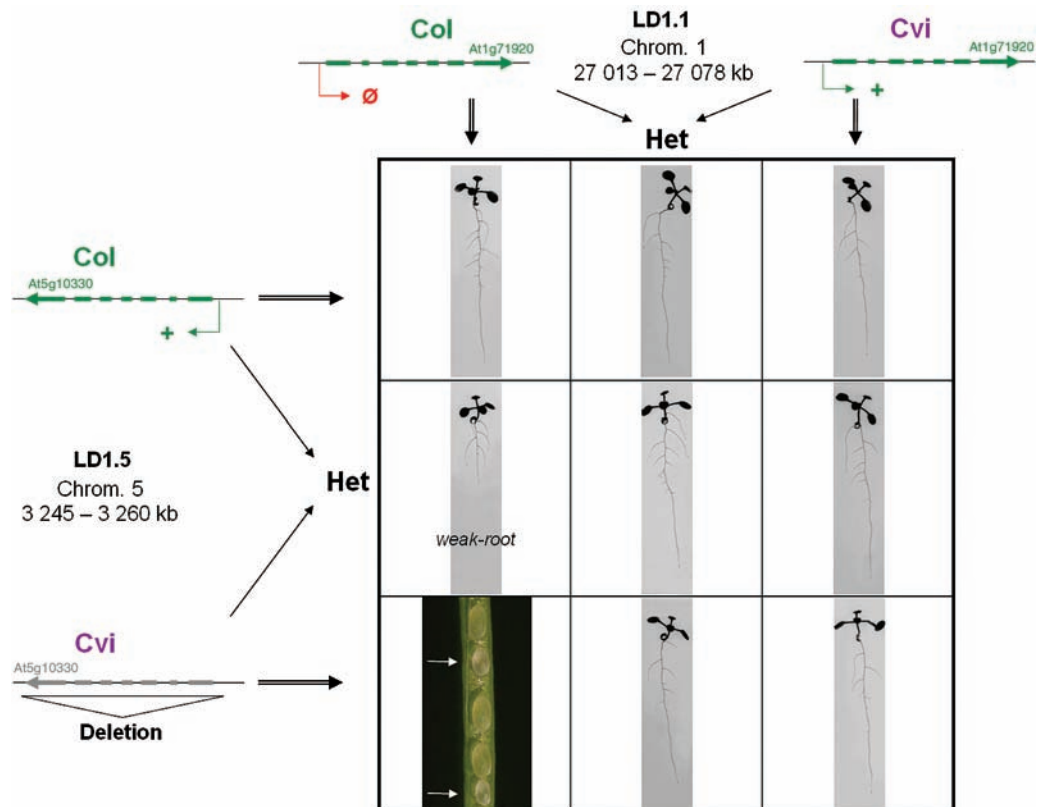
embryo and root development. Moreover, limited root growth in weak-root individuals can be explained by an alteration of the cell production rate because cell elongation remains normal in these plants (fig. S3), which is consistent with the lack of His and its associated defect in root meristem maintenance (21).

One way to prove the link between LD1 and *HPA* genes is to show that LD1 can be complemented by adding exogenous His. Weak-root plants were gradually and quantitatively rescued to the phenotype of the control by growing the plants on increasing concentrations of His in vitro (Fig. 2). Furthermore, watering heterozygous plants from the bolting stage with sufficient amounts of His restored the viability of embryos with incompatible allelic combinations (17), providing complementation of the embryo lethality

phenotype as well. From these experiments, we concluded that the LD1 incompatibility results from a shortage of His in certain genotypes.

To prove the causative role of *HPA* genes on LD1, we performed an allelic (quantitative) complementation test by combining different alleles at LD1.5 and At5g10330 in an identical F1 background. Crossing segregating HIF lines to different mutants (*hpa1* and *emb2196*) revealed how the different alleles at the two duplicate genes interact to qualitatively control embryo development and quantitatively limit primary root growth (Fig. 3). From crosses with the *hpa1* mutant (Fig. 3A), we observed that a Cvi allele at LD1.5 is unable to complement the EMS mutant allele at At5g10330 (whereas a Col allele does). This genotype (Cvi/*hpa1*) leads to embryo lethality, a phenotype even stronger than homozygous *hpa1* mutants (these embryos survive), indicating that the causative Cvi allele at LD1.5 is more deleterious than the EMS mutation (which fits well with the gene being completely deleted in Cvi). The significant interaction between the LD1.5 alleles and *HPA* genotypes argues that *HPA* is involved in LD1 epistasis. Crosses to the *emb2196* mutant led to the same conclusion when Col and Cvi alleles were compared versus the stronger *emb2196* allele (Fig. 3B). However, a heterozygous seedling for *emb2196* mutation (Col/*emb2196* at At5g10330) had a significantly ($P < 0.01$) longer primary root than a heterozygous seedling with a Cvi/Col genotype at LD1.5 (Fig. 3B). This, together with the fact that another T-DNA line in the same gene (SALK_089516) has a shorter root than *emb2196* in the heterozygous state (similar to weak-root plants), indicated that the *emb2196* T-DNA insertion is probably not a complete null allele (in contrast to SALK_089516). From these experiments, we concluded that epistasis at LD1 is most likely explained by allelic variation at *HPA* genes as depicted in Fig. 1 and that the incompatibility observed between Col and Cvi represents an example of intraspecific divergence of a duplicate gene pair.

Similar patterns of segregation distortion involving regions at the bottom of chromosome 1 and top of chromosome 5 were detected in a RIL set derived from the cross between Cvi and Landsberg *erecta* (L*er*), indicating that there may be a similar interaction between these loci (22). Nearly isogenic lines derived from this population (23) confirmed epistasis and incompatibility by showing a specific pattern of segregation and the weak-root phenotype. Both *HPA* genes were expressed in L*er*; however, the chromosome 1 paralog contains a premature stop codon (Fig. 4). This encouraged us to analyze the extent of functional natural variation at those loci and further characterize intraspecific evolution of these genes in *A. thaliana*.

We analyzed 30 accessions derived from distinct natural populations representing most known variation in *Arabidopsis* (24) for the expression of each gene copy, potential deletions,
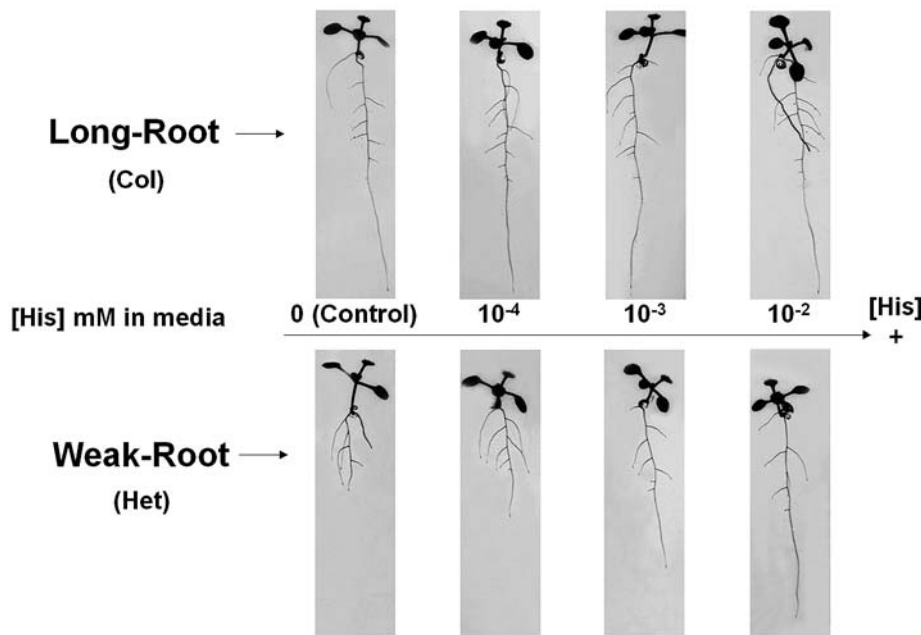

Fig. 2. The weak-root phenotype is quantitatively complemented by exogenous His. Typical phenotypes of descendants from a plant segregating at the LD1.5 locus when grown on media supplemented with different histidine concentrations. Plants were identified on the basis of their genotype at the LD1.5 locus. The complementation of the expected weak-root phenotype was complete when supplied with 10^{-2} mM His.
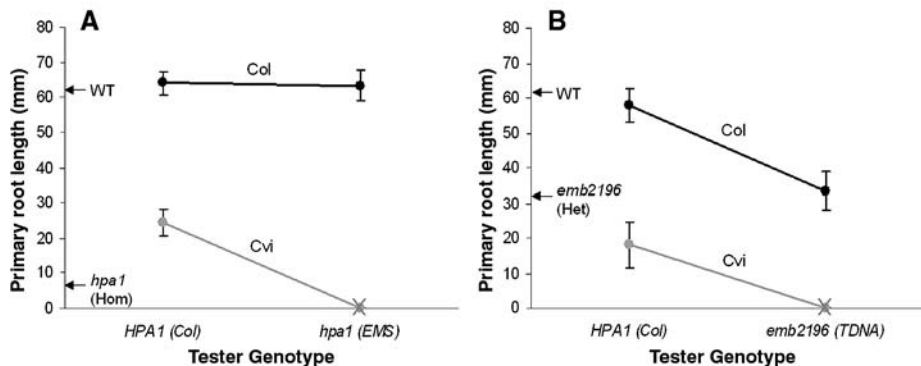
Fig. 3. Allelic complementation of LD1 interaction. Quantitative complementation tests were performed by combining different alleles at LD1 loci in F1 backgrounds. The relative complementation of either (**A**) an EMS mutant allele (*hpa1*) or (**B**) a T-DNA insertion mutant allele (*emb2196*) at At5g10330 by a Col or Cvi allele at LD1.5 was measured through both the length of the primary root (shown in mm) and embryo lethality. Individuals with a phenotype depicted on the *x* axis (null) underwent seed abortion. All plants are Col at LD1.1. Arrows along the *y* axis represent the typical root length of control genotypes [WT indicates wild-type Col plants; *hpa1* (Hom), homozygous *hpa1* mutant plants; and *emb2196* (Het), heterozygous individuals for the *emb2196* mutation]. Each data point represents the mean ± standard error of about 40 plants. The LD1.5 allele × HPA1 genotype interaction term as tested by analysis of variance is significant in crosses to *hpa1* ($P < 0.0001$) and crosses to *emb2196* ($P < 0.05$).

Fig. 4. Divergent evolution of duplicate genes among *A. thaliana* accessions. Groups of accessions are presented according to At5g10330 and/or At1g71920 genotypes and transcript accumulation phenotypes. Accessions (underlined) from each group were crossed to Col and/or Cvi and tested for LD1 incompatibility and compatibility to confirm the loss of function of one or the other duplicate gene. Dash-underlined accessions show conditional incompatibility (table S1). The incompatible groups are circled: green-circled genotypes are incompatible with Cvi; purple-circled genotypes are incompatible with Col. Genotypes not circled are fully compatible with both Col and Cvi.
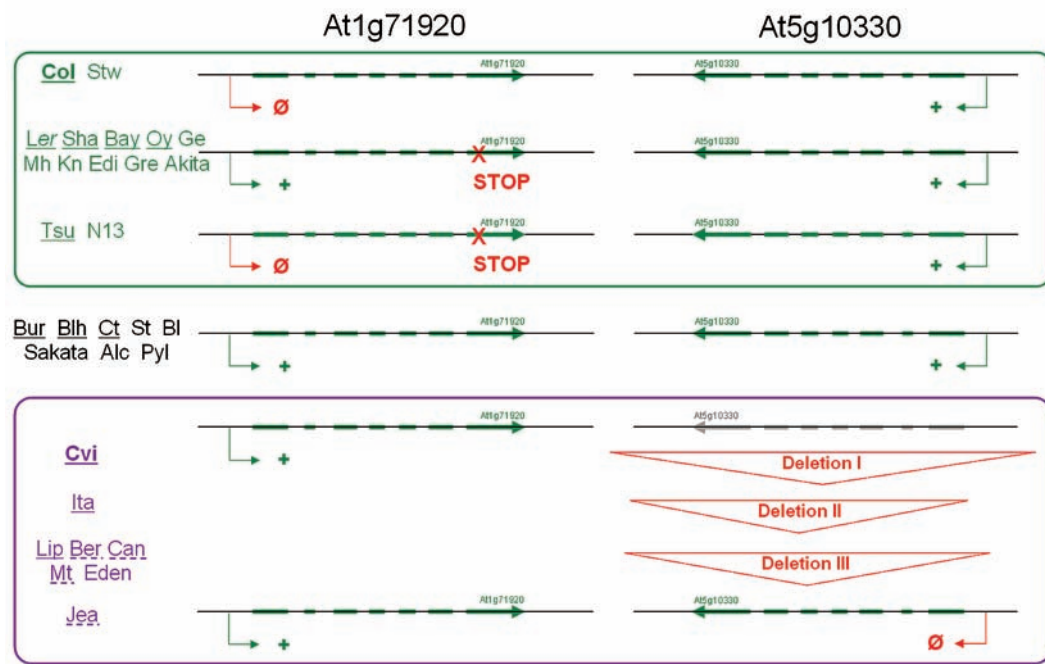
and deleterious mutations. We also looked for LD1.1/LD1.5 genetic incompatibilities (accompanied by segregation of the weak-root phenotype) in 28 of the possible crosses to Col or Cvi (either in RIL sets or F2 populations) to confirm the relevance of variation detected at the nucleotide and/or expression level. Divergent evolution of the *HPA* gene pair was dramatic and widespread (Fig. 4 and table S1) because most (22/30) of the accessions tested have silenced one or the other copy in at least six different ways (early stop, no expression, a combination of both, and different deletions). Cvi- and Col-incompatible groups represent 14 and 8 accessions, respectively, and, on the basis of these data, we estimate that at least one-fourth of all possible crosses among these 30 strains would show *HPA* incompatibility. We observed no particular correlation with geographical population structure (*25*). Most of the four structure groups defined previously (*26*) included strains belonging to the two incompatibility groups as well as compatible accessions.

We confirmed that the rapid and common evolution of duplicate genes provides an important source of Bateson-Dobzhansky-Muller–like epistatic interactions after paralogs have been reciprocally silenced or lost in diverged strains, as proposed (*27*, *28*). Depending on the function and essential nature of the gene, this may result in hybrid and/or F2 defective fitness in the descendants of certain intraspecific crosses and could contribute to reproductive isolation (*29*). Passive gene loss is recognized as the most probable fate of duplicate genes and is especially likely at early stages after small-scale duplication events (*12*, *30*). However, direct evidence of gene loss as a neutral mechanism generating postzygotic isolating barriers within existing spe-

cies with no prior fitness consequences in the parental strains (because only the location of the functional copy changes) has been lacking (*4*, *31*, *32*). Staal *et al*. (*33*) described how a transposition of a resistance gene found in the L*er* strain of *Arabidopsis* was responsible for variation in disease susceptibility in crosses to Col. Similarly, transposition of an essential gene has recently been associated with the sterility of a hybrid between two *Drosophila* species (*29*), and descriptive work on three related yeast species indicated that divergent resolution events after whole-genome duplication may have contributed to their speciation (*34*). Our study extends these observations in demonstrating the link between gene duplication and genetic incompatibility.

References and Notes

1. K. Bomblies, D. Weigel, *Nat. Rev. Genet.* **8**, 382 (2007).
2. J. Mallet, *Trends Ecol. Evol.* **21**, 386 (2006).
3. T. Dobzhansky, *Genetics and the Origin of Species* (Columbia Univ. Press, New York, 1937).
4. J. A. Coyne, H. A. Orr, *Speciation*, M. A. Sunderland, Ed. (Sinauer, Sunderland, MA, 2004).
5. K. Bomblies, D. Weigel, *Curr. Opin. Genet. Dev.* **17**, 500 (2007).
6. A. Kopp, A. Frank, *Genetica* **125**, 55 (2005).
7. N. J. Brideau *et al.*, *Science* **314**, 1292 (2006).
8. H. S. Seidel, M. V. Rockman, L. Kruglyak, *Science* **319**, 589 (2008); published online 10 January 2008 (10.1126/science.1151107).
9. K. Bomblies *et al.*, *PLoS Biol.* **5**, e236 (2007).
10. M. Simon *et al.*, *Genetics* **178**, 2253 (2008).
11. O. Törjék *et al.*, *Theor. Appl. Genet.* **113**, 1551 (2006).
12. M. Lynch, J. S. Conery, *Science* **290**, 1151 (2000).
13. H. A. Orr, J. P. Masly, N. Phadnis, *J. Hered.* **98**, 103 (2007).
14. H. A. Orr, M. Turelli, *Evolution* **55**, 1085 (2001).
15. H. A. Orr, J. P. Masly, D. C. Presgraves, *Curr. Opin. Genet. Dev.* **14**, 675 (2004).
16. O. Loudet, V. Gaudon, A. Trubuil, F. Daniel-Vedele, *Theor. Appl. Genet.* **110**, 742 (2005).
17. Materials and methods are available as supporting material on *Science* Online.
18. R. Muralla, C. Sweeney, A. Stepansky, T. Leustek, D. Meinke, *Plant Physiol.* **144**, 890 (2007).
19. R. C. Moore, M. D. Purugganan, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15682 (2003).
20. I. Tzafrir *et al.*, *Plant Physiol.* **135**, 1206 (2004).
21. X. Mo *et al.*, *Plant Physiol.* **141**, 1425 (2006).
22. C. Alonso-Blanco *et al.*, *Plant J.* **14**, 259 (1998).
23. J. J. Keurentjes *et al.*, *Genetics* **175**, 891 (2007).
24. H. I. McKhann *et al.*, *Plant J.* **38**, 193 (2004).
25. M. Nordborg *et al.*, *PLoS Biol.* **3**, e196 (2005).
26. M. F. Ostrowski *et al.*, *Mol. Ecol.* **15**, 1507 (2006).
27. M. Lynch, A. G. Force, *Am. Nat.* **156**, 590 (2000).
28. C. R. Werth, M. D. Windham, *Am. Nat.* **137**, 515 (1991).
29. J. P. Masly, C. D. Jones, M. A. F. Noor, J. Locke, H. A. Orr, *Science* **313**, 1448 (2006).
30. S. Maere *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5454 (2005).
31. M. Lynch, J. S. Conery, *J. Struct. Funct. Genomics* **3**, 35 (2003).
32. M. Semon, K. H. Wolfe, *Curr. Opin. Genet. Dev.* **17**, 505 (2007).
33. J. Staal, M. Kaliff, S. Bohman, C. Dixelius, *Plant J.* **46**, 218 (2006).
34. D. R. Scannell, K. P. Byrne, J. L. Gordon, S. Wong, K. H. Wolfe, *Nature* **440**, 341 (2006).
35. We thank K. Bomblies and D. Weigel for providing F2 seeds for specific test crosses (Bay-0 × Cvi and Tsu-0 × Cvi), X. Mo for hpa1 seeds, and R. Mercier and F. Roux for discussion and critical reading of the manuscript. Funded partly by an INRA–Département de Génétique et d'Amélioration des Plantes Innovative Project grant to C.C. and O.L. Funding by the University of Nottingham to D.P. and the Biotechnology and Biological Sciences Research Council/ Engineering and Physical Sciences Research Council Centres for Integrative Systems Biology program to M.J.B. Sequences have been deposited in GenBank under accessions FJ436446 and FJ436447.

Divergent Evolution of Duplicate Genes Leads to Genetic Incompatibilities Within A. thaliana

David Bikard, *et al.*

Science **323**, 623 (2009);

DOI: 10.1126/science.1165917