LETTERS

Positive darwinian selection at the imprinted *MEDEA* locus in plants

Charles Spillane[1,2], Karl J. Schmid[3]†, Sylvia Laoueillé-Duprat[2], Stéphane Pien[1], Juan-Miguel Escobar-Restrepo[1], Célia Baroux[1], Valeria Gagliardini[1], Damian R. Page[1], Kenneth H. Wolfe[4] & Ueli Grossniklaus[1]

In mammals and seed plants, a subset of genes is regulated by genomic imprinting where an allele's activity depends on its parental origin. The parental conflict theory suggests that genomic imprinting evolved after the emergence of an embryo-nourishing tissue (placenta and endosperm), resulting in an intragenomic parental conflict over the allocation of nutrients from mother to offspring[1,2]. It was predicted that imprinted genes, which arose through antagonistic co-evolution driven by a parental conflict, should be subject to positive darwinian selection[3]. Here we show that the imprinted plant gene *MEDEA* (*MEA*)[4,5], which is essential for seed development, originated during a whole-genome duplication 35 to 85 million years ago. After duplication, *MEA* underwent positive darwinian selection consistent with neo-functionalization and the parental conflict theory. *MEA* continues to evolve rapidly in the out-crossing species *Arabidopsis lyrata* but not in the self-fertilizing species *Arabidopsis thaliana*, where parental conflicts are reduced. The paralogue of *MEA*, *SWINGER* (*SWN*; also called *EZA1*)[6], is not imprinted and evolved under strong purifying selection because it probably retained the ancestral function of the common precursor gene. The evolution of *MEA* suggests a late origin of genomic imprinting within the Brassicaceae, whereas imprinting is thought to have originated early within the mammalian lineage[7].

Disruption of the imprinted *Arabidopsis MEA* gene, which encodes an Enhancer of zeste [E(z)]-related protein, leads to delayed development and over-proliferation of embryo and endosperm[4,8]. Together with *SWN* and *CURLY LEAF* (*CLF*), *MEA* forms a family of *E(z)*-like genes in *Arabidopsis*[9,10]. To gain insights into their evolutionary relationship, we investigated whether the *Arabidopsis E(z)*-like genes arose via duplication of large genomic blocks[11]. *MEA* (At1g02580) is a recently derived paralogue of *SWN* (At4g02020) located on a block duplication spanning 39 paralogues on chromosome 1 and 41 paralogues on chromosome 4 (Fig. 1a). The block duplication on which the *MEA* and *SWN* paralogues reside arose ~35–85 million years (Myr) ago as a result of a whole-genome duplication within the Brassicaceae lineage[11–13]. In contrast, the *CLF* gene (At2g23380) is located in a genomic region that exhibits no co-linearity with the regions containing *SWN* and *MEA* in either *Arabidopsis* (Fig. 1a) or rice (data not shown).

To investigate further these duplications, we included all known plant *E(z)*-like genes in a phylogenetic analysis (Fig. 1b). The presence of *CLF*-like and *SWN*-like genes in both monocotyledons and dicotyledons indicates a duplication separating *CLF* and the common ancestor of *SWN* and *MEA* before the divergence of these taxa ~200 Myr ago. In agreement with the block duplication data, we found no direct orthologues of *MEA* (as opposed to co-orthologues of both *MEA* and *SWN*) in the available sequences (including expressed sequence tags) from any species other than *Arabidopsis thaliana*. However, we obtained orthologues of both *MEA* and *SWN* in *Arabidopsis lyrata* (Supplementary Fig. 1). All previous phylogenetic analyses of the plant *E(z)*-like genes suggested that *MEA* is a basal outgroup to both *CLF* and *SWN*[6,10,14–17]. In contrast, our data reveal an old duplication between the *CLF* and *MEA/SWN* lineages, followed by a more recent duplication that produced *MEA* and *SWN* (Fig. 1b).

To analyse functional diversification of *E(z)*-like genes in *Arabidopsis*, we studied their expression and function during reproductive development. *MEA* is expressed in the synergids, egg and central cells of the embryo sac before fertilization, and in the embryo and endosperm during seed development[5]. To determine whether *MEA*, *SWN* and *CLF* have overlapping expression patterns in the embryo sac and developing seed, we performed comparative *in situ* hybridization analyses using gene-specific probes (Fig. 2 and Supplementary Fig. 2). Although the three genes have largely overlapping expression patterns, there are important differences: all three transcripts were detected in the synergids and egg cell (Fig. 2a, c, e); however, the expression of *SWN* was strongly reduced and that of *CLF* undetectable in the central cell compared with *MEA*, which showed strong expression (Fig. 2a). This difference was maintained during early seed development when *MEA* was detected in free nuclear endosperm (Supplementary Fig. 2a) but *SWN* and *CLF* were not (Supplementary Fig 2e, i). After fertilization, transcripts of all three genes were detected in the globular embryo and micropylar endosperm, with only *MEA* transcripts detected in the suspensor (Fig. 2b, d, f). The expression of all three genes decreased at the heart stage (data not shown) and they were no longer detectable in embryos of the torpedo stage (Supplementary Fig. 2o, s, w). Thus, in comparison to its paralogues *SWN* and *CLF*, *MEA* has a differential expression domain in the central cell and free nuclear endosperm, and also in the suspensor, both tissues thought to be involved in nutrient transfer to the developing embryo.

Mutations in *MEA* and other members of the *FERTILIZATION-INDEPENDENT SEED* (*FIS*) class of genes show characteristic pre- and post-fertilization phenotypes: endosperm proliferation in the absence of fertilization (*fis* phenotype) and maternal effect seed abortion[4,18–21]. As *MEA* is a maternally expressed imprinted gene, maternal effect seed abortion in *Arabidopsis mea* mutants can occur in heterozygous seeds that have maternally inherited a mutant *mea* allele, yet harbour a wild-type paternal *MEA* allele that is not expressed. To test whether the paternal allele of the *MEA* orthologue in *A. lyrata* is also not expressed in seeds, we analysed the relative expression levels of maternal and paternal *MEA* alleles in seeds

[1]Institute of Plant Biology & Zürich-Basel Plant Science Center, University of Zürich, CH-8008 Zürich, Switzerland. [2]Genetics & Biotechnology Lab, Department of Biochemistry & Biosciences Institute, University College Cork, Cork, Ireland. [3]Department of Genetics and Evolution, Max Planck Institute for Chemical Ecology, Hans-Knöll-Str. 8, D-07745 Jena, Germany. [4]Smurfit Institute of Genetics, University of Dublin, Trinity College, Dublin, Ireland. †Present address: Leibniz-Institute of Plant Genetics and Crop Plant Research, D-06466 Gatersleben, Germany.

generated from crosses between *A. thaliana* and *A. lyrata* (as pollen parent). Similar to the imprinted *MEA* locus in *A. thaliana*, the *MEA* orthologue in *A. lyrata* is not expressed from the paternal allele in developing seeds (Supplementary Table 1). This result is consistent with our findings that a paternally inherited *A. lyrata MEA* allele cannot rescue the *mea* seed abortion phenotype (data not shown), and suggests that *MEA* is imprinted in both of the sister species, *A. thaliana* and *A. lyrata*.

Because *MEA* shows an overlapping expression pattern with *SWN* and *CLF*, we tested whether *swn* or *clf* mutants are impaired in either embryo sac or seed development. It was recently shown that *SWN* and *MEA* have a redundant function with respect to the pre-fertilization *fis* phenotype[22]. In contrast, our analysis of *swn* and *clf* mutants alone and in combination with *mea* showed neither an impairment of post-fertilization seed development nor an enhancement of the *mea* seed abortion phenotype, respectively (Supplementary Table 2). As *swn*;*clf* double mutants also produce normal seeds[6], these results indicate that neither *SWN* nor *CLF* has a role in seed development. Because neither *SWN* nor *CLF* are essential for post-fertilization seed development, we propose that the new post-fertilization role of *MEA* in seed development was acquired within the past ~35–85 Myr.

We further proposed that the protein sequence of MEA evolved rapidly after its origin by selection-driven substitutions of amino acids. In contrast, SWN would have retained the ancestral function and is expected to have evolved under purifying selection. We investigated the neo-functionalization hypothesis by testing whether the ratio $\omega = d_N/d_S$ of non-synonymous (d_N) to synonymous (d_S) divergence[23] is higher for the lineage ancestral to *MEA* than for *SWN* (Fig. 1c). The *E(z)*-like genes from poplar and petunia predate the *SWN–MEA* duplication and were taken as pro-orthologues, *sensu* ref. 24. A two-ratio branch model that estimates a single ω-ratio for the branches leading to the pro-orthologues and to *SWN* (reflecting functional conservation), and a second ω-ratio for *MEA* (allowing functional diversification), provided a significantly better fit to the data than a one-ratio model with a single ω-ratio for the whole phylogeny ($P < 0.0001$; Supplementary Table 3). Three-ratio and free-ratio models were not better than the two-ratio model ($P > 0.05$), suggesting that most variation in selective constraint occurred after the divergence of *SWN* and *MEA*.

Because *E(z)*-like proteins consist of a mosaic of conserved and divergent domains, we analysed which amino acid residues evolved under positive selection during functional diversification using branch-site models (Supplementary Table 4). This analysis revealed no evidence for positive selection in *SWN* ($P = 0.99$) but was highly significant for *MEA* ($P < 0.0001$). The ω-ratio of positively selected codons in the ancestral branch of *MEA* was estimated to be 1.68,

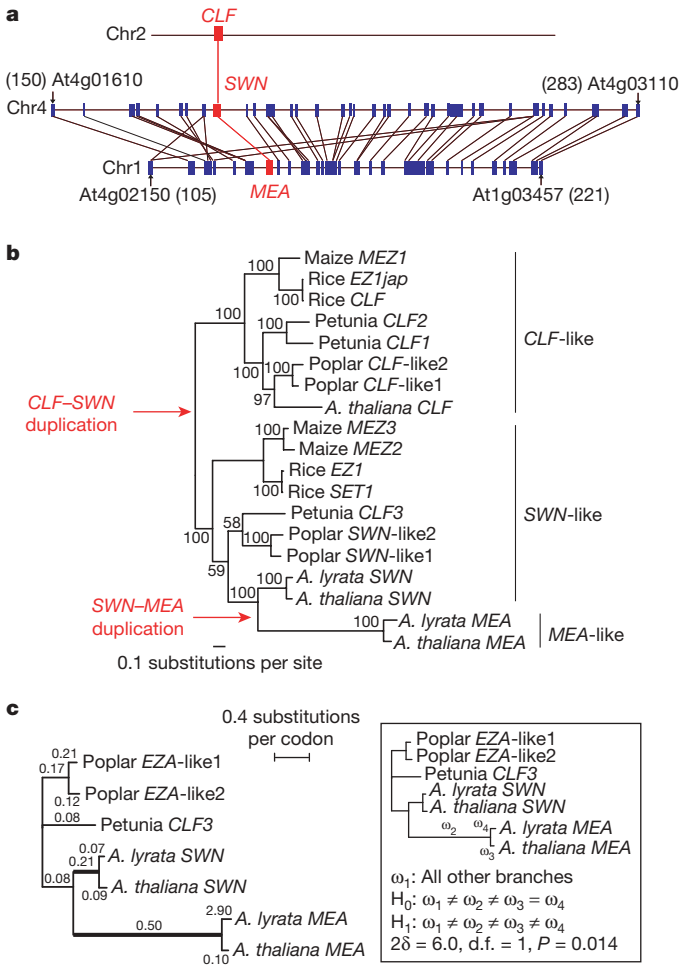

Figure 1 | *MEA* and *SWN* are paralogues. **a**, The imprinted *MEA* gene and its paralogue *SWN* lie on duplicated blocks of 0.457 megabases (Mb) and 0.690 Mb on chromosomes 1 and 4, respectively. The paralogon block spans 39 paralogous genes (blue blocks), including *MEA*, on chromosome 1, and 41 paralogues, including *SWN*, on chromosome 4. *CLF* is located on chromosome 2 in a genomic region exhibiting no co-linearity with the *MEA* and *SWN* paralogon blocks. **b**, Phylogenetic analysis of *E(z)*-like genes in higher plants. The tree was constructed with the protml package and rooted for the *CLF*—*SWN* duplication. **c**, Phylogenetic tree of *MEA*- and *SWN*-like genes from dicotyledonous species. The tree topology was obtained with protml, and the branch lengths (substitutions per codon) were calculated with codeml using Model M0. The numbers above branches indicate the ω-ratio and they were calculated with the free-ratio model. Note that the tree is unrooted.

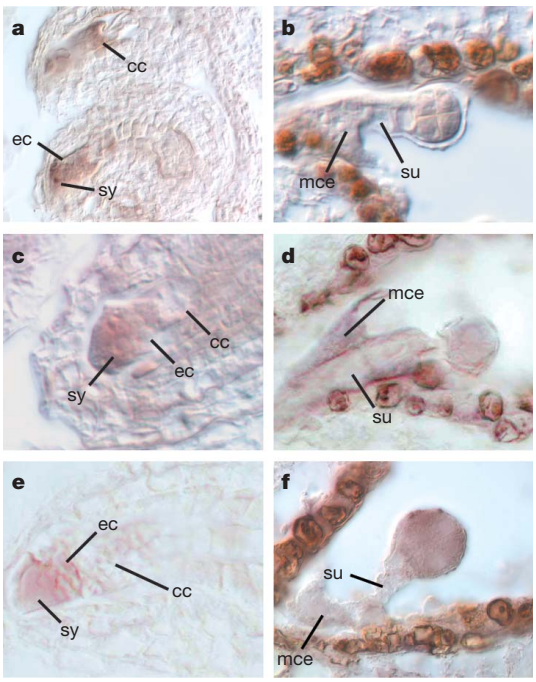

Figure 2 | Spatio-temporal expression patterns of *MEA*, *SWN* and *CLF* in the embryo sac and early seed assayed by *in situ* hybridization. **a**, **b**, Sections probed with anti-sense *MEA*. **c**, **d**, Sections probed with anti-sense *SWN*. **e**, **f**, Sections probed with anti-sense *CLF*. **a**, **c**, **e**, *MEA* (**a**), *SWN* (**c**) and *CLF* (**e**) transcripts accumulate in the synergids (sy) and in the egg cell (ec), whereas only *MEA* is expressed strongly in the central cell (cc). **b**, **d**, **f**, *MEA*, *SWN* and *CLF* transcripts were detected in the globular embryo and the micropylar endosperm (mce). **b**, At the globular stage, *MEA* transcripts are detected in the suspensor (su) in contrast to *SWN* and *CLF*. The strong staining seen in the endothelium is an artefact also observed in sense controls. For sense controls and a more detailed description of the expression patterns of *MEA*, *SWN* and *CLF* in the embryo sac and developing seed, see Supplementary Fig. 2.

which differs from a neutral model with ω fixed to 1.00 ($P = 0.026$). Therefore, the high ω-ratio does not result from relaxed constraints but from positive selection on *MEA*. The 74 codons with a posterior probability of positive selection >0.95 are located throughout the coding sequence (Supplementary Fig. 3), suggesting that positive selection was not restricted to any particular domain of the MEA protein. The numerous insertions and deletions of amino acids may also contribute to the functional divergence.

Within the genus *Arabidopsis*, *MEA*, but not *SWN*, continues to evolve under positive selection. The pairwise ω-ratio of *MEA* ($\omega = 0.75$) in *A. thaliana* and *A. lyrata* is higher than that of *SWN* ($\omega = 0.25$, $P < 0.0001$; Supplementary Table 5). A free-ratio model (Fig. 1c) indicates that *MEA* evolves under positive selection in the branch leading to *A. lyrata* ($\omega = 2.90$) but not in the *A. thaliana* branch ($\omega = 0.10$). A four-ratio model with independent ω-ratios for the *A. thaliana* and *A. lyrata MEA* genes provides a better fit to the data than a three-ratio model with a single ω-ratio for both branches ($2\delta = 6.0$, degrees of freedom = 1, $P = 0.014$; Fig. 1c), supporting positive selection in the *A. lyrata* branch. A test for differences in *SWN* between *A. thaliana* and *A. lyrata* was not significant (data not shown).

These results suggest that *MEA* is involved in a genomic conflict in *A. lyrata*, but not in the *A. thaliana* lineage where similar ω-ratios for *SWN* ($\omega = 0.09$) and *MEA* ($\omega = 0.10$) are observed (Fig. 1c). To test this hypothesis, we analysed intraspecific sequence variation at the *MEA* and *SWN* genes in *A. thaliana*, and at *MEA* in *A. lyrata* and its close, also out-crossing relative *A. halleri* (Supplementary Table 6). In the *A. thaliana* sample, total nucleotide diversity, π, is 1.5 times higher at *MEA* (0.0037) than at *SWN* (0.0022), and lower than the genome-wide average (0.007)[25]. The ratio of non-synonymous to synonymous polymorphisms, π_N/π_S, at *MEA* (0.259) is similar to *SWN* (0.183) and smaller than 1, indicating purifying selection. Several tests of neutral evolution failed to provide evidence for positive selection in *A. thaliana* (Supplementary Table 6); both loci appear to evolve neutrally under similar evolutionary constraints. In *A. lyrata* and *A. halleri*, similar patterns are observed for *MEA* (Supplementary Table 6). The π_N/π_S ratio is 1.25 and 0.58 in *A. lyrata* and *A. halleri*, respectively. Polymorphism levels are in the same range as observed for *A. thaliana*, but low in comparison to other *A. lyrata* genes located on the same chromosome. Tests of neutral evolution are not significant, although there is a slight excess of intermediate frequency polymorphisms in both species (Tajima's $D > 1$).

Positive darwinian selection on *MEA* occurring in the lineage leading to *A. lyrata* but not in *A. thaliana*, and a high π_N/π_S, is consistent with the parental conflict hypothesis for the evolution of imprinting. Within self-fertilizing *A. thaliana*, we expect differential selective pressures between maternal and paternal alleles of genes controlling resource allocation from mother to offspring (such as *MEA*) to be weaker. Because patterns of intraspecific sequence variation do not reject a neutral model in both the inbreeding and out-crossing species, selective pressures on *MEA* may be weak.

In mammals, imprinted gene clusters may have been linked together on one or a few ancestral chromosome(s), arguing for a common mechanistic origin of imprinting early in mammalian evolution[7]. In contrast, imprinting of *MEA* within the *E(z)*-like gene family arose late in the evolution of flowering plants, as *MEA*-like genes are restricted to the Brassicaceae. We propose that *MEA* became imprinted after it arose through a block duplication, possibly because of a need for dosage compensation after it acquired a new function. *SWN* and *CLF* have growth-regulating activity in the seedling, as demonstrated by aberrant growth of *swn;clf* double mutants after germination. Our studies on the evolution, function and expression of the *E(z)*-like genes suggest that *MEA* acquired a new function in regulating growth during seed development. A tight regulation of the *MEA* expression level around fertilization seems to be crucial for normal development[26]. As a result, the level of overlapping *MEA* and *SWN* activity may have required adjustment after the duplication event. A pre-existing imprinting machinery may have been recruited to adjust *MEA* expression levels, leading to the recently evolved regulation of the *MEA* locus by genomic imprinting.

METHODS SUMMARY

Plant material. The *mea-1* and *swn-3* mutants of *A. thaliana* used for single- and double-mutant analyses have been previously described[4,6]. The *swn-4* mutant is the SIGNAL insertion line SALK_109121[27]. *A. lyrata* and *A. halleri* accessions were provided by T. Mitchell-Olds, M. Clauss and R. Oyama.

Expression analyses. *In situ* hybridization on embryo sac and early seed tissues was performed as previously described[5]. Riboprobes designed from the most divergent regions of the *MEA*, *SWN* and *CLF* gene sequences were used. Expression analysis of the *A. lyrata* paternal *MEA* allele in the developing seed was conducted using F_1 seed from crosses between *A. thaliana* and *A. lyrata* (where *A. lyrata* was used as a pollen parent), and quantitative polymerase chain reaction with reverse transcription (qRT–PCR) probes specific for the *MEA* transcript from each species[26].

Sequence analyses. DNA sequencing of *MEA* and *SWN* genes from *A. lyrata* and *A. halleri* was performed by high-fidelity PCR amplifications of overlapping fragments from genomic DNA and sequencing of five independent clones per PCR-amplified fragment. DNA sequence data from *A. thaliana* accessions was based on direct sequencing of high-fidelity PCR-amplified fragments.

Phylogenetic and molecular evolution analyses. Phylogenetic analysis was based on all *CLF*-, *SWN*- and *MEA*-like sequences obtained from BLAST searches of GenBank. Protein sequences were aligned with the CLUSTALW program and a phylogenetic tree constructed with the protml program of the MOLPHY package[28]. The analysis of ω-ratios was conducted with the codeml program of the PAML package[29]. Molecular population genetic analysis was conducted using the DnaSP program[30] based on sequence data from the accessions listed in Supplementary Tables 7 and 8.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 31 March; accepted 5 June 2007.

1. Haig, D. & Westoby, M. Parent-specific gene expression and the triploid endosperm. *Am. Nat.* **134,** 147–155 (1989).
2. Smith, F. M., Garfield, A. S. & Ward, A. Regulation of growth and metabolism by imprinted genes. *Cytogenet. Genome Res.* **113,** 279–291 (2006).
3. McVean, G. T. & Hurst, L. D. Molecular evolution of imprinted genes: no evidence for antagonistic coevolution. *Proc. R. Soc. Lond. B* **264,** 739–746 (1997).
4. Grossniklaus, U., Vielle-Calzada, J. P., Hoeppner, M. A. & Gagliano, W. B. Maternal control of embryogenesis by *MEDEA*, a *Polycomb* group gene in *Arabidopsis*. *Science* **280,** 446–450 (1998).
5. Vielle-Calzada, J. P. *et al.* Maintenance of genomic imprinting at the *Arabidopsis medea* locus requires zygotic DDM1 activity. *Genes Dev.* **13,** 2971–2982 (1999).
6. Chanvivattana, Y. *et al.* Interaction of *Polycomb*-group proteins controlling flowering in *Arabidopsis*. *Development* **131,** 5263–5276 (2004).
7. Walter, J. & Paulsen, M. The potential role of gene duplications in the evolution of imprinting mechanisms. *Hum. Mol. Genet.* **12** (review issue 2), R215–R220 (2003).
8. Kiyosue, T. *et al.* Control of fertilization-independent endosperm development by the *MEDEA Polycomb* gene in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **96,** 4186–4191 (1999).
9. Goodrich, J. *et al.* A *Polycomb*-group gene regulates homeotic gene expression in *Arabidopsis*. *Nature* **386,** 44–51 (1997).
10. Baumbusch, L. O. *et al.* The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Res.* **29,** 4319–4333 (2001).
11. Blanc, G., Hokamp, K. & Wolfe, K. H. A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res.* **13,** 137–144 (2003).
12. Simillion, C., Vandepoele, K., Van Montagu, M. C., Zabeau, M. & Van de Peer, Y. The hidden duplication past of *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **99,** 13627–13632 (2002).
13. Bowers, J. E., Chapman, B. A., Rong, J. & Paterson, A. H. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422,** 433–438 (2003).
14. Springer, N. M. *et al.* Sequence relationships, conserved domains, and expression patterns for maize homologs of the *Polycomb* group genes *E(z)*, *esc*, and *E(Pc)*. *Plant Physiol.* **128,** 1332–1345 (2002).
15. Springer, N. M. *et al.* Comparative analysis of SET domain proteins in maize and *Arabidopsis* reveals multiple duplications preceding the divergence of monocots and dicots. *Plant Physiol.* **132,** 907–925 (2003).
16. Mayama, T., Ohtsubo, E. & Tsuchimoto, S. Isolation and expression analysis of petunia *CURLY LEAF*-like genes. *Plant Cell Physiol.* **44,** 811–819 (2003).

17. Thakur, J. K. *et al.* A *Polycomb* group gene of rice (*Oryza sativa* L. subspecies indica), *OsiEZ1*, codes for a nuclear-localized protein expressed preferentially in young seedlings and during reproductive development. *Gene* **314,** 1–13 (2003).
18. Ohad, N. *et al.* A mutation that allows endosperm development without fertilization. *Proc. Natl Acad. Sci. USA* **93,** 5319–5324 (1996).
19. Chaudhury, A. M. *et al.* Fertilization-independent seed development in *Arabidopsis thaliana. Proc. Natl Acad. Sci. USA* **94,** 4223–4228 (1997).
20. Guitton, A. E. *et al.* Identification of new members of *FERTILISATION INDEPENDENT SEED Polycomb* group pathway involved in the control of seed development in *Arabidopsis thaliana. Development* **131,** 2971–2981 (2004).
21. Köhler, C. *et al. Arabidopsis* MSI1 is a component of the MEA/FIE *Polycomb* group complex and required for seed development. *EMBO J.* **22,** 4804–4814 (2003).
22. Wang, D., Tyson, M. D., Jackson, S. S. & Yadegari, R. Partially redundant functions of two SET-domain *Polycomb*-group proteins in controlling initiation of seed development in *Arabidopsis. Proc. Natl Acad. Sci. USA* **103,** 13244–13249 (2006).
23. Yang, Z. & Bielawski, J. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15,** 496–503 (2000).
24. Bielawski, J. P. & Yang, Z. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J. Mol. Evol.* **59,** 121–132 (2004).
25. Schmid, K. J., Ramos-Onsins, S., Ringys-Beckstein, H., Weishaar, B. & Mitchell-Olds, T. A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics* **169,** 1601–1615 (2005).
26. Baroux, C., Gagliardini, V., Page, D. R. & Grossniklaus, U. Dynamic regulatory interactions of *Polycomb* group genes: *MEDEA* autoregulation is required for imprinted gene expression in *Arabidopsis. Genes Dev.* **20,** 1081–1086 (2006).
27. Alonso, J. M. *et al.* Genome-wide insertional mutagenesis of *Arabidopsis thaliana. Science* **301,** 653–657 (2003).
28. Adachi, J. H. M. MOLPHY Version 2.3: Programs for molecular phylogenetics based on Maximum Likelihood. *Comp. Sci. Monogr.* **28,** 1–150 (1996).
29. Yang, Z. & Nielsen, R. Synonymous and non-synonymous rate variation in nuclear genes of mammals. *J. Mol. Evol.* **46,** 409–418 (1998).
30. Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. DNAsp, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19,** 2496–2497 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Gheyselinck and P. Kopf for the technical support; C. O'Mahony for assistance with artwork and figures; M. O' Connell for comments on the manuscript; and T. Mitchell-Olds, M. Clauss, R. Oyama, J. Goodrich and NASC for seeds. This work was supported by the University of Zürich, a UNESCO fellowship (to J.-M.E.-R.), the EU Network of Excellence 'EPIGENOME', and grants of the Swiss National Science Foundation (to U.G.), the Deutsche Forschungsgemeinschaft and the Max Planck Society (to K.J.S.), and the Science Foundation Ireland (to C.S. and K.H.W.).

Author Information Sequences generated in this study are available from GenBank (accession numbers DQ975464–DQ975465). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to U.G. (grossnik@botinst.uzh.ch) or C.S. (c.spillane@ucc.ie).

METHODS

Germplasm, DNA sequencing, mutants and crosses. The *mea-1* mutant (L*er*-0) used in this study contains a *Ds* insertion in the SET domain of the *MEA* gene (At1g02580) and has been previously described[4]. Heterozygous or homozygous lines of *mea-1* were identified by PCR genotyping using the primer combinations Ds5-1/AS13 (*mea-1* present) and S20/AS13 (*MEA* present) as previously described. The *swn/eza1* allele (*swn-3*) used for the single- and double-mutant analysis of the post-fertilization seed abortion phenotype was a mutant line (SALK_050195 in Col-0 background) from the SIGNAL collection[27] that contains an insertion in exon 15 within the SET domain of the *EZA1*/*SWN* gene[6]. The *swn-4* allele used corresponds to the SIGNAL insertion line SALK_109121 where the T-DNA insertion is located in exon 8 of the *SWN* gene. The *clf-2* allele was obtained from J. Goodrich. To construct double mutants of *mea-1* and *swn-3*, pollen from a *mea-1* homozygous line was used to pollinate *swn-3* heterozygous plants and the F_1 seed progeny selected on MS medium containing kanamycin. F_1 progeny, which were double heterozygotes (*mea-1/MEA; swn-3/SWN*), were identified by PCR genotyping and selfed to generate F_2 progeny segregants. PCR genotyping was used to identify genotypes among the F_2 segregants. The *A. lyrata* (Bish Bash) accession used for crosses and sequencing of the *MEA* and *SWN* orthologues was provided by T. Mitchell-Olds. The *A. lyrata* and *A. halleri* accessions for molecular population genetic studies were provided by M. Clauss and R. Oyama.

***In situ* hybridization.** *In situ* hybridization was performed as described[5,31,32] with modifications. Mature flowers and siliques of *A. thaliana* plants were fixed in 4% paraformaldehyde and embedded in Paraplast Plus (Sigma). Sections of 10-μm thickness were cut with a Leica microtome (Leica RM 2145) mounted on ProbeOnPlus slides (Fischer Biotech). Sections were digested with proteinase K for 30 min at 37 °C, treated with acetic acid anhydride, dried in ethanol, then hybridized with 11-digoxigenin-UTP (DIG)-labelled probes overnight at 55 °C. After washing with 0.2× SSC at 55 °C, the slides were processed for revealing the DIG antigen. This involved blocking with DIG-blocking reagent and BSA, followed by incubation with an ant-DIG antibody conjugated to alkaline phosphatase (Roche Diagnostics), washing with blocking reagent, then colour revealed by incubation in NBT and X-phosphate for periods of 16 to 18 h. Reactions were stopped with TE buffer (10 mM, pH 8.0), then mounted in TE/glycerol (1 : 4 v/v) before viewing. The riboprobes used for *in situ* hydridization were synthesized from RT–PCR products using primers designed on sequence data available. The genes were *CLF* (At2g23380), *SWN* (At4g02020) and *MEA* (At1g02580). These probes were designed to cover the coding region of the genes analysed and to avoid cross hybridization. For synthesis of sense and anti-sense DIG-labelled probes, 350-bp fragments have been cloned in pBluescript SK- vector for each of the genes analysed. For hybridization probe design, the most divergent regions of *MEA*, *SWN* and *CLF* were identified by ClustalX alignment of the coding sequence of each gene. The divergent regions chosen for *in situ* probe construction were: *SWN* (355–606 bp downstream of start codon in messenger RNA); *MEA* (608–955 bp downstream of start codon in mRNA); and *CLF* (33–364 bp downstream of start codon in mRNA). The primer pairs used for amplification of these regions and cloning into the expression vector (pBluescript SK-) for riboprobe synthesis were: *SWN* (*SWN* exon 4F, 5′-GCAGAAATTTGAGGCTAATAG-3′ and *SWN* exon 4R, 5′-CCAGGTAGTGTATGGCGG-3′); *MEA* (*MEA in situ* F, 5′-CGGTTGGGCAGGACTATGG-3′ and *MEA in situ* R, 5′-CTTCTGTCACACTCCTCACC- 3′); *CLF* (*CLF in situ* F, 5′-CACCAGATCGGAGCCACC-3′ and *CLF in situ* R, 5′-GACAGGGACACTAGATCC-3′). Reverse transcription was performed using AMV reverse transcriptase (Clontech) and total RNA (1 μg) extracted from *A. thaliana* siliques using Triazol (GIBCO-BRL).

Expression analysis of *A. lyrata* paternal *MEA* allele in developing seeds. Crosses were conducted between *A. thaliana* and *A. lyrata* as a pollen parent. RNA was extracted at specific time points before or after pollination using trizol, and the accumulation of *MEA* transcripts was measured using quantitative real-time RT–PCR as previously described[26]. Quantitative analyses of transcript levels were carried out using Taqman real-time PCR assays (Applied Biosystem). Three PCR replicates were performed for each cDNA sample, and the specificity and amount of the unique amplification product were determined according to the manufacturer's instructions (Applied Biosystems). To distinguish between maternal (*A. thaliana*) and paternal (*A. lyrata*) *MEA* transcripts, we used probes that specifically recognize the different alleles. In all experiments, transcript levels were normalized to the level of *ACTIN-11*, which is expressed in the gametophyte and zygotic products of the seed (embryo and endosperm) but not in the surrounding maternal tissues[33]. Beyond 4 days after pollination (d.a.p.), *ACTIN-11* levels decrease and cannot be used for normalization (data not shown). The primers used for the real-time assay were: (1) for detection of the *MEA* allele from *A. thaliana* (*MEA-At*), forward 5′-TCTGATGTTCATGGATGGGG-3′; reverse 5′-GGTAGGAAGAACCAATCCGATCT-3′; probe VIC 5′-TCACTCATGATGAAGCTAA-3′ MGB (ABI); (2) for detection of the *MEA* allele from *A. lyrata* (*MEA-Al*), forward 5′-ATCAAGGTTGTGTTTTTAATAAAGAGGC-3′; reverse 5′-CAGCTGGCTACTTTTGATGAAGAC-3′; probe FAM 5′-ACCTTCCAGTTGTTGAGC-3′ MGB (ABI).

DNA sequencing of *MEA* and *SWN* orthologues in *A. lyrata* and *A. halleri*. The sequences of the *MEA* and *SWN* genes from *A. lyrata* (Bish Bash) were initially determined by high fidelity PCR amplification of overlapping fragments from genomic DNA and sequencing of five independent clones (in pGEM) per PCR-amplified fragment. The primer pairs used were designed to be specific to either the *MEA* or *SWN* genes. The *A. lyrata* and *A. halleri* accessions used for sequencing of *MEA* are indicated in Supplementary Table 8. The following overlapping primer pairs were used for amplification of the *MEA* gene from *A. lyrata* and *A. halleri* accessions. *MEA* ORF: MEA-F1/MEA-R1, MEA-F2/MEA-R2, MEA-F3/MEA-R3, MEA-F4/MEA-R4, MEA-F5/MEA-R5, MEA-PF/MEA-PR, MEA-P1/MEA-P2, MEA-P3/MEA-P4. The following overlapping primer pairs were used to amplify the *SWN* ORF from *A. lyrata*: SWN-F1/SWN-R1, SWN-F2/SWN-R2, SWN-F3/SWN-R4, SWN-F4/SWN-R5. The sequences of the primers are listed in Supplementary Table S9.

Phylogenetic analysis. The phylogenetic analysis was based on all *CLF*-, *MEA*- and *SWN*-like sequences obtained from BLAST searches of GenBank. The protein and gene IDs used for the tree construction were: *A. thaliana*, MEA (NP_563658/NM_100139), SWN/EZA1 (AAL90954/AY090293), CLF (AAC23781/AC003040); *A. lyrata*, MEA (bankit835839), SWN/EZA1 (bankit842314); *Zea mays*, MEZ1 (AAM13420/AF443596), MEZ2 (AAM13421/AF443597), MEZ3 (AAM13422/AF443598); *Oryza sativa*, SET1 (AAN01115/AF407010), EZ1 (*O. s. indica*) (CAD18871/AJ421722), CLF (*O. s. japonica*) (NP_910690/NM_185801), EZ1 (*O. s. japonica*) (BAD69169/AP005813); Petunia x Hybrida, PHCLF1 (BAC84950/AB098523), PHCLF2 (BAC84951/AB098524), PHCLF3 (BAC84952/AB098525). The poplar sequences were obtained by searching the poplar genome database at the JGI (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html) (protein IDs: EZA1-like1, 731686; EZA1-like2, 348349; CLF-like1, 719252; CLF-like2, 694432). Protein sequences were aligned with the CLUSTALW program and a phylogenetic tree was constructed with the protml program of the MOLPHY package, which implements a maximum likelihood method[28]. The 'quick add OTUs search' strategy was used in with the JTT substitution matrix and the six best trees were retained. Subsequently, each tree was swapped and re-optimized (local rearrangement search), and branch lengths and local bootstrap probabilities (LBP) were estimated during this last search, leading to the tree with the highest likelihood shown in Fig. 1c.

Evolutionary analysis and tests of natural selection. The ratio $\omega = d_N/d_S$—with d_N as the number of non-synonymous substitutions per non-synonymous site and d_S as the number of synonymous substitutions per synonymous site—was used to test whether protein-coding sequences evolve under positive darwinian selection[34]. The analyses were carried out with the codeml program of the PAML package[29], which uses maximum likelihood estimation of parameters. To test for positive or purifying selection, ω-ratios for different site classes of the coding sequence were estimated. The likelihood of this estimate was then compared with the likelihood of other models with different numbers of parameters. A likelihood ratio test was applied by calculating the test statistic as the twofold difference of the two likelihoods ($2\delta = 2(l_1 - l_2)$); the critical values were looked up in a χ^2 table with the degrees of freedom calculated as the difference of the parameters that were estimated by each model. Two types of models were analysed. Branch models allow different ω-ratios in different branches of the phylogeny and can be used to address whether selection pressures on proteins are variable among species or among paralogues of a gene family. We used branch models with one (Model M0), two, three, four or n (the total number of branches; free-ratio model) ω-ratios[35]. The second type of models analysed were branch-site models[34]. These models allow one to test whether positive selection occurred in a subset of codons in a particular branch of the phylogeny (the 'foreground branch') by assuming two types of codons with $0 < \omega < 1$ and $\omega = 1$ in the entire tree and an additional class of codons with $\omega > 1$ in the foreground branch. After estimating ω-ratios, a bayesian empirical Bayes algorithm was applied to identify amino acid residues with a high posterior probability of $\omega > 1$. Among available branch-site models, we used model A of ref. 35 to test whether functional diversification of newly duplicated genes was driven by positive selection.

Sequence analysis of divergent accessions of *A. thaliana*, *A. lyrata* and *A. halleri*. Sequences of *SWN* and *MEA* genes were amplified from 21 divergent accessions of a worldwide collection of *A. thaliana* (Supplementary Table 7) using overlapping PCR primers (Supplementary Table 9). PCR products were directly sequenced on an ABI 3730xl automated sequencer using dye terminator chemistry. Sequence data were assembled and aligned with an automated sequence analysis pipeline as described[36]. The *MEA* gene was also obtained from

nine individuals originating from geographically distant populations of *A. lyrata* and from ten individuals from the close relative *A. halleri* (Supplementary Table 8). High fidelity PCR amplifications were performed using the Phusion High-Fidelity PCR Kit (FINNZYMES) and the overlapping PCR primers described for the *A. lyrata* sequencing. A first set of PCR amplifications was directly sequenced, and a second set of independent PCR amplifications was cloned before sequencing. Briefly, following an A-tailing step, the PCR products were cloned into the pGEM-T easy vector following the manufacturer's instructions (Promega). For each PCR fragment, two independent clones were sequenced in both directions (Macrogen Inc.). All sequences generated for the *MEA* gene were analysed using the DNASTAR software package (DNASTAR). All polymorphisms were inspected visually. Molecular population genetic analysis was carried out with the DnaSP program[30]. Nucleotide diversity was calculated as the average pairwise nucleotide diversity, π_{tot}, and haplotype diversity as H_d[37]. Several tests of neutral evolution using the polymorphism data where applied. Tajima's D[38] tests whether there is an excess of low- or high-frequency polymorphisms and was calculated with silent (synonymous coding and non-coding) polymorphisms; the H statistic of ref. 39 analyses the frequency spectrum of derived polymorphisms and was also calculated with silent polymorphisms; the McDonald–Kreitman test[40] compares the ratio of non-synonymous to synonymous polymorphisms and fixed differences; the Hudson–Kreitman–Aguade test was used to test whether the polymorphism to divergence ratio between the two genes was significantly different from each other, which is expected if selection acts on one gene but not on the other[41].

31. Coen, E. S. *et al. floricaula*: a homeotic gene required for flower development in *Antirrhinum majus*. *Cell* **63,** 1311–1322 (1990).
32. Jackson, D., Culianez-Macia, F., Prescott, A. G., Roberts, K. & Martin, C. Expression patterns of *myb* genes from *Antirrhinum* flowers. *Plant Cell* **3,** 115–125 (1991).
33. Huang, S., An, Y. Q., McDowell, J. M., McKinney, E. C. & Meagher, R. B. The *Arabidopsis ACT11* actin gene is strongly expressed in tissues of the emerging inflorescence, pollen, and developing ovules. *Plant Mol. Biol.* **33,** 125–139 (1997).
34. Yang, Z. & Nielsen, R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17,** 32–43 (2000).
35. Yang, Z. & Nielsen, R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19,** 908–917 (2002).
36. Schmid, K. J., Ramos-Onsins, S., Ringys-Beckstein, H., Weisshaar, B. & Mitchell-Olds, T. A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics* **169,** 1601–1615 (2005).
37. Nei, M. *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York, 1987).
38. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123,** 585–595 (1989).
39. Fay, J. C. & Wu, C.-I. Hitchhiking under positive Darwinian selection. *Genetics* **155,** 1405–1413 (2000).
40. McDonald, J. & Kreitman, M. Adaptive evolution at the *Adh* locus in *Drosophila*. *Nature* **351,** 652–654 (1991).
41. Hudson, R. R., Kreitman, M. & Aguade, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116,** 153–159 (1987).