Adaptive reduction of male gamete number in the selfing plant *Arabidopsis thaliana*

Takashi Tsuchimatsu1,2,3,4,5,14, Hiroyuki Kakui1,6,7,14, Misako Yamazaki1, Cindy Marona8, Hiroki Tsutsui2,9,10, Afif Hedhly2, Dazhe Meng3,11, Yutaka Sato12, Thomas Städler13, Ueli Grossniklaus2, Masahiro M. Kanaoka9, Michael Lenhard8, Magnus Nordborg3 & Kentaro K. Shimizu1,2,6✉

The number of male gametes is critical for reproductive success and varies between and within species. The evolutionary reduction of the number of pollen grains encompassing the male gametes is widespread in selfing plants. Here, we employ genome-wide association study (GWAS) to identify underlying loci and to assess the molecular signatures of selection on pollen number-associated loci in the predominantly selfing plant *Arabidopsis thaliana*. Regions of strong association with pollen number are enriched for signatures of selection, indicating polygenic selection. We isolate the gene *REDUCED POLLEN NUMBER1* (*RDP1*) at the locus with the strongest association. We validate its effect using a quantitative complementation test with CRISPR/Cas9-generated null mutants in nonstandard wild accessions. In contrast to pleiotropic null mutants, only pollen numbers are significantly affected by natural allelic variants. These data support theoretical predictions that reduced investment in male gametes is advantageous in predominantly selfing species.

---

1Department of Evolutionary Biology and Environmental Studies & Zurich-Basel Plant Science Center, University of Zurich, 8057 Zurich, Switzerland. 2Department of Plant and Microbial Biology & Zurich-Basel Plant Science Center, University of Zurich, 8008 Zurich, Switzerland. 3Gregor Mendel Institute, Austrian Academy of Sciences, Vienna BioCenter, A-1030 Vienna, Austria. 4Department of Biology, Chiba University, Chiba 263-8522, Japan. 5Department of Biological Sciences, University of Tokyo, Tokyo 113-0033, Japan. 6Kihara Institute for Biological Research, Yokohama City University, Yokohama 244-0813, Japan. 7Graduate School of Science and Technology, Niigata University, Niigata 950-2181, Japan. 8Institute for Biochemistry and Biology, University of Potsdam, 14476 Potsdam, Germany. 9Graduate School of Science, Nagoya University, Nagoya 464-8602, Japan. 10JST ERATO Higashiyama Life-Holonics Project, Nagoya University, Nagoya 464-8602, Japan. 11Molecular and Computational Biology, University of Southern California, Los Angeles, CA 90089-0371, USA. 12Department of Genomics and Evolutionary Biology, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan. 13Institute of Integrative Biology, ETH Zurich, 8092 Zurich, Switzerland. 14These authors contributed equally: Takashi Tsuchimatsu, Hiroyuki Kakui. ✉Email: kentaro.shimizu@ieu.uzh.ch
Male gamete numbers, reflected by pollen grain (each containing two sperm cells) numbers in seed plants and sperm numbers in animals, have been studied extensively from agricultural, medical, and evolutionary viewpoints. Evolutionary theory predicts that the breeding system could act as a major selective force on male gamete numbers. In highly promiscuous outcrossing species, a large number of male gametes should be produced because of male–male gamete competition, so reduced male gamete numbers are considered to be deleterious. In contrast, they may be advantageous at lower outcrossing rates because of the high cost of their production, decreasing fitness. In an agricultural context, low pollen number may have been selected during domestication, but may serve as a barrier for hybrid breeding of wheat and other species. In flowering plants, the transition from an outcrossing to a selfing breeding system through loss of self-incompatibility is one of the most prevalent evolutionary trends. Selfing populations or species generally show lower pollen grain numbers per flower (hereafter, pollen number) as well as reduced flower size. There has been a sustained debate on whether the reduced pollen number is a result of adaptive evolution or the accumulation of deleterious mutations owing to reduced selection, but little was known about the genetic basis of pollen number variation to assess molecular signatures of selection on it.

To unravel the genetic basis of quantitative natural variation in pollen number, we here focus on the predominantly selfing plant Arabidopsis thaliana. Studies have shown that the evolution of predominant selfing in A. thaliana occurred much more recently than its evolutionary divergence from outcrossing relatives. Thus, in addition to fixed, genetically based differences from these outcrossing relatives, we expect that variation with regard to pollen number may still be segregating among current accessions. By harnessing the genetic and genomic resources available in A. thaliana, we conduct a genome-wide association study on pollen number variation. We find that natural variants of the RDP1 gene confer variation in pollen number without detectable pleiotropy. Signatures of selection at the top genome-wide association study (GWAS) peaks, including the RDP1 locus, support the theoretical prediction that reduced investment in male gametes should provide an advantage in selfing species.

Results

Genome-wide association study and signatures of selection. To examine variability in pollen number on a species-wide scale, we determined pollen number per flower in 144 natural A. thaliana accessions (Fig. 1a–d; Supplementary Table 1 and Supplementary Data 1) and found approximately fourfold variation (average ~4000) (Fig. 1e). Histological sections of stamens from representative accessions confirmed pollen number variation among accessions (Fig. 1c, d). We also measured the number of oocytes per flower (Supplementary Table 2). We did not find significant correlations between numbers of pollen grains and oocytes (P = 0.5164), although negative correlations have often been reported in between-species comparisons, as expected on theoretical grounds owing to trade-offs in resource allocation to male versus female function. Furthermore, we found that pollen number per flower was not significantly correlated with any of the 107 published phenotypes of flowering, defense-related, ionomic, and developmental traits (Supplementary Table 3; Supplementary Note 1) nor with climate variables, geographic location, or haplogroups across the 144 accessions (Supplementary Tables 4 and 5, Supplementary Fig. 1). These data suggest that variation in pollen number is largely independent of other traits.

To evaluate genome-wide signatures of natural selection on loci associated with gamete numbers, we first performed GWAS for pollen and ovule numbers using a genome-wide single-nucleotide polymorphism (SNP) data set for these lines, which was obtained by imputation based on genome-wide resequencing data and 250 k SNP data (Fig. 1f, h; Supplementary Fig. 2). In total, 68 peaks of association were identified (10-kb windows having SNPs with P < 10⁻⁵), although only one pollen number-associated peak remained significant after Bonferroni correction. Focusing on the identified GWAS peaks, we performed an enrichment analysis to ask whether pollen and ovule number-associated peaks are enriched in long-haplotype regions, which could be owing to partial or ongoing sweeps of segregating polymorphisms. To identify long-haplotype regions, we first calculated the extended haplotype homozygosity (EHH), which measures decay of haplotypes that carry a specified core allele as a function of distance. We then obtained the integrated haplotype score (iHS) statistic for each SNP, which compares EHH between two alleles of the SNP by controlling for the allele frequency of each SNP. We found that 10-kb windows including pollen number-associated loci were significantly enriched in extreme iHS tails (P < 0.05; permutation test; Fig. 1i, j; Supplementary Fig. 3). These loci showed generally high iHS scores, and two of the top five GWAS peaks were outliers of the genome-wide iHS distribution (Supplementary Table 6). The enrichment was robust to changes in sample composition, allele frequency cutoffs, and the use of windows (Supplementary Figs. 4–6; see Supplementary Note 2 for details). Ovule number also showed enrichment, albeit less than pollen number (Fig. 1j). In principle, the iHS enrichment could be confounded by recombination rate and the accuracy of imputation. To deal with such potential confounding factors, we compared these results with the results of an iHS enrichment analysis for GWAS peaks (P < 0.0001) for 107 other phenotypes, as these confounding factors should also influence the enrichment for other traits. We found that the iHS enrichment for pollen number GWAS peaks (P = 0.002 for the top 1% iHS tail; Supplementary Fig. 3) was among the highest, compared with that for many known adaptive traits included in the 107 phenotypes, such as leaf number at flowering time and resistance to Pseudomonas pathogens (Supplementary Table 7). In addition, iHS enrichment of the ovule number GWAS peaks was also significant (P = 0.030 for the top 1% iHS tail; Supplementary Fig. 3). These enrichments support polygenic selection on a considerable number of loci associated with male and female gamete numbers throughout the genome.

Isolation of the REDUCED POLLEN NUMBER1 gene. To further understand the molecular basis of pollen number variation and to examine the nature of the putative targets of selection, we tried to identify the genes underlying pollen number variation; however, the top five peaks of association did not contain any genes with known functions in early stamen or pollen development. To obtain experimental evidence concerning genes underlying pollen number variation, we conducted functional analyses of the genes under the highest pollen number GWAS peak, which explains ~20% of the total phenotypic variance between accessions and satisfies the criterion for genome-wide significance (−log₁₀(P) = 7.60). This region is of particular interest because it also satisfies the criteria for genome-wide significance of the iHS statistic (P = 0.0149; Fig. 1i, Supplementary Fig. 7), suggesting a selective sweep. To test whether the signature of selection in this region might be owing to traits other than pollen number, we examined whether there is an association signal for any of the 107 published phenotypes, ovule number, or variants showing climatic correlations. In the 10-kb window including the SNP of the highest GWAS score for pollen number, we found no genotype–phenotype associations below P < 10⁻⁵ or
Fig. 1 Genome-wide association study of pollen number variation in *Arabidopsis thaliana*. a, b Pollen grains of higher pollen number accession Bor-4 (a) and lower pollen number accession Mz-0 (b) mounted on glass slides for counting. c, d Histological sections of Bor-4 (e) and Mz-0 (d) stamens. At least three independent observations showed similar results (a–d). Scale bars = 50 μm. e Distribution of pollen number variation across 144 natural accessions. f Manhattan plot of the genome-wide association study (GWAS). g Closer view of the region around the significant GWAS peak on chromosome 1 with gene models and coordinates. f, g SNPs with minor allele frequency > 0.15 are shown; horizontal dashed lines indicate the nominal $P < 0.05$ threshold after Bonferroni correction. h Quantile–quantile plot of the GWAS. i Extended haplotype homozygosity (EHH) detected in the RDP1 genomic region. Red and blue lines correspond to the long haplotype and alternative variants, respectively. j Signatures of selection at pollen number-associated loci. Each line indicates a phenotype (red: pollen number, black: ovule number, gray: 107 phenotypes). The x axis quantifies the extreme tails of the integrated haplotype score (iHS) statistic. The pollen and ovule GWAS show significant enrichment (permutation test, $P < 0.05$ cutoff for all iHS statistical tails; Supplementary Fig. 3). k Accessions with the long-haplotype variants (defined by SNP 1-8852112) generally showed lower pollen number ($P = 2.152 \times 10^{-6}$, two-sided t test; population structure-corrected GWAS $P = 2.95 \times 10^{-6}$). Boxplots show center line: median; box limits: upper and lower quartiles; whiskers: not >1.5 times the interquartile range; dots: outliers. Source data underlying e, j and k are provided as a Source Data file.
climatic–SNP correlations below an empirical $P < 0.01$, i.e., there is no significant evidence for selection on traits other than pollen number in this region. We also found that accessions with the long-haplotype variants produced lower pollen numbers than those with alternative haplotype variants ($P = 2.152 \times 10^{-6}$, $t$ test; population structure-corrected GWAS $P = 2.95 \times 10^{-5}$; Fig. 1k), as expected if this haplotype was under selection for reduced pollen number.

Of the three genes in this chromosomal region with the highest GWAS scores (AT1G25250, AT1G25260, and AT1G25270; Fig. 1g), the expression level of AT1G25260, a gene of unknown function, was much higher in an accession from the Nottingham Arabidopsis Stock Centre26. These mutants showed a 32% reduction in pollen number (Fig. 2a; Supplementary Table 8). We therefore refer to AT1G25260 as REDUCED POLLEN NUMBER1 (RDP1). Because both $rdp1-1$ (insertion in the 5' UTR) and $rdp1-2$ (insertion at the end of the coding sequence) (Supplementary Fig. 8). We generated two amorphic (null) frameshift mutants of RDP1 ($rdp1-3$ and $rdp1-4$) using the CRISPR/Cas9 system27,28 in the Col-0 background. These mutants indeed showed an even greater reduction in pollen number, but still produced about half the number of pollen grains of the corresponding wild type (53% for $rdp1-3$; Fig. 2a), suggesting a quantitative nature of the effect of RDP1. Pollen size was slightly increased, in agreement with the well-known negative relationship between pollen number and size, even within the same genotype (Supplementary Fig. 9; Supplementary Table 9)14. The mutant phenotype was complemented by transforming a 4.3-kb genomic fragment of the Col-0 accession encompassing RDP1 (Fig. 2a, Supplementary Fig. 9). In contrast to RDP1 mutants, CRISPR/Cas9 induced null mutants in AT1G25250 or AT1G25270 did not result in any significant change in pollen number (Supplementary Fig. 10). The phenotype of four independent mutants of RDP1 together with successful complementation using the wild-type allele thus demonstrated that RDP1 is involved in the control of pollen number.

Based on phylogenetic analysis, RDP1 is a putative homolog of the yeast mRNA turnover 4 protein (Mrt4p) (Supplementary Fig. 8).
Figs. 11 and 12). The MRT4 gene is nonessential in yeast but null mutants show a phenotype of slightly slower growth. The Mrt4p shares similarity with the ribosome P0 protein and is necessary for the assembly of the P0 protein into the ribosome. The human ribosome P0 gene is reported to have an extra-ribosomal function in cancer by modulating cell proliferation. During anther development, sporogenous cells first divide and differentiate into microsporocytes. Following meiosis of the microsporocyte, four microspores are formed, each of which undergoes two mitotic divisions to form a mature pollen grain containing the male gametes. The null rdp1-3 mutant produced fewer microsporocytes than the wild type (Fig. 2c–e), indicating a reduction in cell numbers before meiosis. Consistent with this, in situ mRNA hybridization experiments detected strong expression of RDP1 in sporogenous cells and the microsporocytes derived from them, but not in microspores (Fig. 2f, g; Supplementary Fig. 13). RDP1 was also expressed in other proliferating cells, including those in inflorescences, floral meristematic regions, and ovules (Supplementary Figs. 8 and 13), supporting a more widespread role in ribosome biogenesis. Furthermore, fusing the RDP1 promoter to the uidA reporter gene encoding β-glucuronidase (GUS) to assess its activity confirmed the RDP1 expression pattern in stamens (Supplementary Fig. 14) and demonstrated marked expression in root tips and young leaf primordia during the vegetative phase; these data are supported by quantitative reverse transcription PCR experiments (Supplementary Fig. 8b). Consistent with RDP1 expression in proliferating tissues, the rdp1-3 null mutant showed pleiotropic phenotypes, including slower vegetative growth and reduced ovule numbers per flower (Supplementary Fig. 15). Because these pleiotropic phenotypes would be deleterious in natural environments, these data indicate that natural alleles of RDP1 are not null variants (see below). In summary, these data suggest that RDP1 is required in proliferating A. thaliana cells, yet the natural variants we have identified predominately affect the proliferation of sporogenous cells in the anthers, consistent with the function of its yeast homolog in cell proliferation.

Natural variants of RDP1 confer pollen number variation. It has been difficult to experimentally determine whether a particular gene has natural alleles with subtle phenotypic effects on quantitative traits. When allelic effects are subtle, transgenic analysis of natural alleles is not sufficiently powerful because the phenotypes of transgenic A. thaliana plants tend to be highly variable as a result of the variation between lines, e.g., owing to different transgene insertion sites. In contrast, a quantitative complementation test can identify responsible genes by testing the effect of natural alleles in a heterozygous state with a null allele if the effects of other loci are small, although this may be confounded by polygenic effects in the genetic background. To conduct such quantitative complementation, we took advantage of the CRISPR/Cas9 technique to generate frameshift null alleles in nonstandard natural accessions, in which no prior mutant was available.

We used Bor-4 and Uod-1, which have high and low pollen number phenotypes, respectively (Fig. 2b; Fig. 3). There was a number of sequence differences between the two accessions in the region encompassing the RDP1 gene from 777 bp upstream of the start codon to 643 bp downstream of the stop codon. We found one non-synonymous and six synonymous substitutions in the coding region, and 62 substitutions and six indel mutations in the non-coding region (Supplementary Fig. 16). Yet, both accessions did not reveal obvious loss-of-function mutations (Supplementary Fig. 16), and rdp1 CRISPR null mutants of each accession showed reduced pollen number compared with the corresponding wild type (P < 2.2 × 10^{-7} for Bor-4; P = 9.84 × 10^{-7} for Uod-1; Fig. 3a, b; Fig. 2h). These results show that both of the naturally occurring variants of the RDP1 gene are not null mutants but rather encode a functional protein. Disruption of RDP1 had a stronger effect on pollen number in Bor-4 than in Uod-1 (analysis of variance (ANOVA) interaction effect P = 1.07 × 10^{-5}; Fig. 2h). This finding supports the notion that the Bor-4 allele has a stronger promotive effect on pollen number than the Uod-1 allele, although other loci in the genetic backgrounds of Bor-4 and Uod-1 may contribute to this difference through epistasis.

To test the allelic effect of RDP1, we utilized a quantitative complementation test that controls for genetic background (Fig. 3). Among F1 plants obtained by crossing heterozygotes for the frameshift mutation in each genetic background, we compared two genotypes: RDP1Bor/rdp1Uod vs. rdp1Bor/RDP1Uod. These F1 genotypes are identical except for the differences at RDP1, where they both carry a frameshift allele but differ with respect to the functional allele; because of the crossing design, any independently segregating off-target effects, resulting from CRISPR/Cas9 mutagenesis, would be equally distributed between the two genotype cohorts of interest. We found that pollen number in plants with RDP1Uod was significantly lower than in plants with RDP1Bor (nested ANOVA, 468 flowers from 26 individuals of RDP1Bor/rdp1Uod and 368 flowers from 20 individuals of rdp1Bor/RDP1Uod; P = 4.85 × 10^{-8}; Fig. 3c). The significant difference cannot be attributed to stochastic individual differences, because the significant difference between plants bearing functional RDP1Bor and RDP1Uod alleles was also observed in an individual-based test using averaged data of each individual separately (P = 0.0331). Thus, in an otherwise identical genetic background, the respective functional RDP1 haplotypes cause differences in pollen number. We also measured rosette leaf size, flowering date, ovule number, dry weight, and seed weight of plants in the two cohorts, but none of these traits showed significant differences between RDP1Bor/rdp1Uod and rdp1Bor/RDP1Uod (Supplementary Fig. 17). Thus, in contrast to the experimentally generated null mutants that showed pleiotropic growth defects, these results indicate that natural allelic differences at RDP1 affect pollen number in the absence of any detectable deleterious pleiotropy. This finding is also supported by no genotype–phenotype associations for other traits, as described above.

Discussion

We here isolated the RDP1 gene underlying natural variation in male gamete numbers. Our study provides evidence for polygenic selection on pollen number-associated loci, including RDP1. Even though RDP1 encodes a ribosome-biogenesis factor that would be required globally for proliferative growth, the naturally selected alleles predominantly confer reduced pollen number. This is analogous to a hypomorphic allele of the human G6PD gene, which encodes an enzyme in the pentose phosphate pathway and features a long haplotype because of selection for malaria resistance.
Thus, in addition to presumably fixed differences to distantly related outcrossing congeners, it is quite conceivable that some underlying loci, not limited to but including RDP1, are still segregating within A. thaliana. These might reflect an ongoing selection process for the further reduction in pollen number, which has been considered a hallmark of the so-called selfling syndrome. Although we note the possibility that partial sweeps of RDP1 and other segregating loci are not directly related to the transition to predominant selfling, our analysis did not find evidence of other selective forces including local adaptation, climate association, or pleiotropic selection on other traits. Therefore, our study supports the theoretical predictions that reduced investment in male gametes is advantageous in predominantly selfling species.

Our work also illustrates that a combination of GWAS and functional analysis using a quantitative complementation test based on the CRISPR/Cas9-based alleles provides a powerful approach to dissect allelic differences underlying quantitative natural variation.

**Methods**

**Pollen and ovule counting for genome-wide association studies.** To perform GWAS, numbers of pollen grains and ovules per flower were counted for 144 and 151 world-wide natural accessions, respectively (Supplementary Tables 1, 2 and Supplementary Data 1). Plants were grown at 21 °C under a 16 h light/8 h dark cycle without vernalization. We grew four plants per accession. Three flower buds per plant were harvested from the main inflorescence, and each flower bud was collected into a 1.5 mL tube and dried at 65 °C overnight. We sampled individual flower buds of young main inflorescences but avoided the first and second flowers of the inflorescence because these flowers tend to show developmentally immature morphologies. We collected flower buds with mature pollen but before the anthers were opened (flower stage 13), and added 30 μL of 5% Tween 20 (Sigma-Aldrich, St. Louis, MO, USA) to each tube. The tubes were sonicated using a Bioruptor (Diagenode, Seraing, Belgium) in high power mode with 10 cycles of sonication-ON for 30 s and sonication-OFF for 30 s so that the pollen grains were released from the anther sacs. After a short centrifugation and vortexing, 10 μL of the solution was mounted on a Neubauer slide. We took three images per sample using a light microscope. The number of pollen grains per image was counted using a computer program developed in our lab.

**Quantitative complementation test of the RDP1 gene.** Violin plots with means and standard errors of means indicated by red bold bars and boxes, respectively. **a**, **b** Pollen number differences between wild-type and homozygous plants of a frameshift allele generated by the CRISPR/Cas9 technique in the bor-4 background (a Numbers of flowers pollen-counted: RDP1bor/RDP1bor, N = 89; rdp1bor/rdp1bor, N = 77) and in the uod-1 background (b Numbers of flowers pollen-counted: RDP1uod/RDP1uod, N = 47; rdp1uod/rdp1uod, N = 43) (same data set with Fig. 2h). **c** The difference in the effect on pollen number by two natural alleles, RDP1bor and RDP1uod. Pollen number of plants with RDP1uod was significantly lower than that of plants with RDP1bor (nested analysis of variance; P = 4.85 × 10−8; Numbers of flowers pollen-counted: RDP1bor/rdp1uod, N = 468 from 26 individuals; rdp1bor/RDP1uod, N = 368 from 20 individuals). The two alleles were compared in the heterozygous state with a frameshift CRISPR/Cas9 allele, with otherwise identical genetic backgrounds. F1 plants were obtained from the cross of two heterozygotes, RDP1bor/rdp1bor and RDP1uod/rdp1uod. Source data are provided as a Source Data file.
particle counting implemented in ImageJ (http://imagej.nih.gov/ij/) and in Fiji (http://fiji.sc/Fiji). We then estimated the total pollen number per flower based on the image size and the total volume. Ovule numbers were counted by dissecting young siliques (3.5 siliques per accession on average) under a dissecting microscope.

Because of limited chamber space, we split the plants into two batches. The two batches were treated under the same conditions in the same chambers, but at different times. We controlled for this potential batch effect for pollen number by setting equal medians and standard deviations for the two batches and used as the GWAS input of pollen number phenotype (Source Data file, Supplementary Table 1).

Sometimes, there were no or very few pollen grains per image. This was mainly in situations where anthers did not open. To eliminate these artefacts, we discarded flowers with pollen counts of <10 per image. We confirmed that such extremely low pollen numbers did not occur in specific accessions, indicating that this is not heritable.

Plant materials and growth conditions for functional analyses. For functional analyses, we mainly used Arabidopsis thaliana wild-type and mutant plants of the Col-0, rdp1-5, and Uod-1 accessions. The T-DNA lines SALK_064854/N666274 (rdp1-1) from the Salk collection41 and Gt-89709/G484369 (rdp1-2) from GABI-Kat collection42 were obtained from the European Arabidopsis Stock Centre46. The T-DNA lines SALK_064854/N666274 (rdp1-1) from the Salk collection41 and Gt-89709/G484369 (rdp1-2) from GABI-Kat collection42 were obtained from the European Arabidopsis Stock Centre46.

We generated single-nucleotide insertion/deletion lines in the Col-0, Uod-1, and Bor-4 accessions using the FAST-CRISPR-Cas9 construct (see the section CRISPR mutant) and designated them rdpi-3 and rdpi-4 (in Col-0), rdpi-5 (in Uod-1), and rdpi-6 (in Bor-4).

Arabidopsis seeds were sown on soil mixed with the insecticide ActaraG (Syngenta Agro, Switzerland) and stratified for 3–4 days at 4°C in the dark. The plants were grown under 16 h of light at 22°C, and 8 h of dark at 20°C, with weekly treatments of insecticide (Kendo Gold, Syngenta Agro) unless noted otherwise (for GWAS, see above).

Statistical analysis. Unless stated otherwise, statistical and population genetic analyses were performed using the statistical software R43. In boxplots, bars indicate the median, boxes indicate the interquartile range, and whiskers extend to the most extreme data point that is not >1.5 times the interquartile range from the box, with outliers shown by dots.

Histological analysis of anthers. For histological analysis, inflorescences were fixed with formaldehyde:acetic acid:70% ethanol (FAA) and dehydrated through an ethanol series. Fixed samples were embedded in Technovit 7100 according to the manufacturer instructions (Heraeus Kulzer GmbH, Wehrheim, Germany). Five-micrometer sections were cut with a microtome (RM2145, Leica, Germany) and stained with toluidine blue before observation under a Leica microscope (DM5000, Leica) equipped with a black-and-white camera (DFC345, Leica).

Correlations with published GWAS results and climatic data. To examine whether pollen and ovule numbers were correlated with any of the other 107 published phenotypes19, or with climate and geographic variables20, Pearson correlations with published GWAS results and climatic data were conducted as described above. We calculated the full length of RDPI and the surrounding genomic sequence that may contain promoter and terminal regions (ca. 4.3 kb, positions 8857,723 to 8853,420 on chromosome 1). The RDPI promoter sequence was amplified by PCR using the following primers: 3550_At1g25260R (5′-GTTTAAATGAGAGAACCCG-3′) and 3551_At1g25260F1 (5′-GTTTTAAATGAGAGAACCCG-3′) to amplify the full length of RDPI and the surrounding genomic sequence that may contain promoter and terminal regions. For the former two positions, we sampled during the first 3 weeks of flowering and excluded the first and second flowers on each branch; this yielded up to 40 flowers per individual. Collecting, suspending, and sonicking of flowers for GWAS were conducted as described above. All pollen solutions were suspended in 10 mL of 1N NaOH for 3 h and then stained with aniline blue solution (0.1% aniline blue, 100 mM NaOH) for 3 h. Anthers at the microspore stage were dissected under a microscope with UV illumination (DM5000, Leica, Germany). Z-stack images were obtained with a confocal microscope (SP5, Leica, Germany).

To expedite pollen number counting, we established a rapid method using a cell counter (CASY TT, OMNI Life Science GmbH, Germany)52. We found that the pollen numbers of flowers on side inflorescences and side branches of the main inflorescence were similar, but those of flowers on the main inflorescence were higher (Supplementary Fig. 18). To obtain a large number of replicates, we sampled flowers from the former two positions. We sampled during the first 3 weeks of flowering and excluded the first and second flowers on each branch; this yielded up to 40 flowers per individual. Collecting, suspending, and sonicking of flowers for GWAS were conducted as described above. All pollen solutions were suspended in 10 mL of 1N NaOH for 3 h and then stained with aniline blue solution (0.1% aniline blue, 100 mM NaOH) for 3 h. Anthers at the microspore stage were dissected under a microscope with UV illumination (DM5000, Leica, Germany). Z-stack images were obtained with a confocal microscope (SP5, Leica, Germany).
were embedded in paraplast using an embedding machine (ASP200, Leica, Germany). RDP1 cDNA was PCR-amplified using the primer pair (At1g25520gF; 5'-tgtgcgaagcgctggagc-3' and At1g25260gR; 5'-taggcagctgtctgaaagtc-3') and cloned into pCR4-TOPO (Thermo Fisher Scientific) vector. Cloned cDNA was used as a template for in vitro transcription using a MAXScript T7 labeling kit (Thermo Fisher Scientific) for hybridization24.

**GUS assays.** Plant samples were incubated in 90% acetone for 20 min at room temperature, washed with 50 mM phosphate buffer containing 0.1% Triton X-100, 2 mM potassium ferrocyanide, and 2 mM potassium ferricyanide, and incubated in the same buffer supplemented with 1 mg mL⁻¹ X-Gluc for 5 h at 37 °C.25

**Phylogenetic analysis.** Multiple sequence alignment was performed using ClustalW implemented in the CLC Workbench (version 7.7). A phylogenetic tree was generated by the neighbor-joining distance algorithm, using the aligned region (amino acid positions 44-153) and a bootstrap value of 1000. Ysh1 of Saccharomyces cerevisiae was used as an outgroup. Accession numbers of used sequences are listed in Supplementary Table 11.

**Selection scan.** For the selection scan, we used the imputed SNP data set that was also used for GWAS. We used 298 accessions (Supplementary Data 1), covering all the accessions used for our GWAS of pollen and ovule numbers and the GWAS of 107 phenotypes reported by Awetl et al. We used the iHS statistics for the selection scan; this statistic compares the EHH between two alleles by controlling for the allele frequency of each SNP. The iHS statistic uses the contrast of EHH values on each SNP; iHS values strongly deviate from zero when one allele has a high haplotype (high EHH) and the other has a short one (low EHH); the R library iHS was used to calculate the iHS statistic. The Arabidopsis lyrata reference genome was also used to infer the ancestral state for each SNP. We performed 1000 times.

**Measuring plant phenotypes.** For measuring rosette leaf size, we took images of plants that included a ruler at 3 weeks after germination. A minimum circumference of circle was drawn manually on the picture using Fiji; then, the area was measured and transformed depending on the scale. The flowering date was counted as days from sowing to flowering. The dry weight of plants was measured using aerial parts, and seed weight was determined by collecting seeds from dried plants. P values for quantitative comparison test (Supplementary Fig. 17) are shown in Supplementary Table 12.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

**Data availability**

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A Reporting Summary for this Article is available as a Supplementary Information file. The data sets generated and analyzed during the current study are available from the corresponding author upon request. RDP1 gene sequence data generated in this article were registered in GenBank (National Center for Biotechnology Information) databases under the following accession numbers: LC164158 (Mr-0), LC164159 (Bor-4), LC504218 (Uod-1), LC164160 (A. lyrata), and LC164161 (Arabidopsis halleri). PO gene sequence data of A. halleri generated in this Article were registered in GenBank databases under the following accession numbers: LC164162 and LC164163. Raw and processed sequencing data for GWAS and population genetic analyses are publicly available at https://doi.org/10.5061/dryad.jhbwobt7z (ref. 166). The source data underlying Figs. 1e, 1j, 1k, 2a, 2e, 2h, and 3, as well as Supplementary Figs. 1, 3–6, 7a, 7d, 8, 9, 10c, 10d, 15, 17, and 18 are provided as a Source Data file.

**Code availability**

Code used for selection enrichment analysis is publicly available at https://github.com/ttsuchimatsu/arabidopsis_pollen_number.

Received: 9 February 2018; Accepted: 19 May 2020;
Published online: 08 June 2020

**References**

1. Harvey, P. H. & May, R. M. Out for the sperm count. Nature 337, 508–509 (1989).
2. Birkheads, T. R. & Moller, A. P. Spenn Competition and Sexual Selection (Academic Press, 1998).
3. Merzenich, H., Zeeb, H. & Blettner, M. Decreasing sperm quality: a global problem? BMC Public Health 10, 24 (2010).
4. Czeizel, A. E. & Rothman, K. J. Does relaxed reproductive selection explain the decline in male reproductive health? A new hypothesis. Epidemiology 13, 113–114 (2002).
5. Kosova, G., Scott, N. M., Niederberger, C., Prins, G. S. & Ober, C. Genome-wide association study identifies candidate genes for male fertility traits in humans. Am. J. Hum. Genet. 90, 950–962 (2012).
6. Shimizu, K. K. & Tsuchimatsu, T. Evolution of selfing: recurrent patterns in molecular adaptation. Annu. Rev. Ecol. Evol. Syst. 46, 593–622 (2015).
7. Darwin, C. The Effects of Cross and Self Fertilisation in the Vegetable Kingdom (John Murray, 1876).
8. Scharf, A. & LENHARD, M. The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. Annu. Bot. 107, 1433–1443 (2011).
9. Barrett, S. C. H. The evolution of plant sexual diversity. Nat. Rev. Genet. 3, 274–284 (2002).
10. Oka, H.-I. & Morishima, H. Variations in the breeding systems of a wild rice, Oryza perennis. Evol. 21, 249–258 (1967).
11. Boeven, P. H. G. et al. Genetic architecture of male floral traits required for hybrid wheat breeding. Theor. Appl. Genet. 129, 2343–2357 (2016).
12. Langer, S. M., Longin, C. F. H. & Wirschum, T. Phenotypic evaluation of floral and flowering traits with relevance for hybrid breeding in wheat (Triticum aestivum L.). Plant Breed. 133, 433–441 (2014).
13. Goldberg, E. E. et al. Species selection maintains self-incompatibility. Science 330, 493–495 (2010).
14. Cruden, R. W. Pollen grains: why so many? Plant Syst. Evol. 222, 143–165 (2000).
15. Willis, J. H. The contribution of male-sterility mutations to inbreeding depression in Mimulus guttatus. Heredity 83, 337–346 (1999).
16. Tsuchimatsu, T. et al. Evolution of self-compatibility in Arabidopsis by a mutation in the male specificity gene. Nature 464, 1342–1346 (2010).
17. Tsuchimatsu, T., Kaiser, P., Yew, C.-L., Bachelier, J. B. & Shimizu, K. K. Recent loss of self-incompatibility by degradation of the male component in allotetraploid Arabidopsis kamchatatica. PLoS Genet. 8, e1002838 (2012).
18. Charnov, E. L. The Theory of Sex Allocation (Princeton University Press, 1982).
19. Awetl et al. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465, 627–631 (2010).
20. Hancock, A. M. et al. Adaptation to climate across the Arabidopsis thaliana genome. Science 334, 83–86 (2011).
21. Shimizu, K. K., Shimizu-Insaduki, R., Tsuchimatsu, T. & Purugganan, M. D. Independent origins of self-compatibility in Arabidopsis thaliana. Mol. Ecol. 17, 704–714 (2008).
22. The 1001 Genomes Consortium. 1.135 genomes reveal the global pattern of polymorphism in Arabidopsis thaliana. Cell 166, 481–491 (2016).
23. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. PLoS Biol. 4, 727 (2006).
24. Tsukof, S. A. et al. Haploype diversity and linkage disequilibrium at human GPD1: recent origin of alleles that confer malarial resistance. Science 293, 455–462 (2001).
25. Horton, M. W. et al. Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. Nat. Genet. 44, 212–216 (2012).
26. Scholl, R. L., May, S. T. & Ware, D. H. Seed and molecular resources for Arabidopsis. Plant Physiol. 124, 1477–1480 (2000).

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-020-16679-7 | www.nature.com/naturecommunications

NEWS & VIEWS | (2020)11:2885 | https://doi.org/10.1038/s41467-020-16679-7 | www.nature.com/naturecommunications
27. Jinne, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821 (2012).
28. Tsutou, H. & Higashiyama, T. pCAMBIA-ITACCHI vectors for highly efficient CRISPR/Cas9-mediated gene knockout in Arabidopsis thaliana. Plant Cell Physiol. 58, 46–50 (2017).
29. Rodriguez-Mateos, M. et al. Role and dynamics of the ribosomal protein P0 and its related trans-acting factor Mrt4 during ribosome assembly in Saccharomyces cerevisiae. Nucleic Acids Res. 37, 7519–7532 (2009).
30. Bhavsar, R. B., Makley, L. N. & Tionis, P. A. The other lives of ribosomal proteins. Hum. Genomics 4, 327–344 (2010).
31. Sanders, P. M. et al. Anther developmental defects in Arabidopsis thaliana male-sterile mutants. Sex. Plant Reprod. 11, 297–322 (1999).
32. Brenninger, H. & Lenhard, M. Control of tissue and organ growth in plants. Curr. Top. Dev. Biol. 91, 185–220 (2010).
33. Turner, T. L. Fine-mapping natural alleles: quantitative complementation to the rescue. Mol. Ecol. 23, 2377–2382 (2014).
34. Stern, D. L. A role of Ultrabithorax in morphological differences between Drosophila species. Nature 396, 463–466 (1998).
35. Wall, Y. Mutational meltdown in selving Arabidopsis lyrata. Evoluton 67, 806–815 (2013).
36. Koch, M. A., Haubold, B. & Mitchell-Olds, T. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae). Mol. Biol. Evol. 17, 1483–1498 (2000).
37. Ossowski, S. et al. The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. Science 327, 92–96 (2010).
38. Shimizu, K. K., Kudo, H. & Kobayashi, M. J. Plant sexual reproduction during climate change: gene function in nature studied by ecological and evolutionary systems biology. Ann. Bot. 108, 777–787 (2011).
39. Tang, C. et al. The evolution of selving in Arabidopsis thaliana. Science 317, 1070–1072 (2007).
40. de la Chaux, N., Tsuchimatsu, T., Shimizu, K. K. & Wagner, A. The predominantly selving plant Arabidopsis thaliana experienced a recent reduction in transposable element abundance compared to its outcrossing relative Arabidopsis lyrata. Mob. DNA 3, 2 (2012).
41. Alonso, J. M. et al. Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science 301, 653–657 (2003).
42. Kleinboelting, N., Huep, G., Kloetgen, A., Viehoever, P. & Weisshaar, B. GABI-Kat SimpleSearch: new features of the Arabidopsis thaliana T-DNA mutant database. Nucleic Acids Res 40, D1211–D1215 (2012).
43. R Core Team. R Foundation for Statistical Computing (The R Foundation, 2013).
44. Cao, J. et al. Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat. Genet. 43, 956–965 (2011).
45. Long, Q. et al. Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. Nat. Genet. 45, 884–890 (2013).
46. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. G3, 816–834 (2010).
47. Segura, V. et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat. Genet. 44, 825–830 (2012).
48. Lei, Y. et al. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. Mol. Plant 7, 1494–1496 (2014).
49. Shimada, T. L., Shimada, T. & Hara-Nishimura, I. A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of Arabidopsis thaliana. Nat. Genet. 39, 1151–1155 (2007).
50. Tedder, A. et al. Evolution of the selving syndrome in Arabis alpina (Brassicaceae). PLoS ONE 10, e0126618 (2015).
51. Peterson, K. M. et al. Arabidopsis homeodomain-leucine zipper IV proteins promote stomatal development and ectopically induce stomata beyond the epidermis. Development 140, 1924–1935 (2013).
52. Vienne-Calzada, J. P., Baskar, R. & Grossniklaus, U. Delayed activation of the paternal genome during seed development. Nature 404, 91–94 (2000).
53. Nordborg, M. et al. The pattern of polymorphism in Arabidopsis thaliana. PLoS Biol. 3, 1289–1299 (2005).
54. Guittet, M. & Vitalis, R. Rail: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics 28, 1176–1177 (2012).
55. Hu, T. T. et al. The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. Nat. Genet. 43, 476–481 (2011).
56. Kim, S. et al. Recombination and linkage disequilibrium in Arabidopsis thaliana. Nat. Genet. 39, 1151–1155 (2007).
57. We thank Matt Horton, Robert Martin, Atsushi Tajima, and Hirokazu Tsukaya for discussions, Karen Thomsen for sample preparation and pollen counting, and the Functional Genomics Center Zurich for technical support. This study was supported by the Swiss National Science Foundation 31003A_182318, 31003A_159767, JST CREST Grant Number JPMICR01603, Japan, to K.K.S., a Syngenta Fellowship through the Zurich Basel Plant Science Center to T.T., T.S., and K.K.S., the University Research Priority Program of Evolution in Action of the University of Zurich to K.K.S. and U.G., MEXT KAKENHI Grant Numbers 16H06469, 16K21727, 26117079 to K.K.S., 16H06467, 17H10583, 18H04813, 19H04851 to T.T., and 16H06465 to M.M.K., JSPS KAKENHI Grant Number 18K05976 to H.K., the Forschungskredit of the University of Zurich and an EMBO long-term fellowship to T.T., a JSPS postdoctoral fellowship for research abroad to H.K. and T.T., and, in part, an Advanced Grant of the European Research Council to U.G.
58. T.S., M.L., and M.N.; All authors discussed the results and commented on the manuscript.
59. The authors declare no competing interests.
60. Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-16679-7.
Correspondence and requests for materials should be addressed to K.K.S.
Peer review information Nature Communications thanks James Prendergast, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.