nature genetics

ARTICLES

https://doi.org/10.1038/s41588-021-00984-y

A *PARTHENOGENESIS* allele from apomictic dandelion can induce egg cell division without fertilization in lettuce

Charles J. Underwood[1,7,10], Kitty Vijverberg[2,8,10], Diana Rigola[1,10], Shunsuke Okamoto[1,3], Carla Oplaat[2,9], Rik H. M. Op den Camp[1], Tatyana Radoeva[1], Stephen E. Schauer[4], Joke Fierens[1], Kim Jansen[1], Sandra Mansveld[1], Marco Busscher[2], Wei Xiong[2], Erwin Datema[1], Koen Nijbroek[1], Evert-Jan Blom[1], Ross Bicknell[5], Andrew Catanach[5], Sylvia Erasmuson[5], Christopher Winefield[6], Arjen J. van Tunen[1], Marcel Prins[1], M. Eric Schranz[2]✉ and Peter J. van Dijk[1]✉

Apomixis, the clonal formation of seeds, is a rare yet widely distributed trait in flowering plants. We have isolated the *PARTHENOGENESIS* (*PAR*) gene from apomictic dandelion that triggers embryo development in unfertilized egg cells. *PAR* encodes a K2-2 zinc finger, EAR-domain protein. Unlike the recessive sexual alleles, the dominant *PAR* allele is expressed in egg cells and has a miniature inverted-repeat transposable element (MITE) transposon insertion in the promoter. The MITE-containing promoter can invoke a homologous gene from sexual lettuce to complement dandelion *LOSS OF PARTHENOGENESIS* mutants. A similar MITE is also present in the promoter of the *PAR* gene in apomictic forms of hawkweed, suggesting a case of parallel evolution. Heterologous expression of dandelion *PAR* in lettuce egg cells induced haploid embryo-like structures in the absence of fertilization. Sexual *PAR* alleles are expressed in pollen, suggesting that the gene product releases a block on embryogenesis after fertilization in sexual species while in apomictic species *PAR* expression triggers embryogenesis in the absence of fertilization.

Apomixis is the fertilization-independent production of seeds resulting in the formation of maternally derived clonal progeny. (Supplementary Note 1)[1]. During his studies, Gregor Mendel noted differences between sexual (pea) and apomictic (hawkweed) modes of inheritance (Supplementary Fig. 1 and Supplementary Note 2)[2–4]. It is now established that apomixis is a rare, yet widespread, reproductive system among flowering plants, occurring in about 0.1% of species spread over 120 genera[5,6]. One of the most iconic and ubiquitous apomicts is the common dandelion, *Taraxacum officinale*, the ecological and evolutionary success of which largely depends on reproduction by apomixis[7].

Apomixis fixes complex/nonadditive genetic traits, such as heterosis, in a single individual and a single step. Despite these advantages, apomixis does not occur in major crops. Breeding of apomictic, high-yielding, hybrid crop varieties would be of immense value to agriculture, in terms of advancing breeding gains and facilitating seed production[8–10]. To achieve this, we require additional knowledge about the genes that govern the native expression of this trait, which reside in apomixis loci.

Apomixis loci are challenging to study: they are absent from well-characterized model systems, recombination tends to be suppressed and nearly all apomicts are polyploid. This makes the positional cloning of apomixis genes difficult[11], and the complete genomic sequence of an apomixis-linked region and its sexual counterpart has yet to be reported. Cytogenetic studies show that apomixis loci are usually hemizygous, dense in transposable elements (TEs) and share structural characteristics with heterochromatic sex chromosomes, which have lost the capacity to recombine and often contain degenerated genes[12–14]. Similarly, over successive clonal generations without outcrossing, it is hypothesized that clones lacking recombination and syngamy will accumulate deleterious mutations and TEs[15–20].

Genetic analysis of apomixis loci is well established in dandelion because polyploid apomictic dandelions can be used to pollinate diploid sexuals[11,21]. Dandelion has a gametophytic-type apomixis, which resembles sexual development because it produces sexual-like female gametophytes. In apomictic triploid dandelions, female meiosis is replaced by diplospory, a cell division without chromosomal recombination and reduction that is functionally identical to a mitotic division[22]. The egg cell is, therefore, genetically identical to the mother plant. In contrast to sexual plants, where egg cells arrest until fertilization by the sperm, apomictic dandelion egg cells develop into an embryo without fertilization via parthenogenesis[23]. Genetic analysis has found two unlinked loci containing the *DIPLOSPOROUS* (*DIP*) and *PARTHENOGENESIS* (*PAR*) genes[24], where each locus dominantly leads to the respective apomixis trait over the two additional recessive haplotypes in the same triploid plant[22,24–26]. In dandelion, autonomous formation of endosperm is

[1]Keygene N.V., Wageningen, the Netherlands. [2]Biosystematics Group, Wageningen University, Wageningen, the Netherlands. [3]Takii & Co. Ltd, Plant Breeding and Experiment Station, Konan Shiga, Japan. [4]Keygene Inc., Rockville, MD, USA. [5]New Zealand Institute for Plant & Food Research, Lincoln, New Zealand. [6]Lincoln University, Lincoln, New Zealand. [7]Present address: Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany. [8]Present address: Naturalis Biodiversity Center, Radboud University, Nijmegen, the Netherlands. [9]Present address: National Reference Centre of Plant Health, National Plant Protection Organization, Wageningen, the Netherlands. [10]These authors contributed equally: Charles J. Underwood, Kitty Vijverberg, Diana Rigola. ✉e-mail: eric.schranz@wur.nl; peter.van-dijk@keygene.com

the third essential component of apomixis, although the genetic control of this is unclear[24].

In the present study, we present the complete sequences of the apomictic haplotype and the two sexual haplotypes of the *PAR* locus from a triploid apomictic plant. We identify the *PAR* gene, which encodes a zinc finger domain protein with an EAR (ethylene-responsive element-binding factor-associated amphiphilic repression, DLNxxP) motif. Expression of *PAR* in egg cells of apomictic dandelion was detected and may be due to the insertion of a nonautonomous TE in its promoter that was present in all of the dandelion apomicts tested. A similar TE is present in the hawkweed apomictic ortholog of *Pilosella piloselloides* (formerly *Hieracium piloselloides*), which is also specifically expressed in the ovule, suggesting parallel evolution. Expression of dandelion *PAR* under the control of an egg-cell-specific promoter in the related, sexual crop lettuce (*Lactuca sativa*) showed that *PAR* can induce the development of ectopic cell divisions to produce haploid embryo-like structures. This shows that *PAR* is sufficient to initiate parthenogenesis, in the sense that egg cells divide without fertilization in a sexual crop species.

Results

Deletion mapping of the *PAR* locus. We screened 2,650 plants derived from γ-irradiated seeds of the triploid dandelion apomictic line A68 for the LOSS OF PARTHENOGENESIS (LOP) phenotype by looking for seed heads that lacked a dark center of viable seeds, which is characteristic of A68 (>13,500 seed heads screened; Fig. 1a). Of *LOP* lines, 23 were identified that lacked from 1 to 19 amplified fragment length polymorphism (AFLP) markers genetically linked to the *PAR* locus and led to the identification of a minimal locus, estimated to be about 300 kb in size (Supplementary Notes 3–5, Supplementary Figs. 2–4 and Supplementary Table 2). The markers in the minimal locus were sequenced and one, PD2 (E55M35-195), was present in a low copy number in the A68 bacterial artificial chromosome (BAC) library (Supplementary Note 5).

The i34 *LOP* mutant still made autonomous endosperm (Fig. 1b), which is consistent with the observation that autonomous endosperm is independent of parthenogenesis[24]. Pollination of i34 with a diploid sexual led to tetraploid offspring (Fig. 1c), and thus triploid *LOP* deletion mutants were still fertile, because they skipped meiosis by diplospory, but lost the ability for parthenogenetic development.

The genomic structure of the *PAR* locus. Using a tiling path from the A68 dandelion BAC library in combination with nanopore (long-read) sequencing of A68, three contigs spanning the whole *PAR* locus were assembled, as expected from a triploid plant with three haplotypes (Fig. 2a). The sequence of one haplotype was markedly different from the others and contained two copies of the PD2 AFLP marker associated with the *PAR* locus and was thus identified as the dominant apomictic (Apo) haplotype. The other haplotypes, which do not confer parthenogenesis and represent the recessive 'sexual' haplotypes, were named sex1 and sex2. The *PAR* locus was defined as the region delimited by a ring finger (*A_g15*) and a Hikeshi (*A_g485*) gene, shared by the three haplotype contigs (DNA sequences: Supplementary Data 1–3). This genomic interval is syntenic with lettuce chromosome 8 (170.78–172.08 Mb), the first part in a direct orientation with 171.61–172.12 Mb, the second part in inverted orientation (170.78–171.46; Fig. 5a and Supplementary Fig. 18).

The Apo haplotype has 95 predicted open reading frames (ORFs), whereas sex1 and sex2 have only 50 and 48 predicted ORFs, respectively (Supplementary Table 4). The Apo haplotype (386 kb) is 1.8× larger than the two sex haplotypes (215 kb and 212 kb). Approximately 32% of the size difference is accounted for by the higher occurrence of predicted TE ORFs in the Apo haplotype versus sex (48 versus 12 and 8, respectively) (Supplementary Table 3). Apo-specific PCR markers were tested on a broad panel of

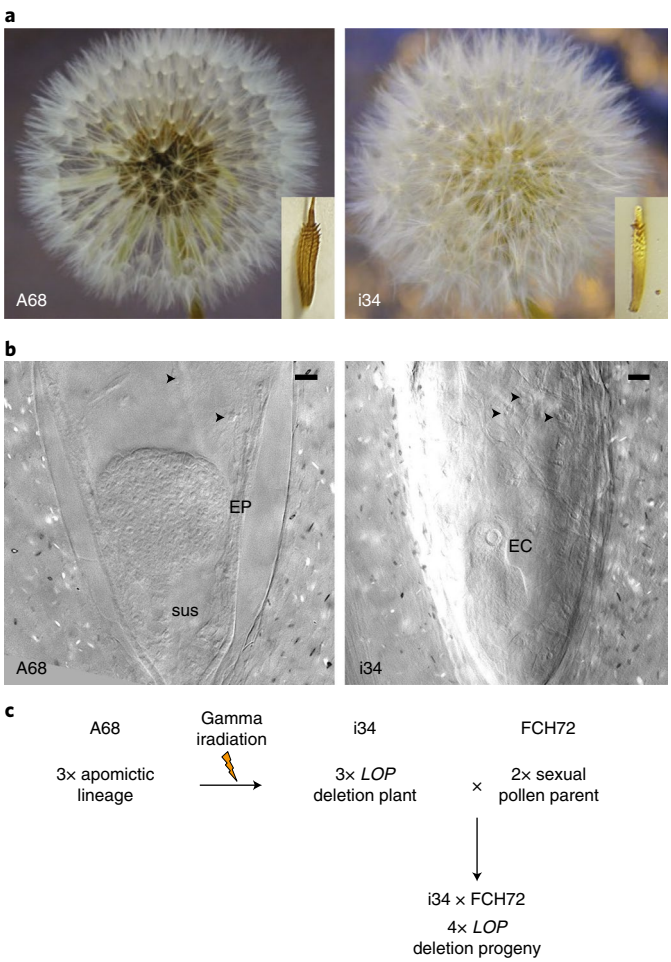


Fig. 1 | Deletion mutant seed-head phenotypes. a, Left: a seed head of the apomictic wild-type A68, with an inset of a single seed. Right: a seed head of the *LOP* deletion plant i34 with an inset of a single seed. Note the absence of the dark-brown seeds in the center. **b**, Differential interference contrast microscopy of developing cleared seeds, 1 d after anthesis. The arrows show the nuclei of autonomous endosperm. EC, egg cell; EP, embryo proper; SUS, suspensor. Left: a developing embryo of apomictic wild-type A68 (observed in >100 samples). Right: an egg cell of deletion line i34, arrested in development (observed in >50 samples). Scale bar, 20 μm. **c**, The genetic relationships between the different genotypes used in the present study. Diploid, triploid and tetraploid plants are indicated with 2x, 3x and 4x, respectively.

apomictic and sexual dandelions, which showed that the core (markers LD2, LD3, LD3', LD4) of the Apo haplotype is present in all apomicts, but absent in all sexuals (Fig. 2c and Supplementary Note 7). Alignments of 20 genes in the Apo haplotype showed frequent signs of gene conversion across all three alleles (Supplementary Data 4, Supplementary Fig. 5, Supplementary Note 8 and Supplementary Table 5). This explains why the number of synonymous substitutions per synonymous site (*Ks*) is sometimes greater between two sex alleles than between the Apo allele and the sex alleles (Fig. 2b). The ratio of *Ka* (number of nonsynonymous substitutions per nonsynonymous site):*Ks* for single-copy genes was in all cases much lower than 1 (maximum of 0.29). This suggests that these are functional genes still under purifying selection.

Identification of the *PAR* gene. To find the causal gene for parthenogenesis, 13 candidate genes were selected by putative function and their apomictic alleles were specifically targeted by CRISPR

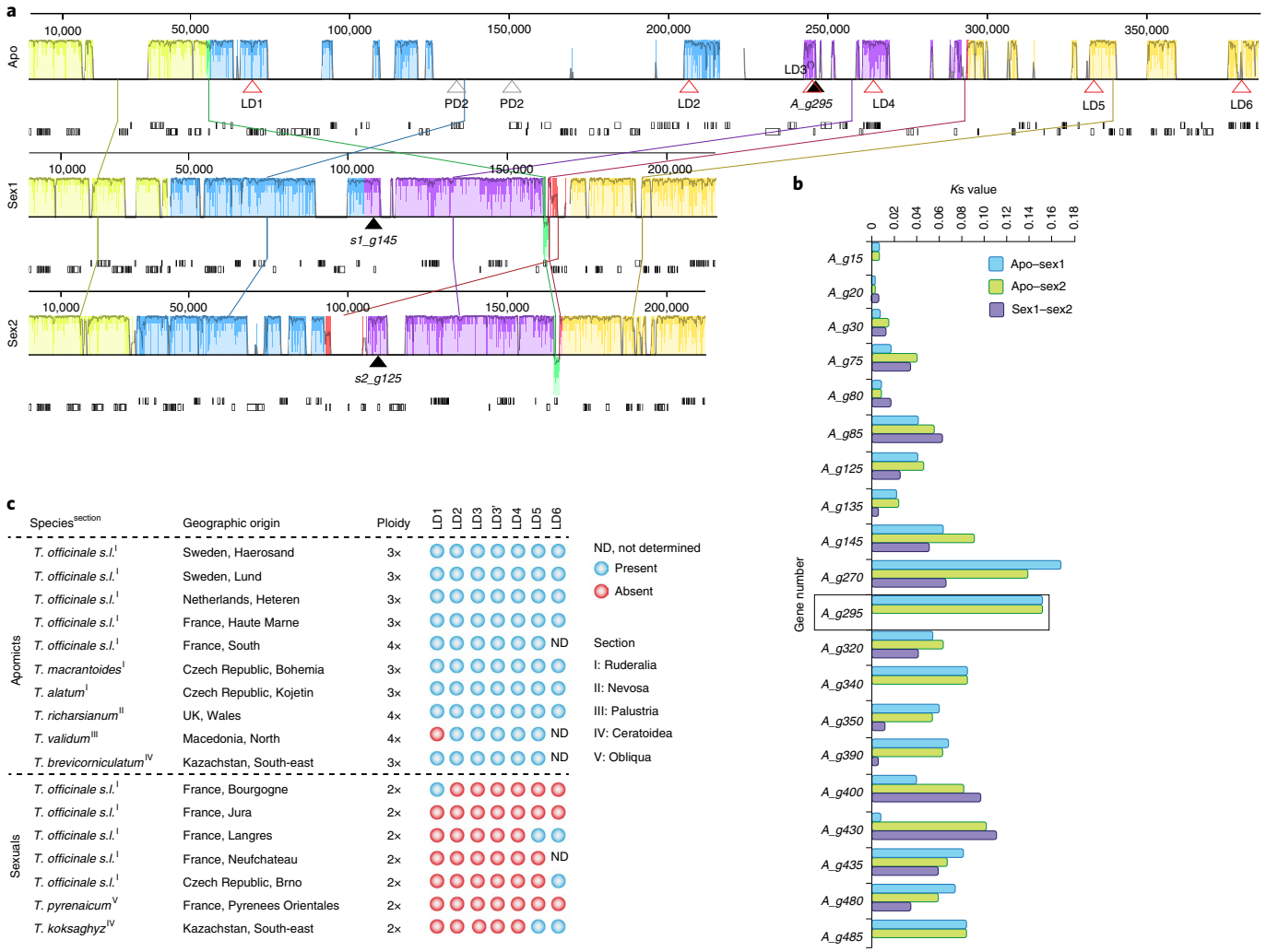


Fig. 2 | Structural characteristics of the *PAR* locus and LD analysis. a, Mauve alignment of the three haplotypes of the *PAR* locus. The top panel shows the Apo haplotype, the middle the sex1 haplotype and the lower the sex2 haplotype. The blocks with the same color show homologous regions. Inside each block a similarity profile is shown, which corresponds to the average level of conservation in that region of the genome sequence. Regions that are white have sequences specific to a particular haplotype. Annotated, predicted ORFs are shown in the tracks below the haplotypes as white boxes, with those transcribed from the reverse strand shown downward. Arrowheads mark the locations of genetic markers and genes. The LD3 and LD3′ marker locations partially overlap. **b**, The *Ks* across the 3 alleles of 20 genes in the *PAR* locus. The blue column represents the *Ks* between the Apo and sex1 haplotypes, the green column the *Ks* between the Apo and the sex2 haplotypes, and the purple column the *Ks* between the two sexual haplotypes. **c**, A linkage disequilibrium (LD) panel. Shown are *PAR* haplotype-specific markers in different diploid sexuals and polyploid apomicts from different species (Supplementary Note 6) and geographic origins. Sections: I, *Ruderalia*; II, *Nevosa*; III, *Palustria*; IV, *Ceratoidea*; V, *Obliqua*.

(clustered regularly interspaced short palindromic repeats)–Cas9 guide RNAs (gRNAs) in the apomictic dandelion line A68 (Supplementary Tables 3 and 6). For three genes (*A_g90*, *A_g295* and *A_g320*) two different gRNAs were used. Similar to the *LOP* deletion mutants, it was expected that *LOP* mutants would not produce seeds without crossing. Pollination of a triploid *LOP* mutant by a diploid sexual would be expected to produce viable tetraploid seeds, and offspring, because the diplospory gene is present and the egg cells are therefore unreduced, and triploid.

In regenerated plants, only the presence of gRNA9 or gRNA10, both targeting the zinc finger gene *A_g295*, correlated with a seed-head phenotype equivalent to the *LOP* deletion mutants (Figs. 1a and 3a–c). Regenerated plants containing either gRNA9 or gRNA10, which had wild-type or *LOP* seed-head phenotypes, were further characterized. Seeds collected from at least three independent seed heads from ten CRISPR–Cas9 *LOP* plants (eight plants with gRNA9 and two plants with gRNA10) and seven plants with unaltered phenotypes (five plants with gRNA9 and two plants with gRNA10) were tested for germination. Seeds collected from the CRISPR–Cas9 *LOP* plants were not capable of germination, whereas those from plants with unaltered phenotypes germinated normally (Fig. 3b and Supplementary Tables 7 and 8). Sequencing of gRNA targets revealed that all eight of the CRISPR–Cas9 *LOP* plants harboring gRNA9 that failed to produce viable seeds had mutations in *A_g295* (Fig. 3d and Supplementary Table 7). The five plants with unaltered phenotypes that contained gRNA9 had no mutations in *A_g295* (Supplementary Table 7). In addition, the two CRISPR–Cas9 *LOP* plants that had gRNA10 also had mutations in *A_g295* (Fig. 3e and Supplementary Table 8). Pollination of seven different lines that carried mutations in *A_g295* with a diploid sexual line (FCH72) yielded a large number of viable seeds, as was previously observed with the *LOP* deletion mutants (Figs. 1c and 3f,g and Supplementary Tables 7 and 8). The offspring of these crosses were all shown to be tetraploid (Fig. 3h and Supplementary

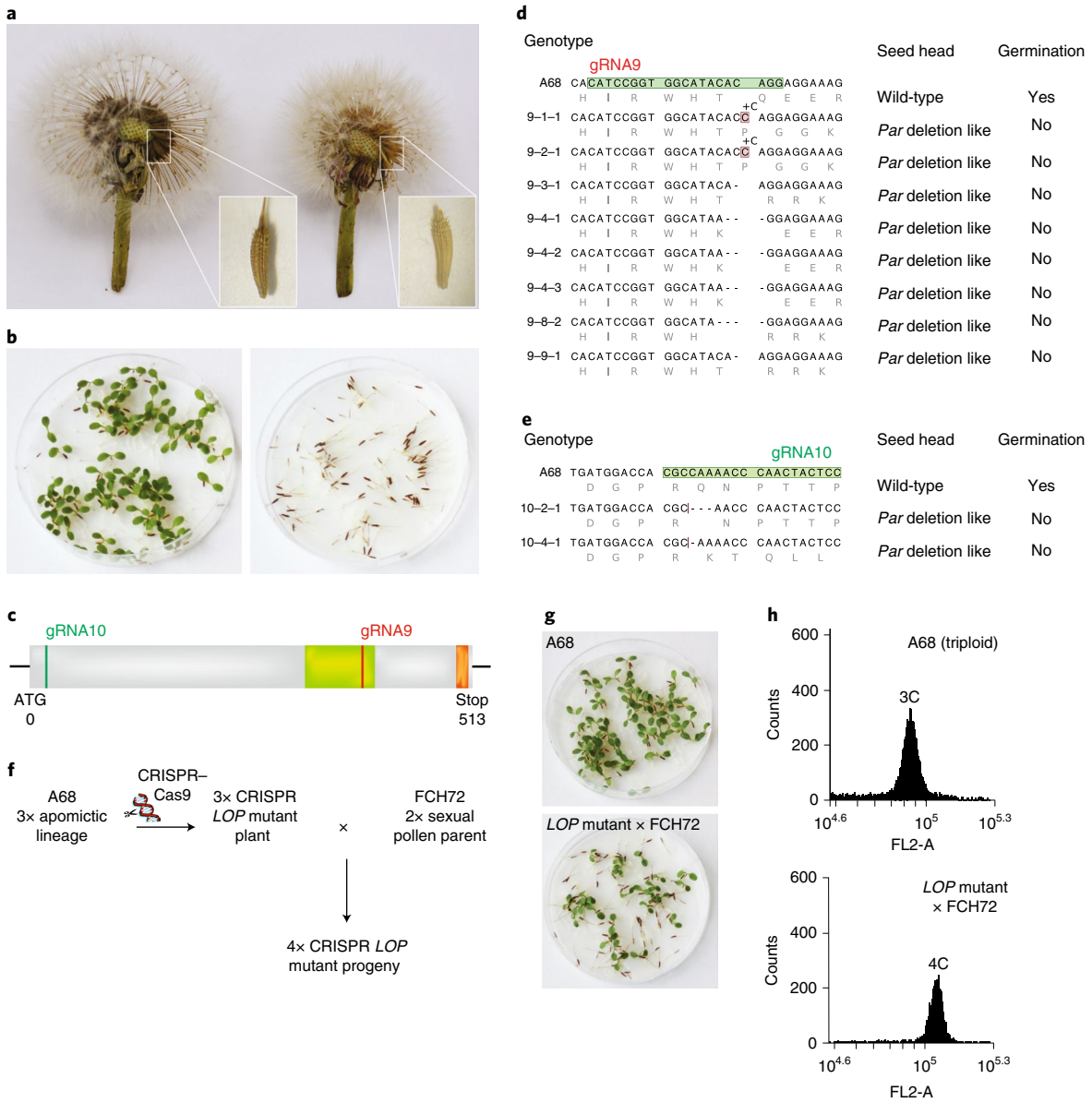


Fig. 3 | Identification of the *PAR* gene by CRISPR–Cas9 mutagenesis. a, A seed head from wild-type apomictic dandelion A68 (left) and CRISPR–Cas9 *LOP* mutant 9-9-1 (right). A representative individual seed is shown in the inset. **b**, Germination of A68 (left) and CRISPR–Cas9 *LOP* mutant (right) seed heads. **c**, Gene *A_g295* with gRNA targets indicated; the green and orange boxes mark the zinc finger and the EAR domains, respectively. **d**, Genotypes at gRNA9, target of eight CRISPR–Cas9 *LOP* mutants, derived from five independent calli. Plants are named as guide number–callus number–plant number. **e**, Genotypes at gRNA10 target of two CRISPR–Cas9 *LOP* mutants. **f**, The crossing scheme used to confirm reversion to sexuality of the CRISPR–Cas9 *LOP* mutants. **g**, Germination of A68 and *LOP* mutant 9–4–2 × FCH72 progeny. **h**, Flow cytometry analysis of plant ploidy of A68 and *LOP* mutant 9–4–2 × FCH72 progeny.

Tables 7 and 8). As *A_g295* mutants were not parthenogenetic and required pollination for seed production (thereby reverting to sexuality), we concluded that *A_g295* encodes the *T. officinale PARTHENOGENESIS* (*PAR*) gene (Supplementary Fig. 6 and Supplementary Data 5).

Transcriptomic analysis of the *PAR* locus. To determine expression of *PAR* and linked genes in the Apo haplotype during female gametophyte development, transcriptome analysis was performed on laser-assisted microdissected tissues (egg-cell apparatus, central cell (CC) and whole gametophyte) in three genotypes (apomictic A68, *LOP* line i34 × FCH72 and sexual FCH72) at different developmental stages (Fig. 4a,b, Supplementary Figs. 7 and 8, and Supplementary Note 9). Of 100 predicted ORFs from the 386-kb Apo haplotype (9 kb of each flank), 47 were expressed in the samples (Supplementary Tables 11–13). Genes with high expression in the Apo haplotype compared with the sex haplotypes are presented in Fig. 4c. Of the 47 genes, 2 showed unique expression of the Apo allele: *A_g295* (*PAR*) and *A_g490* (Supplementary Fig. 9). *A_g490*, a small ORF with unknown function, was highly expressed during early gametophyte development, well before cellularization and egg cell/CC fate determination, and is therefore unlikely to be responsible for parthenogenesis. *PAR* showed specific expression from the apomictic allele in the mature egg cell apparatus, and not from the sexual alleles. The spatiotemporal expression pattern of *PAR* was confirmed by in situ hybridization (Supplementary Fig. 10),

providing strong evidence that *PAR* is expressed in the expected cell and at the expected time to be causal for parthenogenesis. Reverse transcription PCR (RT-PCR) showed that the sexual *par* allele is highly expressed in mature pollen (male gametophyte) of the diploid sexual plant FCH72, compared with somatic leaf tissue (Supplementary Fig. 11).

The *PAR* gene and conserved upstream MITE. *PAR* encodes a 170-amino acid protein with a C2H2 zinc finger (ZF) domain of the rare *Arabidopsis* K2-2 class[27] (Supplementary Fig. 12), with predicted nucleic acid-binding activity, and a C-terminal EAR motif, a plant-specific transcriptional repression domain[28,29] (Fig. 5b,c). We identified *PAR* homologs in available Asteraceae genomes and found that they commonly have deletions/insertions around the ZF domain and that this gene has evolved very rapidly within Asteraceae (Supplementary Fig. 13). *PAR* is a single-copy gene in both dandelion and lettuce (*Ls_8X112340*, '*Lssex*'). Homologs in species outside the Asteraceae could be identified by searching for genes with a single C2H2 ZF domain of the rare K2-2 type and one or two C-terminal EAR motifs (Supplementary Figs. 14 and 15 and Supplementary Note 17).

The dominant dandelion *PAR* allele has a single putative start codon whereas the sexual *par-1* and *par-2* alleles have two putative start codons (adding 57 amino acids to the N terminus). The lettuce ortholog has only the second start codon, so the shorter protein is not specific for parthenogenesis. The striking difference between the dandelion alleles is a 1,335-bp insertion in the upstream promoter region. The insertion is 110 bp upstream from the *PAR* start codon, with a 9-bp terminal inverted repeat (TIR, CAGGGCCGG and CCGGCCCTG) and an 8-bp target site duplication (TSD, ACTGCTAC). The insertion has no homologies with sequences in the National Center for Biotechnology Information (NCBI) GenBank and does not have an ORF, all consistent with being a new nonautonomous MITE. The LD3 and LD3′ PCR markers that amplify part of the MITE are present in all apomicts and absent in all sexuals (Fig. 2c and Supplementary Fig. 16). Re-sequencing of five apomictic and two sexual accessions revealed that the insertion of the MITE in all apomicts was the same, suggesting a common origin (Supplementary Fig. 17). The PAR protein differs at five amino acids from the sexual dandelion proteins: Gly20Pro, Pro22Gln, Tyr40Asn, Tyr83Ile and Lys99Asn. Similar to the MITE, the re-sequencing showed 100% concordance of Gly20Pro and Pro22Gln variants with apomixis, but other variation could be excluded as being causal.

Allele sequence divergence at synonymous sites between the Apo and sex haplotypes allowed us to estimate when parthenogenesis evolved in dandelions. Allele sequence divergence is reduced by gene conversion, although this is retarded by triploidy and still substantially higher than in sexual plant species (Supplementary Note 8). Averaging over the *PAR* gene and four flanking genes, we calculated an average *Ks* of 0.084 ± 0.012 (s.e.m.) for the core of the *PAR* locus. Assuming a plant substitution rate of 5.35×10^{-9} per site per year[30], the minimum age of the *PAR* locus is 7.85 million years (95% confidence interval: 2.20 million years). Since then *PAR* most probably spread, together with *DIP*, through *Taraxacum* spp. by backcrossing with sexuals.

Evidence of a similar *PAR* locus in *P. piloselloides*. The *LOP* locus controlling parthenogenesis in *P. piloselloides*[31] was mapped to a 650-kb interval, using deletion mapping, BAC walking and a targeted polyhaploid mapping approach (Supplementary Notes 12 and 13). The genome of a diploid, apomictic clone of *P. piloselloides* was sequenced and a contig containing the *Pilosella LOP* locus identified (Supplementary Note 13). The *Pilosella LOP* locus was found to be collinear to the *Taraxacum PAR* locus (Supplementary Fig. 18) and to include a single sequence that is orthologous to the *Taraxacum PAR* gene (Fig. 5a). The dominant *PpPAR* allele was identified from BAC clones specific to that allele and a recessive allele was identified by homology from the genomic sequence. Remarkably, the dominant *Pilosella* allele contained a 1,282-bp-long MITE located 137 bp upstream of the *PpPAR* start codon, indicating that very similar, yet independent, insertions have occurred in the evolution of the dominant, apomictic *PAR* alleles of both *Taraxacum* and *Pilosella* spp. (Fig. 5b–e). As in *Taraxacum* sp., the recessive *Pilosella* allele sequence lacks this feature. It is notable that the MITE length is different between the species (1,335 bp in *Taraxacum* sp. and 1,282 bp in *Pilosella* sp.), the location of the elements also differs (110 bp upstream of the start codon in *Taraxacum* sp. and 137 bp in *Pilosella* sp.) and the sequences internal to the inverted repeats of the two elements share no homology. However, the TIR sequences indicate that they do belong to the same family of *hAT*-derived nonautonomous transposons (Supplementary Data 6 and Supplementary Note 16). The 200-bp promoter region upstream of the start codon into which the two MITEs are inserted is highly conserved in sexual *Taraxacum*, *Pilosella* and *Lactuca* spp. (Supplementary Fig. 19 and Supplementary Table 15).

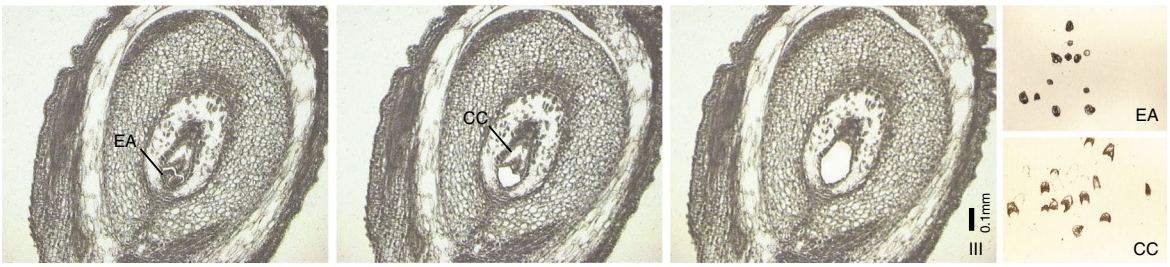
Association mapping and ovule gene expression studies were performed to further study the *PpPAR* gene. A panel of 13 apomictic and 5 sexual *Pilosella* spp. and varieties were analyzed for the presence of the MITE. As in *Taraxacum* spp., the MITE was present in all apomicts and absent in all sexual biotypes (Supplementary Table 16). RNA-sequencing (RNA-seq) reads of microdissected *P. piloselloides* ovules[32], from eight different developmental stages (meiosis until torpedo embryos), were mapped on to the dominant and recessive allele sequences of *P. piloselloides*. Most of the reads that aligned to the dominant allele were found in the samples taken between late embryo sac development and globular embryos, the stage where parthenogenesis would be expected to occur (Supplementary Table 17 and Supplementary Note 14). Conversely, relatively few reads aligned to the recessive allele sequence of *Pppar*, none of which was found during the period when parthenogenesis should occur. As expected, no reads derived from a mutant lacking the *PpPAR* allele mapped to the *PpPAR* allele. We concluded that the transcript mapping was specific to the *PAR* gene of *P. piloselloides*, and confirmed that the *PpPAR* gene was the only allele detected during the stages when parthenogenesis occurred.

Functional evidence that *PAR* induces *PARTHENOGENESIS*. To test whether the *Taraxacum PAR* promoter plays a role in the control of parthenogenesis, we tested whether it could be combined with a *PAR* homolog from a sexual species to induce parthenogenesis. The *Taraxacum PAR* promoter was used to drive the expression of a homologous gene (*Lssex*) from lettuce, a related species and

Fig. 4 | Comparative transcriptomics of female gametophytic cell types of Apo, *LOP* and Sex *T. officinale*. **a**, Example of laser-assisted microdissection of the EA (egg cell and synergids) and CC from a mature gametophyte (stage III) and of EAs (right, upper) and CCs (lower) isolated. **b**, Overview of the 27 samples analyzed, each representing ~150–300 isolates from ~40–75 whole cells of one flower head. **c**, Expression pattern of 30 genes (*A_gene number* indicated on left) in the Apo haplotype (left panel, blue) and their homologs in the sex haplotype (right panel, orange, averaged over the two sex alleles), showing the unique expression of *A_295* (the *PAR* gene) in the Apo haplotype. The expression represents the number of reads mapped to these genes in each of the 27 samples (see **b**; indicated above the panels). The genes presented include those that showed relatively high expression (>33%) in the sum of Apo samples compared with the sum of *LOP* and sex samples (see Supplementary Note 9), and in addition all genes targeted by CRISPR–Cas9 ([6]). The genes in gray italics indicate transposons.

important vegetable crop (Fig. 6a). This construct was transformed into a self-incompatible, tetraploid *Taraxacum* CRISPR–Cas9 *LOP* mutant incapable of seed production (derived from the crosses described in Fig. 3f). As *PAR* is dominant, testing was performed on the primary transformed plants (T_0). Remarkably the *PAR::Lssex* construct led to seed production and tetraploid (due to presence

a



b

Sample ID	Sample type	Stage		Genotype	Apo	*LOP*	Sex
				Ploidy	2*n* = 3×	2*n* = 4×	2*n* = 2×
WG-I	Whole gametophyte	Functional megaspore (FMS): 4 nuclei	Number		1	1	1
WG-II		8 nuclei (N): cellularized embryo sac			1	1	1
EA-II	Egg apparatus				1	1	1
EA-III		Mature embryo sac			3	3	3
CC-III	Central cells				3	3	3

c

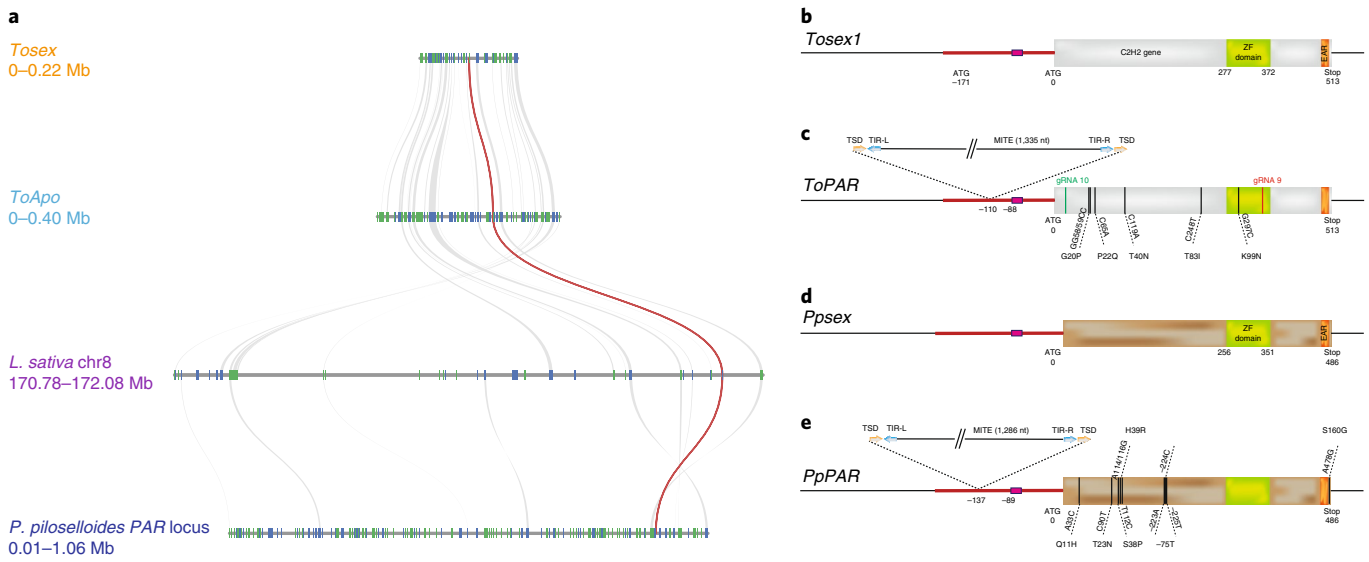


Fig. 5 | Microsynteny between *Taraxacum* and *Pilosella PAR* regions and the structure of the *PAR* and sex alleles. a, Microsynteny of the Apo and the sex1 haplotypes of *T. officinale*, the homologous region on linkage group 8 of *L. sativa* (lettuce) and *P. piselloides PAR* locus the red line connects *PAR* homologs between regions and the gray lines connect additional anchor syntelogs between regions. Blue indicates the positive strand and green the negative strand. **b**, *Taraxacum* sexual *par-1* allele. **c**, *Taraxacum PAR* allele. **d**, *Pilosella* sex allele. **e**, *Pilosella PAR* allele. For **b** to **e** the boxes indicate the ORF with the C2H2 zinc finger motif in green and the EAR motif in red. The first ATG in the *Taraxacum par-1* allele indicates a potential start codon of a longer ORF. Amino acid substitutions in the *PAR* alleles, compared with the sexual alleles, are indicated. The black lines mark the 5′-upstream regions. The red line is the conserved upstream region, with the MYB DNA-binding domains represented as a small red box. The size of the MITEs in the *PAR* alleles is not drawn to scale. Salmon-colored arrows represent an 8-bp target site duplication and purple a 9-bp terminal inverted repeat. Note that the *Pilosella* MITE is inserted 27 bp further upstream compared with the *Taraxacum* MITE.

of the dominant diplospory gene) offspring in four independent transformants (Supplementary Table 18). This demonstrates that the *Taraxacum PAR* promoter can invoke a lettuce gene to induce parthenogenesis. No genetic polymorphisms specific to the *ToPAR* coding sequence (when compared with sex1 and sex2) were found in the lettuce gene, ruling out coding sequence polymorphisms as being causal for parthenogenesis. Upstream of the dandelion *PAR* ATG start site, the MITE insertion is the first genetic polymorphism unique to the apomictic allele when compared with three dandelion sexual alleles (Supplementary Fig. 20). Besides the MITE insertion only three *PAR*-allele-specific SNPs are found across the four *Taraxacum* promoters in the 350 bp upstream of the ATG (when the MITE is excised from the PAR allele). Taken together this provides strong evidence that the functional difference between the dandelion alleles of the *PAR* gene is caused by the MITE insertion in the promoter. A different construct, where the *PAR* gene is expressed under the egg-cell-specific *Arabidopsis* EC1 (*pEC1::PAR*) promoter, can also lead to complementation of the CRISPR–Cas9 *LOP* mutant, consistent with the hypothesis that egg-cell expression of *PAR* can cause parthenogenesis (Fig. 6a and Supplementary Table 19).

Next, we questioned whether *PAR* would be sufficient to induce parthenogenesis in lettuce. To this end a *pEC1::PAR* construct was introduced into lettuce (Fig. 6b). Seven independent T_0 lines containing the transgene were evaluated for the occurrence of parthenogenesis, which would be expected to occur dominantly. To prevent fertilization, flower buds were decapitated before anthesis, ensuring that stigmas, styles and anthers were removed. In all evaluated transgenic lines, embryo-like structures with ectopic cell divisions were seen with an average penetrance of at least 15% (where 50% would be the maximal penetrance of single-insertion lines) (Fig. 6b,d and Supplementary Table 20), whereas in nontransformed controls only the unfertilized egg-cell apparatus was observed (Fig. 6c). Occasionally, multiple embryos in one embryo sac (known as polyembryony; Fig. 6e) were found in *pEC1::PAR* lines, along with some autonomous endosperm initiation. The embryo-like structures aborted at later developmental stages and did not result in viable seeds. This is consistent with a previous report where pollination of lettuce by sunflower led to haploid, globular embryos that degenerate due to insufficient support from abnormal, nonsexual, endosperm[33]. The seven independent T_0 lines were used to pollinate wild-type lettuce, and the resulting F_1 offspring were analyzed for parthenogenesis. From all seven lines, F_1 offspring containing the transgene gave embryo-like structures after emasculation, further validating that the construct could dominantly trigger egg-cell division.

In the absence of fertilization, parthenogenetic haploid egg cells of the diploid sexual lettuce should develop into haploid embryos. To assess whether the observed lettuce embryo-like structures are haploid, their ploidy was analyzed in pools of ~15 embryos sacs by flow cytometry. Embryo sacs from nontransformed controls had diploid (embryo) and triploid peaks (endosperm) (Fig. 6f), whereas a clear haploid peak was found in the embryo sac pools from decapitated flower buds (Fig. 6g). The latter, and the absence of a triploid sexual endosperm peak, indicate that the embryo-like structures are a result of *PAR*-induced parthenogenesis in lettuce.

In nonemasculated transgenic lines, embryo-like structures in advanced stages could already be seen before completion of male gametogenesis (that is, before fertilization, Supplementary Fig. 21). The seed set of nonemasculated transgenic lines was highly reduced compared with nontransformed controls, which is an expectation of failed embryo and endosperm development. Germinated offspring were all diploid and probably the products of sexual reproduction. However, polyembryonic seeds were found in multiple transgenic lines, with twin seedlings germinating from a single seed, something never observed in nontransformed controls. The observed twin seedlings grew slowly and either they died before true leaf formation, or those that did finally form true leaves were diploid. These diploid twins could be of sexual or parthenogenetic origin,

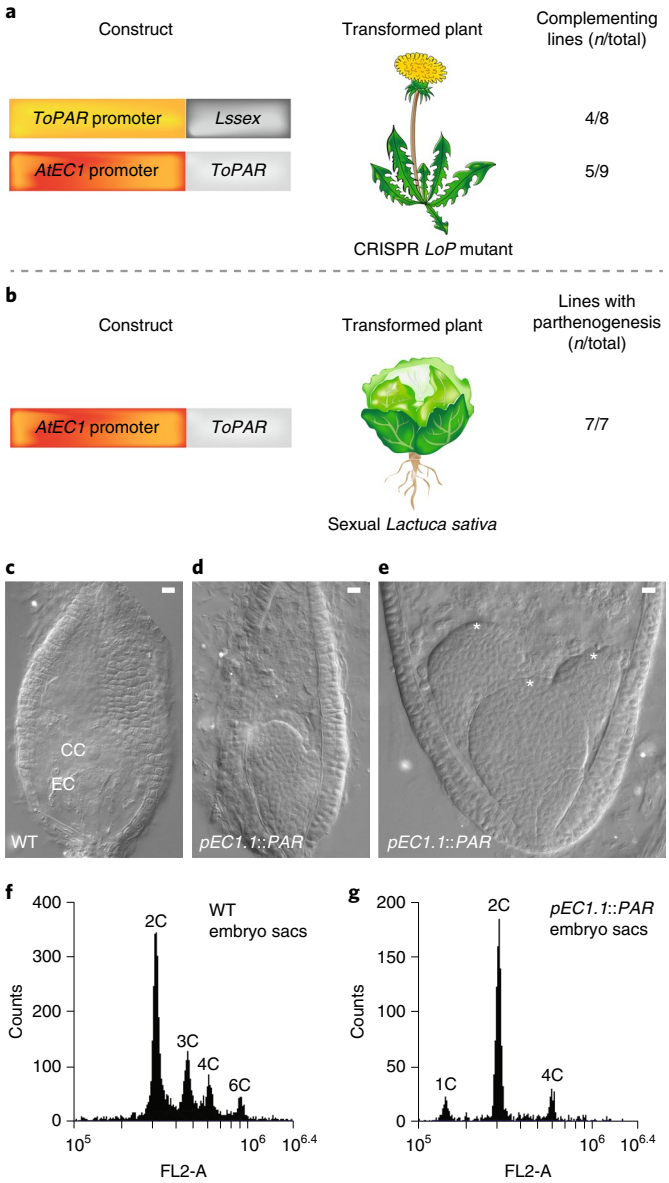


Fig. 6 | Complementation and transformation experiments in the *Taraxacum* CRISPR–Cas9 *LOP* mutant and sexual lettuce. a, Different promoter–gene constructs used for complementation of the *Taraxacum LOP* mutant and the number of successfully complemented lines. Shown are the *ToPAR* promoter and the sexual homolog from lettuce (*Lssex*) and the *Arabidopsis* egg cell (EC1.1) promoter with the *Taraxacum PAR* gene (for detailed results, see Supplementary Table 18). **b**, Transformation of lettuce with the *Taraxacum ToPAR* gene driven by *Arabidopsis* egg cell EC1.1 promoter (for detailed results, see Supplementary Table 19). **c**,**d**,**e**, Cleared ovules in decapitated inflorescences of nontransformed lettuce and lettuce transformed with a *pEC1.1::PAR* construct. Scale bar, 20 µm. **c**, Embryo sac from control nontransformed lettuce 75 h after decapitation. Unfertilized egg cell (EC) and CC nuclei are visible (common, observed in >100 samples). WT, wild type. **d**, Embryo sac with developing embryo-like structures 75 h after decapitation (common, observed in >80 samples). **e**, Embryo sac with multiple embryo-like structures (rare, observed in >10, but <20 samples). Asterisks show individual embryo-like structures. **f**, Flow cytometry analysis of embryo sacs from control nontransformed lettuce 5 d after self-pollination. **g**, Flow cytometry analysis of embryo sacs from transgenic lettuce carrying the *pEC1.1::PAR* construct 5 d after decapitation.

in the latter case followed by spontaneous haploid genome doubling, which is often observed in crop species[34].

Discussion

The *PAR* gene is necessary for embryo development in the absence of fertilization in dandelion and is sufficient for parthenogenetic division of egg cells in lettuce. The assembly of the *PAR* locus showed that the accumulation of TEs expanded the Apo haplotype compared with the sexual haplotypes. This can be explained by differences in evolutionary history: the Apo haplotype has been fixed in successive clonal generations since its origin (roughly between 5 and 10 million years BP (years before the present)), whereas the sex haplotypes were recently introduced by backcrossing from sexual gene pools[20]. Similar to sex chromosomes, the Apo haplotype has not undergone recombination for an extended time. In contrast to sex chromosomes, the genes in the Apo haplotype appear functional with no premature stops and low rates of nonsynonymous substitution.

How has sexual development been modified to allow parthenogenesis in apomictic dandelion? First and foremost, it has been coopted by ectopic expression of the *PAR* gene in the female gametophyte (Fig. 4c). From the complementation of *Taraxacum LOP* mutants it can be concluded that the *Taraxacum PAR* promoter driving the sexual lettuce gene can induce parthenogenesis. There are two *PAR* homologs in *Arabidopsis* spp.: *DAZ3* (DUO1 activated zinc finger 3 (ZF3); At4g35700) and *DAZ3like*/*TREE1* (transcriptional repressor of EIN3-dependent ethylene-response 1; At4g35610)[27,35–37]. Similar to *PAR*, these encode small proteins (275 and 271 amino acids, respectively) with a single ZF of the rare K2-2 class, but with two, rather than one, EAR motifs in the C terminus. The EAR motif suggests that they are repressors, which has recently been demonstrated in *Arabidopsis* spp.[36]. *DAZ3* and *TREE1* both directly interact with *EIN3* (ethylene-insensitive 3). *TREE1* directly binds a DNA motif found in the promoters of hundreds of genes downregulated in response to ethylene, and *daz3 tree1* double mutants fail to repress these genes in response to ethylene[36]. *DAZ3* and *TREE1* are among the highest expressed genes in *Arabidopsis* sperm cells[38,39], but are not expressed in the vegetative pollen cell[37]. Both gene promoters contain a MYB DNA-binding motif (AACCGC), required for the binding of the DUO1 transcription factor that controls sperm cell development[35]. *DAZ3* expression is not detected in egg cells, messenger RNA is detected in zygotes 14 h after fertilization, but not 24 h after fertilization, suggesting that, after delivery from the sperm, it is rapidly degraded well before the first embryo division[40].

In apomictic dandelion, the *PAR* gene is expressed in egg cells, whereas, in sexual dandelion, *par* genes are highly expressed in pollen, similar to *DAZ3* and *TREE1* in *Arabidopsis* spp. DNA sequences 5′ of the sexual *par* genes in dandelion, hawkweed and lettuce are highly conserved up to about 200 bp upstream of the ATG (Supplementary Fig. 19 and Supplementary Table 15) and contain a core MYB DNA-binding motif (TAACCGCC), which probably serves as a DUO1-binding site. In both the *ToPAR* and *PpPAR* alleles, the MITE insertions are located just 5′ of the conserved TAACCGCC motif, and the insertions themselves are more than six times longer than the conserved region. Despite the insertions, the sequences 5′ and 3′ of the MITE insertion have high identity with the respective sexual alleles, indicating functional constraint. It is well established that transposon insertions can act as controlling elements of genes and can cause phenotypic changes in plants[41–44]. We conclude that MITE insertions in the promoter of the *PAR* alleles of both species probably caused a shift in the mode of reproduction to parthenogenetic embryo development. The position of the MITE upstream of *PAR* could impact expression as *cis*-regulatory elements in two ways: either by introducing a cryptic female gametophyte-specific enhancer element within the MITE or

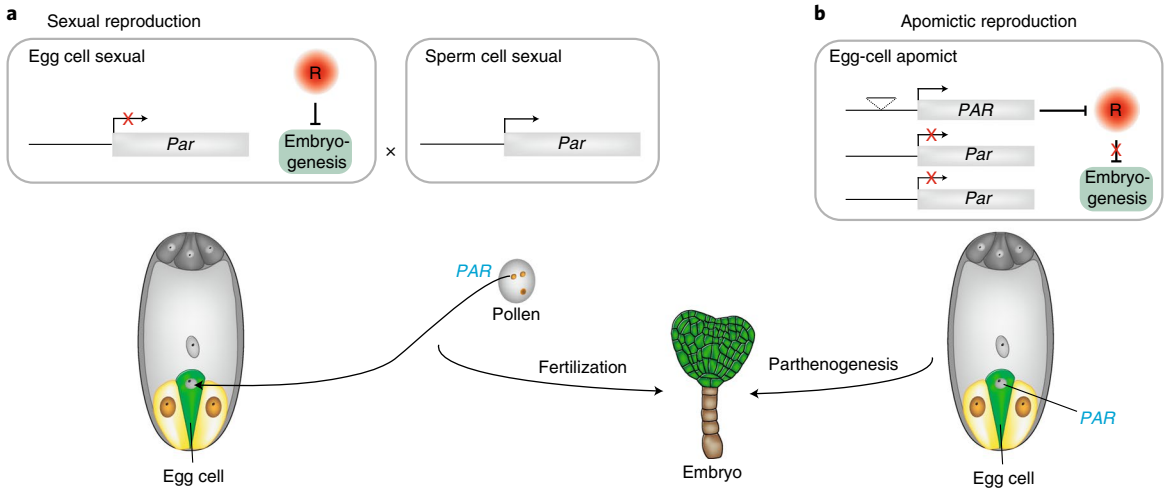


Fig. 7 | A model of the possible function of PAR in *Taraxacum* spp. as a repressor of an unknown gene that suppresses the embryo developmental program in the egg cell. a, In sexual dandelions the maternal copy of *par* is not expressed in the egg cell, whereas *par* is expressed in dandelion pollen. *Arabidopsis* homologs (*DAZ3* and *TREE1*) are highly expressed in *Arabidopsis* sperm, suggesting delivery by male gametes on fertilization. We propose that, in sexual reproduction, paternal PAR protein suppresses inhibitors of the embryo developmental program. **b**, In polyploid apomicts the *PAR* allele is expressed, whereas the recessive *par* alleles are silent, in the egg cell. The PAR protein represses inhibitors of embryogenesis in the egg cell, meaning that fertilization is not required.

by disrupting a repressive regulatory element that normally keeps *PAR* off in the female gametophyte.

Although both dandelion and hawkweed contain a similar *hAT*-derived large MITE upstream of the dominant *PAR* allele, the insertion sites and internal sequences differ considerably. This implies that the two MITE insertions were independent ancient events. If so, this represents a remarkable case of parallel evolution of regulatory sequences[45–47]. Although both species are within the Cichorioideae subfamily of the Asteraceae, they occur in different subtribes that split between 13.5 and 20.8 million years ago[30,34]. They also differ fundamentally in the way female meiosis is avoided to achieve functional apomixis (Supplementary Note 1). Despite this, two species seem to have recruited the same gene and *cis*-regulatory element to achieve parthenogenesis.

Over the last century, possible applications of apomixis in agriculture have been recognized[9,48,49] and recent innovations[50,51] suggest that synthetic apomixis on the field is becoming a realistic prospect. Expression of *PsBBML* in egg cells has been observed in the monocot apomict *Pennisetum squamulatum*, and *PsBBML* is an efficient haploid inducer in pearl millet, rice and maize[52,53], but only at low frequencies in dicot tobacco[54]. Artificial expression of *OsBBM1*, a paternally expressed gene and the closest rice homolog of *PsBBML*, in rice egg cells can efficiently trigger parthenogenesis[51]. In the present study, we identified a distinct pathway that can trigger parthenogenetic division of egg cells in dicot dandelion. We propose that expression of dandelion *PAR* in the egg cell of apomicts mimics its deposition by sperm cells in sexual plants and that its presence in egg cells leads to the repression of inhibitors of embryogenesis, triggering cell division without fertilization (Fig. 7). Akin to the work on *BBM* genes, driving the expression of a *PAR* gene in dandelion egg cells can bring about parthenogenesis. Thus, there are clear similarities in the organization of parthenogenesis in *P. squamulatum* and dandelion, yet the underlying genes, and their respective roles as activator and repressor, are different. To date, parthenogenesis remains a limiting factor in dicot crops for the implementation of synthetic apomixis. Therefore, the modulation of *PAR* expression, and that of homologous genes in sexual crop species, could potentially contribute to the generation of apomictic crops.

Online content

Any methods, additional references, Nature Research reporting summaries, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00984-y.

Received: 19 November 2020; Accepted: 3 November 2021;
Published online: 06 January 2022

References

1. Nogler, G. A. *Embryology of Angiosperms* 475–518 (Springer Berlin, 1984). https://doi.org/10.1007/978-3-642-69302-1_10
2. Van Dijk, P. J. & Ellis, T. H. N. The full breadth of Mendel's genetics. *Genetics* **204**, 1327–1336 (2016).
3. Mendel, G. Über einige aus künstlicher Befruchtung gewonnenen Hieracium-Bastarde. *Verhandlungen des. naturforschenden Vereines Br. ünn* **8**, 26–31 (1870).
4. Bicknell, R., Catanach, A., Hand, M. & Koltunow, A. Seeds of doubt: Mendel's choice of *Hieracium* to study inheritance, a case of right plant, wrong trait. *Theor. Appl. Genet.* **129**, 2253–2266 (2016).
5. Van Dijk, P. J. and Vijverberg, K. *Regnum Vegetabile* vol. 143 (Koeltz Scientific Books, 2005).
6. Mogie, M. *The Evolution of Asexual Reproduction in Plants* (Chapman & Hall, 1992).
7. van Dijk, P. J. Ecological and evolutionary opportunities of apomixis: insights from *Taraxacum* and *Chondrilla*. *Philos. Trans. R. Soc. London B Biol. Sci.* **358**, 1113–1121 (2003).
8. Spillane, C., Curtis, M. D. & Grossniklaus, U. Apomixis technology development—virgin births in farmers' fields? *Nat. Biotechnol.* **22**, 687–691 (2004).
9. Vielle Calzada, J. P., Crane, C. F. & Stelly, D. M. Apomixis: the asexual revolution. *Science* **274**, 1322–1323 (1996).
10. Van Dijk, P. J., Rigola, D. & Schauer, S. E. Plant breeding: surprisingly, less sex is better. *Curr. Biol.* **26**, R122–R124 (2016).
11. Ozias-Akins, P. & van Dijk, P. J. Mendelian genetics of apomixis in plants. *Annu. Rev. Genet.* **41**, 509–537 (2007).
12. Wang, J. et al. Sequencing papaya X and Y h chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc. Natl Acad. Sci. USA* **109**, 13710–13715 (2012).
13. Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124 (2013).
14. Charlesworth, D., Charlesworth, B. & Marais, G. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95**, 118–128 (2005).

15. Muller, H. J. The relation of recombination to mutational advance. *Mutat. Res.* **1**, 2–9 (1964).
16. Felsenstein, J. The evolutionary advantage of recombination. *Genetics* **78**, 737–756 (1974).
17. Kondrashov, A. S. Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440 (1988).
18. Nuzhdin, S. V. & Petrov, D. A. Transposable elements in clonal lineages: lethal hangover from sex. *Biol. J. Linn. Soc.* **79**, 33–41 (2003).
19. Arkhipova, I. & Meselson, M. Deleterious transposable elements and the extinction of asexuals. *BioEssays* **27**, 76–85 (2005).
20. Van Dijk, P., de Jong, H., Vijverberg, K. & Biere, A. *Lost Sex* 475–493 (Springer Netherlands, 2009). https://doi.org/10.1007/978-90-481-2770-2_22
21. Hand, M. L. & Koltunow, A. M. G. The genetic control of apomixis: asexual seed formation. *Genetics* **197**, 441–450 (2014).
22. Van Dijk, P. J. & Bakx-Schotman, J. M. T. Formation of unreduced megaspores (diplospory) in apomictic dandelions (*Taraxacum officinale*, s.l.) is controlled by a sex-specific dominant locus. *Genetics* **166**, 483–492 (2004).
23. Vijverberg, K., Ozias-Akins, P. & Schranz, M. E. Identifying and engineering genes for parthenogenesis in plants. *Front. Plant Sci.* **10**, 128 (2019).
24. Van Dijk, P. J., Camp, R. O., Den & Schauer, S. E. Genetic dissection of apomixis in dandelions identifies a dominant parthenogenesis locus and highlights the complexity of autonomous endosperm formation. *Genes* **11**, 961 (2020).
25. Vijverberg, K., Milanovic-Ivanovic, S., Bakx-Schotman, T. & van Dijk, P. J. Genetic fine-mapping of DIPLOSPOROUS in *Taraxacum* (dandelion; Asteraceae) indicates a duplicated DIP-gene. *BMC Plant Biol.* **10**, 154 (2010).
26. Van Dijk, P. J., Rigola, D., Prins, M. W. & Van Tunen, A. J. Diplospory gene (Patent application no. WO2017039452 A1) (2017).
27. Englbrecht, C. C., Schoof, H. & Böhm, S. Conservation, diversification and expansion of C2H2 zinc finger proteins in the *Arabidopsis thaliana* genome. *BMC Genom.* **5**, 39 (2004).
28. Kagale, S. & Rozwadowski, K. Small yet effective: the ethylene-responsive element binding factor-associated amphiphilic repression (EAR) motif. *Plant Signal. Behav.* **5**, 691–694 (2010).
29. Yang, J. et al. PlantEAR: functional analysis platform for plant EAR motif-containing proteins. *Front. Genet.* **9**, 590 (2018).
30. De La Torre, A. R., Li, Z., Van De Peer, Y. & Ingvarsson, P. K. Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. *Mol. Biol. Evol.* **34**, 1363–1377 (2017).
31. Catanach, A. S., Erasmuson, S. K., Podivinsky, E., Jordan, B. R. & Bicknell, R. Deletion mapping of genetic regions associated with apomixis in *Hieracium*. *Proc. Natl Acad. Sci. USA* **103**, 18650–18655 (2006).
32. Bräuning, S., Catanach, A., Lord, J. M., Bicknell, R. & Macknight, R. C. Comparative transcriptome analysis of the wild-type model apomict *Hieracium praealtum* and its loss of parthenogenesis (lop) mutant. *BMC Plant Biol.* **18**, 206 (2018).
33. Piosik, Zenkteler, E. & Zenkteler, M. Development of haploid embryos and plants of *Lactuca sativa* induced by distant pollination with *Helianthus annuus* and *H. tuberosus*. *Euphytica* **208**, 439–451 (2016).
34. Verzegnazzi, A. L. et al. Major locus for spontaneous haploid genome doubling detected by a case–control GWAS in exotic maize germplasm. *Theor. Appl. Genet.* **134**, 1423–1434 (2021).
35. Borg, M. et al. The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in *Arabidopsis*. *Plant Cell* **23**, 534–549 (2011).
36. Wang, L., Ko, E. E., Tran, J. & Qiao, H. TREE1-EIN3–mediated transcriptional repression inhibits shoot growth in response to ethylene. *Proc. Natl Acad. Sci. USA* **117**, 29178–29189 (2020).
37. Taimur, N. *Genetic and molecular mechanisms mediating sperm cell formation in Arabidopsis thaliana* (PhD Thesis, Univ. of Leicester) https://leicester.figshare.com/ndownloader/files/18288044 (2014).
38. Borges, F. et al. Comparative transcriptomics of arabidopsis sperm cells. *Plant Physiol.* **148**, 1168–1181 (2008).
39. Borg, M. et al. Targeted reprogramming of H3K27me3 resets epigenetic memory in plant paternal chromatin. *Nat. Cell Biol.* **22**, 621–629 (2020).
40. Zhao, P. et al. Two-step maternal-to-zygotic transition with two-phase parental genome contributions. *Dev. Cell* **49**, 882–893.e5 (2019).
41. Underwood, C. J., Henderson, I. R. & Martienssen, R. A. Genetic and epigenetic variation of transposable elements in *Arabidopsis*. *Curr. Opin. Plant Biol.* **36**, 135–141 (2017).
42. Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86 (2016).
43. McClintock, B. Controlling elements and the gene. *Cold Spring Harb. Symp. Quant. Biol.* **21**, 197–216 (1956).
44. Ong-Abdullah, M. et al. Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature* **525**, 533–537 (2015).
45. Christin, P. A., Weinreich, D. M. & Besnard, G. Causes and evolutionary significance of genetic convergence. *Trends Genet.* **26**, 400–405 (2010).
46. Stern, D. L. The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751–764 (2013).
47. Rosenblum, E. B., Parent, C. E. & Brandt, E. E. The molecular basis of phenotypic convergence. *Annu. Rev. Ecol. Evol. Syst.* **45**, 203–226 (2014).
48. Hanna, W. W. & Bashaw, E. C. Apomixis: its identification and use in plant breeding 1. *Crop Sci.* **27**, 1136–1139 (1987).
49. Karpechenko, G. D. Experimental polyploidy and haploidy. *Teor. Osn. Sel. Rastenii, Obqaja Sel.* **I**, 397–434 (Sel'khozgiz, 1935).
50. d'Erfurth, I. et al. Turning meiosis into mitosis. *PLoS Biol.* **7**, e1000124 (2009).
51. Khanday, I., Skinner, D., Yang, B., Mercier, R. & Sundaresan, V. A male-expressed rice embryogenic trigger redirected for asexual propagation through seeds. *Nature* **565**, 91–95 (2019).
52. Conner, J. A., Mookkan, M., Huo, H., Chae, K. & Ozias-Akins, P. A parthenogenesis gene of apomict origin elicits embryo formation from unfertilized eggs in a sexual plant. *Proc. Natl Acad. Sci. USA* **112**, 11205–11210 (2015).
53. Conner, J. A., Podio, M. & Ozias-Akins, P. Haploid embryo production in rice and maize induced by PsASGR-BBML transgenes. *Plant Reprod.* **30**, 41–52 (2017).
54. Zhang, Z., Conner, J., Guo, Y. & Ozias-Akins, P. Haploidy in tobacco Induced by PsASGR-BBML transgenes via parthenogenesis. *Genes* **11**, 1072 (2020).

Methods

Plant materials. The apomictic triploid dandelion clone that was used in the present study (A68) was collected in a meadow in Heteren, the Netherlands. Other dandelion materials used in the present study are listed in Supplementary Table 1. Lettuce plants used for transformation were of the iceberg type, (cv. Legacy, Takii & Co. Ltd). Plants were grown in the greenhouse (16 h day:8 h night). The tetraploid, apomictic *P. piloselloides* wild-type 'R35' was used in this study along with polyhaploids derived from R35 (Supplementary Note 12).

Deletion irradiation mutagenesis and mapping. To induce deletions, three batches of 2,000 seeds of the triploid apomict A68 received a dose of, respectively, 250, 300 or 350 Gy of γ-radiation from a cobalt-60 source at Synergy Health Ede B.V. Ede, the Netherlands. These dosages were based on a median lethal dose (LD_{50}) test series (Supplementary Note 4). Plants with the *LOP* phenotype were test-crossed with a diploid FCH72 pollen donor and the ploidy level of the progeny was determined by flow cytometry (Ploidy Analyser, Partec) as described[55]. Chimerism of *LOP* deletion mutants was removed by in vitro regeneration from leaf explants as described[56]. Bulked deletion analysis (Supplementary Fig. 4) found two new *PAR* deletion AFLP fragments[57], which were converted into sequences according to a published method[58]. These sequences were used to screen the A68 BAC library (Supplementary Note 6). For the mapping of *Pilosella PpPAR* to the *PAR* locus, see Supplementary Note 13.

***Taraxacum* sequencing.** DNA was extracted from young leaf material of the apomictic A68 line according to a CTAB (cetyltrimethylammonium bromide) DNA extraction protocol[59]. One-half of the DNA was sized on the BluePippin system (Sage Science) to remove small fragments, and sequenced on three ONT PromethION flow cells. The nonsized DNA was sequenced on two ONT PromethION flow cells. Over the five ONT PromethION flow cells, a total of 140.2 Gb was sequenced (175× coverage of the 800-Mb estimated haploid genome size of *T. officinale*). The pore version used was R9.4.1 and the PromethION release version was 19.05.1.

In addition, A68 DNA was used to prepare an Illumina overlap library. DNA was sheared using a Covaris M220 (Covaris, Inc.) to 450 bp. The sheared DNA was sized using a Pippin prep system (Sage Science) and used to prepare a PCR-free Illumina sequencing library. The library was sequenced on the Illumina NovaSeq6000 platform using a SP flow cell and 2×250 bp protocol.

Genome assembly and annotation. A selection of the longest sequence reads that together represent 90× haploid genome coverage were assembled using Minimap2 (v.2.11-r797, with settings –m 1600, -K2G, -I8G)[60] and Miniasm (v.0.2-r137-dirty, with settings –R –c 2 –m 500 –s 4000;)[61]. A consensus sequence was generated through three iterations of Racon (v.1.3.1, default settings)[62], using all ONT sequence reads. The Illumina read pairs from the overlap library were merged using FLASH (v.1.2.11, default settings)[63]. The merged reads were subsequently used to polish the consensus through three rounds of BWA mem (v.0.7.17; default parameters)[64] and Pilon (v.1.22, default settings)[65], resulting in the final assembly.

From the final assembly the contigs matching the *PAR* locus were identified by homology to the previously sequenced BACs (Supplementary Note 5). For these contigs, structural gene prediction and transcript alignments were integrated with new gene prediction and supplied to EvidenceModeler v.1.1.1 (default settings), which is able to provide consensus predictions based on gene predictions from first principles[66]. The genes from the gene prediction results served as the input for the functional annotation pipeline. First, the predicted proteins were merged with all plant entries from the Uniprot database and clustered with the CD-HIT tool[67]. To facilitate the CD-HIT cluster analysis, protein sequences from each CD-HIT cluster were aligned using CLUSTAL omega[68] and processed by TrimAl[69]. Last, RaxML[70] was used to reconstruct phylogenetic trees. Next, BLAST hits using the NCBI NR protein database were obtained as well as protein domains that are predicted using the InterProScan software[71]. To enable a fast way of identifying the putative function of a gene, the Automated Human Readable Descriptions pipeline (see Supplementary URLs) was applied to the protein set. For *Pilosella* sequencing, see Supplementary Note 14.

Haplotype maps, *Ks* and *Ka* calculations and phylogenies. Haplotype synteny maps were generated with Mauve software[72]. For 20 genes in the *PAR* locus, the DNA sequences of the three alleles were aligned using the Muscle Codon function of MEGA-X v.10.0.5 (ref. [73]). The aligned sequences were imported into DnaSP6 (ref. [74]) v.6.12.03 and nucleotide diversity (π, Jukes & Cantor), *Ka* and *Ks* were calculated. Homologous protein sequences from different species were retrieved by BLAST searches in NCBI. Proteins were aligned using the Muscle function of MEGA-X and phylogenetic trees were constructed with the same program. The divergence time was calculated as $T = Ks/(2r)$, where r is the number of synonymous substitutions per synonymous site per year.

LD mapping. A total of five primer pairs was designed distributed over the 400-kb Apo haplotype with an additional two in the upstream region of *A_g295* (Supplementary Table 10). PCRs were performed on 3 plants, each from a panel of 17 dandelion populations representing 5 sections and a wide distribution range over Europe and Asia (Fig. 2c and Supplementary Table 1). Plants were germinated and a subset kept in the greenhouse under natural, frost-free conditions. DNAs were isolated from ~5 mg of either seedlings or leaf tips using the CTAB method as previously described[75]. PCR reactions followed the latter and products were analyzed on 1.5% agarose gels.

CRISPR–Cas9 gene editing. CRISPR–Cas9 constructs used were based on an *Arabidopsis* codon-optimized Cas9 variant driven by a parsley ubiquitin promoter and a BASTA/PPT plant selection gene[76], and an *Arabidopsis* U6 promoter drove the expression of the gRNAs[77]. Each guide was designed to be allele specific so that it targeted the gene on the apomictic haplotype, but not on the counterpart sexual haplotypes (Supplementary Table 6 shows all gRNA sequences). The constructs were transformed into the *T. officinale* apomictic line A68 by *Agrobacterium tumefaciens* infection of leaf explants. For genotyping of mutations induced by CRISPR–Cas9, gene-specific primers were designed and used to amplify the gRNA target regions (Supplementary Table 9). Amplicons were barcoded per plant and sequenced on the Illumina MiSeq platform. The sequencing data were aligned to reference sequences in CLC bio and variants were identified to determine mutations induced by CRISPR–Cas9. The identified mutations and corresponding seed-head/crossing phenotypes for plants with gRNA9 and gRNA10 are described in Supplementary Tables 7 and 8, respectively.

Comparative transcriptomics of apomictic, *PARTHENOGENESIS*-deleted and sexual dandelion female gametophytic cell types. A total of 27 samples was analyzed, including triplicates of EA and CC samples from mature embryo sacs and singletons of EA and whole gametophyte (WG) samples from young embryo sacs, each with three genotypes: the triploid apomict A68 (Apo), tetraploid *PAR*-deleted A68-i34×diploid sexual FCH72 (*LOP*) and FCH72 (sex). Tissue embedding of dandelion buds and laser-assisted microdissection of the different cell types was carried out following a published protocol[78] with minor modifications (Supplementary Note 9).

A total of 150–200 (young) to 200–300 (mature embryo sac) cell sections were isolated from one flower head per sample, corresponding to 40–75 whole cells (Fig. 4a,b and Supplementary Fig. 7). PicoPure isolated RNAs were linearly amplified according to the CEL-seq and CEL-seq2 protocols[79,80], which combine poly(A)-based amplification in the first round with random hexamer-primed PCR amplification in the second round (Supplementary Table 10). Two samples were pooled before amplification, resulting in 10–11 µl from 5–14 ng µl^{-1} of 3′-biased short fragment libraries (Supplementary Fig. 8). These were used to prepare equimolar pools for 125-nt Paired End Illumina HiSeq2500 sequencing over a total of four lanes. Around 20–25 million raw R1-reads and raw R2-reads (MRds) were delivered as FASTQ files per library (Supplementary Table 11) and are deposited in the European Nucleotide Archive (ENA) Sequence Read Archive (SRA) database with accession no. PRJEB40645. Raw reads were trimmed, cleaned and analyzed in CLC Genomics Workbench v.12 and v.20 (QIAGEN), with only R2-reads used for mapping (Supplementary Note 9). A total of ~16 MRds was used for WG and CC samples and ~24 MRds for EA samples, and the results were interpreted without normalization. Forward mapping with high stringency (similarity fraction = 0.98) was performed to detect allele specificity, while allowing a low level of sequence errors.

Mappings were performed on the ~400-kb sequences of the Apo haplotype combined with the 2 ~215-kb sequences of the 2 sex haplotypes, covering a total of 100 genes in the Apo haplotype (Supplementary Table 12). Unique mapping output was interpreted for: (1) genes expressed in Apo samples, defined as present in at least two of the nine samples with a total of at least three reads; (2, 3) genes relatively highly (>33%) expressed in the 9 Apo samples compared with the total of 27 samples for the Apo haplotype (2) or Apo and 2 sex haplotypes (3); and (4) unique (U) or allele-specific (A) expression in Apo samples only (Fig. 4c and Supplementary Table 12).

In situ hybridization. In situ hybridizations were performed on 10- to 12-µm sections of paraffin-embedded dandelion buds, using digoxigenin UTP-labeled gene sequences of *A_g295* and *A_g490* as probes and the dandelion homologs of egg cell 1.1 (ref. [81]) (Tof-*EC*1.1) as a positive control (Supplementary Note 10). The results are presented in Supplementary Fig. 10.

RT-PCR analysis of gene expression in pollen grains. Total RNA from pollen grains and leaves was isolated using Maxwell RSC Plant RNA kit (Promega). RNA was assessed for quality on a chip-based capillary electrophoresis (Agilent, 2100 Bioanalyzer). Thereafter, complementary DNA was synthesized from 0.4 µg of total RNA with an iScript cDNA synthesis kit (BioRad) following the manufacturer's protocol. RT-PCR reactions were performed using Phusion Flash Master Mix (Thermo Fisher Scientific) and the cycling conditions were: 98 °C for 30 s; 40 cycles of 98 °C for 10 s, 56 °C for 30 s, 72 °C for 15 s; and 72 °C for 5 min. *Eukaryotic initiation factor 4a* (*EIF4a*) was used as a reference gene, along with genomic DNA and water controls. Two biological and three technical replicates were performed.

Sequencing of the MITE insertion site. Five apomictic (*T. richardsianum*, *T. albidum*, *T. brevicorniculatum*, *T. brachyglossum* and *T. gratum*) and two sexual (*T. cylleneum* and *T. koksaghyz*) species, belonging to different sections, were grown

and their reproduction mode confirmed by phenotyping. DNA was extracted from leaves using QIAGEN DNeasy Plant DNA extraction kit and quantified with Qubit (Invitrogen, Thermo Fisher Scientific). DNA was sheared using a Covaris M220 (Covaris, Inc.) to 580–600 bp and PCR-free Illumina sequencing libraries were prepared. Libraries were sequenced across eight lanes on the Illumina HiSeq2500 platform at KeyGene, using the run protocol 126+6+126 (V4 chemistry). The variable genome size of the *Taraxacum* germplasm (see Supplementary Table 14) was taken into account by normalizing the library sample concentrations to obtain an average of 17× coverage per allele. The raw reads of each *Taraxacum* sp. were mapped using BWA mem[64] (v.0.7.17-r1194-dirty) with default settings to the following sequence references: APOhaplotype_contig, sex1haplotype_Contig and sex2haplotype_Contig. Reads were kept if the alignment coverage was at least 95%, the alignment similarity was at least 95% and the alignment coverage of its paired mate was at least 95%. Finally, SAMtools (v.1.1) was used to convert the *PAR* MITE-selected sequence region output into a sorted BAM file for visual inspection of the data. Using CLC Genomic workbench 12.02, the retrieved reads were first assembled within each species, then the consensus sequences of the germplasm set were aligned.

Synteny maps of the *Taraxacum*, *Pilosella* and *Lactuca PAR* regions. BLASTP (v.2.2.31+) (e value: 1×10^{-5}) was applied to compare genes from the sex1 locus and *L. sativa* LG8 with genes from the *APO* locus[72], and genes from the *PpPAR* locus with genes from *L. sativa* LG8. The top BLASTP hits were extracted for subsequent synteny detection. GFF files were modified and merged in before synteny search. McScanX (collinear block size = 5, maximum gaps = 25) was used to scan the synteny (putative homologous chromosomal regions) between *L. sativa* LG8 and *APO* haplotype, between the dandelion sex1 haplotype and *APO* haplotype separately[73], and between *L. sativa* LG8 and *PpPAR* locus. Next, the syntenic blocks containing the target *PAR* gene (syntenlog: syntenic homolog) were selected to generate the microsynteny plot using the JCVI toolkit[74,75]. For comparison without synteny, best BLASTP hits were additionally added to the microsynteny plot to illustrate the homologous relationship.

Complementation transformation *T. officinale*. To generate the *pEC1.1::PAR* complementation construct for transformation to dandelion, the *PAR* gene (genomic coding sequence) was synthesized so that it was immune to the gRNA by using alternative codons but maintaining the original translated protein sequences; 1,000 bp of the 3′-regulatory sequence of *PAR* was included. The EC1.1 promoter from *A. thaliana*[81] and the gRNA-immune *PAR* gene were subcloned into pDONR P4-P1r and pDONR221 vector, respectively. Thereafter, the pEC1.1 and *PAR* entry clones were recombined into the destination vector pK7m24GW,3 via an LR reaction. To generate the *PAR::Lssex* construct, the *PAR* promoter (−1,900 bp to −1 bp) was cloned into pDONR P4-P1r, and the lettuce-coding sequence fused with 1,000 bp of the 3′-regulatory sequence of *PAR* was synthesized and subcloned into pDONR221. An LR reaction recombined these sequences into pK7m24GW,3.

The *T. officinale* apomictic transgenic line 9–4–2 of the CRISPR–Cas9 *LoP* mutants (Fig. 3d) was crossed with FCH72 to produce seeds (as shown in Fig. 3f). The resulting tetraploid CRISPR–Cas9 *LoP* mutant offspring was transformed with the *pEC1.1::PAR and PAR::Lssex* construct by *A. tumefaciens* infection of leaf explants. Primary transgenic lines were tested for complementation construct presence by PCR and ploidy was confirmed by flow cytometry. Plants were phenotyped by evaluating seed heads for seed setting. A subset of the seeds was germinated, checked for the presence of the *DIP* gene by PCR. and ploidy of a subset of plants was tested by flow cytometry. The *DIP* gene was detected by PCR amplification with Phusion Flash Master Mix (Thermo Fisher Scientific). The cycling conditions were: 98 °C for 30 s; 35 cycles of 98 °C for 10 s, 56 °C for 20 s, 72 °C for 1 min; and 72 °C for 5 min. The primers used for genotyping are listed in Supplementary Table 10.

Transformation of lettuce with *PAR*. To generate a *pEC1.1::PAR* construct for transformation to lettuce, the genomic coding sequence (CDS) of *PAR*, including 250 bp of the 3′-regulatory region of *PAR*, was subcloned into pDONR221. Thereafter, the pEC1.1 (see Complementation transformation T. officinale) and *PAR* entry clones were recombined into the destination vector pK7m24GW,3 via an LR reaction. The *pEC1.1::PAR* construct was introduced into lettuce (cv Legacy, Takii & Co. Ltd.) via *Agrobacterium* sp.-mediated transformation[82]. Analysis of *PARTHENOGENESIS* in transformed plants was performed 75 h after decapitation of the flower buds[83]. Isolated embryo sacs were directly collected in a drop of clearing/mounting solution (chloral hydrate:water:glycerol (w:v:v) 8:3:1) and without further manipulation seen by Nomarski differential interference contrast microscopy. Embryo ploidy was analyzed in pools of 10–15 embryo sacs, which were isolated from developing seeds approximately 5 d after decapitation (or 5 d after self-pollination for the wild-type control). Ploidy measurements and data analysis were performed on BD Accuri C6 flow cytometer (Supplementary Note 17).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Transcriptome data of 27 laser-assisted microdissected female gametophyte tissue samples of *T. officinale* are deposited in the ENA SRA database under accession no. PRJEB40645. Illumina HiSeq2500 paired end sequencing datasets of *T. richardsianum*, *T. albidum*, *T. brevicorniculatum*, *T. brachyglossum*, *T. gratum*, *T. cylleneum*, and *T. koksaghyz* are deposited in the ENA SRA database under accession no. PRJEB40739. The *T. officinale* ONT PromethION sequence reads, and the Illumina NovaSeq6000 reads from the overlap library are deposited in the ENA SRA database under accession no. PRJEB48186.

Code availability

Automated Human Readable Descriptions: https://github.com/groupschoof/AHRD

References

55. Van Baarlen, P., De Jong, H. J. & Van Dijk, P. J. Comparative cyto-embryological investigations of sexual and apomictic dandelions (*Taraxacum*) and their apomictic hybrids. *Sex. Plant Reprod.* **15**, 31–38 (2002).
56. Collins-Silva, J. et al. Altered levels of the *Taraxacum kok-saghyz* (Russian dandelion) small rubber particle protein, TkSRPP3, result in qualitative and quantitative changes in rubber metabolism. *Phytochemistry* **79**, 46–56 (2012).
57. Vos, P. et al. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**, 4407–4414 (1995).
58. Brugmans, B., van der Hulst, R. G. M., Visser, R. G. F., Lindhout, P. & van Eck, H. J. A new and versatile method for the successful conversion of AFLP markers into simple single locus markers. *Nucleic Acids Res.* **31**, e55–e55 (2003).
59. Doyle, J. & Doyle, J. Isolation of DNA from fresh tissue. *Focus* **12**, 13–15 (1990).
60. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
61. Li, H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**, 2103–2110 (2016).
62. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
63. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
64. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
65. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
66. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
67. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
68. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
69. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
70. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
71. Hunter, S. et al. InterPro: the integrative protein signature database. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkn785 (2009).
72. Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403 (2004).
73. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
74. Rozas, J. et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).
75. Vijverberg, K., Van der Hulst, R. G. M., Lindhout, P. & Van Dijk, P. J. A genetic linkage map of the diplosporous chromosomal region in *Taraxacum officinale* (common dandelion; Asteraceae). *TAG Theor. Appl. Genet.* **108**, 725–732 (2004).
76. Fauser, F., Schiml, S. & Puchta, H. Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant J.* **79**, 348–359 (2014).
77. Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D. G. & Kamoun, S. Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**, 691–693 (2013).

78. Wuest, S. E. et al. Arabidopsis female gametophyte gene expression map reveals similarities between plant and animal gametes. *Curr. Biol.* **20**, 506–512 (2010).
79. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-seq: single-cell RNA-seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).
80. Hashimshony, T. et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-seq. *Genome Biol.* **17**, 77 (2016).
81. Sprunck, S. et al. Egg cell—secreted EC1 triggers sperm cell activation during double fertilization. *Science* **338**, 1093–1097 (2012).
82. Curtis, I. S., Power, J. B., Blackhall, N. W., De Laat, A. M. M. & Davey, M. R. Genotype-independent transformation of lettuce using *Agrobacterium tumefaciens*. *J. Exp. Bot.* **45**, 1441–1449 (1994).
83. Koltunow, A. M., Bicknell, R. A. & Chaudhury, A. M. Apomixis: molecular strategies for the generation of genetically identical seeds without fertilization. *Plant Physiol.* **108**, 1345–1352 (1995).

Acknowledgements

We thank the following: T. Gerats (KeyGene) for essential vision and support; S. van Liere, S. Lecoulant, D. Valkenburg, A. Bergsma, M. Frescatada, W. van Rengs and C. Schol (all of KeyGene) for technical assistance; H. Schneiders, H. van der Poel, R. Hulzink, A. Wittenberg and A. Janssen (all of KeyGene) for DNA-sequencing and bioinformatics; S. Keshtkar, M. V. Boekschoten, H. Hackert, I. van der Meer, L. Berke, T. Zhao, H. van der Geest and J. van Haarst (all of Wageningen University & Research (WUR)) and J. A. Mol (Utrecht University) for advice and access to facilities; R. van den Bulk (WUR) for supporting institutional cooperation; R. Mank and KeyGene greenhouse staff for plant culturing; P. Bundock and M. de Both for critical reading and discussion of the manuscript; and J. Kirschner and J. Štěpánek (Institute of Botany, Průhonice), R. Vašut (Palacký University Olomouc) and R. van der Hulst (Solynta) and K. Verhoeven (Netherland Institute of Ecology) for supplying plant materials and taxonomic information. The *pUBI::Cas9* construct was a kind gift of H. Puchta. The Dutch Research Council Applied and Engineering Sciences grant (no. 13700 (ParTool) to K.V., C.O., M.B. and M.E.S.). The elucidation of the *Pilosella LOP* locus was supported by a basic science grant from the New Zealand Foundation for Science and Technology. The formation of the *Pilosella* BAC library was supported in part by the Arizona Genome institute.

Author contributions

P.J.V.D., K.V., R.H.M.O.d.C., C.J.U., D.R., S.O., M.P., A.V.T., R.B. and M.E.S. conceived and designed the project. C.J.U., K.V., D.R., S.O., C.O., T.R., J.F., K.J., S.M., M.B., R.B., A.C., S.E., C.W. and P.J.V.D. generated the experimental data. K.V., D.R., C.J.U., R.H.M.O.d.C., T.R., S.E.S., S.M., R.B., P.J.V.D. and M.E.S. analyzed the experimental data. D.R., K.V., S.M., W.X., E.D., K.N., E-J. B., C.J.U., P.J.V.D., R.B., M.E.S. and C.W. conducted the bioinformatic analysis. P.J.V.D., M.E.S., S.E.S., R.B. and C.J.U. wrote the manuscript with contributions from K.V., T.R., R.H.M.O.d.C. and D.R. All authors critically read and edited the manuscript.

Competing interests

D.R., P.J.V.D., R.H.M.O.d.C., M.P., A.J.V.T., T.R., J.F., K.J., S.M., E.D., K.N. and E-J.B. are employees of KeyGene N.V. C.J.U. is a former employee of KeyGene N.V. S.E.S. is an employee of KeyGene Inc. These authors declare that they are bound by confidentiality agreements that prevent them from disclosing their competing interests in this work. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00984-y.

Correspondence and requests for materials should be addressed to M. Eric Schranz or Peter J. van Dijk.

Peer review information *Nature Genetics* thanks Venkatesan Sundaresan, Thomas Dresselhaus and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

nature research

Corresponding author(s): Peter J. van Dijk & M. Eric Schranz

Last updated by author(s): Oct 28, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	As detailed in full in the online methods, for Taraxacum genome sequencing Illumina HiSeq 2500, NovaSeq6000 and ONT PromethION (release version 19.05.1 and pore version R9.4.1) were used. The Taraxacum female gametophytic cell type transcriptome libraries were 125 nt. Paired-End Illumina HiSeq2500 sequenced. Hieracium sequencing is described in the Supplementary Notes.
Data analysis	For Taraxacum genome assembly and annotation the following software was used: Minimap2 (v2.11-r797, with settings –m 1600, -K2G, -I8G); Miniasm (v0.2-r137-dirty, with settings –R –c 2 –m 500 –s 4000); Racon (version 1.3.1, default settings); FLASH (version 1.2.11, default settings); BWA mem (version 0.7.17; default parameters); Pilon (version 1.22, default settings); EvidenceModeler version 1.1.1 (default settings); CD-HIT (version 4.6); CLUSTAL omega (version 1.1.0); TrimAl (version 1.4.rev11); RaxML (version 7.0.4) ; BLAST-NCBI; InterProScan (version 5.30-69.0) and Automated Assignment of Human Readable Descriptions (AHRD, version 3.3.3, - https://github.com/groupschoof/AHRD). Hieracium sequence assembly is described in the Supplementary Notes. For analysis of Taraxacum re-sequencing data the following software was used: BWA mem (version 0.7.17-r1194-dirty, with default settings) SAMtools (version 1.1) and CLC genomic workbench (version 12.02). Synteny maps of the PAR Taraxacum haplotypes were made with MAUVE (version 2.3.1); alignments of coding sequences were made with the Muscle codon function of MEGA-X version 10.0.5. the same program was used for the construction of Maximum Likelihood phylogenetic trees; Ks, Ka and π (Jukes & Cantor) were calculated using DnaSP version 6.12.03. Synteny maps of the Taraxacum and Lactuca PAR regions were made with McScanX (version 2012) and visualized using JCVI toolkit (version 1.0.6). Taraxacum female gametophytic cell type transcriptome library sequences were processed and mapped by using CLC Genomic Workbench

versions 12 and 20. Mapping was done in the Apo- and two Sex-haplotypes of apomictic dandelion accession A68 described in this study.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Short sequence reads generated during the current study were deposited at the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home). Raw 125 nt. Paired-end Illumina HiSeq2500 sequencing reads including 27 Taraxacum female gametophytic cell type samples were deposited in the ENA SRA database under accession number PRJEB40645. These data are now public.

Illumina HiSeq 2500 paired-end sequencing datasets of T. richardsianum, T. albidum, T. brevicorniculatum, T. brachyglossum, T. gratum, T. cylleneum, and T. koksaghyz were deposited in the ENA SRA database under accession number PRJEB40739. These data are now public.

The T. officinale ONT PromethION sequence reads, and the Illumina NovaSeq6000 reads from the overlap library are deposited in the ENA SRA database under accession number PRJEB48186.

Supplemental Data Sets have been submitted with the manuscript as TXT files: 1. The sequence of the PAR haplotype; 2. The sequence of the sex1 haplotype; 3. The sequence of the sex2 haplotype; 4. The sequences of the 20 genes of the Ks and Ka analyses; 5. The sequences of the Taraxacum and Hieracium PAR alleles; 6. The sequences of the Taraxacum and Hieracium MITEs.

All other data generated or analysed during this study are included in this published article (and its supplementary information files).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For Linkage Disequilibrium analysis, five different Taraxacum species (sections) were analysed comprising 10 apomictic and 7 sexual accessions. These accessions represented a wide taxonomic and geographic range and are representative of the Taraxacum genus. To investigate the conservation of the MITE insertion site in Taraxacum, the PAR promoter regions of two sexual and six apomictic species (sections) were resequenced. The transcriptomes of female gametophytes of three genotypes were compared (apomictic accession A68, the sexual accession FCH72 and the hybrid between the PAR deletion line A68-i34 and FCH72). The sample size was chosen so that one sexual, one apomictic and one apomictic mutant (reverting to sexuality) were included.
Data exclusions	No data were excluded.
Replication	CRISPR/Cas9 targeted mutagenesis: Six independent gRNA9 with a DNA footprint in the A_295 allele had the Loss-of-Parthenogenesis phenotype. Two independent gRNA10 with a DNA footprint in the A_295 allele had a LoP phenotype. At least three seed-heads were germinated per transformation line or control line and each replication gave consistent results. Transcriptomics female gametophytes: Nine samples for each of the three genotypes were compared in this study. Biological replicates from the same genotype and sample type gave consistent results. Lettuce transformations: Seven independent lettuce T0 lines containing the transgene pEC1::PAR construct were evaluated for the occurrence of parthenogenesis and gave consistent results. Briefly, per line at least three, but often more than six decapitated flower buds were screened for the presence of, and contained, autonomous embryo-like structures.
Randomization	In all experiments plants were grown in consistent conditions, with appropriate controls grown side-by-side. Location of plants on the greenhouse tables was random.
Blinding	The forward genetic screen for the Par locus was a blind, unbiased study that led to the identification of the Par locus. Further work focused on finding the causal gene by application of CRISPR/Cas9 and functional genomic approaches.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging