Article

Harbinger transposon insertion in ethylene signaling gene leads to emergence of new sexual forms in cucurbits

Received: 12 December 2023

Accepted: 28 May 2024

Published online: 07 June 2024

Check for updates

Hsin-Ya Huang $\mathbb{O}^{1,3}$, Siqi Zhang $\mathbb{O}^{1,3}$, Fadi Abou Choucha $\mathbb{O}^{1,3}$, Marion Verdenaud¹, Feng-Quan Tan¹, Clement Pichot¹, Hadi Shirazi Parsa¹, Filip Slavkovic \mathbb{O}^{1} , Qinghe Chen¹, Christelle Troadec \mathbb{O}^{1} , Fabien Marcel¹, Catherine Dogimont \mathbb{O}^{2} , Leandro Quadrana¹, Adnane Boualem \mathbb{O}^{1} & Abdelhafid Bendahmane \mathbb{O}^{1}

In flowering plants, the predominant sexual morph is hermaphroditism, and the emergence of unisexuality is poorly understood. Using Cucumis melo (melon) as a model system, we explore the mechanisms driving sexual forms. We identify a spontaneous mutant exhibiting a transition from bisexual to unisexual male flower, and identify the causal mutation as a Harbinger transposon impairing the expression of Ethylene Insensitive 2 (CmEIN2) gene. Genetics and transcriptomic analysis reveal a dual role of CmEIN2 in both sex determination and fruit shape formation. Upon expression of CmACS11, EIN2 is recruited to repress the expression of the carpel inhibitor, CmWIP1. Subsequently, EIN2 is recruited to mediate stamina inhibition. Following the sex determination phase, EIN2 promotes fruit shape elongation. Genome-wide analysis reveals that Harbinger transposon mobilization is triggered by environmental cues, and integrates preferentially in active chromatin, particularly within promoter regions. Characterization of a large collection of melon germplasm points to active transpositions in the wild, compared to cultivated accessions. Our study underscores the association between chromatin dynamics and the temporal aspects of mobile genetic element insertions, providing valuable insights into plant adaptation and crop genome evolution.

Flowering plants exhibit a remarkable diversity of reproductive strategies, ranging from asexual to crossbreeding, each strategy influencing the genetic landscape and evolutionary trajectory of plant populations. One crucial aspect of this reproductive diversity lies in the development of unisexual flowers, a phenomenon intricately linked to sex determination mechanisms¹. In dioecious species, plants bear either male or female flowers. In monoecious species, separate male and female flowers develop on the same individual. Other intermediary systems also exist. Andromonoecious plants bear both male and bisexual flowers on the same plants, and

gynoecious and androecious plants bear only female and male flowers, respectively^{1,2}.

Two models have been proposed to explain the transition to unisexuality. The first model, known as sex allocation theory, posits that ecological and physiological factors favor directing resources toward a single sexual function rather than both^{3,4}. The second model relies on population genetic arguments, suggesting that dioecy becomes advantageous in the presence of inbreeding depression^{5,6}. Consistent with these concepts, in monoecious species, the female function is prioritized when resources are ample, while the male

¹Université Paris-Saclay, CNRS, INRAE, Université Evry, Institute of Plant Sciences Paris-Saclay (IPS2), 91190 Gif-sur-Yvette, France. ²INRAE, Génétique et Amélioration des Fruits et Légumes (GAFL), 84143 Montfavet, France. ³These authors contributed equally: Hsin-Ya Huang, Siqi Zhang, Fadi Abou Choucha. 🖂 e-mail: Abdelhafid.bendahmane@inrae.fr

function is expressed more often under adverse conditions, or when the female function is already fulfilled⁷. Male plants excel in pollen donation, potentially attributed to their capacity to produce a greater quantity of male flowers and pollen⁸. Besides, pollen development is highly sensitive to adverse temperatures⁹. Producing more male flowers could serve as a compensatory mechanism, mitigating the impact of the adverse conditions on male function.

In spite of our progress in decoding the genetic pathway leading to the development of unisexual flowers¹⁰⁻¹⁸, the molecular mechanisms integrating environmental cues driving the emergence of new sexual morphs are still obscure. The tapestry of evolutionary processes is woven with the intricate threads of genetic innovation. Among the most enigmatic contributors are transposable elements (TEs), ubiquitous mobile genetic elements that traverse the genomes of organisms across the tree of life¹⁹. Based on their mobility, TEs can be classified as DNA transposons, which use a cut-and-paste mechanism for their mobilization, and retro-transposons, which jump via a reverse-transcribed RNA intermediate. These two classes are further divided into TE superfamilies and families based on particular sequences, such as the presence of specific terminal repeats or conserved protein domains¹⁹. For instance, Harbinger transposons are DNA transposons which usually encode a DDE transposase and a SANT/Myb/trihelix domain-containing DNA-binding protein²⁰. The role of transposons in adaptation and genome evolution is complex and multifaceted. Transposons can cause genetic instability by disrupting the functions of essential genes. Transposons can also contribute to genome evolution by creating advantageous genetic modifications²¹.

In the present work, we investigated how sexual morphs emerge in Cucurbitaceae, a large plant family that displays different sexual morphs. We focused our investigation on a monoecious species, Cucumis melo (melon). Male flowers are prevalent in melon, while female flowers develop on the initial nodes of the branches. During early developmental stages, flowers are initially bisexual. Subsequently, sex determination occurs through the developmental arrest of either stamen or carpel primordia^{2,18}. This process is controlled by the monoecy pathway implicating the andromonoecious (M), androecious (A), and gynoecious (G) genes. The G gene codes for the transcription factor CmWIP1, a C2H2 zinc finger transcription factor, which functions as a master regulator in the process of sex determination in cucurbits¹⁶. The A and M genes encode for two aminocyclopropane-1carboxylic acid synthase enzymes, CmACS7 and CmACS11, respectively^{14,15}. CmACS11 expression in lateral branches suppresses the expression of the male-promoting gene CmWIP1. CmACS7 expression in carpel primordia of female flowers inhibits stamina development through a non-autonomous mechanism¹⁷.

Through characterization of a spontaneous sex transition mutant, we identified a *Harbinger* transposon, *AndroPIF*, inserted in the *Ethylene Insensitive 2* (*EIN2*). We investigated the role of *EIN2* in sex determination and fruit development, and the role of TEs in the emergence of sexual morphs in plants.

Results

A *Harbinger* transposon insertion in the EIN2 gene is responsible for androecy in CAM106 line

In andromonoecious melon, male flowers develop in inflorescences at every node of the main stem, and in distal nodes of the lateral branches. Hermaphrodite flowers develop at the first two nodes of lateral branches. During seed propagation of an andromonoecious Charentais line, hereafter named CAM_WT, we identified a spontaneous mutant line that is androecious, CAM106, where hermaphroditic flowers are converted to male unisexual flowers (Fig. 1a–c). Furthermore, male inflorescences develop in clusters of about 15 flowers, compared to 6 in the CAM_WT parental line (Fig. 1d). Androecy in melon can arise via the inactivation of *CmACS11* or the melon *CRABS* coding regions of the two genes in CAM106 but found no mutations. pointing to a new locus controlling sex transition in cucurbits. To characterize the inheritance of such mutation, we crossed CAM106 to an andromonoecious line, CAM Sister, and phenotyped 10 F1 hybrids and their F2 progenies. CAM Sister, is genetically related to CAM WT but not identical (see material and methods). We found all the F1 hybrids and romonoecious, indicating that the CAM106 mutation is recessive. Further phenotyping of 200 F2 plants showed 3:1 segregation of andromonoecy versus androecy, consistent with a single-locus recessive mutation leading to androecy (Supplementary Fig. S1a, b). We refer hereafter to this novel androecious mutation as a3. To identify the causal mutation in the CAM106 line, we sequenced bulkedgenomic DNA from andromonoecious, and androecious segregant plants, and determined the delta-SNP index. We found a single genomic region on the right end of chromosome 8 completely associated with androecy (Fig. 1e). Fine mapping in a population of 1500 segregants plants, further delimited a3 to a single gene, CmEIN2, encoding the melon ortholog of the *ethylene-insensitive protein* 2^{22} (Fig. 1f). Sequence analysis revealed an insertion of a DNA transposon of the P Instability Factor (PIF)/Harbinger superfamily^{8,20}, henceforth termed AndroPIF at the 5' UTR of CmEIN2 (Fig. 1f, Supplementary Figs. S1c and S2). Genotyping of plants harboring recombination events in the vicinity of *a3*, using a marker specific to the transposon insertion showed complete association of the transposon insertion in CmEIN2 and androecy (Supplementary Fig. S1d).

CLAW gene $(CmCRC)^{15,18}$. So, we sequenced the promoter and the

Transposable elements are exposed to epigenetic modifications, with consequences on the expression of nearby genes, especially through the genomic spreading of DNA methylation²³. We examined the DNA methylation status of CAM_WT parental line and CAM106, using whole genome bisulfite sequencing. As expected, strong DNA methylation of *AndroPIF* was observed. However, weak methylation spreading was detected in *AndroPIF* flanking sequences (Supplementary Fig. S3). To test if *AndroPIF* insertion was interfering with *CmEIN2* transcript accumulation, we measured the expression of *CmEIN2*, using quantitative real-time PCR. We found a weak to no expression of *CmEIN2* in leaves and flowers of CAM106, indicating that *a3* is likely a loss-of-function allele of *CmEIN2* (Fig. 1g).

Induced mutations in EIN2 lead to androecy in melon

Since there is no efficient transformation protocol in cucurbits, to further validate the role of CmEIN2 in androecy, we used TILLING (Targeting Induced Local Lesions IN Genomes) collections from monoecious melon and screened for mutations in CmEIN2. Sixtytwo induced mutations were identified, including two splicing, ein2-SP1 (G1175A) and ein2-SP2 (G1246A), and 17 missense mutations. As expected, missense mutations had no effect on the plant sexual phenotype (Supplementary Table 1), consistent with earlier reports on *EIN2* mutagenesis in *Arabidopsis*²². We selected the two spicing mutants for further analysis. ein2-SP1 is altered in the splicing donor site of intron 4, while ein2-SP2 harbors a mutation in the splicing acceptor site of intron 4 (Supplementary Fig. S4a). Transcript analysis revealed introns 3 and 4 retention (30%), intron 4 retention (47%) or exon 4 skipping (23%) for ein2-SP1 mutant; similarly, ein2-SP2 mutation leads to intron 4 retention (27%) or to splice variant with a 4 bp deletion in exon 4 (73%). For both mutants, all the splicing forms lead to premature stop codons and thus are likely lossof-function alleles (Supplementary Fig. S4b). We backcrossed the mutant lines harboring the splicing mutations to the wild type and followed the segregation of the mutations with the phenotypes. As predicted, plants homozygous for the splicing mutations were androecious, i.e., female flowers being transformed into male flowers (Fig. 2a, b; Supplementary Fig. S5). Interestingly, in plants heterozygous for ein2-SP1 and ein2-SP2 mutations, 20-40% of female flowers were transformed into bisexual flowers (Fig. 2c, d;



Fig. 1 | A *Harbinger* transposon insertion in *EIN2* gene is responsible for androecy. a Sexual phenotype of andromonoecious melon (CAM_WT), and androecious mutant (CAM106). Scale bar = 1 cm. b Lateral branches showing a hermaphrodite flower in CAM_WT, and a cluster of male flowers in CAM106. c Schematic of sexual morphs. d Number of hermaphroditic or male flowers per node on the main stem (MS), or at the 1st node of lateral branches (LB). Data are mean \pm SE. (n = 8 biological replicates). e Cloning of the *a3* locus. Δ SNP index between andromonoecious and androecious bulks, calculated with a 1-Mb sliding

window. Arrows indicate candidate genes on chromosome 8. **f** Structure of the *CmEIN2* gene and the position of *AndroPIF* insertion. The *ein2_SP1* and *ein2_SP2* splicing mutations are shown by arrowhead. **g** Quantitative real-time PCR (qPCR) of *CmEIN2* in leaf, hermaphrodite (H) and male (M) flowers of CAM_WT, and CAM106. *P*-values from an unpaired two-tailed Student's t-test are indicated. Data are mean \pm SE. (n = 3 biological replicates). Source data are provided as a Source Data file.

Supplementary Fig. S5), highlighting also the role of *CmEIN2* in stamen inhibition in female flowers. This phenotype was not observed in CAM106 mutant because CAM_WT parental line is andromonoecious developing male and hermaphrodite flowers (Fig. 1a).

Since *EIN2* is a key regulator of ethylene signaling²⁴, we examined the ethylene response phenotype of 3 days old etiolated seedlings of CAM106-AndroPIF insertion mutant and *ein2-SP1* and *ein2-SP2* splicing mutants. Dark-grown melon seedlings treated with ethylene display a typical "triple response" consisting of inhibition of root development and hypocotyl elongation, radial swelling of the hypocotyl, and apical hook formation²⁵. We found CAM106, *ein2-SP1* and *ein2-SP2* mutants to be insensitive to ethylene (Fig. 2e), supporting that the *AndroPIF* insertion and the splicing mutations are likely *CmEIN2* loss-of-function alleles. Altogether, the characterization *CmEIN2*-AndroPIF and the *CmEIN2*-splicing mutants validate the role of *CmEIN2* in the development of carpel-bearing flowers, both in andromonoecious and monoecious plants.

CmEIN2 is required both for carpel development and stamina inhibition

To integrate the function of CmEIN2 in the monoecy sex determination pathway, we generated mutant combinations with CmWIP1, CmACS7, and CmACS11, and examined the sex morphs of these double mutants in relation with the expression of sex genes (Fig. 3a). In melon, the inappropriate sexual organs are arrested at stage 6 after the initiation of both carpel and stamen primordia¹⁸. In monoecious plants, expression of CmACS11 in flower buds of lateral branches represses the expression of the male-promoting gene CmWIP1, leading to female flower development. Since loss of function of CmEIN2 leads to the development of male instead of female flowers in lateral branches, we first analyzed the expression of CmACS11, CmACS7, and CmWIP1. Compared to WT, the expression of CmACS11 was not compromised. In contrast, the repression of CmWIP1 expression was relieved, leading to male plants, and indirectly to repression of CmACS7 that relies on carpel development for its expression (Fig. 3a, b). To investigate the role of CmEIN2 in stamina inhibition, we crossed ein2 mutant to wip1

-Ethylene

+Ethylene



Fig. 2 | **Induced mutations in** *CmEIN2* **lead to androecy in melon. a** Sexual phenotype of a monoecious plant (Mono_WT), and ein2 splicing mutants (–/–). **b** Sexual phenotype of pistillate flowers of WT and *ein2* heterozygous mutants (+/–). Sg, stigma; st, stamen. Scale bars in (**a**, **b**) = 0.5 cm. **c** Schematic of sexual morphs. **d** Relative ratio of female, or hermaphrodite flowers, per total carpel-

bearing flowers. Data are mean \pm SE. (n = 3 biological replicates). **e** Triple response phenotype of melon genotypes in absent (–Ethylene), and presence of ethylene (+Ethylene). The photography is taken 7 days after germination. Scale bars = 1 cm. Source data are provided as a Source Data file.

ein2

gynoecious line to generate *ein2* and *wip1* homozygotes double mutants, and analyzed the expression of the stamina inhibitor *CmACS7*. Compared to *wip1* mutant, we found *wip1* and *ein2* homozygous double mutant fully hermaphroditic, despite the expression of the stamina inhibitor, *CmACS7* (Fig. 3c).

Ethylene has been associated with fruit shape development²⁶. To investigate the role of *CmEIN2* in fruit shape we measured the ovary shape index (OSi) and the fruit shape index (FSi). The OSi is determined as the ratio of the ovary length by the ovary width. FSi is determined as the ratio of the fruit length by the fruit diameter at mature stage. We found that OSi and FSi are significantly higher in Charentais Mono line (WT) compared to wip1/ein2 mutant. Conversely, the ovaries and fruits of wip1/ein2 mutants were flatter, phenocopying the one of $acs^{7^{16}}$ mutant, while the double mutants showed an additive effect (Fig. 3d-f). Collectively, our findings highlight the dual function of EIN2 in both sex determination and fruit shape. In lateral branches, upon the expression of CmACS11, EIN2 is recruited to repress the expression of the carpel inhibitor, CmWIP1. Subsequent to the expression of CmACS7 in carpel primordia, EIN2 is recruited in the stamina inhibition processes. Following the sex determination phase, EIN2 promotes fruit shape elongation.

EIN2 signaling plays contrasting roles in carpel development and stamina inhibition

To gain insight into the gene network associated with CmEIN2 roles in stamina and carpel development, we examined the transcriptome of male, female, and hermaphrodite flower buds at stage 6. Male and female flower buds were collected from ein2 and wip1 loss-of-function mutants, respectively. Hermaphrodite flower buds were collected from wip1 and ein2 double mutant. Pairwise comparisons of the transcriptomes of male versus female, male versus hermaphrodite, and female versus hermaphrodite flowers, identified 8343, 5492 and 491 differentially expressed genes (DEG), respectively (Supplementary Fig. S6a, Supplementary Data 1). We grouped the differentially expressed genes by their expression patterns into 4 clusters: stamen suppression (S–), stamen development (S+), carpel suppression (C–), and carpel development (C+). Cluster S- and Cluster S+ denote genes that are upregulated, and genes that are downregulated, respectively, in female flowers compared to staminate flowers. Conversely, Cluster C- and Cluster C+ denote genes that are upregulated, and genes that are downregulated, respectively, in male flowers compared to pistillate flowers (Fig. 3g, Supplementary Fig. S6b). Gene Ontology (GO) term enrichment analysis revealed cluster S- enriched in GO terms of biological process related to response to ethylene and negative regulation of signaling (Fig. 3g). Conversely, cluster S+ is enriched in GO terms associated to stamen formation, reproductive shoot system development, and positive regulation of transcription (Fig. 3g, Supplementary Fig. S7). Consistent with this, RT-qPCR analysis showed the upregulation of ethylene responsive factor 1 (ERF1) and EIN3 Binding F-Box Protein (EBF1) associated with response to ethylene (GO: 0009723), and the downregulation of PISTILLATA (PI) and bHLH91 related to stamen development (GO: 0048443), in female, compared to staminate flowers (Supplementary Fig. S6c, d). Cluster C+ was found enriched in GO terms related to regulation of histone acetylation, gametophyte development, and mitotic cell cycle, while cluster C- is enriched in GO terms related to response to abscisic acid, water deprivation, and protein dephosphorylation (Fig. 3g, Supplementary Fig. S7). Alongside, RT-qPCR analysis validated the downregulation of the carpel marker genes, CRC and SEEDSTICK (STK) (GO: 0035065), and the upregulation of abscisic acid (ABA) associated genes, ABA RESPONSIVE ELEMENTS-BINDING FACTOR 3 (ABF3) and Cytochrome P450 (CYP707A10) (GO: 0009737) in male, compared to pistillate flowers (Supplementary Fig. S6e, f). In summary, our RNA-seq analysis reveals two contrasting indirect roles of CmEIN2 in carpel development, and in stamina inhibition. In developing carpel, through inhibition of the carpel repressor, CmWIP1 (Fig. 3b), EIN2 promotes the expression of genes associated with cell division and female gametophyte development, including the carpel identity gene, CRC (Fig. 3g, Supplementary Fig. S6e). In inhibited stamina, EIN2 promotes the expression of genes associated with negative regulation and signaling, and the repression of the stamina-promoting gene, Pl (Fig. 3g, Supplementary Fig. S6c).

Transposons of the *PIF/harbinger* family are highly active transposons in melon

To determine if *AndroPIF*, and related *PIF/Harbinger* elements, are active transposons in the melon genome, we analyzed two biological replicates of pooled genomic DNA of 50 CAM_WT plants, including CAM106 control line. To identify *AndroPIF* insertions, we performed transposon display followed by illumina sequencing (TEDseq²⁷; Supplementary Fig. S8). We detected an average of 34 read coverage per *PIF/Harbinger* insertion loci (PIL), in both replicates, indicating the high sensitivity of TEDseq (Supplementary Fig. S9a). In total, we found 332



Fig. 3 | *CmEIV2* is required both for carpel development and stamina inhibition. a Sexual morphs of melon plants across combinations of alleles at different sex loci. b qPCR analysis of *CmACS11*, *CmACS7* and *CmWIP1* in Mono_WT female(\mathfrak{P}) or male(\mathfrak{F}) flowers, and in male flowers (\mathfrak{F}) of CAM106(*ein2*) collected from main stems (MS), or lateral branches (LB). c qPCR analysis of *CmACS7* and *CmEIN2* in pistillate flowers of *wip1*,and *wip1/ein2* double mutants. P-values from an unpaired two-tailed Student's t-test are indicated in (**b**, c). Data in (**b**, c) are mean ± SE. (n = 3biological replicates). **d**, **e** Ovary shape index(d) and fruit shape index(e) of different melon genotypes. Different letters indicate significant differences based on a one-way ANOVA with Tukey's multiple comparisons test. Data in (**d**, **e**) are mean ± SE. n is indicated for number of biological replicates in (**d**, **e**). **f** Ovary and

fruit phenotypes of the genotypes shown in (**d** and **e**). Scale bar in the flower panel = 1 cm. Scale bar in the fruit panel = 5 cm. **g** Gene-wise hierarchical clustering heat map of differentially expressed genes (DEGs, adjusted *P* value < 0.05) between female (F, *wip1* mutant), hermaphrodite (H, *wip1/ein2* double mutant), and males (M, *ein2* mutant) flowers. DEGs (adjusted *P* value < 0.05) showing segregation into 4 clusters: stamen suppression (S–), stamen development (S+), carpel development (C+) and carpel suppression (C–). The *z*-score scale represents mean-subtracted regularized log-transformed read counts. GO terms specifying the biological process are shown at the right of the panels. Source data are provided as a Source Data file.

PILs, out of which 282 were common between CAM and Mono WT, and 50 were specific to CAM (Fig. 4a, b; Supplementary Data 2). Further analysis of the 50 CAM PILs identified two insertions as ancient ones, validated by broken-paired reads in the CAM WT parental line, resulting in 48 new AndroPIF insertions. These new insertions were distributed over the 12 melon chomosomes, with preferential insertion bias in gene rich regions (Fig. 4c). Furthermore, 83% of novel insertions locate in promoter regions, compared to 58% of the common ones (Fig. 4d). Local sequence analysis of AndroPIFs insertion sites revealed the consensus motif 'MWYTWARWK' (Fig. 4e), similar to that previously reported for PIF transposons^{28,29}. Open chromatin, H3K9ac histone mark, and low CHG/CHH DNA methylation are key features associated with active chromatin³⁰. We observed a distinct pattern wherein active chromatin marks exhibit a positive correlation with the insertion of new PILs (Fig. 4f, h). In contrast, common AndroPIF insertions, indicative of ancient insertions, are consistently located in closed chromatin states, highlighting the temporal dynamics of chromatin from active to inactive, or reduced counter selection of insertions within inactive chromatin regions (Fig. 4g, h, Supplementary Fig. S9).

PIF/harbinger SANT1 protein binds to PIF TIR sequences

AndroPIF, like most *PIF/Harbinger* transposons encodes two proteins: a nuclease and a DNA binding protein, SANT1 (Supplementary Fig. S2)²⁰. Physical interaction between the two proteins is necessary for *PIF/Harbinger* transposition^{28,31}. To further characterize *AndroPIF*, and related *PIF/Harbinger* transposon insertions in CAM melon line, we mapped genome wide DNA sequences binding to *AndroPIF* SANT1 protein, using amplified DNA affinity purification sequencing (amp-DAP-seq)³². A total of 1521 SANT1 binding peaks were detected in the CMiso reference genome (Supplementary Data 3). Of this, 60,4% (n = 919) overlapped sequences annotated as PIF, 21.5% (n = 327) were



linked to sequences annotated as non-TE, and 18% bound sequences annotated as TE, but not-PIF (Supplementary Fig. S10a, b). Sequence analysis of SANT1 insertion sites in melon genome revealed two consensus motifs. Motif1'TTYYTGTTTCTTGTT' mapped at peak summits, while motif 2 'AGGGCCCGTTTGGTAACGTT' was located at the flanking sides of the summits (Supplementary Fig. S10c-d). Further sequence annotation identified part of motif 2 as the TIR sequence 'GGGCCCGTTTG' of *AndroPIF*, pointing to the role of SANT1 in binding to the transposon extremities. Consistent with this, we systematically observed two SANT1 ampDAP-seq peaks (double peaks, Supplementary Fig. S10e, f) mapping to loci annotated as PIF, including the original copy of *AndroPIF* on chromosome 5, also captured by TEDseq (60,4% in Supplementary Fig. S10b). However, we found loci showing single peaks, corresponding to degenerate transposons that lost part of the TIR sequence, or random melon sequences that have similarities to TIR motif (18,5%, and 21.5% in Supplementary Fig. S10b). Several TEs have been shown to move following exposure to various environmental stressors^{33,34}, we analyzed whether the expression of *AndroPIF TRANSPOSASE* and *SANT1* genes are induced under heat stress conditions. CAM-WT plants were grown under control conditions (28 °C day/22 °C night), and shifted to heat stress conditions, for two hours at 42 °C. The expression of the *TRANSPOSASE* and *SANT1* genes were then

Fig. 4 | *AndroPIF* transposon activity in melon genome. a Venn plot of annotated *PIF/Harbinger* insertion Loci (PIL) in Mono genome (Ref PIL), and PIFs identified by TEDseq in poll of 50 CAM plants. b Sequence coverage of PILs identified by TEDseq. c Genomic position of PIL identified by TED-seq. PIL common, represent PILs that are common between CAM and Mono_WT. New PIL indicates new insertions observed in 50 plants descending from CAM parental line. d Distribution in percent of *AndroPIF* insertions in genes, and in intergenic regions. *P*-value from a Z-test are indicated and applied only to regions with non-zero values for both groups.
e Target site insertion preference of new PILs, compared to *mPing* insertion sites, in Arabidopsis and rice^{28,29}. f-h, Boxplots of H3K9ac (f), ATACseq(g), and DNA methylation(h) (CHG, CHH, and CG) levels in bin sizes of 100 bp in ±3 kb sequences flanking insertion sites of common PIFs, new PIFs, and random regions. For each boxplot, the lower and upper bounds of the box indicate the first (Q1) and third (Q3) quartiles, respectively, the center line indicates the median, and the whiskers

monitored by qPCR. We found the expression of both genes inducible by heat stress, pointing toward *AndroPIF* mobility triggered by abiotic factors (Fig. 4j, k).

PIF/Harbinger display contrasted transposition in wild versus cultivated melons

Crop domestication involved recurrent selection of desirable traits, leading to a severe loss in sequence diversity. We investigated whether sequence diversity due to PIF/Harbinger transposon insertions has been also constrained in cultivated compared to wild melon accessions. Positions of PIF/Harbinger polymorphic insertions in the genomes of 480 cultivated (group CM, CA, and CAF in Fig. 5a) and 81 wild (group WM, WAF, and WA in Fig. 5a) accessions³⁵, collected from different geographic areas, was determined using short-read resequencing, and TEFLoN method^{36,37}. Integration of PIF polymorphic insertion sites, SANT1 binding peak pairs, and search for TIR sequences resulted in the identification of 183 full-length AndroPIF copies (Supplementary Data 4, 5). The copies identified in the Mono-WT genome³⁸ are denoted as reference insertions, while others are labeled as non-reference insertions (see Supplementary Fig. S11). Phylogenetic tree construction based on the presence/absence of reference AndroPIF elements exhibits a clustering structure akin to that generated by SNP polymorphism (Fig. 5a, Supplementary Data 6)³⁵. In contrast, the tree generated based on non-reference insertions displays a mixed clustering pattern, suggesting that transposition of AndroPIF elements contributed to increased diversity in these populations. We found most PIF/Harbinger reference insertions (169 out of 182) shared between cultivated and wild accessions (Fig. 5b), indicating that they are ancestral. In contrast, out of the 894 non-reference-insertion, 561 are found in single accessions, pointing toward recent transposition events (Fig. 5c). We call this new category of PIF/Harbinger transposons "non-reference-newinsertions". Distribution of non-reference-new-insertions in melon germplasms revealed active transpositions in wild melon accessions compared to melon breeding materials (Fig. 5d, e). As in CAM WT, nonreference-new-insertions were overrepresented in promoter regions (65.14%), compared to reference insertion (57.09%), indicating that integration preference has shaped the genome landscape of AndroPIF insertions (Fig. 5f).

Discussion

We investigated how new sex determination morphs could arise. During the propagation of an andromonoecious melon line, we encountered a spontaneous mutant, sexually converted from andromonoecy to androecy. This intriguing phenotype prompted us to study the causal mutation. We identified a *PIF/Harbinger* transposon as the origin of the sex transition, disrupting the expression of the *Ethylene Insensitive 2* (*CmEIN2*) gene. Through genetic and expression analyses, we unveiled a dual function of *CmEIN2* in sex determination and fruit shape. Upon the expression of *CmACS11* in carpel primordia, of flowers developing on lateral branches, CmEIN2 is recruited to extend to a maximum of 1.5 times the interquartile range (Q3–Q1). Number of bin is 60 for each box. Asterisks mark statistically significant differences (two-sided Wilcoxon rank-sum test, **P < 0.01, ***P < 0.001, NS. P > 0.05). i Genome browser view of SANT1 DAPseq peaks, TEDseq reads, and DNAseq, mapped on Mono genome. *AndroPIF* original locus in Mono genome, and in CAM_W (left), *AndroPIF* insertion in *EIN2* in CAM106 line (middle), and a new PIL in CAM_W progeny plant (right). Color scheme in DNAseq track: blue and purple, pair-end reads; yellow, highly repetitive reads; green and red, discordant reads. Soft-clipped reads are shown as sequences with red background. **J**, **k**, qPCR of SANT1 (**j**) and Transposase (**k**) of *AndroPIF* in young leaf of CAM-WT under heat stress (42 °C/ 32 °C) and control (28 °C/22 °C) conditions. *P*-values from an unpaired two-tailed Student's t-test are indicated. Data are mean ± SE. (n = 3 biological replicates). Source data are provided as a Source Data file.

suppress the expression of the carpel inhibitor, *CmWIP1*. Subsequently, the expression of *CmACS7*, in developing carpel primordia, produce ethylene that is perceived in the stamina primordia, likely through a spatially differentially expressed ethylene receptors¹⁷. In stamina primordia, CmEIN2 is then recruited to mediate stamina inhibition, leading to female flower development. If *CmEIN2* is inactivated, the repression of *CmWIP1* expression is relieved leading to androecy. Inactivation of both *CmEIN2* and *CmWIP1* lead to hermaphrodite flower development (Fig. 6).

Transposons are powerful drivers of genome evolution and plasticity and contribute to genetic variability, potentially under stress conditions³⁴. We thus tested whether *PIF/Harbinger* mobility is induced by adverse conditions. Structurally, *PIF/Harbinger* transposons are predicted to encode two proteins, a transposase and a DNA-binding protein. We found the expression of both genes activated by heat stress, pointing toward triggering of *PIF/Harbinger* mobility under adverse conditions. Similar stress-responsive transcription or movement has been reported in several TEs^{33,39}. In rice, the DNA transposon mPing can be activated in response to cold and salt stress²⁷; in tobacco, Tnt1 element can be induced by different biotic and abiotic stress³⁴.

Given the sensitivity of TEs to environmental threats, our findings suggest that androecy may arise as a result of transposition triggered by stress into *CmEIN2*. Ethylene is a key feminizing hormone in cucurbits^{40,41}. Consistent with this scenario, we found *CmEIN2* required for the transcriptional repression of the male-promoting gene *CmWIP1*, leading to female flower development (Fig. 6). Under adverse conditions, fruit set could be costly, inactivation of *CmEIN2* relieves *CmWIP1* from the repression, and consequently leading to androecy. By producing only male flowers, androecious plants can have an advantage in promoting outcrossing.

The role of transposon mobilization in shaping sexual reproduction presented here could be a general mechanism. Besides *AndroPIF* transposon leading to androecy, we previously showed that the DNA transposon, *Gyno-hAT* is at the origin of male to female sex transition, resulting in gynoecious plants¹⁶. In species harboring sex chromosomes, TEs also tend to accumulate in genomic regions containing sex genes, underscoring their significant role in the emergence and development of sex chromosomes⁴².

In addition to promoting female flower development, ethylene is required for inhibition of stamina in pistillate flowers (Fig. 6). We introduced *ein2* mutation in monoecious genetic background. We found plants heterozygous for *ein2* mutation having 20–40% of female flowers converted to bisexual, pointing to the quantitative role of ethylene in sex determination (Fig. 2d). Furthermore, ethylene is a key regulator for fruit elongation in melon. Consistent with this, fruits developing on plant harboring *ein2* mutation are significantly affected in both ovary and fruit shape indexes. In summary, our work proposes a model in which *CmEIN2* play a central role both in the control of unisexual flowers in monoecious plants as well as the emergence of male and female unisexual plants.



Fig. 5 | *PIF/Harbinger* transposition in wild and cultivated melon accessions. **a** Phylogenetic tree constructed using reference and non-reference *PIF/Harbinger* insertion sites identified using genomic re-sequencing data of 567 melon accessions, and 7 outgroup relatives (Zhao, 2019). AF, African; WAF, wild African; CAF, cultivated African; WM, wild melo; CM, cultivated melo; WA, wild agrestis; CA, cultivated agrestis. **b**, **c** Histogram and venn plots of reference (**b**) and nonreference (**c**) *AndroPIF* insertions identified in wild (WA and WM) and cultivated (CM and CA) accessions. **d** Venn plot of new insertions identified in wild and cultivated accessions. Number of accessions is indicated for each catalog, **e** Number of

The transcriptome analysis of flower buds collected from female (*wip1/ElN2*), hermaphrodite (*wip1/ein2*), and male (*ein2*) plants, revealed 4 clusters of gene expression patterns. We found S- cluster enriched in GO terms such as ethylene response genes, while S+ cluster contains several TFs, such as *PI*, implicated in stamen development. We also found C+ cluster enriched in GO terms related to female gametophyte development, and mitotic cell cycle, while C- cluster enriched in GO terms related to abscisic acid, water deprivation, and protein dephosphorylation. These results corroborate that ethylene signaling negatively regulates stamen development, while simultaneously exerting a positive effect on carpel development.

Annotating *PIF/Harbinger* elements in plant genomes presents several challenges due to their high diversity in sequences, sizes, and organizations. Beside in silico annotation, we used SANT1 ampDAPseq analysis as a tool to identify and annotate *PIF/Harbinger* transposons in melon genomes. In loci displaying ampDAPseq double peaks, we systematically found either a new or an ancient PIF element. However, in loci showing single ampDAPseq peaks, the accuracy of the method is

new-non-reference insertions per accession, in wild and cultivated accessions. *P*-values are calculated based on a two-sided Wilcoxon rank-sum test. For each boxplot, the lower and upper bounds of the box indicate the first (Q1) and third (Q3) quartiles, respectively, the center line indicates the median, and the whiskers extend to a maximum of 1.5 times the interquartile range (Q3–Q1). Number of accessions is indicated for each catalog in (**d**). **f** Distribution, in percents of *AndroPIF* insertions in genes, and intergenic regions. *P*-values for both groups.

weak, as the binding could be due to degenerate transposon or random sequences that has similarities to TIR sequences. Nevertheless, when we combined transposon annotation with SANT1 ampDAPseq analysis, the approach proved highly efficient in identifying new transpositions as well as ancient and degenerate transpositions.

Following the annotation of *PIF/Harbinger* elements, we classified them into three categories, refence PILs, non-reference PILs and new PILs. We found most reference and non-reference PILs reflect the genetic relationship between cultivated and wild melon accessions, while new PILs corresponded to recent insertions. Interestingly, we found the new PILs more frequent in wild melon accessions compared to breeding materials, pointing toward their roles in plant adaptations. The movement of transposons in wild and cultivated plants is of particular interest because it can provide new genetic variation for plant breeding⁴³⁻⁴⁵. The majority of novel insertions are located in promoter regions. It will be interesting to explore their influence on the modulation of gene expressions and, as a result, on plant adaptation.



Fig. 6 | **Model of sex-determination pathway in melon integrating** *EIN2* **function.** The model integrates *ACS11, WIP1,* and *ACS7* sex-determination genes. Pink box indicates carpel primordia and blue box indicates stamen primordia. Spatial expression of *ACS11, ACS7,* or *WIP1* is restricted to carpel primordia. *EIN2* is expressed both in the carpel and stamina primordia. The chemical symbol indicates ethylene gas. Red cross indicates no ethylene produced by ACS7 or ACS11. Ethylene

Our work reveals a central role of *PIF/Harbinger* class of transposons in the emergence of sexual morphs in flowering plants. As *PIF/ Harbinger* are present in organisms spanning the tree of life¹⁹, our study will likely inspire future investigations beyond cucurbits.

Methods

Plant material

The CAM106 spontaneous mutant was discovered during the propagation of the andromonoecious line CAM_WT (Cucumis melo L. var. cantalupensis). To identify the causal mutation, we crossed CAM106 to CAM_Sister, an andromonoecious EMS-mutagenized mutant line derived from CAM_WT, and phenotype the F2 segregant plants. The plants were cultivated in a greenhouse with a daytime temperature of 27 °C, a nighttime temperature of 21 °C, and a photoperiod of 16 hours of light. The plants were assessed for flower sex type and fruit shape.

Bulk segregant analysis

Genomic DNA was extracted from CAM106 X CAM_Sister F2 population. For bulk segregation analysis, equal amounts of genomic DNA from 10 individual plants with either andromonoecious or androecious phenotypes were pooled to create the andromonoecious and androecious pools. Subsequently, DNA libraries were prepared, using the Next Ultra DNA Library Prep Kit, and 150 bp paired-end reads sequenced on HiSeq 2000 Illumina platform. Reads were aligned to the reference Charentais Mono melon genome CMiso1.1³⁸ (here referred to as Mono_WT). The SNP-index across all loci was calculated, and the delta (Δ) SNP-index was determined by subtracting the SNPindices of the two bulks at each locus (SNP-index_mutant – SNPindex_WT).

Triple response assay

Melon seeds were germinated in a 10-litter wet box in the dark for 7 days at 27 °C, either with or without 50 μ M ethylene. Ethylene solution at 50 μ M concentration was prepared by diluting 1 M ethephon (Sigma) in 5 mM Na₂HPO3 buffer, as described previously²⁵.

Sexual morph phenotyping and epistasis analysis

Monoecious melons carry only male flowers in the main stem, female flowers develop on the first nodes (L1) of lateral branches. The following flowers on L2 and L3 nodes of the lateral branches can be female depending upon plant age and environment. Melon also exhibits other sex morphs, andromonoecious, gynoecious, androecious and hermaphrodite. Androecious plants bear only male flowers.



produced by ACS7 in the carpel is perceived in the stamina primordia likely through a spatially differentially expressed ethylene receptors¹⁷. Monoecious plants develop male and female flowers. *AndroPIF* insertion in *EIN2* leads to Androecy. *Ein2/wip1* double mutant is hermaphrodite. *EIN2*-mediated repression of *WIP1* depends on the expression of *ACS11*. *EIN2*-mediated repression of stamen requires the expression of *ACS7*.

Gynoecious and hermaphrodite plants develop only female and hermaphrodite flowers, respectively. Andromonoecious plants develop separate male and hermaphrodite flowers. To position EIN2 function in the monoecy pathway we generated double mutants, through crossing of the single mutant and genotyping F2 segregant plants. The loss-offunction alleles leading to gynoecy (*wip1*), androecy (*acs11*) or andromomonoecy (*acs7*) used in the epistasis analysis are described in Boualem et al¹⁴., Martin et al¹⁶., and Boualem et al¹⁵., respectively^{14–16}. Single, and double homozygous plants for *wip1*, *acs11*, *acs7*, and *ein2* mutations were phenotyped for sexual transition of the flowers on the male stem and lateral branches.

Quantitative RT-PCR

Total RNA was extracted from frozen leaves or flowers. First-strand cDNA was synthesized from 2 pg of total RNA with the Superscript® III reverse transcriptase (Invitrogen). Primer design was performed using the Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Primer sequences used are listed in Supplementary Table 2. Polymerase chain reactions were performed with the Bio-Rad CFX96 Real-time PCR apparatus, with qPCR MasterMix Plus for SYBR® Green I (Eurogentec, France). Gene expression is normalized to the expression levels of housekeeping genes: *CmActin2*. The qRT-PCR results were analyzed using a $\Delta\Delta$ Ct methodology.

RNA-seq assay

Flower buds at developmental stage 6 were sampled, and total RNA was extracted using the NucleoSpin RNA Plus kit (Macherey Nagel). The quality and concentration of the RNA were determined using Agilent RNA chips (Bioanalyser 2100). Three biological replicates for female (*wip1*), hermaphrodite (*wip1/ein2*), and male (*ein2*) flowers were used for library preparation. RNA-seq was performed on the Illumina NovaSeq 6000 with 150 bp paired-end reads. The reads were aligned to the reference genome CMiso1.1 by STAR (v2.7.5c)⁴⁶, and the mapped reads were assigned to genes with featureCount (v2.0.3)⁴⁷. Differentially expressed gene calling was conducted using DiCoExpress⁴⁸, with a p-value adjusted <0.05. GO enrichments were completed by ClusterProfiler⁴⁹ with customized GO annotation.

Whole-genome bisulfite sequencing

Genomic DNA of young leaves of CAM_WT and CAM106 were extracted with CTAB. The gDNA was used to generate the libraries and treated by bisulfite. Sequencing was done on Illumina NovaSeq 6000 with 150 bp paired-end reads. The reads were analyzed by MethylStar pipeline⁵⁰. The differentially methylated sites (DMRs) were identified by methylkit⁵¹ and DMRcaller⁵².

TE display sequencing

TEDseq was performed on a DNA pool of 50 CAM WT seedlings, and CAM-106 DNA, used a control. Genomic DNA was extracted using the DNeasy Plant Mini Kit (QIAGEN). Sequencing libraries were prepared from 1µg of pooled DNA using the NEBNext[®] Ultra[™] II DNA Library Prep Kit. PCR enrichment was carried out for 20 cycles, using Illumina P7 Primer with index, and P5 primers PIF-P1 or PIF-P2 (Supplementary Table 2; Supplementary fig. S8). Nested PCR enrichment was then carried out for 12 cycles, using PIF-P3 or PIF-P4. This process was independently repeated twice. Pair-end sequencing was carried out using the Illumina Miseq, generating 150 bp reads. To identify new TE insertions, soft-clipped and discordant reads that partially mapped to the reference sequence of AndroPIF were extracted. The reads were subsequently remapped to the CMisoV1.1 reference genome using Bowtie2 v2.3.5 with the arguments 'very-sensitive-local'53. Putative insertions were retained if they were supported by more than one set of 3 soft-clipped and discordant reads in at least one repeat at the same locus. TE target site analysis was performed with a Rscript, using packages 'dplyr', 'readxl', 'knitr', 'ggplot2', 'ggseqlogo' and 'Biostrings'. For each identified region, reads overlapping putative site and TE sequence were retrieved and trimmed with Cutadapt⁵⁴, and were aligned with Bowtie2 on CMisoV1.1 genome. The exact positions of the insertion sites were calculated by strand specificity and CICAR flag. A sequence of 6 nucleotides flanking the insertion site were extracted to calculate consensus matrix and motifs. The level of epigenetic marks around the TE insertion sites were calculated by Deeptools3.5.5 using the data described previously^{17,55,56} (accession number: PRJNA383830).

DNA affinity purification sequencing

AndroPIF SANT1coding sequence was isolated from leaf RNA by RT-PCR and cloned into pIX-HALO by Gateway recombination according to the manufacturer's protocol (Invitrogen). The primers used in cloning of SANT1 are listed in the Supplementary Table 2. Expression of HALO-SANT1 fusion protein and enrichment of DNA targets were performed as described previously³². The DNA used for DNA library preparations was extracted from the Charentais Mono line (WT). AmpDAP-seq was employed, utilizing amplification to eliminate DNA methylation. Library preparations followed the manufacturer's protocol for the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (Illumina). The final enriched DNA library was sequenced on a NextSeq500 as 2×75 nt reads. The control HALO protein AmpDAP-seq data were generated previously¹⁸ (accession numbers: SRX15631305, SRX15631306).

Transposable elements insertion detection in melon germplasm The genomic sequences of the 567 *C. melo* accessions, and the 7 rela-

ted species (Supplementary Data 6) were retrieved from previous study³⁵. The genome wide identification of *AndroPIF* related elements was performed using the TEFLoN method and the McClintock pipeline³⁶.

TEs were subsequently annotated manually, using SANT1 amp-DAPseq peaks, and the search of TIR sequences 'GGGCCCGTTTG', allowing 2 bp mismatches. CMisoV1.1 genome was used a reference genome for the mapping of the germplasm sequences.

The gff3 results from McClintock pipeline were converted to vcf format with the code present(REF0/0)/deletion(1/1) for reference insertion, and absence(0/0)/insertion(1/1) for non-reference insertion. For phylogenetic tree building, Euclidean distance was computed for bitwise matrix and the Neighbor Joining method was used for clustering. R packages 'vcfR', 'poppr', 'ggtree', and 'dplyr' were used for the analysis.

Heat stress assays

Germinated seeds were grown for 3 weeks in the greenhousse at 27 °C daytime, 21 °C nighttime, and a photoperiod of 16 hours of light. For the heat stress assay, the plants were transferred to a growth chamber with a temperature set at 42 °C. Total RNA samples were extracted at 0 hours, and 2 hours after the initiation of the heat stress.

Statistical analyses

All experiments were carried out in at least three biological replicates. Differences in mean for RT-qPCR data were assessed using an unpaired two-sided Student's t-test. For genomic distribution in percent of TE insertions in promoter, genes, or in intergenic regions, a two-sided Wilcoxon rank-sum test was performed and applied only to regions with non-zero values for both groups. For each boxplot, the lower and upper bounds of the box indicate the first (Q1) and third (Q3) quartiles, respectively, the center line indicates the median, and the whiskers extend to a maximum of 1.5 times the interquartile range (Q3–Q1). All the specific statistic methods were described in the figure legends.

Reporting summary

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

Data availability

Sequencing data generated in this study have been deposited in the NCBI Sequence Read Archive database under the accession number PRJNA1085522, including RNA-Seq, Bisulfite-Seq, TEDseq, and AmpDAP-seq. Previously published data used in this study are SRX15631305, SRX15631306, and PRJNA383830. Source data are provided with this paper.

References

- Barrett, S. C. The evolution of plant sexual diversity. Nat. Rev. Genet. 3, 274–284 (2002).
- 2. Ming, R., Bendahmane, A. & Renner, S. S. Sex chromosomes in land plants. *Annu. Rev. Plant Biol.* **62**, 485–514 (2011).
- Charnov, E. L., Bull, J. J. & Maynard Smith, J. Why be an hermaphrodite? *Nature* 263, 125–126 (1976).
- 4. Pannell, J. R. & Jordan, C. Y. Evolutionary transitions between hermaphroditism and dioecy in animals and plants. *Annu. Rev. Ecol. Evol. Syst.* **53**, 183–201 (2022).
- Charlesworth, D. & Charlesworth, B. Population genetics of partial male-sterility and the evolution of monoecy and dioecy. *Heredity* 41, 137–153 (1978).
- 6. Charlesworth, B. & Charlesworth, D. A model for the evolution of dioecy and gynodioecy. *Am. Nat.* **112**, 975–997 (1978).
- 7. Dai, C. & Galloway, L. F. Male flowers are better fathers than hermaphroditic flowers in andromonoecious Passiflora incarnata. *N. Phytol.* **193**, 787–796 (2012).
- Huang, S. Q. Flower dimorphism and the maintenance of andromonoecy in Sagittaria guyanensis ssp. lappula (Alismataceae). *N. Phytol.* **157**, 357–364 (2003).
- 9. Hedhly, A. Sensitivity of flowering plant gametophytes to temperature fluctuations. *Environ. Exp. Bot.* **74**, 9–16 (2011).
- Akagi, T. et al. A Y-encoded suppressor of feminization arose via lineage-specific duplication of a cytokinin response regulator in kiwifruit. *Plant cell* **30**, 780–795 (2018).
- Akagi, T., Henry, I. M., Tao, R. & Comai, L. A Ychromosome–encoded small RNA acts as a sex determinant in persimmons. *Science* **346**, 646–650 (2014).
- 12. Müller, N. A. et al. A single gene underlies the dynamic evolution of poplar sex determination. *Nat. Plants* **6**, 630–637 (2020).
- 13. Harkess, A. et al. Sex determination by two Y-linked genes in garden asparagus. *Plant cell* **32**, 1790–1796 (2020).

- Boualem, A. et al. A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. *Science* **321**, 836–838 (2008).
- 15. Boualem, A. et al. A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. *Science* **350**, 688–691 (2015).
- 16. Martin, A. et al. A transposon-induced epigenetic change leads to sex determination in melon. *Nature* **461**, 1135–1138 (2009).
- Rashid, D. et al. Ethylene produced in carpel primordia controls CmHB40 expression to inhibit stamen development. *Nat. Plants* 9, 1675–1687 (2023).
- 18. Zhang, S. et al. The control of carpel determinacy pathway leads to sex determination in cucurbits. *Science* **378**, 543–549 (2022).
- Wells, J. N. & Feschotte, C. A field guide to eukaryotic transposable elements. Annu. Rev. Genet. 54, 539–561 (2020).
- Kapitonov, V. V. & Jurka, J. Harbinger transposons and an ancient HARBI1 gene derived from a transposase. DNA Cell Biol. 23, 311–324 (2004).
- 21. Fedoroff, N. V. Transposable elements, epigenetics, and genome evolution. *Science* **338**, 758–767 (2012).
- Alonso, J. M., Hirayama, T., Roman, G., Nourizadeh, S. & Ecker, J. R. EIN2, a bifunctional transducer of ethylene and stress responses in *Arabidopsis*. Science **284**, 2148–2152 (1999).
- Cui, X. & Cao, X. Epigenetic regulation and functional exaptation of transposable elements in higher plants. *Curr. Opin. Plant Biol.* 21, 83–88 (2014).
- 24. Li, W. et al. EIN2-directed translational regulation of ethylene signaling in Arabidopsis. *Cell* **163**, 670–683 (2015).
- Guzman, P. & Ecker, J. R. Exploiting the triple response of Arabidopsis to identify ethylene-related mutants. *Plant cell* 2, 513–523 (1990).
- 26. Boualem, A. et al. Ethylene plays a dual role in sex determination and fruit shape in cucurbits. *Curr. Biol.* **32**, 2390–2401.e2394 (2022).
- 27. Va, D. et al. Transposon display identifies individual transposable elements in high copy number lines. *Plant J.* **13**, 121–129 (1998).
- Yang, G., Zhang, F., Hancock, C. N. & Wessler, S. R. Transposition of the rice miniature inverted repeat transposable element mPing in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **104**, 10962–10967 (2007).
- Naito, K. et al. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* 461, 1130–1134 (2009).
- Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* 20, 207–220 (2019).
- Sinzelle, L. et al. Transposition of a reconstructed Harbinger element in human cells and functional homology with two transposon-derived cellular genes. *Proc. Natl Acad. Sci. USA* **105**, 4715–4720 (2008).
- Bartlett, A. et al. Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat. Protoc. 12, 1659–1672 (2017).
- Makarevitch, I. et al. Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet.* 11, e1004915 (2015).
- Seidl, M. F. & Thomma, B. P. Transposable elements direct the coevolution between plants and microbes. *Trends Genet.* 33, 842–851 (2017).
- Zhao, G. et al. A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. *Nat. Genet.* 51, 1607–1615 (2019).
- Nelson, M. G., Linheiro, R. S. & Bergman, C. M. McClintock: an integrated pipeline for detecting transposable element insertions in whole-genome shotgun sequencing data. *G3-Genes Genom. Genet.* 7, 2763–2778 (2017).
- Adrion, J. R., Song, M. J., Schrider, D. R., Hahn, M. W. & Schaack, S. Genome-wide estimates of transposable element insertion and deletion rates in Drosophila melanogaster. *Genome Biol. Evol.* 9, 1329–1340 (2017).

- 38. Pichot, C. et al. Cantaloupe melon genome reveals 3D chromatin features and structural relationship with the ancestral *Cucurbitaceae* karyotype. *Iscience* **25**, 103696 (2022).
- 39. Pietzenuk, B. et al. Recurrent evolution of heat-responsiveness in Brassicaceae COPIA elements. *Genome Biol.* **17**, 1–15 (2016).
- Switzenberg, J. A., Little, H. A., Hammar, S. A. & Grumet, R. Floral primordia-targeted ACS (1-aminocyclopropane-1-carboxylate synthase) expression in transgenic Cucumis melo implicates fine tuning of ethylene production mediating unisexual flower development. *Planta* 240, 797–808 (2014).
- 41. Martínez, C. & Jamilena, M. To be a male or a female flower, a question of ethylene in cucurbits. *Curr. Opin. Plant Biol.* **59**, 101981 (2021).
- 42. Hobza, R. et al. Impact of repetitive DNA on sex chromosome evolution in plants. *Chromosome Res.* **23**, 561–570 (2015).
- Castanera, R., Morales-Díaz, N., Gupta, S., Purugganan, M. & Casacuberta, J. M. Transposons are important contributors to gene expression variability under selection in rice populations. *Elife* 12, RP86324 (2023).
- 44. Domínguez, M. et al. The impact of transposable elements on tomato diversity. *Nat. Commun.* **11**, 4058 (2020).
- 45. Carpentier, M.-C. et al. Retrotranspositional landscape of Asian rice revealed by 3000 genomes. *Nat. Commun.* **10**, 24 (2019).
- 46. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
- Lambert, I., Paysant-Le Roux, C., Colella, S. & Martin-Magniette, M.-L. DiCoExpress: a tool to process multifactorial RNAseq experiments from quality controls to co-expression analysis through differential analysis based on contrasts inside GLM models. *Plant Methods* 16, 1–10 (2020).
- Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS J. Integr. Biol.* 16, 284–287 (2012).
- Shahryary, Y., Hazarika, R. R. & Johannes, F. MethylStar: A fast and robust pre-processing pipeline for bulk or single-cell whole-genome bisulfite sequencing data. *BMC Genom.* 21, 1–8 (2020).
- Akalin, A. et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* 13, 1–9 (2012).
- Catoni, M., Tsang, J. M., Greco, A. P. & Zabet, N. R. DMRcaller: a versatile R/Bioconductor package for detection and visualization of differentially methylated regions in CpG and non-CpG contexts. *Nucleic Acids Res.* 46, e114 (2018).
- 53. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
- 54. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
- 55. Latrasse, D. et al. The quest for epigenetic regulation underlying unisexual flower development in *Cucumis melo. Epigenetics Chromatin* **10**, 1–17 (2017).
- 56. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160 (2016).

Acknowledgements

The authors thank Olivier Martin for the critical reading of the manuscript. We thank Vincent Rittener for technical assistance, and IPS2 glasshouse staffs and the research facilities provided by GAFL and the experimental unit, for plant handling. We thank the IPS2 bioinformatics team for providing help and computing resources. This work was supported by European Research Council(grant ERC-NectarGland Project, No.101095736 to A.Be.), Agence National de la Recherche (grant NEC-TAR, No. ANR-19-CE20-0023 to A.Be.), Inititative d'Excellence Paris-Saclay (grant Lidex-3P, No. ANR-11-IDEX-0003-02 to A.Be. and A.Bo.).

Article

Author contributions

The conceptualization of the project was undertaken by A.Be. and A.Bo. The methodology. involved the efforts of H.H., S.Z., F.A.C., F.T., C.P., H.S.P., C.T., F.S., Q.C., and F.M. The investigation was conducted by A.Be., H.H., S.Z., F.A.C., M.V., A.Bo., C.D., and L.Q. The visualization was handled by H.H., S.Z., and M.V. Funding was acquired by A.Be., and project administration was led by A.Be. Supervision of the project was provided by A.Be., A.Bo., and C.D. The original draft of the writing was done by A.Be. and S.Z., and the review and editing were carried out by A.Be., S.Z., A.Bo., L.Q., and C.D.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-49250-9.

Correspondence and requests for materials should be addressed to Abdelhafid Bendahmane.

Peer review information *Nature Communications* thanks Pierre Bourguet and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/ licenses/by/4.0/.

© The Author(s) 2024

	Nb. Plant	Andromonoecious	Androecious	chi-square	p-value
CAM_WT	10	10	0		
CAM106	10	0	10		
CAM106 x CAM_Sister	200	147	53	0 088889	0 7656
F2 progeny	200	147	53	0.066669	0.7656





d

Sexual morphs	Nb. Plant	Plants homozygous for AndroPIF insertion
CAM_WT (Andromonoecious)	10	0%
CAM106 (Androecious)	10	100%
Andromonoecious F2 plants	150	0%
Androecious F2 plants	57	100%

Figure S1. Sexual morphs observed in androecious *CAM106* spontaneous mutant and andromonoecious CAM_WT line.

a Segregation analysis of sexual morphs in CAM_WT x CAM106 F2 progeny. Nb, total number of plants. **b** Graphic presentation of flower sexual types observed in CAM_WT and in CAM106 lines. **c** Screenshot of CLC Genomics Workbench software showing the insertion of *AndroPIF* into 5' UTR of *CmEIN2*. Discordant reads are shown in green and red color while paired reads are shown in blue and purple. d, PCR screening for *AndroPIF* insertion in F2 plants, described in a, and plants harboring recombination linked to *a3*.



Figure S2. AndroPIF is a class II DNA transposon encoding 2 ORFs, SANT1 and Transposase.

a Structural representation and annotation of the *AndroPIF* transposon inserted at the *CmEIN2* locus. Picture is scaled. The 3 bp Target Site Duplication (TSD), and the Terminal Inverted Repeats(TIR) are shown. **b** Amino-acids sequence comparison of *AndroPIF* transposase, and ALP1 and HARBI1 proteins of *M. truncatula* (Mt), *O. sativa* (Os), *V. vinifera* (Vv), and *A. lyrata* (Al). The conserved regions are underlined and the conserved DDD residuals are highlighted by blue dots. **c** Amino-acids sequence comparison of *AndroPIF* SANT1/Myb protein, and homologous proteins from *M. truncatula*, *V. vinifera*, *S. bicolor* (*Sb*), and *O. sativa*. The conserved W residuals are highlighted by green dots.



	Sequence	Cytosine	% of methylation	
	context	number	CAM_WT	CAM106
500 hrs	CG	10	0.00%	0.00%
before	CHG	19	0.00%	0.76%
Delote	CHH	131	0.15%	0.34%
500 hno	CG	10	0.00%	28.57%
500 bps in	CHG	32	0.00%	65.91%
	CHH	99	0.00%	40.91%
500 bps after	CG	12	1.11%	10.53%
	CHG	26	0.53%	0.43%
	CHH	127	0.21%	0.16%

Figure S3. DNA methylation analysis of *AndroPIF* and *CmEIN2* **5'UTR in CAM_WT and CAM106 mutant a** Bisulphite sequencing of 1kb region overlapping *AndroPIF* and *CmEIN2* 5'UTR. The percentage of methylated cytosine is indicated by vertical bars. **b** Percentage of methylated cytosine in CGN, CHG and CHH sequence context for the CAM_WT and CAM106. Data on 500bps before, within and after *AndroPIF* insertion site is shown.

Genotype	Mutation location	Type of transcript	Percentage of transcript	Predicted protein length (aa)
WT		-	100%	1291
		Intron 3&4 retention	30%	133
ein2-SP1	G1175A	Intron 3 retention	47%	172
	_	Exon 4 skipping	23%	1220
		Intron 3 retention	27%	172
ein2-SP2	G1246A	Cryptic acceptor site	73%	172





Figure S4. Transcript forms of CmEIN2 splicing mutant detected by reverse transcription PCR analysis

a Splicing forms of *ein2-SP1* and *ein2-SP2* transcripts in percent. Transcript variations were detected by reverse transcription PCR analysis of leaf mRNAs. The splicing forms were validated by Sanger sequencing, and quantification was carried out using imageJ analysis. **b** Schematic representations of the splicing forms, and predicted protein sequences. Red arrow indicated the mutations located in donner (*ein2-SP1*) and acceptor site (*ein2-SP2*).

а



Figure S5. Sexual morphs observed in *CmEIN2* TILLING mutants.

Graphic presentation of flower sexual types observed in wildtype monoecious melon (**a**), heterozygous ein2-SP1 (+/-) mutant (**b**), homozygous ein2-SP1 (-/-) mutant (**c**), heterozygous ein2-SP2 (+/-) mutant (**d**), and homozygous ein2-SP2 (-/-) (**e**) mutant.



Figure S6 Transcriptome analysis of female, hermaphrodite and males flowers

a Number of differentially expressed genes (DEGs) in pairwise comparison groups. F, female (*wip1*), H, hermaphrodite (*wip/ein2* double mutant), M, males (acs7/*ein2* double mutant) flowers. **b** Venn diagram showing overlap between the pairwise comparison groups. **c-f** qPCR analysis of selected candidate genes associated with stamen development (**c**), S+; stamen inhibition (**d**), S-; carpel development (**e**), C+; and carpel inhibition (**f**), C-. Data are mean \pm SE, *P < 0.05, **P < 0.01, two-way ANOVA (with Tukey's post hoc test). M, Male flowers; H, hermaphrodite flowers; F, Female flowers. n= 3 biological replicates.



Figure S7. Differential gene expression analysis between female (*wip1*), hermaphrodite (*wip1/ein2*) and male (*acs7/ein2*) flowers. a Heat maps of E- ethylene-related genes differentially expressed in female compared to hermaphrodite and male flowers. The *z*-score scale represents mean-subtracted regularized log-transformed read counts. b Heat maps of E+ genes differentially expressed in female compared to hermaphrodite and male flowers. c Heat maps of C+ (up) and C- (down) ethylene-related DEGs expressed in male compared to hermaphrodite and female flowers. d Heat maps of C+ DEGs expressed in male compared to hermaphrodite and female flowers. f Heat maps of C- ABA related DEGs expressed in male compared to hermaphrodite and female flowers. f Heat maps of C- ABA related DEGs expressed in male compared to hermaphrodite and female flowers.



Figure S8. Process of TEDseq analysis of AndroPIF related elements

a Schematic diagram of primer design for AndroPIF capture TEDseq. b TEDseq library construction for AndroPIF detection in melon genomes. c TEDseq pipeline for AndroPIF detection.

replication	Percent of reads mapping to androPIF reference sequence	Percent of split-clipped and discordant- pair reads mapping to Mono genome	Number of detected PILs	Number of PILs identified in rep1 and rep2	Reads of AndroPIF inserted in CmEIN2
rep1	71.40%	95.58%	367		10
rep2	70.24%	97.14%	496	341	24

ATACseq signal b * Low High PIL_commor PIL_denovo 4 * 2 4 A X ¥ * A X 4 1 4 1 X 4 * XXX . * 1 . 1 X * 1 * A X * 1 * X Δ. 10 С H3K9ac CG ATACseq CHH CHG 0.04 0.02 8 0.02 60 0.01 0.00 0.00 0.00 40 -0.01 -0.02 -0.02 2 -0.0 20 CK PIL new -0.03 -0.04 3.0Kb -3.0 3.0Kb -3.0 ò -3.0 3 0Kh ò ò 3.0 6 3.0Kb 3.0Kb -3.0 New PILs Common PILs random



a TED-seq data summary. AndroPIF insertion in EIN2 5'UTR was used as a positive control. Rep1 and rep2, are two independent assays using a pool of 50 CAM plants plus CAM106 as control. **b** Genomic positions of PILs and ATACseq signal density. **c** Level of H3K9ac, ATACseq, and DNA methylation in CHG, CHH, and CG context, in \pm 3kb sequences flancking insertion sites of Common PIFs, new PIFs, and random genomic regions (random).



Figure S10. Data summary of SANT1 DAPseq analysis

a RPKM scatter plot, showing correlation between R1 and R2 SANT1 ampDAPseq assays. Spearman correlation coefficient is indicated on the plot. **b** Sequence annotation of SANT1-binding sites displayed in percent. **c,d** Top motifs enriched in SANT1 binding sites, and CentriMo plots highlighting the position of the motifs relative to the summit of SANT1 binding peaks. **e,f** Heatmap (e) and metaplot (f) showing SANT1 binding signals relative to Reference PIF loci



b

	SANT1 peak	Number of PIF element			
TIR sequences	pairs (total=302)	2 borders supported by annotations	1 border supported by annotations	Not supported by annotations	
TIR perfect much	54	54	0	0	
TIR 1 mismatch	126	126	0	0	
TIR 2 mismatch	184	182	1	1	
TIR 3 mismatch	238	234	3	1	
TIR 4mismatch	277	260	14	3	
TIR Not found	25	17	4	4	



Figure S11. *AndroPIF* elements polymorphism analysis based on reference and non-reference TE insertions a Schematic diagram of reference and non-reference *PIF/Harbinger* insertions. b Summary of the number of reference *androPIF* elements identified by *in silico* analysis, *or* manually annotated, by searching TIR sequences within SANT1 DAPseq peaks. The analysis allow 1 to 4 mismatches in 'GGGCCCGTTTG', androPIF reference TIR sequence. A total of 182 *androPIF* elements were identified, and classified as full length PIF. c Number of new insertion per accession in different melon botanical groups.

Supplementary Table S1: EMS-induced TILLING mutations in CmEIN2

Mutations in gDNA	Mutations in cDNA	Type of Mutations	AA changes	Phenotypes
G1175A	-	Splicing		Androecious
G1246A	-	Splicing		Androecious
C2518T	C1324T	Missense	L442F	Monoecious
C2570T	C1376T	Missense	T459I	Monoecious
G2854A	G1660A	Missense	E554K	Monoecious
G3118A	G1924A	Missense	G642R	Monoecious
C3155T	C1960T	Missense	L654F	Monoecious
G3159A	G1964A	Missense	S655T	Monoecious
G3490A	G2296A	Missense	D766N	Monoecious
G3601A	G2407A	Missense	V803I	Monoecious
G3643A	G2449A	Missense	D817N	Monoecious
G3713A	G2519A	Missense	G840D	Monoecious
C3865T	C2671T	Missense	L891F	Monoecious
C3971T	C2777T	Missense	S926F	Monoecious
G4417A	G3223A	Missense	E1075K	Monoecious
C4536T	C3341T	Missense	S1114F	Monoecious
C5282A	C3521T	Missense	P1174L	Monoecious
G5486A	C3725T	Missense	A1242V	Monoecious
G5515A	G3754A	Missense	E1252K	Monoecious

Supplementary Table S2: List of primers used in this study.

Name	Sequence	Purpose
Chr8_3820_F	GTCTAATTATTAATAGTAAACAATACCATG	Positional cloning
Chr8_3820_R	GTAAAATTGGTGAAATGGTGTCGTATG	Positional cloning
Chr8_1725_F	GCTTTACGAAAGAAAAGGATAGCAAGC	Positional cloning
Chr8_1725_R	GTGTATAAAACATCTTGGAAGGGTGAG	Positional cloning
Chr8_2871_F	GATCGTTTGTCAAAGTTACACGATCATTTAG	Positional cloning
Chr8 2871 R	GTCCCTTACTGGGTATATCTTATACTCAC	Positional cloning
 Chr8_6571_F	GTGAGCACCTTCCCCGCTAGTAATCCGAGTG	Positional cloning
Chr8 6571 R	CAATTCTCCCATTTCCACTCTGCTTTC	Positional cloning
Chr8 4840 F	CATCTGACTTTAATGGACCAAGAAAGC	Positional cloning
Chr8 4840 R	CGACGATCCATTATAGTGGAGGATCCAT	Positional cloning
Chr8 7220 F	TGGTGTTTGTGTGCTAAATGGTTGCTTCTTAGT	Positional cloning
Chr8_7220_F		Positional cloning
Chr8_7220_K		Positional cloning
EIN2-qPCR-5UTR -F		Quantitative RT-PCR
CmACS7 aPCR F2	CCTTACTATCCTGGATTTGACAGAGA	Quantitative RT-PCR
CmACS7 qPCR R2	TGGAGCGTTGGATCGTTGCTC	Quantitaitve RT-PCR
Cm_GENE-A_F	GAAAGAGGAGATTGATGTTGGAGT	Quantitaitve RT-PCR
Cm_GENE-A_R	TTCAAAAGATGCCTCATATCCAC	Quantitaitve RT-PCR
CmWIP1F	TAGGGCTTCCAACTCCTTCCTCTT	Quantitaitve RT-PCR
CmWIP1R	CTTGCAATTGATGGGTGTGATCTTCTTG	Quantitaitve RT-PCR
CmAct2_qPCR_F	ATTCTTGCATCTCTAAGTACCTTCC	Quantitaitve RT-PCR
CmAct2_qPCR_R	CCAACTAAAGGGAAATAACTCACC	Quantitaitve RT-PCR
SANT_exon12_2-F	GCAATGGAATGAGCTTTTTAAGGTG	Quantitaitve RT-PCR
SANT_exon2-R	CTTCAACACCAACGTTTGTGTTGCTT	Quantitaitve RT-PCR
Tpase_exon1-F	CTCTAGCTTTTAACAAGGTTTTGAGG	Quantitaitve RT-PCR
Tpase_exon12-R	CCATCAATAGCACCAATACAATCCTT	Quantitative RT-PCR
attb1 SANT1 F1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCA	DAPseq
	TGTCACAAAAATCAACTGAACAAC	212.004
atth2R SANT1 R1	GGGGACCACTTTGTACAAGAAAGCTGGGTCCT	DAPsea
auozk_SANTI_KI	AAAGACTATTTATGAAATCTCTC	DAI seq
	AATGATACGGCGACCACCGAGATCTACACTCT	
AndroPIE-P1	TTCCCTACACGACGCTCTTCCGATCTCAATGAA	TFDsea
	TTAATGATCGTTGAAGAT (Adaptor sequence is	TEDSeq
	underlined)	
	AATGATACGGCGACCACCGAGATCTACACTCT	
AndroPIF-P2		TEDseq
	UATATIKI VAATCAAATTI	
	AATGATACGGCGACCACCGAGATCTACACTCT	
AndroPIF-P3		TEDseq
	AATGATACGGCGACCACCGAGATCTACACCTAC	TEDara
AnaroPIF-P4		TEDseq
Illumina P7 sequencing	5'-CAAGCAGAAGACGGCATACGAGAT XXXXXX GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC	TEDaga
index primer	-s-T-3' (XXXXXX is specific sequencing index)	TEDsey