



**HAL**  
open science

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

Hsin-Ya Huang, Siqu Zhang, Fadi Abou Choucha, Marion Verdenaud, Feng-Quan Tan, Clement Pichot, Hadi Shirazi Parsa, Filip Slavkovic, Qinghe Chen, Christelle Troadec, et al.

### ► To cite this version:

Hsin-Ya Huang, Siqu Zhang, Fadi Abou Choucha, Marion Verdenaud, Feng-Quan Tan, et al.. Harbinger transposon insertion in ethylene signaling gene leads to emergence of new sexual forms in cucurbits. *Nature Communications*, 2024, 15 (1), pp.4877. 10.1038/s41467-024-49250-9 . hal-04661928

**HAL Id: hal-04661928**

<https://hal.inrae.fr/hal-04661928v1>

Submitted on 25 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# 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<sup>1,3</sup>, Siqi Zhang<sup>1,3</sup>, Fadi Abou Choucha<sup>1,3</sup>, Marion Verdenaud<sup>1</sup>, Feng-Quan Tan<sup>1</sup>, Clement Pichot<sup>1</sup>, Hadi Shirazi Parsa<sup>1</sup>, Filip Slavkovic<sup>1</sup>, Qinghe Chen<sup>1</sup>, Christelle Troadec<sup>1</sup>, Fabien Marcel<sup>1</sup>, Catherine Dogimont<sup>2</sup>, Leandro Quadrana<sup>1</sup>, Adnane Boualem<sup>1</sup> & Abdelhafid Bendahmane<sup>1</sup>✉

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 staminal 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 mechanisms<sup>1</sup>. 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<sup>1,2</sup>.

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<sup>3,4</sup>. The second model relies on population genetic arguments, suggesting that dioecy becomes advantageous in the presence of inbreeding depression<sup>5,6</sup>. Consistent with these concepts, in monoecious species, the female function is prioritized when resources are ample, while the male

<sup>1</sup>Université Paris-Saclay, CNRS, INRAE, Université Evry, Institute of Plant Sciences Paris-Saclay (IPS2), 91190 Gif-sur-Yvette, France. <sup>2</sup>INRAE, Génétique et Amélioration des Fruits et Légumes (GAFL), 84143 Montfavet, France. <sup>3</sup>These authors contributed equally: Hsin-Ya Huang, Siqi Zhang, Fadi Abou Choucha.

✉ e-mail: [Abdelhafid.bendahmane@inrae.fr](mailto:Abdelhafid.bendahmane@inrae.fr)

function is expressed more often under adverse conditions, or when the female function is already fulfilled<sup>7</sup>. Male plants excel in pollen donation, potentially attributed to their capacity to produce a greater quantity of male flowers and pollen<sup>8</sup>. Besides, pollen development is highly sensitive to adverse temperatures<sup>9</sup>. 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 flowers<sup>10–18</sup>, 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 life<sup>19</sup>. 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 domains<sup>19</sup>. For instance, *Harbinger* transposons are DNA transposons which usually encode a DDE transposase and a SANT/Myb/trihelix domain-containing DNA-binding protein<sup>20</sup>. 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 modifications<sup>21</sup>.

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<sup>21,18</sup>. 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 cucurbits<sup>16</sup>. The *A* and *M* genes encode for two aminocyclopropane-1-carboxylic acid synthase enzymes, *CmACS7* and *CmACS11*, respectively<sup>14,15</sup>. *CmACS11* expression in lateral branches suppresses the expression of the male-promoting gene *CmWIP1*. *CmACS7* expression in carpel primordia of female flowers inhibits staminal development through a non-autonomous mechanism<sup>17</sup>.

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

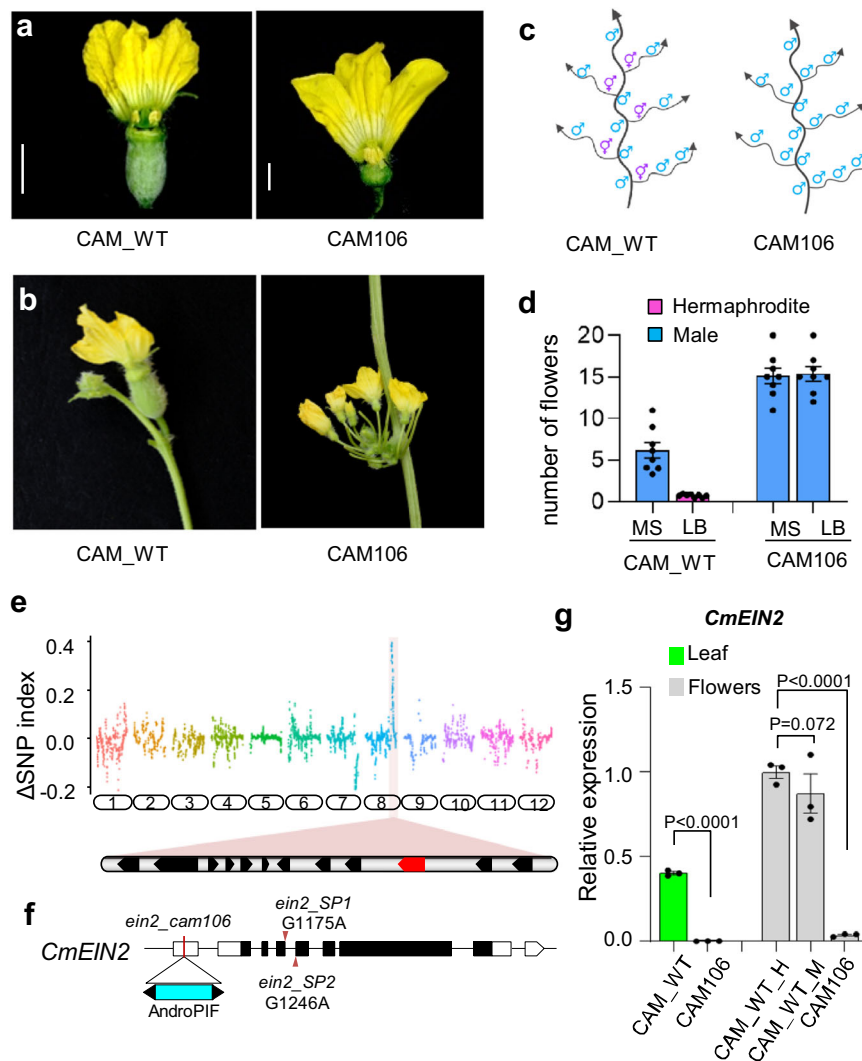
In andromonoecious melon, male flowers develop in inflorescences at every node of the main stem, and in distal nodes of the lateral branches. Hermaphrodite flowers develop at the first two nodes of lateral branches. During seed propagation of an andromonoecious Charantais line, hereafter named CAM\_WT, 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*

*CLAW* gene (*CmCRC*)<sup>15,18</sup>. So, we sequenced the promoter and the 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 andromonoecious, 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 bulked-genomic 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 segregant plants, further delimited *a3* to a single gene, *CmEIN2*, encoding the melon ortholog of the *ethylene-insensitive protein 2*<sup>22</sup> (Fig. 1f). Sequence analysis revealed an insertion of a DNA transposon of the *P Instability Factor (PIF)/Harbinger* superfamily<sup>8,20</sup>, 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).

Transposable elements are exposed to epigenetic modifications, with consequences on the expression of nearby genes, especially through the genomic spreading of DNA methylation<sup>23</sup>. 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*. Sixty-two induced mutations were identified, including two splicing, *ein2-SPI* (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*<sup>22</sup>. We selected the two splicing mutants for further analysis. *ein2-SPI* 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-SPI* 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 loss-of-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-SPI* 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.  $\Delta$  SNP index between andromonoecious and androecious bulks, calculated with a 1-Mb sliding

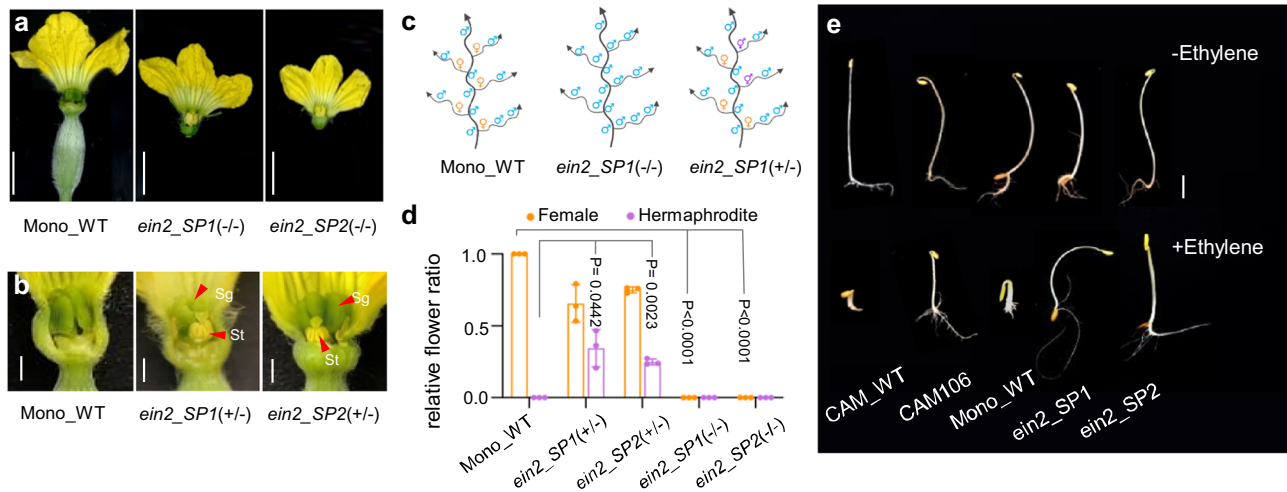
window. Arrows indicate candidate genes on chromosome 8. **f** Structure of the *CmEIN2* gene and the position of *AndroPIF* insertion. The *ein2\_SP1* and *ein2\_SP2* splicing mutations are shown by arrowhead. **g** Quantitative real-time PCR (qPCR) of *CmEIN2* in leaf, hermaphrodite (H) and male (M) 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 signaling<sup>24</sup>, 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 formation<sup>25</sup>. 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 staminal 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 primordia<sup>18</sup>. 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 staminal inhibition, we crossed *ein2* mutant to *wip1*



**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.

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

Ethylene has been associated with fruit shape development<sup>26</sup>. 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 *acs7*<sup>6</sup> 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, *CmWIPI*. Subsequent to the expression of *CmACS7* in carpel primordia, EIN2 is recruited in the stamena inhibition processes. Following the sex determination phase, EIN2 promotes fruit shape elongation.

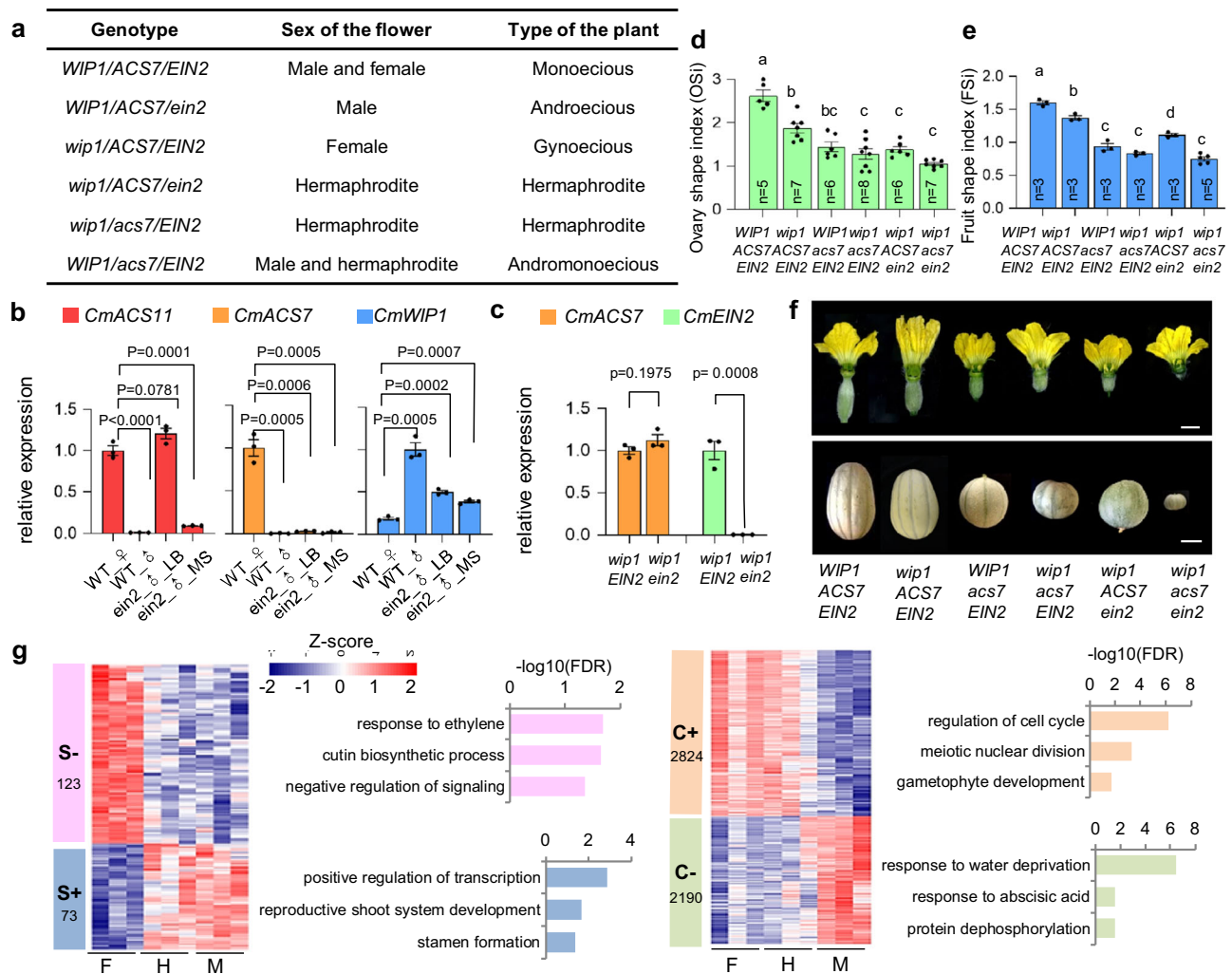
### EIN2 signaling plays contrasting roles in carpel development and stamena inhibition

To gain insight into the gene network associated with *CmEIN2* roles in stamena 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: stamena suppression (S-), stamena 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 stamena 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 stamena 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 stamena inhibition. In developing carpel, through inhibition of the carpel repressor, *CmWIPI* (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 stamena, EIN2 promotes the expression of genes associated with negative regulation and signaling, and the repression of the stamena-promoting gene, *PI* (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 (TEDseq<sup>27</sup>; 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 | *CmEIN2* is required both for carpel development and stamena 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 (♀) or male (♂) flowers, and in male flowers (♂) 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. ( $n = 3$  biological 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

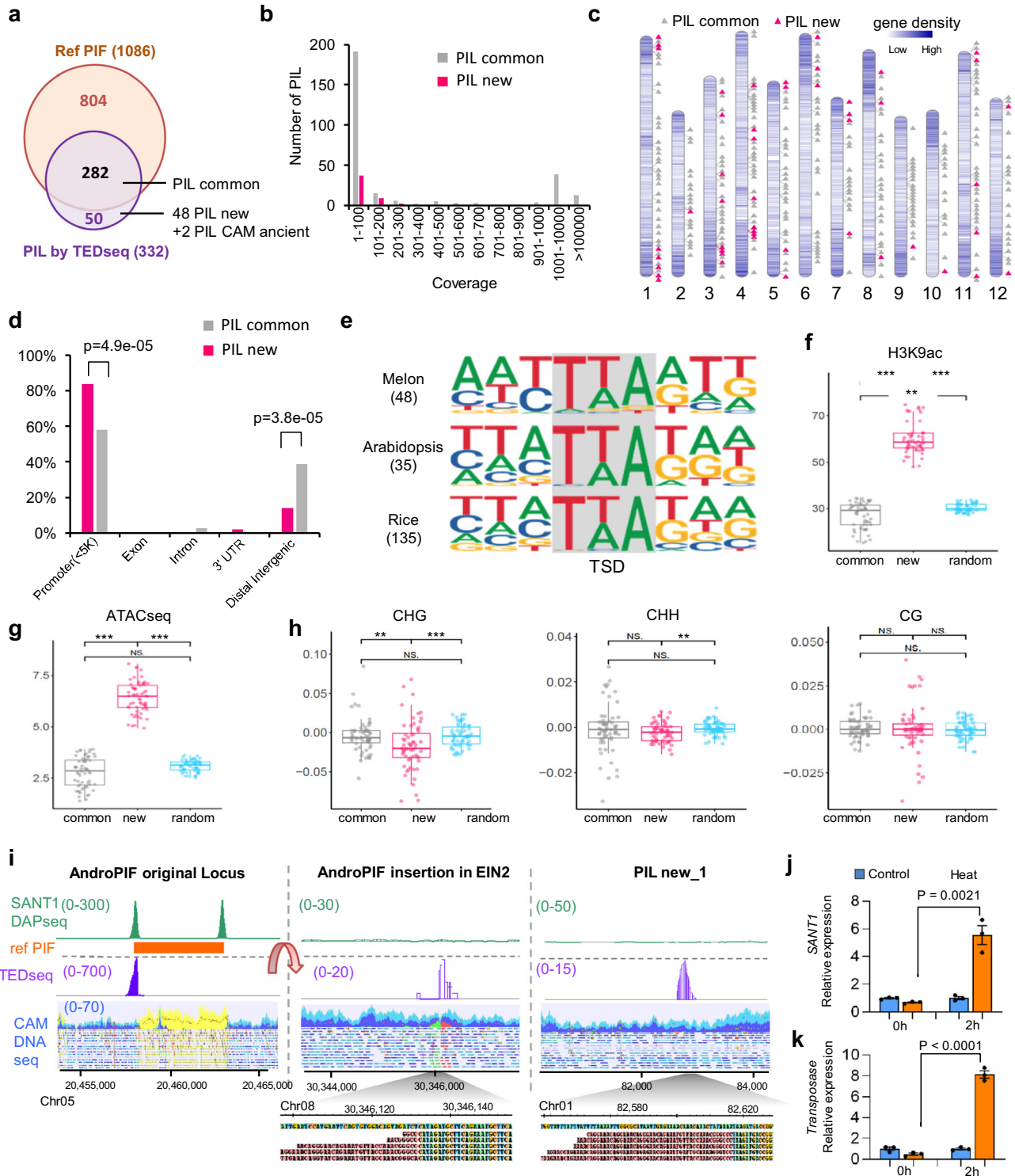
fruit phenotypes of the genotypes shown in (d and e). Scale bar in the flower panel = 1 cm. Scale bar in the fruit panel = 5 cm. **g** Gene-wise hierarchical clustering heatmap of differentially expressed genes (DEGs, adjusted  $P$  value < 0.05) between female (F, *wip1* mutant), hermaphrodite (H, *wip1/ein2* 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 processes 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 chromosomes, 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 *AndroPIF* insertion sites revealed the consensus motif 'MWYTWARWK' (Fig. 4e), similar to that previously reported for PIF transposons<sup>28,29</sup>. Open chromatin, H3K9ac histone mark, and low CHG/CHH DNA methylation are key features associated with active chromatin<sup>30</sup>. 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* SANTI protein binds to PIF TIR sequences

*AndroPIF*, like most *PIF/Harbinger* transposons encodes two proteins: a nuclease and a DNA binding protein, SANTI (Supplementary Fig. S2)<sup>20</sup>. Physical interaction between the two proteins is necessary for *PIF/Harbinger* transposition<sup>28,31</sup>. To further characterize *AndroPIF*, and related *PIF/Harbinger* transposon insertions in CAM melon line, we mapped genome wide DNA sequences binding to *AndroPIF* SANTI protein, using amplified DNA affinity purification sequencing (amp-DAP-seq)<sup>32</sup>. A total of 1521 SANTI 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 SANTI1 insertion sites in melon genome revealed two consensus motifs. Motif1 'TTYTGTTCTTGT' mapped at peak summits, while motif 2 'AGGCCCGTTTGTAACGTT' 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 'GGGCCGTTG' of *AndroPIF*, pointing to the role of SANTI1 in binding to the transposon extremities. Consistent with this, we systematically observed two SANTI1 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<sup>33,34</sup>, we analyzed whether the expression of *AndroPIF* *TRANSPOSASE* and *SANTI* 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 *SANTI* 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<sup>28,29</sup>. **f–h**, Boxplots of H3K9ac (f), ATACseq(g), and DNA methylation(h) (CHG, CHH, and CG) levels in bin sizes of 100 bp in  $\pm 3$  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

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 SANTI 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 SANTI (**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  $\pm$  SE. (*n* = 3 biological replicates). Source data are provided as a Source Data file.

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) accessions<sup>35</sup>, collected from different geographic areas, was determined using short-read resequencing, and TEFLON method<sup>36,37</sup>. Integration of PIF polymorphic insertion sites, SANTI 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 genome<sup>38</sup> 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)<sup>35</sup>. 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-new-insertions”. Distribution of non-reference-new-insertions in melon germplasm revealed active transpositions in wild melon accessions compared to melon breeding materials (Fig. 5d, e). As in CAM\_WT, non-reference-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 *CmACSII* in carpel primordia, of flowers developing on lateral branches, *CmEIN2* is recruited to

suppress the expression of the carpel inhibitor, *CmWIPI*. 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 receptors<sup>17</sup>. In stamina primordia, *CmEIN2* is then recruited to mediate stamina inhibition, leading to female flower development. If *CmEIN2* is inactivated, the repression of *CmWIPI* expression is relieved leading to androecy. Inactivation of both *CmEIN2* and *CmWIPI* lead to hermaphrodite flower development (Fig. 6).

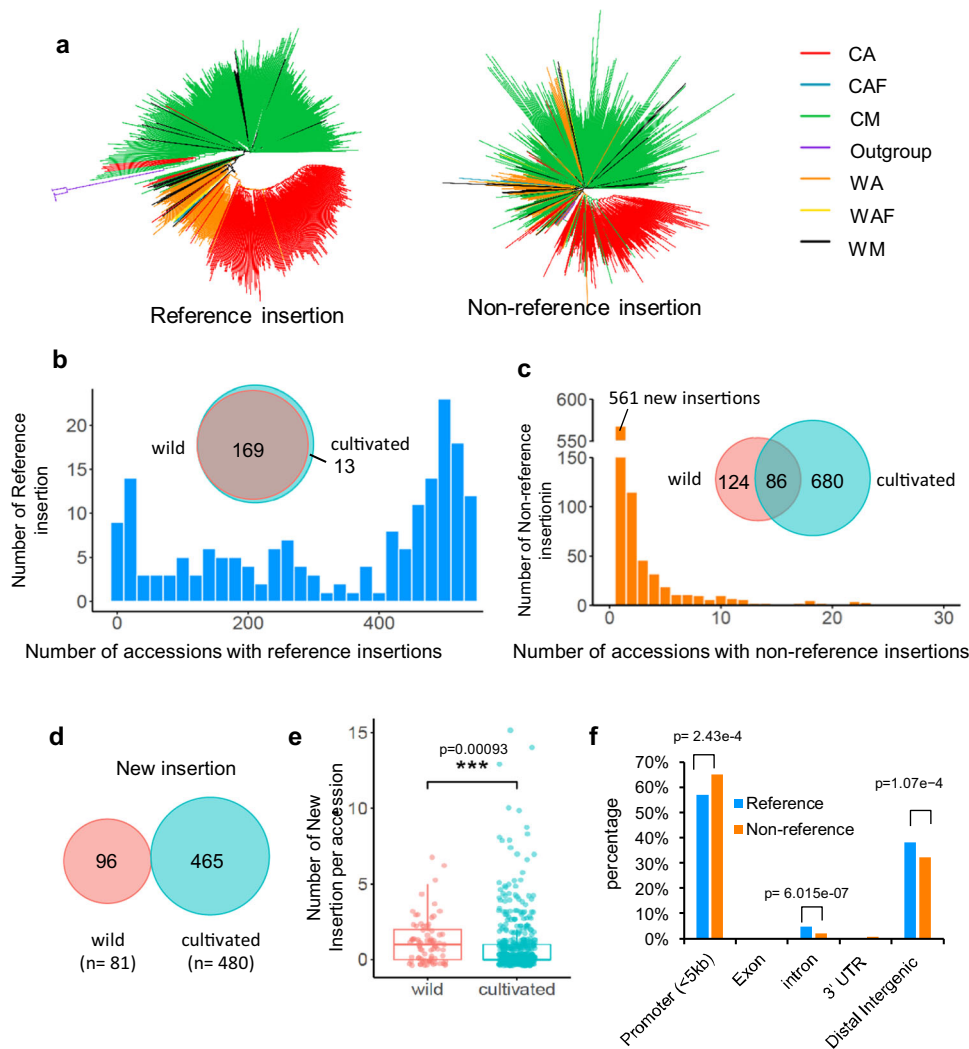
Transposons are powerful drivers of genome evolution and plasticity and contribute to genetic variability, potentially under stress conditions<sup>34</sup>. We thus tested whether *PIF/Harbinger* mobility is induced by adverse conditions. Structurally, *PIF/Harbinger* transposons are predicted to encode 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<sup>33,39</sup>. In rice, the DNA transposon *mPing* can be activated in response to cold and salt stress<sup>27</sup>; in tobacco, *Tnt1* element can be induced by different biotic and abiotic stress<sup>34</sup>.

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<sup>40,41</sup>. Consistent with this scenario, we found *CmEIN2* required for the transcriptional repression of the male-promoting gene *CmWIPI*, leading to female flower development (Fig. 6). Under adverse conditions, fruit set could be costly, inactivation of *CmEIN2* relieves *CmWIPI* 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 gynodioecious plants<sup>16</sup>. 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 chromosomes<sup>42</sup>.

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; CA, cultivated African; WM, wild melo; CM, cultivated melo; WA, wild agrestis; CAF, cultivated agrestis. **b**, **c** Histogram and venn plots of reference (**b**) and non-reference (**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

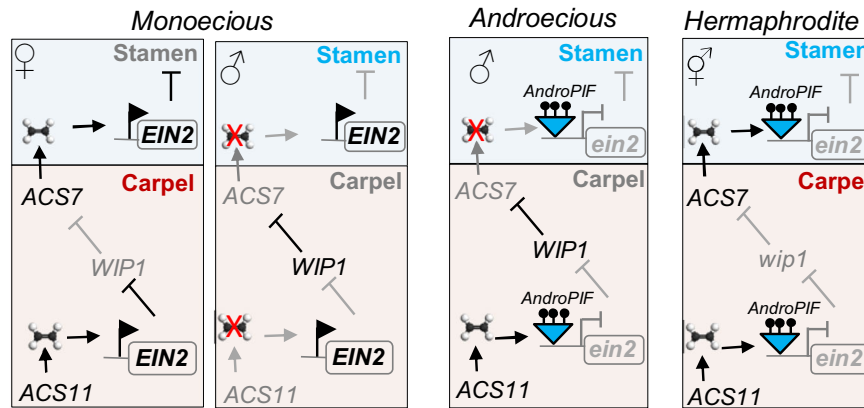
new-non-reference insertions per accession, in wild and cultivated accessions. *P*-values are calculated based on a two-sided Wilcoxon rank-sum test. For each boxplot, the lower and upper bounds of the box indicate the first (Q1) and third (Q3) quartiles, respectively, the center line indicates the median, 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 from a Z-test are indicated and applied only to regions with non-zero values for both groups.

The transcriptome analysis of flower buds collected from female (*wip1/EIN2*), 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 SANTI 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

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 SANTI 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, reference 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 breeding<sup>43–45</sup>. 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 staminal primordia. The chemical symbol indicates ethylene gas. Red cross indicates no ethylene produced by *ACS7* or *ACS11*. Ethylene

produced by *ACS7* in the carpel is perceived in the staminal primordia likely through a spatially differentially expressed ethylene receptors<sup>17</sup>. 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*.

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 life<sup>19</sup>, 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 CMiso1.1<sup>38</sup> (here referred to as Mono\_WT). The SNP-index across all loci was calculated, and the delta ( $\Delta$ ) SNP-index was determined by subtracting the SNP-indices of the two bulks at each locus (SNP-index\_mutant - SNP-index\_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  $\mu$ M ethylene. Ethylene solution at 50  $\mu$ M concentration was prepared by diluting 1 M ethephon (Sigma) in 5 mM Na<sub>2</sub>HPO<sub>3</sub> buffer, as described previously<sup>25</sup>.

### 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, gynoeceous, androecious and hermaphrodite. Androecious plants bear only male flowers.

Gynoeceous 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-of-function alleles leading to gynoecey (*wip1*), androecy (*acs11*) or andromonoecy (*acs7*) used in the epistasis analysis are described in Boualem et al<sup>14</sup>, Martin et al<sup>16</sup>, and Boualem et al<sup>15</sup>, respectively<sup>14–16</sup>. 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  $\mu$ g of total RNA with the Superscript<sup>®</sup> III reverse transcriptase (Invitrogen). Primer design was performed using the Primer3 software ([http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](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<sup>®</sup> 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)<sup>46</sup>, and the mapped reads were assigned to genes with featureCount (v2.0.3)<sup>47</sup>. Differentially expressed gene calling was conducted using DiCoExpress<sup>48</sup>, with a p-value adjusted <0.05. GO enrichments were completed by ClusterProfiler<sup>49</sup> 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 MethyStar

pipeline<sup>50</sup>. The differentially methylated sites (DMRs) were identified by methylkit<sup>51</sup> and DMRcaller<sup>52</sup>.

### TE display sequencing

TEseq 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 1 µg 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’<sup>53</sup>. 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 Cutadapt<sup>54</sup>, 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<sup>17,55,56</sup> (accession number: PRJNA383830).

### DNA affinity purification sequencing

AndroPIF SANTI coding 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 SANTI are listed in the Supplementary Table 2. Expression of HALO-SANTI fusion protein and enrichment of DNA targets were performed as described previously<sup>32</sup>. 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 previously<sup>18</sup> (accession numbers: SRX15631305, SRX15631306).

### Transposable elements insertion detection in melon germplasm

The genomic sequences of the 567 *C. melo* accessions, and the 7 related species (Supplementary Data 6) were retrieved from previous study<sup>35</sup>. The genome wide identification of *AndroPIF* related elements was performed using the TEFLon method and the McClintock pipeline<sup>36</sup>.

TEs were subsequently annotated manually, using SANTI ampDAPseq peaks, and the search of TIR sequences ‘GGGCCGTTTG’, 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 greenhouse 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, TEseq, 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).
- 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).
- 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).
- Charlesworth, B. & Charlesworth, D. A model for the evolution of dioecy and gynodioecy. *Am. Nat.* **112**, 975–997 (1978).
- 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).
- 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 Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science* **346**, 646–650 (2014).
- Müller, N. A. et al. A single gene underlies the dynamic evolution of poplar sex determination. *Nat. Plants* **6**, 630–637 (2020).
- Harkess, A. et al. Sex determination by two Y-linked genes in garden asparagus. *Plant cell* **32**, 1790–1796 (2020).

14. 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).
17. 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).
19. Wells, J. N. & Feschotte, C. A field guide to eukaryotic transposable elements. *Annu. Rev. Genet.* **54**, 539–561 (2020).
20. Kapitonov, V. V. & Jurka, J. Harbinger transposons and an ancient HARB1 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).
22. 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).
23. 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).
25. 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).
28. 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).
29. Naito, K. et al. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* **461**, 1130–1134 (2009).
30. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
31. 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).
32. Bartlett, A. et al. Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat. Protoc.* **12**, 1659–1672 (2017).
33. Makarevitch, I. et al. Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet.* **11**, e1004915 (2015).
34. Seidl, M. F. & Thomma, B. P. Transposable elements direct the coevolution between plants and microbes. *Trends Genet.* **33**, 842–851 (2017).
35. 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).
36. 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).
37. 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. Pietzenek, B. et al. Recurrent evolution of heat-responsiveness in Brassicaceae COPIA elements. *Genome Biol.* **17**, 1–15 (2016).
40. Switzberg, 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).
43. 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).
47. 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).
48. 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).
49. 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).
50. 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).
51. Akalin, A. et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, 1–9 (2012).
52. 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 NECTAR, No. ANR-19-CE20-0023 to A.Be.), Initiative d'Excellence Paris-Saclay (grant Lidex-3P, No. ANR-11-IDEX-0003-02 to A.Be. and A.Bo.).

## 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