

Graph-based pangenomes and pan-phenome provide a cornerstone for eggplant biology and breeding

Received: 31 May 2025

Accepted: 30 September 2025

Published online: 11 November 2025

 Check for updates

Luciana Gaccione¹, Laura Toppino², Marie Bolger³, Maximilian Schmidt⁴, Maria Rosaria Tassone², Maria Sulli⁵, Emily Idahl⁶, David Alonso⁷, Giuseppe Aprea⁵, Paola Ferrante⁵, Véronique Lefebvre⁸, Hatice Filiz Boyaci⁹, Roland Schafleitner¹⁰, Stefano Gattolin¹¹, Richard Finkers^{12,13}, Matthijs Brouwer¹², Arnaud Bovy¹², Jaime Prohens¹², Ezio Portis¹, Sergio Lanteri¹, Giuseppe Leonardo Rotino², Giovanni Giuliano^{5,14}✉, Björn Usadel^{3,6}✉ & Lorenzo Barchi¹✉

Eggplant (*Solanum melongena* L.) is a major Solanaceous crop of Asian origin, but genomic resources remain limited compared to related species. Here, a core collection of 368 accessions spanning global diversity of *S. melongena* and wild relatives is phenotyped for agronomic, disease resistance and fruit metabolomic traits and resequenced. Additionally, 40 chromosome-level assemblies of *S. melongena*, its progenitor *S. insanum* and the allied species *S. incanum* enable the construction of two graph-based pangenomes, capturing broad genetic variation. We demonstrate the power of these datasets by identifying major loci controlling prickliness and resistance to *Fusarium oxysporum* f. sp. *melongenae*, driven by SVs affecting the *LONELY GUY 3* gene and a resistance gene cluster, respectively, as well as a mutation in a *GDSL-like esterase/lipase* gene altering the levels of dicaffeoyl-quinic acids. These findings provide a cornerstone for pangenome-assisted breeding, enabling detailed analyses of genetic diversity, domestication history, and trait evolution in eggplant.

Eggplant (*Solanum melongena* L.) is a member of the Solanaceae family and a widely cultivated crop, with over 59 Mt produced globally¹. China and India are the top global producers, while Egypt, Türkiye, and Italy lead the production in the Mediterranean region. Unlike most Solanaceous crops, which are native to the New World^{2–6}, eggplant originates in Asia and belongs to the subgenus

Leptostemonum, also known as the “spiny Solanum” group⁷, phylogenetically distant from other *Solanum* crops like tomato and potato. *Solanum melongena*, its direct wild progenitor *S. insanum*, and the sister closely related species *S. incanum* form a well-supported clade, providing a rich source of genetic diversity for research and breeding efforts^{7–9}.

¹DISAFA, Plant Genetics, University of Turin, Grugliasco, TO, Italy. ²CREA, Research Centre for Genomics and Bioinformatics, Montanaso Lombardo, LO, Italy. ³IBG-4, CEPLAS, Forschungszentrum Jülich, Jülich, Germany. ⁴Plant Breeding Department, Geisenheim University, Geisenheim, Germany. ⁵ENEA, Casaccia Res Ctr, Roma, Italy. ⁶Faculty for Mathematica and Natural Sciences, Institute for Biological Data Sciences, Heinrich Heine University, Düsseldorf, Germany. ⁷Universitat Politècnica de València, Valencia, Spain. ⁸INRAE, GAFI, Montfavet, France. ⁹Faculty of Agriculture, University of Recep Tayyip Erdogan, Rize, Türkiye. ¹⁰The World Vegetable Center, Tainan City, Taiwan. ¹¹National Research Council of Italy, Institute of Agricultural Biology and Biotechnology (IBBA-CNR), Milan, Italy. ¹²Plant Breeding, Wageningen University & Research WUR, Wageningen, The Netherlands. ¹³Present address: GenNovation B.V., Wageningen, The Netherlands. ¹⁴Present address: Pandora Research Srls, Rome, Italy. ✉e-mail: giovannigiuliano1@gmail.com; b.usadel@fz-juelich.de; lorenzo.barchi@unito.it

Our previous studies on 3400 eggplant accessions provided a snapshot of the worldwide genetic diversity of this species, and supported the hypothesis of independent domestication events in two distinct centers: Southeast Asia and the Indian subcontinent, where both anthropic and environmental selection acted on different genomic regions^{8,10}. A first pangenome, constructed using short-read resequencing of 23 representative eggplant accessions, and one accession each of *S. insanum* and *S. incanum*, added 51.5 Mb and 816 genes to the reference genome; this study also identified 53 selective sweeps related to key domestication traits such as fruit color, prickliness, and fruit shape¹¹.

The advent of graph-based pangenomes and super-pangenomes, representing single or multiple species, respectively^{12,13}, has provided deeper insights into the evolutionary history, domestication, and genetic relationships within and between species/taxa. These reference-unbiased frameworks enable more accurate comparisons by identifying core, dispensable, and private genomes, as well as a better understanding of the structural variation at nucleotide-level resolution¹⁴.

In this study, we describe the construction of a core collection (CC) of 368 accessions, including 321 accessions that capture the worldwide genetic and phenotypic diversity of cultivated eggplant and 47 that represent its wild relatives. The full collection is phenotyped in multiple locations and resequenced using short reads. Long-read sequencing is performed on 33 *S. melongena*, three *S. insanum*, and four *S. incanum* accessions (Fig. 1a) to generate chromosome level-assemblies, which are integrated into a reference-unbiased graph-based pangenome. A second pangenome graph, limited to the 33 *S. melongena* accessions, is used for pangenome-wide association (Pan-GWA) analysis to identify genes/quantitative trait loci (QTLs) associated with key agronomic and metabolic traits, as well as in responses to biotic and abiotic stresses. We provide examples how these pangenomes and pan-phenome can be used to explore eggplant evolution and demography, and to identify the genomic basis of important traits, such as prickle development, resistance to *Fusarium* wilt and variation in fruit chlorogenic/isochlorogenic acids content.

Results

Construction of a worldwide eggplant core collection

CCs are essential for simplifying phenotyping and phenotype-genotype association studies, by reducing redundancy and maximizing diversity, thus identifying selection targets and enhancing the effectiveness of breeding efforts¹⁵. We have previously described and genotyped a worldwide collection of 3412 worldwide accessions from 105 countries^{8,16}. Using the genotyping data, we established a CC of 368 accessions (see “Methods”; Supplementary Fig. 1), of which 321 represent the worldwide genetic diversity of cultivated eggplant, and 47 are wild relatives from the primary, secondary, and tertiary gene pools (Supplementary Data 1). The latter comprised 13 accessions of *S. insanum* (the direct progenitor of cultivated eggplant) and 7 of the sister species *S. incanum*.

Genome sequencing and chromosome-scale genome assemblies

The 368 accessions were sequenced using short reads (2 × 150 bases), providing an average coverage of 20× per accession (Supplementary Data 1). Additionally, 40 accessions (33 of *S. melongena*, including GPE001970 (67/3), previously used to generate the v4.1 of the reference genome¹¹, three of *S. insanum* and four of *S. incanum*) representing the genetic diversity of the species (Fig. 1b) were sequenced using Oxford Nanopore Technology (ONT), generating almost 66 Gb of data per individual, with an overall read N50 of 32.2 kb (Supplementary Data 2). The genomes of each of the 40 accessions were assembled and polished, yielding an average genome size of 1.13 Gb, and contig N50 up to 83 Mb (Supplementary Data 3 and Fig. 1c).

To further enhance assembly quality, six *S. melongena* assemblies (including GPE001970) and two each of *S. insanum* and *S. incanum* were subjected to high-throughput chromosome conformation capture technique (Hi-C) scaffolding (Supplementary Data 4 and Supplementary Fig. 2). The remaining 30 assemblies were scaffolded using as a guide an accession from the same species (see “Methods”). After the scaffolding steps, on average 99.54% of the contig sequences were anchored on the 12 eggplant chromosomes. The average genome sizes were 1.106 Gb (*S. incanum*), 1.115 Gb (*S. insanum*), and 1.137 Gb (*S. melongena*) and the average scaffold N50 were 94.67 Mb (*S. incanum*), 90.76 Mb (*S. insanum*), and 93.43 Mb (*S. melongena*) (Supplementary Data 5).

The quality and completeness of the reference genomes was assessed using several parameters: Benchmarking Universal Single-Copy Orthologs (BUSCO) scores¹⁷ ranged between 98.3 and 98.7% (98.7% for version 5 of the reference GPE001970 genome vs 96.9% in the previously published v4.1)¹¹. All 40 assemblies reached the “reference” level (LAI > 10) of the LTR Assembly Index (LAI)¹⁸, with 16 reaching a “gold standard” level (LAI > 17), while the average QV score¹⁹ was 49.26 (less than 1 error in 63,131 bases, Supplementary Data 5).

Repetitive elements covered 72.8% (GPE013320; *S. insanum*) to 75.3% (GPE008640; *S. incanum*) of the total genome sequences, with long terminal repeat retrotransposons (LTR-RTs) being the predominant class (Supplementary Data 6 and Fig. 1d). *Solanum insanum* and *S. incanum* exhibited a higher percentage of LTR-RTs than *S. melongena*, compensated by a lower percentage of DNA transposons (Supplementary Data 6 and Fig. 1d).

Gene prediction was conducted by combining Helixer and BRAKER3 (see “Methods”). The number of protein-coding gene models per genome ranged from 30,886 to 33,449, with an annotation completeness using BUSCO ranging between 95.3% and 97.2% (Supplementary Data 5).

Phylogeny and synteny analyses

To clarify the phylogenetic relationships, we identified 69 single-copy and 9323 low-copy gene orthogroups, respectively, across all 40 accessions and 6 other plant species (see “Methods”). The two approaches produced similar phylogenetic tree structures (Supplementary Fig. 3a, b), with *S. melongena* accessions clustering together, and *S. insanum* and *S. incanum* positioned as a separate group. Minor topological differences were observed in both the single-copy and low-copy trees; in the former, a *S. incanum* accession (GPE008640) was positioned in the *S. melongena* group, reflecting an ambiguous genetic background for this accession. Consequently, we excluded *S. incanum* GPE008640 from the species-level comparison, adopting the resulting species trees for the subsequent analyses (Supplementary Fig. 4a, b). In both trees, *S. insanum* and *S. incanum* accessions appeared intermingled. These patterns likely reflect a combination of evolutionary processes, notably hybridization²⁰, introgression, and incomplete lineage sorting, leading to shared genetic variation and blurred species boundaries in the reconstructed trees. The estimated divergence times obtained using the single-copy tree, revealed that *S. incanum* diverged from the common ancestor of *S. melongena* and *S. insanum* approximately 0.75 million years ago (Mya), at the end of the Mid-Pleistocene climatic transition²¹, while the split between *S. insanum* (the direct progenitor) and *S. melongena* dated around 0.29 Mya (Pleistocene) (Fig. 2a).

The analysis of gene family evolution highlighted how domestication and ecological adaptation shaped distinct gene family trajectories across the “Eggplant” clade (Fig. 2b, Supplementary Fig. 5a, b, and Supplementary Data 7). In *S. melongena*, expansions were mainly associated with stress adaptation (gene silencing, desiccation tolerance, cuticular wax biosynthesis), while contractions affected sugar transport and circadian metabolism. *Solanum insanum* showed expansion in detoxification and terpenoid biosynthesis, with

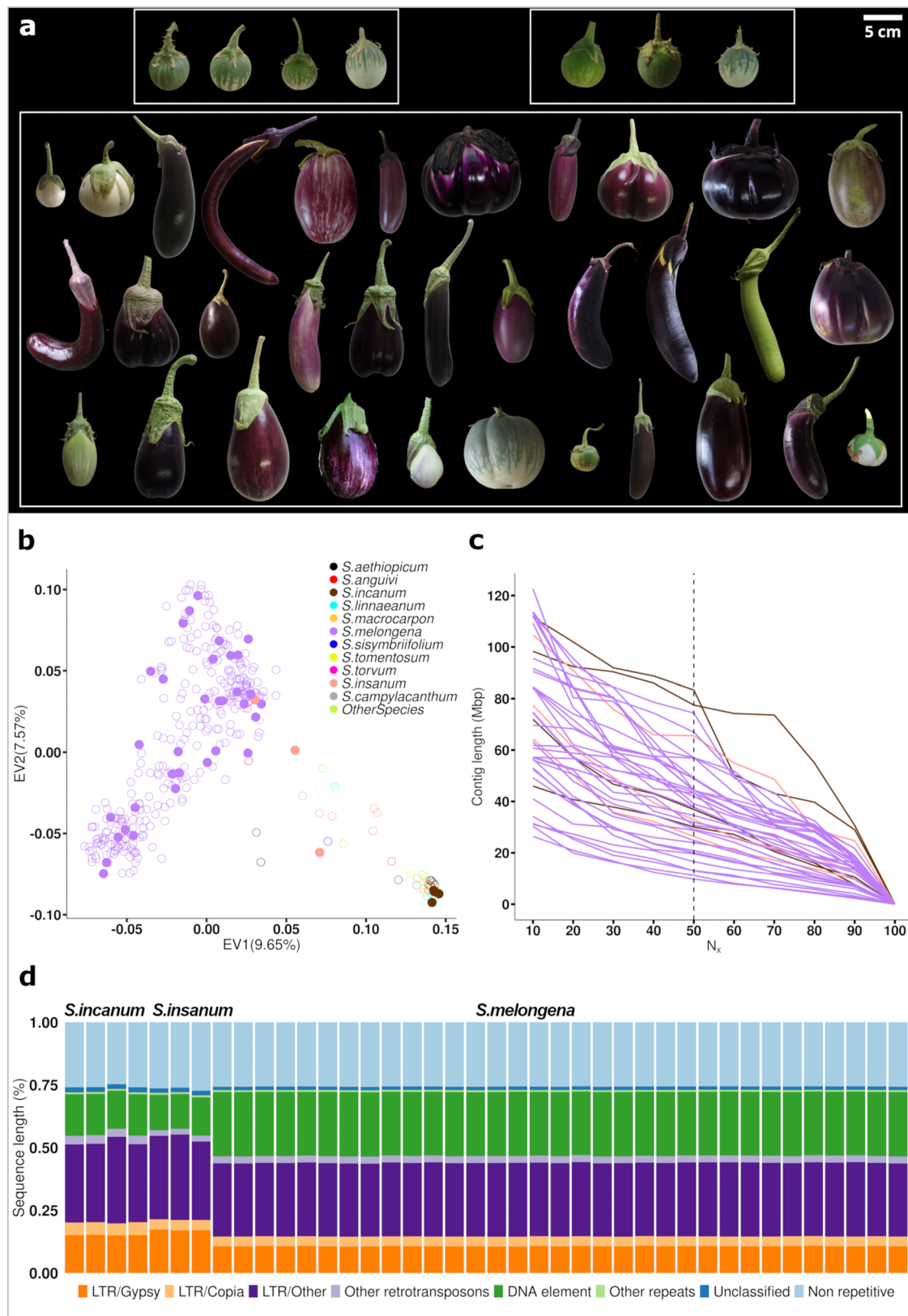


Fig. 1 | 40 eggplant reference genomes. **a** Fruit phenotypes of the 40 chromosome-scale accessions. In clockwise direction: fruits of the four *S. incanum*, the three *S. insanum*, and the 33 *S. melongena* accessions. *Solanum incanum* and *S. insanum* pictures are not in scale. **b** Principal component analysis (PCA) of component 1 (EV1) vs component 2 (EV2) for the 368 eggplant accessions in the core

collection; the 40 accessions with reference genomes are represented by solid dots; dots are colored by species. **c** Assessment of genome contiguity by contig N_{50} . The N_x value represents x% of the total contigs length that is covered by the shortest contig length. **d** Fractions of transposable elements, low-complexity (simple repeats), and non-repetitive sequences in the 40 reference genomes.

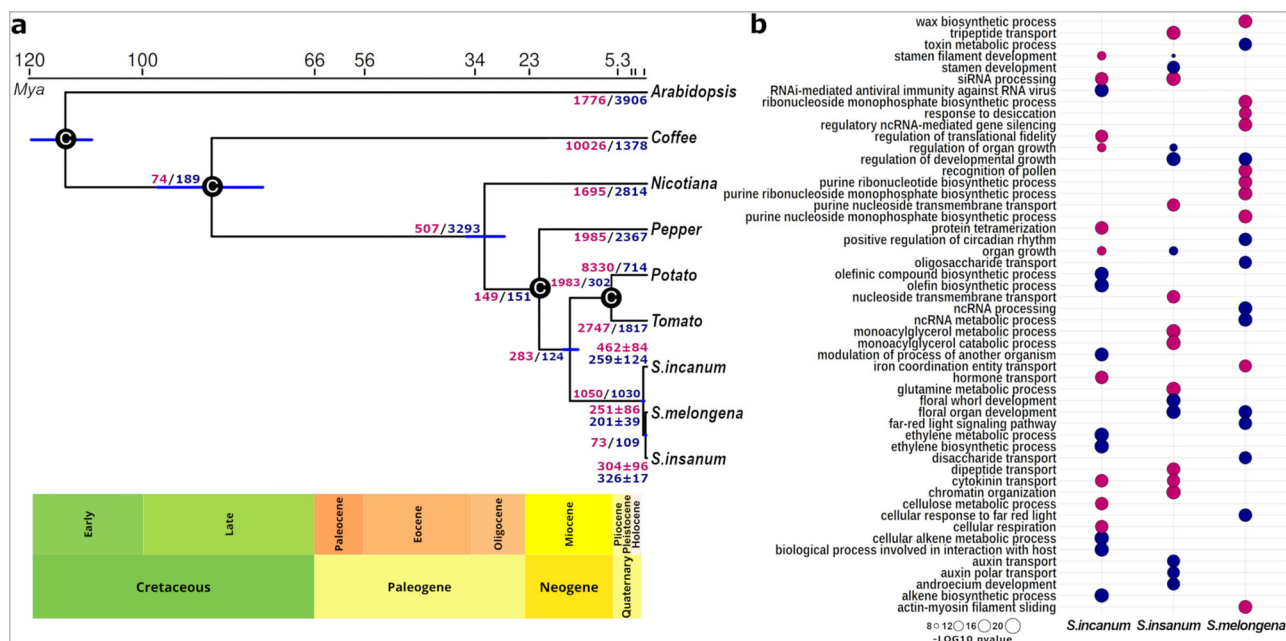


Fig. 2 | Divergence times of eggplant and related species and gene family evolution. **a** Time-calibrated single-copy species tree illustrating the estimated divergence times among *S. melongena*, *S. insanum*, *S. incanum*, and six other plant species. Calibration points are marked as black dots. Gene family expansions and contractions are indicated by magenta and blue values, respectively, along the branches. **b** GO enrichment analysis of the biological processes for expanded and

contracted gene families in *S. melongena*, *S. insanum*, and *S. incanum*. The enriched GO categories were determined using the hypergeometric test, followed by the Benjamin–Hochberg correction to obtain adjusted *P* values for multiple testing. Circle size represents the statistical significance ($-\log_{10} p$ -value), magenta color represents gene family expansions and blue represents contractions.

contraction of growth-related pathways. In *S. incanum*, gene family expansions predominantly involved growth and reproductive programs, whereas contractions were enriched in ethylene metabolism and host–pathogen responses. These patterns suggest that domestication reshaped gene family content towards stress adaptation in *S. melongena*, metabolic plasticity in *S. insanum*, and structural investment in *S. incanum*.

Synteny analysis of the 40 eggplant genome assemblies revealed strong overall collinearity, with limited structural variations such as inversions on chromosomes 1 and 10, and a few translocations (e.g., between chromosomes 7 and 9) within *S. melongena* (Fig. 3a and Supplementary Data 8). More rearrangements were observed in *S. insanum* and *S. incanum*, including a large translocation between chromosomes 1 and 5 in GPE008290 (*S. incanum*). In addition, expansion or contraction of syntenic regions is evident in some accessions, suggesting segmental duplications or deletions.

Paranome (i.e., paralogous within a genome) analysis identified 4456, 4757, and 4818 tandem duplications in *S. melongena*, *S. insanum*, and *S. insanum*, respectively (Fig. 3b). Tandem duplicates were consistently enriched for genes involved in secondary metabolite biosynthesis, with lineage-specific differences observed: lipid localization and transport terms were reduced in *S. incanum* and *S. insanum*, while phenylpropanoid metabolism was enriched (Supplementary Data 9). Conversely, terpene-related pathways were enriched in *S. melongena*. Segmental duplications were more abundant (7925, 7398, and 7810 for *S. melongena*, *S. insanum*, and *S. incanum*) and enriched across all three species for core functions such as cell growth, transcriptional activation, and translation. On the other hand, developmental processes were particularly enriched in *S. incanum* and *S. insanum*, while in *S. melongena*, there was a peculiar enrichment for energy metabolism and abscisic acid response (Supplementary Data 9).

A graph-based pangenome of the eggplant clade

To access the genetic diversity of *S. melongena* and its allied species, and avoid any bias towards a specific reference, a reference-free

pangenome graph (“PG-SMA”) was constructed with PGGB¹⁴ using the 40 chromosome-level assemblies (Fig. 4a). PG-SMA exhibited a total length of 2.94 Gb, i.e., over twofold that of our previously published reference-based pangenome (1.21 Gb)¹¹. It exhibited 141.8 M nodes, 194.7 M edges, an average node degree of 2.75 (Supplementary Data 10) and a total of 83,147 structural variants (SVs; Fig. 4a, b). The Jaccard similarities of the embedded paths in PG-SMA graph structure were used to investigate the phylogenetic relationship among the 40 accessions. In agreement with the results from both single-copy and low-copy species trees, *S. incanum* and *S. insanum* accessions appeared intermingled rather than forming clearly separated groups (Supplementary Fig. 6).

The PG-SMA variation graph revealed a predominance of soft-core nodes, highlighting a substantial dispensable genome content (~40%) as well as core genomic sequences (~50%, Supplementary Fig. 7). The PG-SMA growth curves modeled on protein families indicated a closed pangenome with a γ -value of 0, reaching essentially a plateau at 40 genomes (Fig. 4c). In contrast, the same curves modeled on genome sequences had a γ -value of 0.16, suggesting that additional genomic sequences remain to be characterized (Supplementary Fig. 8).

Population structure of cultivated eggplant and its allied species

Mapping of short reads from the resequenced CC accessions on the PG-SMA graph yielded 34,576,891 polymorphic sites. After quality filtering, 12,893,056 SNPs and 12,597 SVs were retained for phylogenetic and population structure analyses.

The maximum likelihood (ML) tree obtained using the SNP markers identified two main branches, including: (i) *S. melongena* accessions; (ii) *S. insanum*, *S. incanum*, and all the other species (Fig. 5a). Five *S. insanum* accessions grouped within *S. melongena*, likely due to feral escapes and/or admixture events²². Within the broader group, three sub-clusters were identified: (i) the “Eggplant” clade, which includes *S. insanum*, *S. incanum*, *S. linnaeanum*, and *S. campylacanthum* Hochst. ex A. Rich.; (ii) the “Anguivi” grade, comprising *S. aethiopicum* and its wild progenitor *S. anguivi*, *S. macrocarpon*, along with *S. humile* Lam.

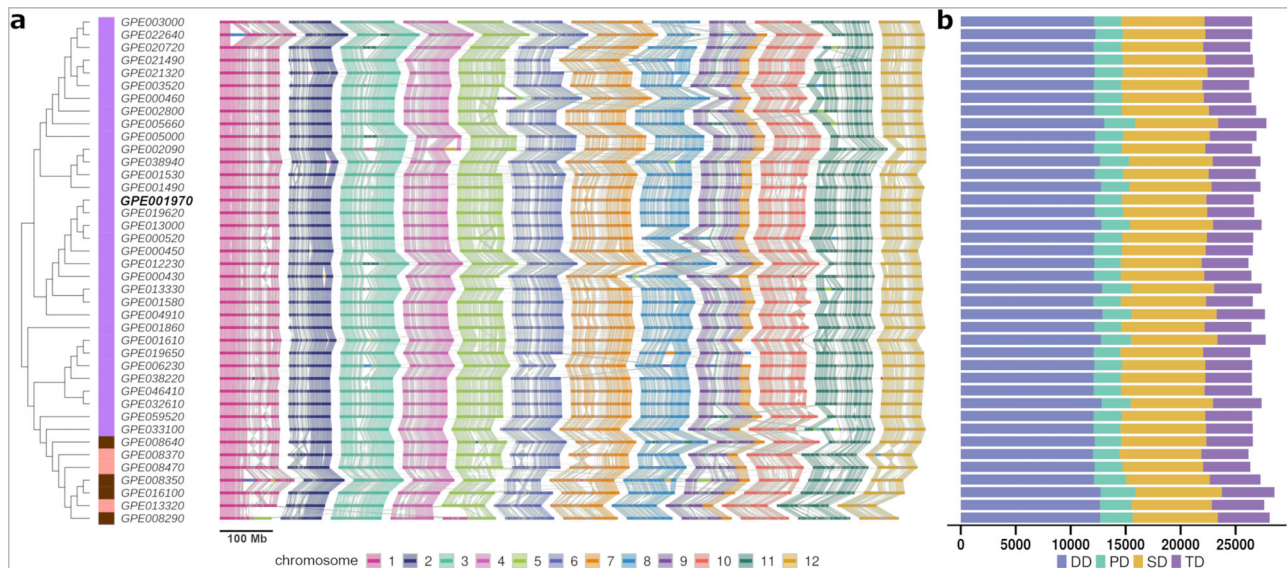


Fig. 3 | Synteny and gene duplications across the 40 eggplant reference genomes. a Synteny map of the 40 chromosome-scale eggplant accessions. On the left is shown the species tree of the 40 eggplant accessions (*S. melongena* in purple, *S. insanum* in pink, and *S. incanum* in brown) constructed based on low-copy (1 to 5 genes) orthogroups. On the right is reported the synteny map, arranged according to the species tree, based on genome similarities across the eggplant accession

panel. Synteny blocks are represented by color-coded bands that connect various genomes, with each color corresponding to a distinct chromosome according to the first accession reported at the top of the species tree. The scale bar represents 100 Mb. **b** Counts of duplicated genes across the 40 eggplant accessions, classified into dispersed duplications (DD), proximal (PD), tandem (TD), and segmental (SD). Bar order corresponds to the species tree shown in (a).

and *S. tomentosum* L.; and (iii) a cluster consisting of East African (*S. tectense* Klotzsch and *S. schimperianum* Hochst. ex A. Rich), Madagascar (*S. myoxotrichum* Baker, *S. pyracanthos* Lam.), and “Lasiocarpa” clade species (*S. sisymbriifolium* and *S. torvum*). These results were consistent with previous reports^{8,16,23,24}, and confirmed coherent phylogeny obtained using different genetic markers (SNPs identified with SPET and GBS methods) for *S. melongena* and its wild relatives.

In a PCA based on SNPs (Fig. 5b, d), the first two components explained 28.66% and 6.05% of the total genetic variance, respectively. *Solanum melongena* accessions formed a compact cluster, while its wild progenitor, *S. insanum*, showed a broader dispersion with several accessions clustering near or within *S. melongena*, reflecting their genetic affinity. *Solanum incanum* partially overlapped with other wild species, such as *S. linnaeanum* and *S. campylacanthum*, while accessions of the “Anguivi” grade clustered closely together. Tertiary genepool species like *S. torvum* and *S. sisymbriifolium* were distinctly divergent. A SV-based PCA (Fig. 5c, e) showed a different structure, despite a lower explained variance (EVI: 8.89%, EV2: 6.35%). *Solanum melongena* accessions dispersed broadly along both components, whereas the “Eggplant” clade and “Anguivi” grade species clustered distinctly, with limited overlap. *Solanum insanum* occupied an intermediate position, consistent with its progenitor status, with some accessions again mapping near or within *S. melongena*, while *S. incanum* formed a separate cluster. The differences of clustering in the two PCAs is expected, given the different nature (SNPs vs SVs) and numbers (23,600 vs 151 after filtering) of the variants used to produce them. Geographical analysis (Fig. 5d, e) revealed two major clusters corresponding to accessions from Southeast Asia and the Indian subcontinent, both closely related to *S. insanum*, confirming the identification of these regions as primary domestication centers for *S. melongena*⁸.

A model-based clustering analysis using the pruned SNPs from PG-SMA, with membership proportions exceeding 0.70, was used to infer the genetic ancestry of *S. melongena* groups originating from diverse geographical regions and their allied species, setting the optimal number of ancestral populations (K) to 6 (Fig. 5f and Supplementary

Fig. 9a, b). *Solanum incanum* and *S. insanum* exhibited minimal admixture, highlighting their distinct genetic identities (Fig. 5f and Supplementary Fig. 9c). The latter showed an additional genetic substructure based on its geographical origin: *Solanum insanum* from Southeast Asia and India are clearly separated, while those from Africa (Madagascar) are an admixture of the two (Fig. 5f), possibly resulting from past hybridization in Madagascar of *S. insanum* introductions from different Asian provenances²⁵. Southeast Asian *S. insanum* genetic signatures remain detectable in local *S. melongena* germplasm, whereas Indian eggplants largely lack genomic traces of Indian *S. insanum*, potentially reflecting post-domestication selection linked to regional phenotypic preferences¹⁰. These patterns align with PCA results and further support separate domestication events in Southeast Asia and the Indian subcontinent⁸. Chinese and Japanese accessions predominantly reflect Southeast Asian ancestry, while accessions from the Middle East and Africa display signals from both domestication centers (Fig. 5f), also observed in European (EEU, SEU, NCU) and American (NAM, MAC) accessions. In contrast, wild species retained mainly unique, divergent components, highlighting their value for conservation and future eggplant breeding efforts.

Population dynamics and inference of demographic history

To investigate the demographic history and population structure of *S. melongena*, we integrated multiple population genetic approaches using both SNP and SV data. Pairwise fixation index (F_{ST}), a metric for evaluating genetic differentiation between populations and lineage-specific genetic drift, revealed the highest genetic differentiation between *S. melongena* and its wild relative *S. incanum*, followed by *S. insanum* (Fig. 6a and Supplementary Fig. 10a). Among *S. melongena* groups, accessions from Southeast Asia, India, and Japan showed the lowest F_{ST} values relative to *S. insanum*, indicating a close genetic affinity consistent with recent domestication events⁸. Outgroup f_3 -statistics, using *S. insanum* as the outgroup, supported Southeast Asia as the primary domestication center, with this group showing the lowest shared drift with other *S. melongena* populations (Fig. 6b and Supplementary Fig. 10b). D -statistics further confirmed

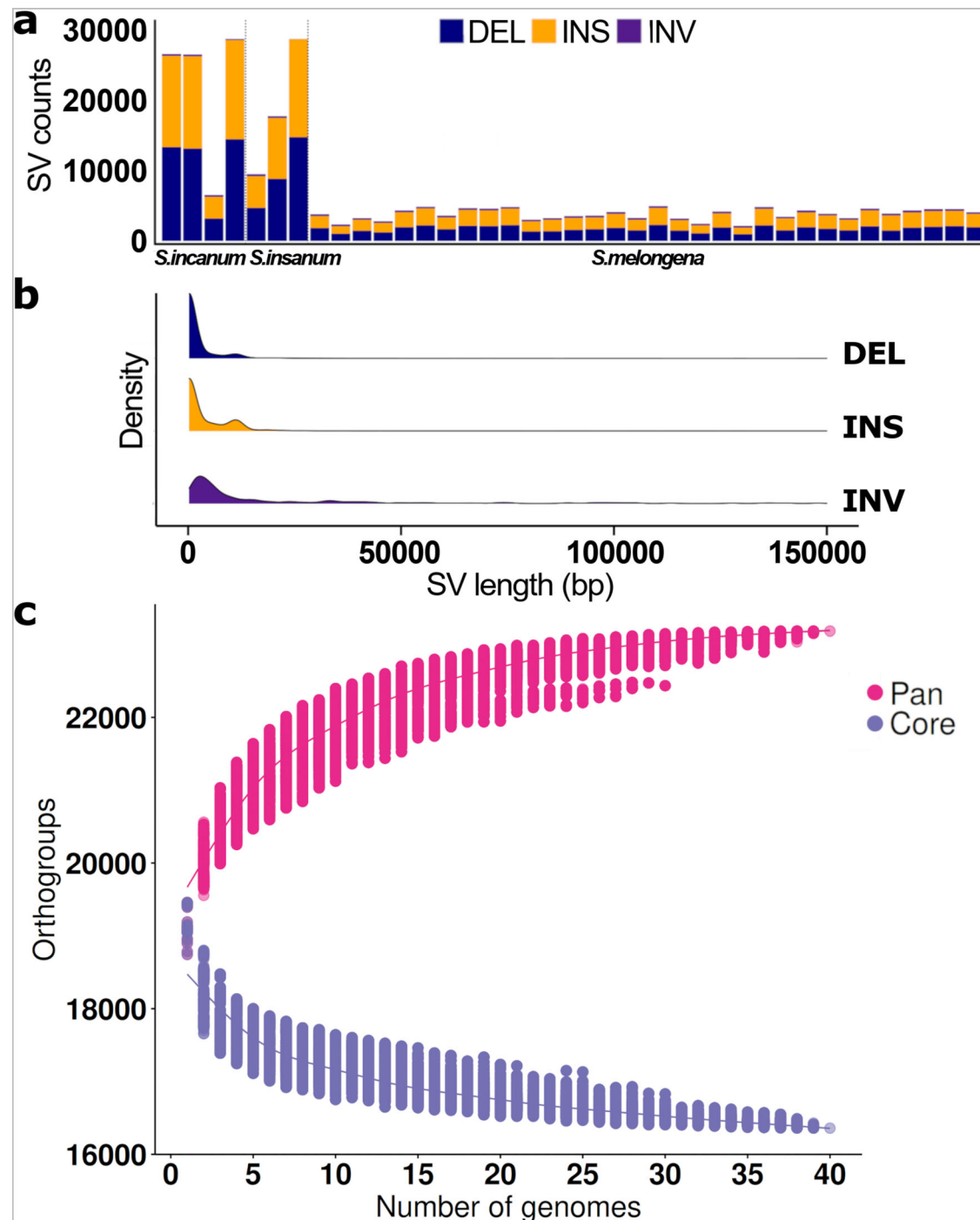


Fig. 4 | Graph-based pangenome of *S. melongena* and its allied species.

a Number of structural variants (SVs) in each accession of the *PG-SMA* graph.

b Length distribution of SVs in *PG-SMA* graph. **c** Accumulation curves showing pan

and core gene families as a function of the number of genomes incorporated into the *PG-SMA* graph. Dots represent permutations of genome addition order; solid lines indicate model fits. Source data are provided as a Source Data file.

asymmetric gene flow from *S. insanum* into Southeast Asian, Indian, and Japanese accessions, while European varieties showed limited introgression (Fig. 6c and Supplementary Fig. 10c).

The domestication and migration history were also explored using TreeMix²⁶ and the OptM package²⁷, with *S. incanum* as outgroup and grouping accessions according to their geographical origin. OptM determined that the most plausible scenario included two migrations (Fig. 6d). The first involved gene flow from *S. insanum* into Southeast Asian *S. melongena*⁹. The second suggested back-introgression from Japanese/Korean accessions into *S. insanum*, indicating bidirectional

gene flow across generations. Residual matrices (Supplementary Fig. 11) suggested an additional admixture event between Chinese and European accessions. Most observations confirm our previous study conducted on a much larger set of accessions using SPET genotyping, with the exception of the back-introgression of Japanese/Korean accessions into *S. insanum*, which was discovered due to the high resolving power of whole genome resequencing, and is consistent with the *S. melongena*-*S. insanum* admixture observed in other geographical areas.

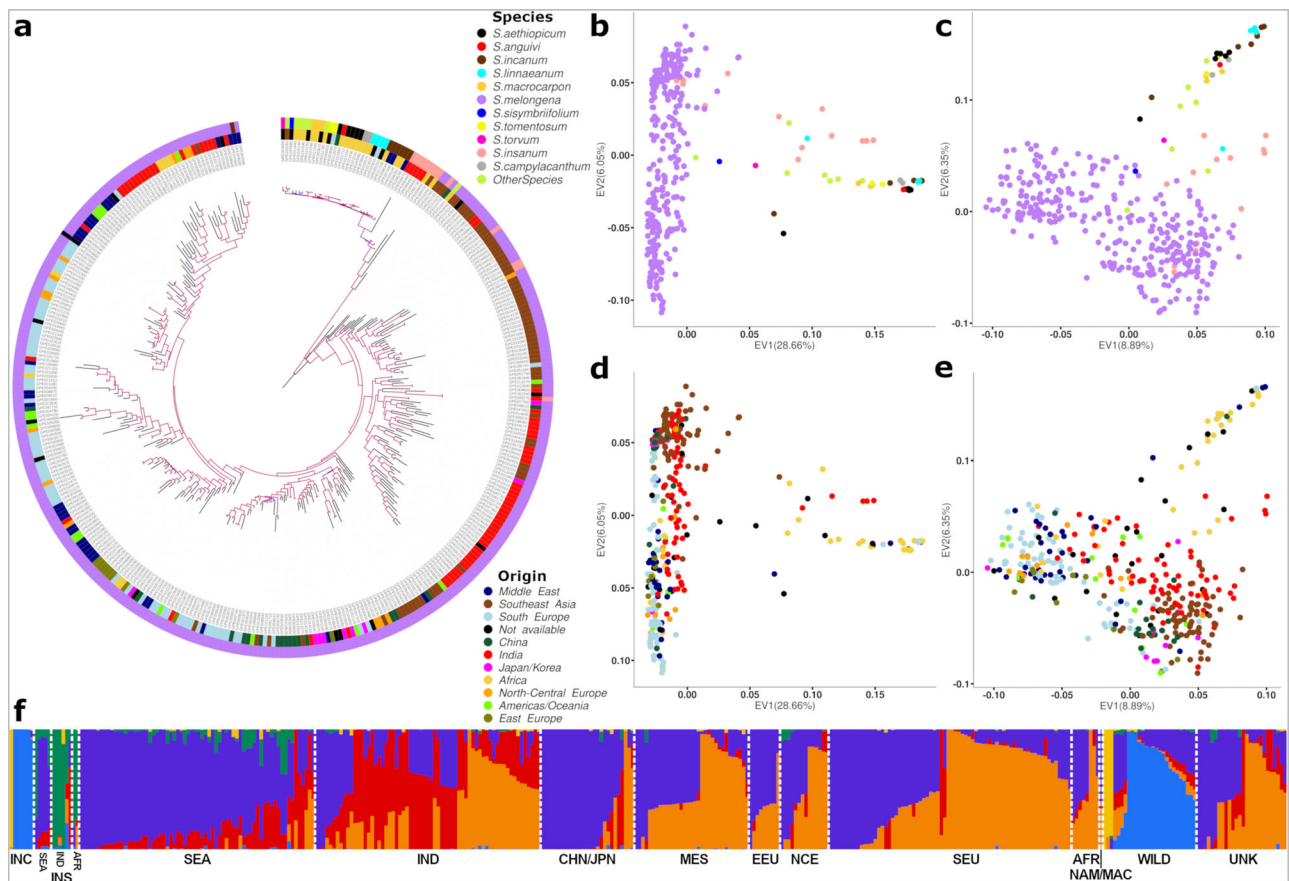


Fig. 5 | Phylogeny and population structure of *S. melongena* and wild relatives. **a** Maximum likelihood (ML) tree using the SNP dataset from *PG-SMA* graph. The colors in the outer circle correspond to species, in the inner circle to geographical origin. Branch support is indicated by bootstrap values, ranging from electric blue (43) to magenta (100). **b, d** Principal component analysis (PCA) using SNPs colored according to species (**b**) and origin (**d**). **c, e** PCA obtained using SVs colored according to species (**c**) and origin (**e**). **f** snmf clustering analysis using *PG-SMA*

graph SNPs at $K = 6$. Accessions are grouped according to geographical origin (INC *S. incanum*; INS *S. insanum* from: SEA Southeast Asia, IND India, AFR Africa; *S. melongena* from: MES Middle East, SEA Southeast Asia, SEU South Europe, UNK Not available, CHN/JPN China and Japan, IND India, AFR Africa, NCE North and Central Europe, NAM Americas and Oceania, EEU East Europe). White dashed lines separate the groups according to geographical origin. The full snmf analyses, with $K = 2-20$, are shown in Supplementary Fig. 7.

Demographic modeling suggests that a reduction in effective population size (N_e) for both *S. insanum* and *S. incanum* already occurred ~40–45 thousands years ago (kya), during a glacial maximum, while a more recent bottleneck (~15 kya) was observed in *S. melongena* accessions from Southeast Asia and India (Fig. 6e), in agreement with previous studies^{8,25} and compatible with the timing suggested for the domestication of eggplant.

A pan-phenome of *S. melongena*

The CC was cultivated in a randomized block design in three different locations (Antalya, Türkiye; Montanaso Lombardo, Italy; Valencia, Spain, Supplementary Fig. 12) and phenotyped for 46 agronomic, 10 biotic/abiotic stress, and 162 fruit metabolomic (80 peel and 81 flesh) traits. In the present study, we focused on a selected set of representative traits from the broader phenotyping dataset, including fruit calyx and leaf prickliness as an example of agronomic variation, resistance to *Fusarium* wilt as a key biotic stress trait, and the contents of chlorogenic and isochlorogenic acids in fruit peel and flesh as metabolic features.

To facilitate alignment of the reads of the 321 *S. melongena* accessions, polymorphism identification and Pan-GWA on this species, a second pangenome graph (“*PG-SM*”) was built with Minigraph-Cactus²⁸, using only the 33 *S. melongena* accessions, with the GPE001970 genome sequence as the reference. The metrics of this

pangenome graph are reported in Supplementary Data 10. Both graphs were used to carry out Pan-GWA studies using selected traits (see “Methods”; Supplementary Data 11–16).

Eggplant Pan-GWA analysis clarifies the genetic mechanisms regulating prickliness

Eggplant is unique among Solanaceous crops, showing prickles on leaves, stems, and fruit calyxes, that provide adaptive advantages, including herbivore deterrence and water retention, but also complicate harvesting, storage, and transport of fruits. For this reason, prickles have been selected against in modern varieties, although some regions, close to the domestication centers, still prefer prickly varieties for their better perceived organoleptic qualities²⁹. Using SNP- and short indel-based Pan-GWA analyses, we confirmed the *LONELY GUY (LOG) 3* gene (*SMEL5_06g025740*) on chromosome 6 as the strongest candidate gene associated with prickle formation on both leaves and calyxes (Fig. 7a, Supplementary Figs. 13a and 14a, and Supplementary Data 11 and 12). Recently, a splice-site mutation in *LOG3* (A “prickleless” > G “prickly”) was identified as a key variant causing the prickleless phenotype³⁰. However, in our genome this mutation is located in the second intron at 90,042,124 bp (Fig. 7b). Our Pan-GWA analysis reported an identical SNP (A/G) at a slightly different position 90,042,191 bp, annotated as a splice-site variant at the end of exon 3. Examination of the *PG-SMA* graph showed that in all species examined,

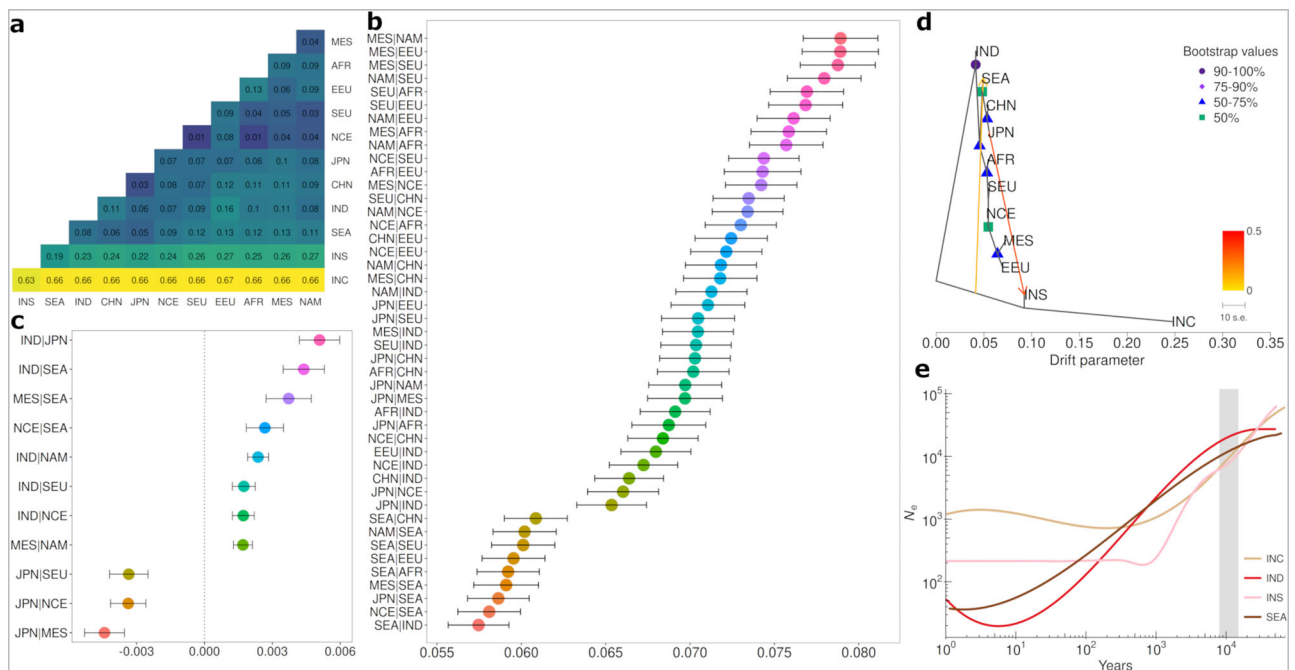


Fig. 6 | Eggplant domestication from *Solanum insanum* based on PG-SMA SNPs.

a F_{ST} values of *S. melongena* from different geographical areas, *S. insanum* and *S. incanum*. **b** Outgroup f_3 -statistic for all possible admixture populations using *S. insanum* as outgroup. The higher the value, the more recently the two populations diverged. The error bars indicate ± 1 standard error (SE). A total of 1,249,951 SNPs were used to compute the f_3 -statistic. **c** D -statistic to detect admixture according to W, X, Y, Z model. World regions of origin are coded as: MES Middle East, SEA Southeast Asia, SEU South Europe, CHN China, IND India, NAM Americas and Oceania, JPN Japan, AFR Africa, NCE North and Central Europe, EEU East Europe.

The error bars indicate ± 1 standard error (SE). A total of 1,249,951 SNPs were used to compute the D -statistic. **d** Maximum likelihood (ML) tree with bootstrap support values with migration events from TreeMix, indicated by colored arrows. The color scale shows the migration weight. The scale bar shows 10 times the average standard error of the estimated entries in the sample covariance matrix. **e** SMC++ plot showing changes in effective population size (N_e) of *S. melongena* accessions grouped according to their origin (SEA: Southeast Asia, IND: India) through relative time. Gray bar delimits the time of N_e decrease (8–15 kya).

the “A” allele is responsible for the prickly phenotype, while prickless accessions harbor the “G” allele (Fig. 7b). Indeed, the splice-site mutation leads to the loss of exon 3 (30 bp), resulting in a 10 amino acid deletion, which however occurs only in a subset of accessions carrying the “G” allele, indicating possible context-specific alternative splicing or other genetic mechanisms, in agreement with previous observations³⁰. Accessions carrying the “G” allele showed strong reduction of *LOG3* gene expression (Supplementary Fig. 15a). We found no evidence for the alternative splicing observed by Satterlee et al.³⁰, possibly due to different sensitivity of the methods used (RNA-Seq vs RT-PCR) or genotype- or tissue-specific differences. Interestingly, heterozygous accessions at this position exhibit enhanced prickliness, suggesting hyperdominance of the prickly “A” allele. We also confirmed that in accession GPE003520 the prickless phenotype is associated to a 226 bp deletion that disrupts exon 6 of *LOG3*, although it carries the “A” allele (Fig. 7b).

The presence of *S. incanum* accessions in the PG-SMA pangenome helped decipher additional mechanisms involving *LOG3* and controlling prickliness. As expected, the *S. incanum* accession (GPE008290) exhibiting a highly prickly phenotype (Fig. 7c) contains a functional *LOG3* gene (Fig. 7d). On the other hand, accession GPE008350 (*S. incanum*) is prickless (Fig. 7c) in spite of the fact that it carries the wild-type “A” allele. Inspection of the pangenome revealed a translocation of the region containing *LOG3* (about 22 Mb) from one arm of chromosome 6 to the heterochromatic, pericentromeric region of chromosome 1 (Fig. 7d), which is probably causing the prickless phenotypes. To validate this translocation, we aligned ONT reads from GPE008350 (carrying the putative rearrangement) to the Hi-C-scaffolded *S. incanum* GPE008290 and GPE016100 genomes. The analysis confirmed the presence of the translocation, observable by

regions of zero coverage coinciding with the translocation breakpoints on chromosomes 1 and 6 (Supplementary Fig. 16).

The molecular basis of *Fusarium* resistance in eggplant

We analyzed the resistance to *Fusarium oxysporum* f. sp. *melongenae* (*Fom*; Fig. 8a, b), a pathogen of significant economic importance in eggplant cultivation. Two strong associations on chromosomes 2 and 11 were identified, co-localizing with previously identified regions, conferring partial resistance to *Fom*^{31,32} (Fig. 8c, d, Supplementary Figs. 13b and 14b, and Supplementary Data 13 and 14). The QTL on chromosome 11 contained a cluster of *RPPI3-like* disease resistance genes, known for their role in plant immune responses^{31–35}. On chromosome 2, a QTL was identified at 44.7 Mb, in correspondence with a 260 bp insertion (Fig. 8c, d, f, Supplementary Figs. 13b and 14b, and Supplementary Data 13 and 14). This genomic region contains an *AT-hook motif nuclear-localized (AHL) protein 10-like* gene (*SMEL5_02g002450*) located at 24.78 kb upstream of the insertion and involved in plant stress resistance. In tomato, overexpression of *CaATL1* (*Capsicum annuum* AHL gene) significantly enhanced resistance to oomycete and bacterial pathogens^{36,37}. Additional minor QTLs were identified in single-environment analyses on chromosomes 6, 7, 9, and 11 (Supplementary Data 13 and 14, and Supplementary Figs. 13b and 14b).

The pangenome graph revealed an insertion in the first exon of *SMEL5_11g023850*, encoding a *RPPI3-like* gene, present exclusively in *Fom*-susceptible accessions and absent in resistant accessions such as the reference line “67/3” (Fig. 8e). An additional deletion of about 31 kb was identified in the region, affecting a gene (*SMEL5_11g023860*) encoding a ginkbilobin-like antifungal protein and a second *disease resistance protein RPPI3-like* gene (*SMEL5_11g023870*) (Fig. 8e). This

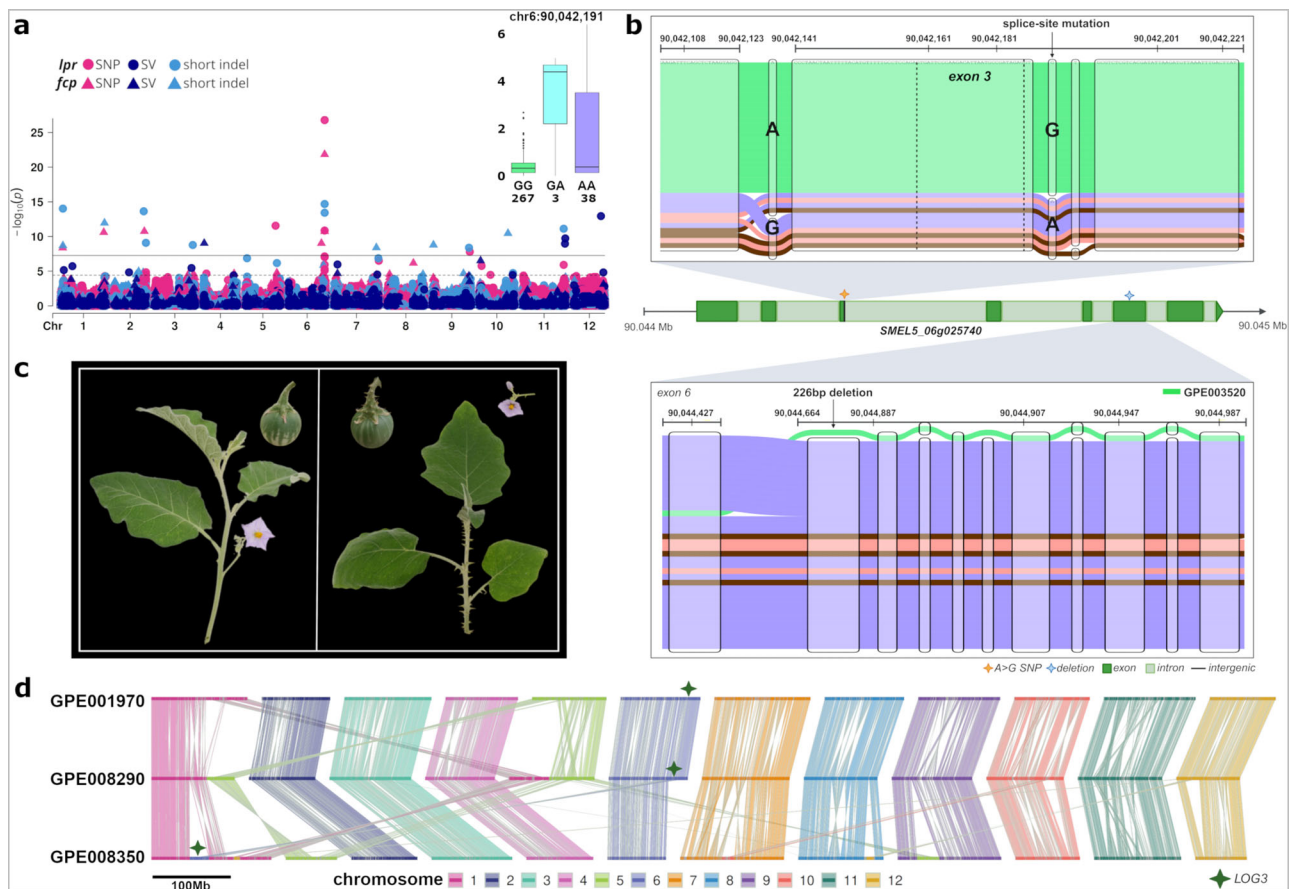


Fig. 7 | Mutations influencing leaf and calyx prickliness in eggplant. **a** Multi-environment Manhattan plot from the *PG-SM*-based Pan-GWA analyses for leaf and calyx prickliness, together with the corresponding box plot of the number of accessions carrying distinct alleles for the splice-site mutation SNP in *LOG3* (GG, accessions with homozygous reference type of allele; AA, accessions possessing homozygous alternative allele; GA, accessions possessing both alleles). In Manhattan plots, the colors refer to SNP (magenta), SV (dark blue), short indel (light blue), while the shape refers to leaf prickles (*lpr*: solid dot) and fruit calyx prickles (*fcp*: triangle). Genome-wide thresholds are marked using dashed (minimum) and continuous (maximum) lines, and estimated based on *q*-values. For box plots, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Data beyond the end of the whiskers are displayed as black dots. **b** Sequence Tube Map representation of the

paths embedded in the *PG-SMA* graph highlighting the splice-site mutation SNP in *SMEL5_06g025740* (*LOG3*) in and around exon 3 (top), as well as the GPE003520 deletion in exon 6 (bottom). *Solanum incanum* and *S. insanum* accessions are represented as brown and pink paths, respectively, while *S. melongena* accessions are colored based on the allele at the locus (purple—prickly; green—prickleless). Gene structure created in BioRender. Gaccione, L. (2025) <https://BioRender.com/pf32ps3>. **c** Phenotypes of the prickless GPE008350 (left) and prickly GPE008290 (right) *S. incanum* accessions. **d** Synteny map of the translocations of *LOG3* in *S. incanum* (GPE008350) with respect to *S. melongena* (GPE001970) and *S. incanum* GPE008290. Synteny blocks are represented by color-coded bands that connect various genomes, with each color corresponding to a distinct chromosome. The scale bar represents 100 Mb.

deletion showed a peculiar pattern, since it contains a relic fragment of about 2400 bp corresponding to a portion of the second exon of *SMEL5_11g023870* (Supplementary Fig. 15b, c). Transcriptomic evidence highlighted altered expression of *SMEL5_11g023870* in the accessions carrying the 31 kb deletion (Supplementary Fig. 15b). The presence of this deletion is strongly (but not exclusively) associated with the susceptible state in the three species included in the graph. This was also confirmed by the PanSel analysis, which identified a divergent bin subjected to selective pressure (i.e., selective sweep) at this genomic position (Supplementary Data 17). Further analysis of the *PG-SMA* graph revealed that the deletion on chromosome 11 is present in 5 out of the 7 *S. insanum* and *S. incanum* accessions used for graph construction, suggesting that this polymorphism is ancient (Fig. 8f). The chromosome 2 insertion is consistently present in our *PG-SMA* *S. incanum* and *S. insanum* accessions, with the exclusion of GPE008370 and GPE008470 (Fig. 8f). The former lacked both QTLs and was thus susceptible to the pathogen, while the latter, although lacking the insertion in chromosome 2, showed the presence of the *RPPL3-like* gene on chromosome 11 and was fully resistant to *Fusarium* (Fig. 8e, f).

The molecular basis of chlorogenic acid metabolism in eggplant Chlorogenic acid (caffeoyl-quinic acid; CQA) and its derivatives isochlorogenic acid A and B (3,5- and 4,5-DiCQA, respectively), are abundant phenolic compounds present in eggplant fruits, responsible both for their antioxidant properties and for their browning through oxidation by polyphenol oxidases. CQA is the major compound, accounting for more than 90% of these compounds³⁸. Pan-GWA analyses for CQA highlighted the presence of eight weak QTLs, of which the strongest, on chr 6, (*p*-value 2.34×10^{-11} ; PVE 45%; Supplementary Data 15 and 16), did not contain any evident candidate genes. We acknowledge that environmental conditions (temperature, light exposure, and water availability) in the two cultivation locations may have influenced the biosynthesis of chlorogenic acids, as previously reported^{39,40}. In contrast, we identified a major QTL for isochlorogenic acids content in both fruit peel and flesh on chromosome 4 (4.88–5.55 Mb), with highly significant marker-trait associations (*p*-values from 1.2×10^{-19} to 1.15×10^{-76} ; PVE up to 74%; Fig. 9a, b, Supplementary Figs. 13c, d, and 14c, d, Supplementary Data 15 and 16). Within this region, three *GDSL lipase-like* genes were

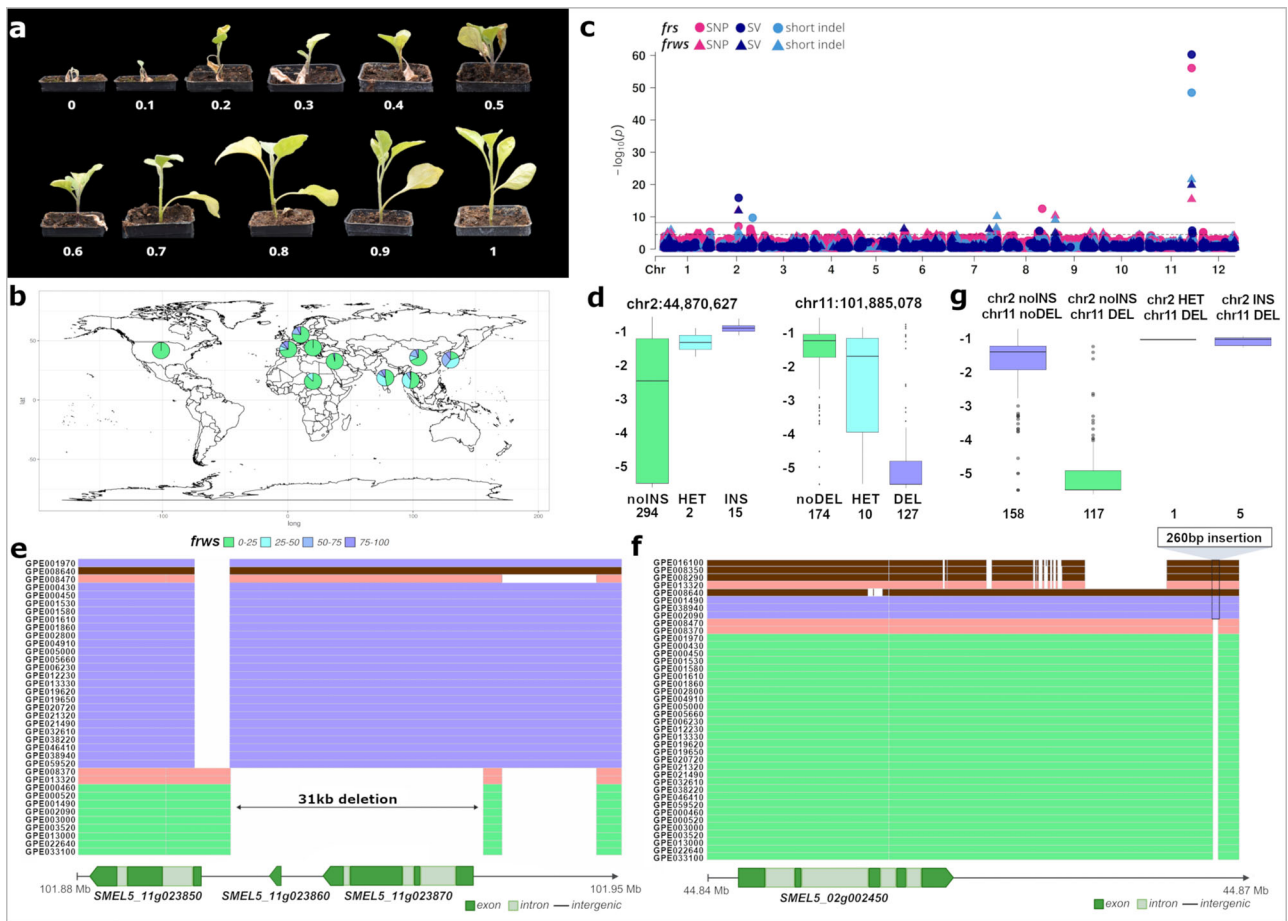


Fig. 8 | Deciphering *Fusarium* wilt resistance in eggplant. **a** Visual representation of the degree of symptoms, ranging from 0 to 1. **b** Distribution of *Fusarium* wilt resistance (*frws*) according to the geographic origin of the eggplant accessions in the core collection. Color scale ranges from 0 (fully sensitive) to 100 (fully resistant). **c, d** Multi-environment Manhattan plot from the *PG-SM*-based Pan-GWA analyses for *Fusarium* wilt (**c**) together with the corresponding box plots (**d**) of the number of accessions carrying or not the insertion (INS) on chromosome 2 and deletion (DEL) on chromosome 11 (HET, accessions possessing both alleles). In Manhattan plots, the colors refer to SNP (magenta), SV (dark blue), short indel (light blue), while the shape refers to *frs* (solid dot) and *frws* (triangle). Genome-wide thresholds are marked using dashed (minimum) and continuous (maximum) lines, and estimated based on *q*-values. For box plots, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Data beyond the end of the whiskers are displayed as black dots. For *frs* and *frws* parameters, see “Methods”. **e** 1-D representation of the

paths embedded in the *PG-SMA* graph within the *RPPI3-like* gene cluster on chromosome 11, highlighting the deletion in correspondence of *SMEL5_11g023860* and *SMEL5_11g023870*. *Solanum incanum* and *S. insanum* accessions are represented as brown and pink paths, respectively, while *S. melongena* accessions are colored based on their resistance status (purple—resistant; green—susceptible). Genes structure created in BioRender. Gaccione, L. (2025) <https://BioRender.com/pf32ps3>. **f** 1-D representation of the paths embedded in *PG-SMA* graph within the genomic region, including *SMEL5_02g002450* on chromosome 2 highlighting the insertion. Gene structure created in BioRender. Gaccione, L. (2025) <https://BioRender.com/pf32ps3>. **g** Phenotype distribution of *Fusarium* resistance according to the combination of both major QTLs on chromosome 2 and chromosome 11 and their sample size (see **d** for description). The central mark of box plots indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Data beyond the end of the whiskers are displayed as black dots.

found, including *SMEL5_04g005040*, which harbored a high-impact SNP introducing a premature codon in exon 2, and thus resulting in a non-functional protein (Fig. 9c). Recently, a GDSL lipase-like enzyme from *Ipomoea batatas* was reported to catalyze the conversion of CQA to 3,5-DiCQA⁴¹. The inspection of the *PG-SM* graph revealed two distinct haplotypes among *S. melongena* accessions: those with high isochlorogenic acids (both 3,5- and 4,5-DiCQA) levels consistently carried the same haplotype, while those with low levels carried an alternative one. Consistent with their metabolic profiles, most accessions of wild relatives (*S. insanum*, *S. incanum*), which naturally accumulate high levels of 3,5- and 4,5-DiCQA, predominantly harbored the high-accumulation haplotype in the *PG-SMA* graph; only *S. incanum* GPE008640 displayed lower accumulation and carried the alternate haplotype. Thus, it appears that *SMEL5_04g005040* is the major determinant of isochlorogenic acid content in eggplant and its wild relatives.

Discussion

Despite its global importance, eggplant has historically lacked the genomic resources available for other Solanaceae crops such as tomato, potato, and pepper^{42–45}. Here, we present a comprehensive pangenome resource based on 368 accessions, encompassing both cultivated varieties and wild relatives. This resource offers new insights into eggplant’s domestication, diversity, and the genetic basis of agronomically important traits.

We selected 40 accessions from the CC representing *S. melongena* and its closely related species (*S. insanum*, *S. incanum*) based on geographic distribution and breeding potential (Fig. 1a). High-quality chromosome-level assemblies, including the reference line “67/3” (GPE001970, v5), facilitated the construction of two graph-based pangenomes: a reference-free multiple sequence alignment-based pangenome (*PG-SMA*) and a reference-based pangenome limited to *S. melongena* (*PG-SM*). *PG-SMA* is more complex, capturing more

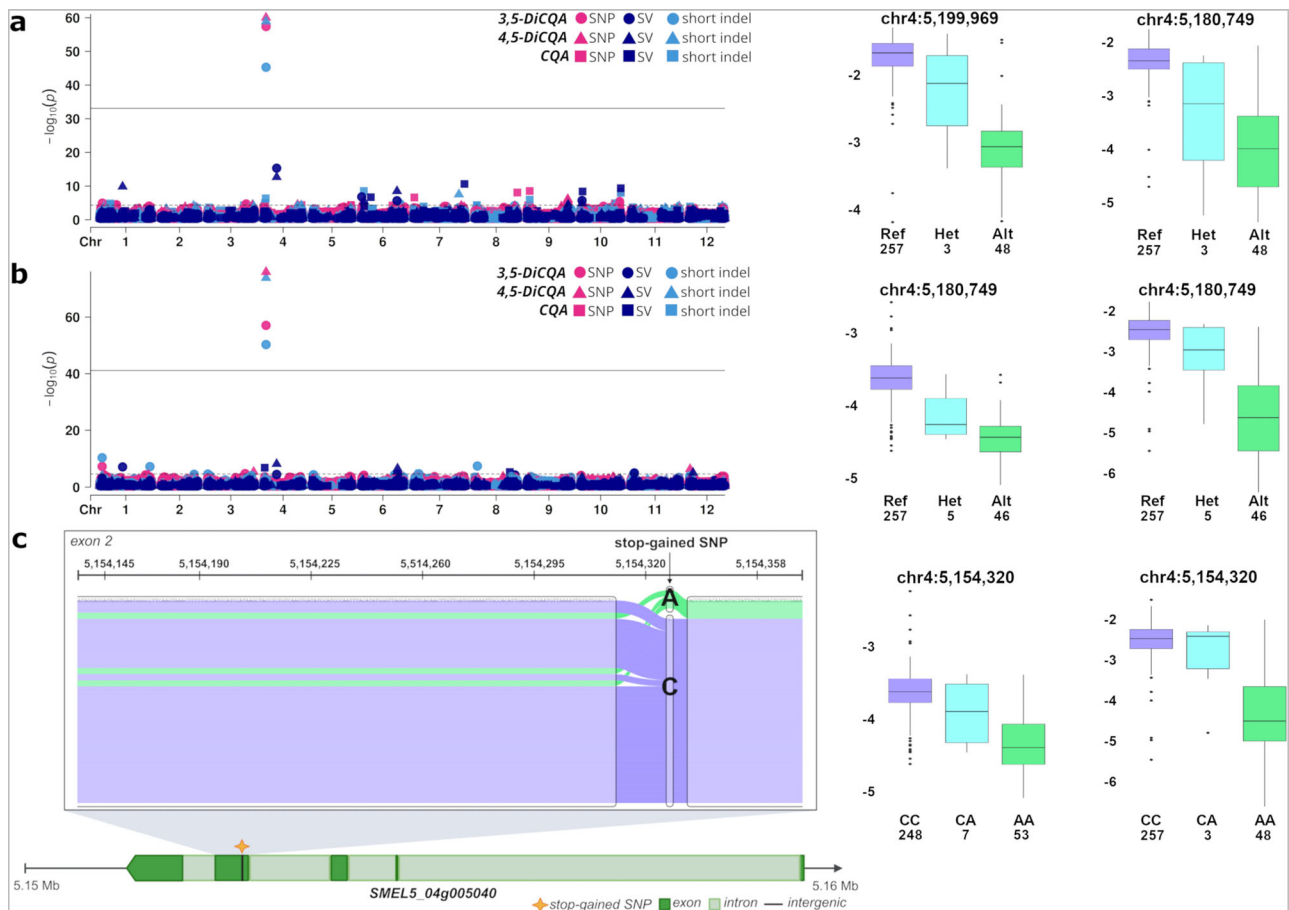


Fig. 9 | Pan-GWA results for chlorogenic acid and its derivatives. **a, b** Multi-environment Manhattan plots from the *PG-SM*-based Pan-GWA analyses for chlorogenic acid (caffeoyl-quinic acid; CQA), 3,5- and 4,5-dicafeoyl-quinic acids (3,5-DiCQA and 4,5-DiCQA) for peel (**a**) and flesh (**b**) tissues, together with the corresponding box plots (left 3,5-DiCQA and right 4,5-DiCQA) in accessions carrying distinct alleles (REF, accessions with homozygous reference type of allele; ALT, accessions possessing homozygous alternative allele; HET, accessions possessing both alleles) and samples size. In Manhattan plots, the colors refer to SNP (magenta), SV (dark blue), short indel (light blue), while the shape refers to 3,5-DiCQA (solid dot), 4,5-DiCQA (triangle), and CQA (square). Genome-wide thresholds for GWAS are marked using dashed (minimum) and continuous (maximum) lines, and estimated based on q -values. For box plots, the central mark indicates the

median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Data beyond the end of the whiskers are displayed as black dots. **c** Sequence Tube Map representation of the paths embedded in *PG-SM* graph within the second exon of *SMEL5_04g005040* highlighting the stop-gained SNP and the corresponding box plots in accessions carrying distinct alleles (for peel and flesh, respectively), and samples size. For box plots, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Data beyond the end of the whiskers are displayed as black dots. *Solanum melongena* accessions carrying or not the stop-gained variant are represented as green and purple paths, respectively. Gene structure created in BioRender. Gaccione, L. (2025) <https://BioRender.com/pf32ps3>.

extensive structural variation and nucleotide-level diversity. Additionally, it relies on all-versus-all sequence alignments of chromosome-level diploid assemblies, thus overcoming the limitations of reference bias¹³. Compared to traditional SV calling using a linear reference, graph-based SV identification does not suffer from issues of reliability but rather from a reduced sensitivity (ability to genotype all the SVs embedded in the graph) when genotyping with short reads⁴³. Indeed, 15,600 SVs were confirmed by short read alignment to *PG-SMA* out of the 83,147 SVs present in the variation graph (see section “A graph-based pangenome of the eggplant clade”). This is an interesting aspect, that could be explored in future studies focused on comparing different SVs identification and genotyping procedures.

Analysis of *PG-SMA* clarified the domestication history and population structure. PCA and ML tree analyses revealed a good separation between *S. melongena* on one side, and *S. insanum*/*S. incanum* on the other, with different tree topologies suggesting potential hybridization/introgression and/or incomplete lineage sorting events for the latter two species. This ambiguity could be resolved by future studies comprising a larger number of the two species.

Genetic differentiation among cultivated eggplants correlated strongly with geographic origin. Southeast Asian and Indian sub-continent varieties formed distinct genetic clusters, with Southeast Asia exhibiting significant continuity with *S. insanum*, supporting its role as a primary domestication center. Indian accessions showed evidence of an independent domestication event, followed by gene flow to the Middle East, Africa, and Europe. Gene flow from *S. insanum* to Southeast Asian *S. melongena* further supports the hypothesis of dual domestication centers. These findings demonstrate the usefulness of SNPs and SVs from pangenome graphs in phylogenetic studies.

In addition to their applications in studying genetic variation, evolution and domestication of crop species, pangenome graphs find application also in GWA studies, because they capture a larger number of variants than single reference genomes^{43,45,46}. In addition to *PG-SMA*, we constructed a second *S. melongena* pangenome graph (*PG-SM*). This graph, in combination with a *S. melongena* pan-phenome, was leveraged for Pan-GWA studies. Mapping of short reads from the CC on the *PG-SM* graph resulted after filtering (see Pan-GWA studies in “Methods”) in 2.13 M SNPs and 4.5k SVs, vs the 1.2 M SNPs and 366 SVs

obtained after mapping to the GPE001970 reference accession, thus confirming the higher capacity of the graph to capture variants absent from a single linear reference.

Eggplant exhibits unique agronomic traits among Solanaceae, such as fruit bitterness/browning (due to the accumulation of chlorogenic/isochlorogenic acids) and leaf and fruit calyx prickliness (due to its classification in the “spiny Solanum” subgenus). A Pan-GWA investigation of the genetic basis of prickliness in the “Eggplant” clade confirmed *LOG3*, a previously described cytokinin biosynthetic gene within the *prickleless* (*pl*) locus⁴⁷, as the major determinant of prickly development in eggplant. Our findings differ slightly from those of Satterlee et al.³⁰, in that the splice-site mutation abolishing prickly formation is in position 90,042,191 bp, and the deletion in exon 6 of *LOG3* that independently leads to a pricklyless phenotype is 226-bases long. In addition, we uncovered in the wild sister species *S. incanum*, an interchromosomal translocation of *LOG3* associated with a pricklyless phenotype. Taken together, these results highlight the complex genetic architecture of prickliness in eggplant and demonstrate how large structural variations, beyond SNPs, may contribute to trait diversity.

Cultivated eggplant is highly susceptible to various diseases, particularly soil-borne fungal wilt caused by *Fusarium oxysporum* f. sp. *melongenae* (*Fom*)⁴⁸. This pathogen invades the plant through the roots, leading to progressive wilting from the lower to the upper leaves. Resistance traits and genes against *Fusarium* wilt have been identified in *S. melongena*^{31,32,49} as well as in its progenitors and wild relatives^{50,51}. Our analysis identified a major QTL on chromosome 11 within an *RPPI3*-like resistance gene cluster as a key determinant of *Fusarium* wilt resistance. A large 31 kb deletion within this cluster represented the key element, since it resulted in the deletion of an *antifungal protein ginkbilobin-like* gene (*SMEL5_11g023860*) and the *disease resistance protein RPPI3-like* gene (*SMEL5_11g023870*; Fig. 8e). Although many *RPPI3-like* genes were found in this region, our findings suggested that *SMEL5_11g023870* is the key gene that recognizes *Fom* effectors and triggers immune responses. *RPPI3* genes are known to confer resistance to fungi in species as different as *Arabidopsis*, grape, and wheat^{35,52,53} as well as resistance to heat in maize⁵⁴. Thus, it will be interesting to verify if the Chr11 QTL is involved in resistance to multiple stresses. Interestingly, a second QTL on chromosome 2 containing an *AT-hook motif nuclear-localized (AHL) protein 10-like* gene compensated for susceptibility caused by the chromosome 11 deletion in the *RPPI3-like* gene cluster, thus providing an alternative source of adaptation to the pathogen (Fig. 8g). This suggests a multi-locus control of resistance, supporting resistance pyramiding strategies in breeding programs.

Eggplant berries are particularly high in phenolic compounds, predominantly anthocyanins (in the peel), as well as chlorogenic (caffeoyl-quinic) acid and its derivatives in the flesh. These compounds have been shown to offer benefits ranging from anti-inflammatory effects to improvements in type 2 diabetes and cardiovascular health⁵⁵. In our study, chlorogenic acid (CQA) and its derivatives (3,5- and 4,5-DiCQA) were found to vary among accessions, with 3,5-DiCQA and CQA variations being highly significant ($p < 0.001$ in one-way ANOVA). Notably, variation in 3,5- and 4,5-DiCQA production correlated with a high-impact (stop-gained) mutation in a *GDSL-like esterase/lipase* gene (*SMEL5_04g005040*) on chromosome 4, primarily found in domesticated accessions (Fig. 9c). Wild relatives predominantly retained the functional haplotype, indicating that some accessions lost the ability to produce these isomers during domestication, through direct selection or linkage drag with other selected traits. These findings may thus provide potential genetic targets for breeding eggplants enriched in health-promoting phenolics through introgression or gene-editing approaches.

A comparison of the Pan-GWA conducted on the intraspecific *PG-SM* and the interspecific *PG-SMA* graphs revealed QTLs in comparable

positions, although the *PG-SMA* GWA exhibited slightly higher p -values (Supplementary Data 11–16), suggesting that the intraspecific graph performs better, even in the presence of introgressions from wild/allied species. This observation, and the shorter time needed for running GWA analyses on single traits on *PG-SM*, makes this the preferred graph for future studies.

In conclusion, the present study illustrates the power of CCs representative of the worldwide genetic diversity of a crop, and of the construction and use of graph pangenomes and pan-phenomes of such CCs for dissecting the genetic architectures of complex traits, ushering an era of pangenome-assisted breeding for enhanced stress resistance and quality.

Methods

Eggplant worldwide core collection

The eggplant CC was constructed starting from a collection of 3412 worldwide accessions, whose population structure had been characterized by SPET^{8,16}. An Euclidean distance matrix was computed using the function `dist` from the stats R package⁵⁶ and used to generate a CC of 368 accessions (Supplementary Data 1) with the `coreCollection` package (v0.9.5)⁵⁷, using the Accession-to-Nearest-Entry “A-NE” criterion⁵⁸, the “split” adjusted group method, the “randomDescent” algorithm, and default values for the remaining parameters. A total of 88 preselected accessions (10 wild and 78 cultivated) were included in the CC (Supplementary Fig. 1).

Short reads sequencing of the eggplant core collection

Genomic DNA was extracted from young fresh leaves of the CC using the SILEX protocol⁵⁹. DNA integrity was evaluated with agarose electrophoresis, DNA quality through the 260/280 and 260/230 nm ratios from NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, Delaware, USA) and concentration with a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Library preparation from high-quality DNA samples and sequencing were performed at BGI Group. Libraries were prepared using a DNA short-insert protocol for 150 bp paired-end reads and re-sequenced at 20× coverage on the DNBseq BGI platform. In addition, published genome sequencing data from 24 *S. melongena* accessions, one *S. insanum* and one *S. incanum* accession^{9,11,60} were retrieved from the NCBI Sequence Read Archive (SRA) database.

Oxford Nanopore sequencing and genome assemblies

For genome assemblies, a total of 40 accessions, including the reference line “67/3” (GPE001970), were selected based on their origin and passport data from the accessions of the CC. High-molecular-weight (HMW) DNA was extracted using the Macherey Nagel NucleoBond HMW Kit, starting with 1 × g of flash frozen leaf material per accession ground in a Retsch MM400 ball mill with a pre-cooled 50 mL stainless steel insert with a single 25 mm steel ball grinding for 15 s at 30 HZ. DNA extraction was then done following the “Lysis with liquid nitrogen and mortar/pestle” section of the Nucleobond Kit manual, followed by size selection using the Circulomics (now Pacific Biosciences) Short Read Eliminator XL Kit. Library preparation was carried out following ONT guidelines and sequenced on PromethION R9.4.1 flow cells. Base calling was done using Guppy (v6.1.1)⁶¹ and the plant super accurate (SUP) model. Primary assemblies were then generated using Nextdenovo (v2.5.2)⁶² with default parameters and polished using Medaka (v2.0.0)⁶³. Plastidial and mitochondrial sequences identified within the unplaced sequences of each genome assembly were then removed based on their alignment to available eggplant plastidial and mitochondrial genome¹¹, using `minimap2` (v2.26)⁶⁴.

For ten accessions (six *S. melongena*, two *S. insanum*, and two *S. incanum*), Hi-C libraries were constructed following Omni-C protocol and sequenced on BGI-DIPSEQ platform with 150 bp paired-end reads. Cleaned reads were aligned to the draft sequence of each assembly

using the Arima pipeline⁶⁵ and Ychs (v1.2.1)⁶⁶ was used for the scaffolding step to achieve chromosome-scale genome sequences. Finally, the order and orientation errors in each chromosome were manually curated based on Hi-C contact maps using Juicebox⁶⁷. The chromosomes orientation was adjusted based on genome collinearity with the eggplant reference line “67/3” version 4.1¹¹ using ntSynt (v1.0.2)⁶⁸. The remaining 30 assemblies were anchored and oriented to chromosomes by the reference-guided tool RagTag (v.2.1.0)⁶⁹ with the scaffold command and the GPE001970, GPE008470, and GPE016100 genome sequences as backbone to anchor the *S. melongena*, *S. insanum*, and *S. incanum* assemblies, respectively.

Assembly quality assessment

The N50 contig length, BUSCO, QV, and LAI values were calculated independently for each genome. To assess the completeness of each genome BUSCO (v5.5.0)¹⁷ was used with the following parameters: “--lineage solanaceae_odb10 --mode genome”. Merquy (v1.3)⁹ was used to estimate QV scores. The LTR Assembly Index (LAI)¹⁸, assessed using the LTR_retriever pipeline⁷⁰ with default parameters, was measured to obtain standardized quality index values for each eggplant genome. In particular, all LTR-RT candidates were initially obtained using LTRharvest within GenomeTools (v1.6.5)⁷¹ using the following parameters: “-mintsd 4 -maxtsd 6 -motif TGCA -motifmis 1 -similar 85 -vic 10 -seed”. LTR-RT candidates were identified using LTR_FINDER with the following parameters: “-D 15000 -d 1000 -L 7000 -l 100 -p 20 -C -M 0.85”. The LAI scores were calculated using LTR_retriever based on the identified LTR-RTs.

Gene prediction and TE annotation

Extensive de novo TE Annotator (EDTA; v2.1.0)⁷² was employed to de novo detect transposable elements (TEs) in *S. melongena*, *S. insanum*, and *S. incanum*, selecting one representative accession per species: GPE001970 (the reference line), GPE008740, and GPE08290, respectively. Subsequently, libraries generated by EDTA were used to annotate TEs across the 40 genomes via RepeatMasker (v.3.2.9)⁷³.

For structural annotation, an integrated approach using both Helixer (v0.3.3)⁷⁴ and the BRAKER3 (v3.0.8)⁷⁵ pipeline was used. First, Helixer identified gene models using a deep-learning ab-initio strategy with just the genomes as input. For the BRAKER3 pipeline, RNA-seq reads for the three species were retrieved from SRA archive (SRP078398, PRJNA526115) and aligned to the assembled genome sequences using HISAT2⁷⁶ (v.2.2.1) with the “--dta” parameter. BRAKER3 was subsequently applied to train the ab initio prediction model from AUGUSTUS⁷⁷ (v.3.4.0) and collected high-quality RNA-seq hints using the Hidden Markov Model (HMM) from GeneMark-ETP⁷⁸ (v.3.67) with the parameter “--nocleanup --softmasking”. Finally, AGAT (v1.3.3)⁷⁹ was used to combine the annotations from Helixer and BRAKER3, using Helixer gene models as primary annotations. Gene functions were inferred by aligning proteins to the Viridiplantae NCBI Reference Sequence Database using diamond⁸⁰ blastp algorithm. InterProScan⁸¹ (v.5.34-73.0) was then used to identify potential protein domains and Gene Ontology. Protein domain information from Pfam were also extracted by enabling the “-appl Pfam” parameter. For each gene of each accession, the functional description was then assigned based on the best hit. Following the annotation, BUSCO (v5.5.0)¹⁷ completeness scores were assessed.

Phylogeny and synteny analyses

OrthoFinder (v2.5.5)⁸² was used with default parameters for phylogenetic orthology inference on the chromosome scale accessions together with 6 additional plant species (*S. lycopersicum*, *S. tuberosum*, *Capsicum annuum*, *Nicotiana sylvestris*, *Coffea arabica*, *Arabidopsis thaliana*). The protein sequences of the resulting single-copy (69 gene families) and low-copy orthologs, which ranged from one to five gene copies per species for each orthologous group, were aligned using

MAFFT (v7.526) with default parameters⁸³. The multiple protein alignments were then trimmed with trimAl (v1.5)⁸⁴ and the corresponding gene trees for each single-copy and low-copy orthologs groups were built using FastTree (version 2.1.11)⁸⁵. The resulting gene trees were used for the phylogenetic tree construction with ASTRAL-Pro3 (v1.20.3.6)^{86,87}, using *Arabidopsis thaliana* as outgroup. Divergence times were inferred using MCMCTree within PAML (v4.10.7)⁸⁸ and the single-copy species tree. This analysis was carried out selecting the GTR substitution model^{89,90} (model = 7) and the independent rates clock model (clock = 2), where rates follow a log-normal distribution. The numbers of Markov chain iterations were set to burnin = 500,000, sampfreq = 100, and nsample = 2000. To constrain the divergence time, four calibration time-points were chosen from the TimeTree 5 database⁹¹: 111.4–123.9 Ma for *A. thaliana* and *C. arabica*, 72.5–97.6 Ma for *C. arabica* and *N. sylvestris*, 16.1–22.7 Ma for *S. lycopersicum* and *C. annuum*, 6.1–9 Ma for *S. lycopersicum* and *S. tuberosum*. Two independent MCMCTree runs were performed to confirm the accuracy of the divergence time estimation. The evolutionary species tree and their divergence time were visualized using MCMCTreeR (v1.1)⁹². The CAF5 (v1.1)⁹³ tool with default parameters was used to predict expansions and contractions within the gene families, using as input files the species tree with the divergence times and the results from gene family cluster analysis. GO over-representation analysis of the expanded and contracted gene families for each accession of *S. melongena*, *S. insanum*, and *S. incanum* was then carried out using the enrichGO function from the clusterProfiler R package (v3.22)⁹⁴.

The 40 chromosome-level sequences were then used to identify multi-genome synteny blocks using ntSynt (v1.0.2)⁶⁸. Ribbon plots were then obtained using ntSynt-viz (v1.0.0)⁹⁵ and the input assemblies were ordered from top-to-bottom based on the species tree structure derived from gene trees of low-copy orthogroups. The 40 proteomes from the three eggplant species were analyzed to identify the full set of paralogous gene pairs (i.e., paranome) within each genome, and gene duplication events were then determined using doubletrouble (v1.7.1)⁹⁶. The gene duplication events were classified in four categories: segmental, tandem, proximal or dispersed duplications. GO over-representation analysis of each duplication category was carried out using the enrichGO function from the clusterProfiler R package⁹⁴.

Pangenome graph of *S. melongena* and its allied species

The graph-based reference-unbiased pangenome of eggplant and its allied species (*PG-SMA*) was built from the 40 chromosome scale genomes. Initially, chromosome sequences were extracted from each assembly and named according to the PanSN specification (e.g., “GPE001970#0#1” referring to chromosome 1). Following best practices for Pangenome Graph Builder (PGGB, v0.5.4)¹⁴ pipeline, the assemblies were partitioned into 12 communities, each corresponding to a chromosome, based on whole-genome alignments performed using mashmap⁹⁷. A separate pangenome graph was then constructed for each chromosome. This process involved wfmash⁹⁸ (v0.9.1) for aligning the assemblies, seqwish⁹⁹ (v0.7.6) to generate a graph from the alignments, smoothxg¹⁰⁰ (v0.6.5) and gffaffix¹⁰¹ (v0.1.3) for graph refinement, and odgi (v0.8.6)¹⁰² for computing statistics and visualizing the graph. Given the importance of key parameters at each stage and their impact on the resulting pangenome graph, different parameter settings were explored to determine the optimal configuration. To estimate the divergence across assemblies, mash dist (v2.3)¹⁰³ was used and a minimum percentage identity (-p) of 93 was used during the graph construction process. Additionally, different combinations of segment length (-s) and minimum match length (-k) were tested to optimize the graph construction parameters. Specifically, combinations of s = {10,000; 20,000; 30,000} and k = {19, 23, 47} were evaluated for the community associated with chromosome 1. The outcomes were assessed by examining graph statistics, including length, node count, and maximum degree, using the odgi¹⁰² stats

command, in accordance with the PGGB guidelines. The complete overview of the graph statistics for each parameter setting is shown in the Supplementary Data 18. Finally, the parameters “-s 20,000 -k 23 -p 93” were selected for the graph construction using target poa lengths (-l) of 3000, 5000, and 7000 for the first, the second and the third smoothxg pass, respectively (Supplementary Data 19). In the final step, a unified pangenome variation graph was obtained incorporating the connected component for each community. This was achieved using the odgi squeeze command with the “-O” option, which compacts the node ID space for each connected component before squeezing. The process produced a single graph file with the extension “.og,” which can be easily converted to other graph formats, such as GFA and VG, using odgi and vg toolkit (v1.61.0)¹⁰⁴. The odgi stats and degree commands with default parameters were used to produce the final statistics of the *PG-SMA* graph. For postprocessing and optimal visualization, the 2D graph visualization was drawn using odgi layout with default parameters and odgi draw command with parameters “-C -w1000”. Gene annotations from genomes into the *PG-SMA* pangenome were projected to the graph structure using vg annotate as described by Novak et al.^{104,105}. A single combined GAF file representing the annotations as paths through the pangenome was obtained.

Pangenome graph of *S. melongena*

The Minigraph-Cactus pipeline (v2.7.2)²⁸ was employed to construct a *S. melongena* pangenome (*PG-SM*), incorporating the 33 *S. melongena* chromosome-level assemblies reported in Supplementary Data 2. The line “67/3” (GPE001970) was selected as reference, since it is the line used for the previously genome assemblies^{9,11}. Odgi stats and degree commands were used to produce the final statistics of the *PG-SM* graph (Supplementary Data 10). For postprocessing and visualization, the odgi layout and odgi draw command were used as done for the *PG-SMA* visualization. A GAF file containing the genes projections was obtained, as previously outlined for the *PG-SMA* pangenome.

Pangenomes statistics and estimation of pangenomes size and growth

To count and annotate SNPs, insertions and deletions embedded in the *PG-SMA* graph, vclength and vcflength tools from the vcflib (v1.0.10)¹⁰⁶ package were used. In parallel, INVPG-annot¹⁰⁷ was employed to identify and annotate inversions from pangenome graph bubbles.

The pangenome size and growth ratio of the *PG-SMA* was assessed using Panacus (v.0.3)¹⁰⁸, which estimates pangenome openness by applying the binomial formula. Panacus hist and growth commands with parameters “-l1,1,1,1 -q 0,0,1,0.95 -s -a” were used to calculate the cumulative bases based on quorum (minimum fraction of accessions sharing a graph feature after accession are sequentially added to the growth histogram), and proportion of conserved ($\geq 90\%$ of accessions), variable ($< 90\%$ of accessions) and unique sequences (1 accession) in the pangenome graph. Curve fitting of the pangenome was provided with Panacus.

The Graph evaluation toolkit (Gretl, v.0.1.1)¹⁰⁹ was also used to analyze the *PG-SMA* graph for bootstrapping different combinations of accessions of the graph up to the maximum number of accessions. For each iteration, the number of nodes and the sequence length were computed. In parallel, PanGP (v1.0.1)¹¹⁰ was used to model the pangenome based on gene families identified using Orthofinder⁸².

To analyze the relationship among the eggplant genomes, we used odgi similarity command to extract a sparse similarity and distance matrices for paths of the *PG-SMA* graph. The paired Jaccard distances were used for hierarchical clustering to construct the phylogenetic relationship among the eggplant genomes. The latter was then compared to the species tree inferred from single-copy orthogroups by obtaining a co-phylogenetic plot using the R package phytools¹¹¹.

Variants calling using the *PG-SMA* graph

To align short reads to *PG-SMA*, the graph (.og file from PGGB pipeline) was converted to GFA format using the odgi view command. Because the *PG-SMA* graph structure does not contain a reference sequence like *PG-SM*, the reference chromosomes were set using vg gbwt command with the arguments “--ref-paths GPE001970”. The *PG-SMA* graph was then simplified by deconstructing the variations (SNPs, indels, and SVs) from the graph structure based on GPE001970 coordinate system. Small variants and short/large SVs were detected by vg deconstruct from snarls genotyping the input assemblies relative to the selected reference. Vcfbub and vcfwave¹⁰⁶ were then used to keep only top-level and < 1 Mb variants. These variants and the genome sequence of GPE001970 were finally used as inputs to the command construct from vg to build the simplified reference-based *PG-SMA* graph structure. The resulting graph (GFA format) was converted to VG format using vg convert and the vg giraffe indices were finally created from the output with “vg autoindex -w giraffe” command. Subsequently, short reads from the 368 accessions of the CC were aligned to the graph using vg giraffe with default parameters, and the results were output in GAM format. The read support was extracted from the alignment files using vg pack command and the variant genotyping was carried out using the following command for each accession: “vg call --threads number_of_threads N -k accession_on_graph.pack -r graph.snarls -s accession_id -a graph.xg”. The “-a” flag is required to merge all the samples. Variant information of SNPs, short indels, and SVs from each accession was merged into a single VCF file using bcftools (v1.21)¹¹² merge with the parameter “-m all” parameter. SNPs and large SVs (≥ 50 bp) were then split into two separate VCF files and used for the downstream analyses.

Variants calling using the *PG-SM* graph

By default, Minigraph-cactus can output the vg giraffe indices, that were used to align short reads from the 368 accessions of the CC to the *PG-SM* graph using vg giraffe. Variant calling was then carried out as described for *PG-SMA*. Subsequently, SNPs, short indels, and large SVs (≥ 50 bp) were extracted and separated into three distinct VCF files for downstream Pan-GWA analyses.

Genomic diversity and population structure

The genomic diversity and the genetic relationships were analyzed using the polymorphisms derived from the genotyping of the CC based on the *PG-SMA* graph. The polymorphic sites were filtered to remove those with a depth < 10 , a mean depth > 50 , genotype quality < 20 , a missing rate $> 10\%$, while retaining only biallelic variants. A PCA was performed with SNPrelate (v1.20.1)¹¹³ on both the SNPs and SVs after pruning and retaining sites with a minor allele frequency (MAF) less than 0.05 with plink2 (v2.0.0)¹¹⁴, using the parameters “--indep-pairwise 50 10 0.1 --maf 0.05”. The same variant panels were used to generate a dendrogram representation of the population’s structure in the ML framework using IQ-TREE2 (v2.1.3)¹¹⁵. To ascertain the support for each branch, the ultrafast bootstrap method was employed, and the tree layout was rendered using the online tool iTOL (v7)¹¹⁶. The snmf function implemented in LEA R package (v1.4.0)¹¹⁷ was used to infer individual admixture coefficients using the pruned SNPs dataset (“plink2 --indep-pairwise 50 10 0.1”) with the following parameters: number of subpopulations (K) ranging from K = 1 to 20, 20-fold cross-validation (CV). Each K was run with 20 replicates and the outputs were submitted to pophelper (v2.3.1)¹¹⁸. The best K was selected by combining the cross-entropy criterion with the Cattell’s method¹¹⁹ (Supplementary Fig. 9a, b). Individuals were tentatively assigned to one of the K populations if its membership coefficient (qi) in that group was ≥ 0.70 .

Population dynamics and inference of demographic history

ADMIXTOOL 2 (v2.0.4)¹²⁰ was used to calculate pairwise Weir and Cockerham’s weighted F_{ST} ¹²¹ for *S. melongena*, *S. insaanum*, and *S.*

incanum accessions using both the SNPs and the SVs data that were genotyped based on the *PG-SMA* pangenome graph and keeping sites with a MAF > 0.05. Using the *S. insanum* genotypes as an outgroup, the *f3*-statistic was then measured within ADMIXTOOL 2 to investigate the shared drift between two populations according to their geographic origin. The *D*-statistic was also calculated using the *f4* function implemented in ADMIXTOOL 2 to assess the direction of gene flow.

Using the SNPs dataset, TreeMix (v1.13)²⁶ was employed to investigate population splits and admixtures without requiring prior hypotheses on the presence or absence of gene flow. OptM (v0.1.6)²⁷ was employed to estimate the optimal number of migration edges to add to the tree (from 1 to 7) and BITE¹²² was used to carry out 500 bootstrap replicates for the optimal migration events assumed and to visualize the consensus trees with bootstrap values and migration edges.

The effective population size (N_e) for *S. incanum*, *S. incanum*, and *S. melongena* across two centers of domestication—India and South-east Asia—was estimated using SMC++ (v1.15.4)¹²³, which uses a coalescent HMM to leverage information on linkage disequilibrium (LD) and the site frequency spectrum from unphased genomic data. To reduce confounding due to gene flow, samples used in this analysis were restricted to the individuals with low admixture, selecting only those assigned to a specific K population. The SNPs panel of the selected accessions was filtered for a missing rate less than 0.1 and the resulting VCF file was converted to SMC format using the *vcf2smc* command in SMC++ and by treating repetitive sites identified by RepeatMasker (v3.2.9)⁷³ as missing to avoid potential bias. Population size history was estimated using the *estimate* command with the default parameters. A mutation rate of 2.35×10^{-8} per site per generation was used and a generation time was set to 1 year, as reported by Barchi et al.¹¹.

Prickles evaluation

Plants of the 321 *S. melongena* accessions of the CC were evaluated among others for leaves (*lpr*) and calyx (*fcp*) prickles in trials carried out by CREA - Italy (from June to October 2019), UPV in Valencia - Spain (from June to October 2020) and BATEM - Türkiye (from May to October 2019; Supplementary Data 20). For each accession, leaf and calyx prickles were scored on a zero (no prickles) to nine (many strong prickles) scale, from three different blocks.

Fom resistance assay

To assess the resistance to *Fom*, seeds from the CC eggplant accessions were sown in 104 hole plastic trays filled with pasteurized peat and grown in greenhouse during spring season at 25/12 °C day/night temperature at Batı Akdeniz Agricultural Research Institute (BATEM, Antalya, Türkiye) and the Council for Agricultural Research and Economics (CREA, Montanaso Lombardo, Italy). Additional lines were included in the panel for standardized phenotypic assessment of disease severity: a fully susceptible (“Tall/1” – GPE000510), a fully resistant (“305E40” – GPE001980) and a partially resistant (“67/3” – GPE001970) line. In the *Fom* inoculation assay, 12 plantlets for each accession were singularly evaluated, with an average of 8135 plantlets assessed per experiment.

The inoculation with *Fom* was conducted according to the dip-root method as reported by Cappelli et al.¹²⁴. Plantlets, at the 2–3 true leaf stage, were gently removed from the tray and their roots washed under running tap water, then immersed for 15 min in a conidial suspension of *Fusarium oxysporum* f. sp. *melongenae* at a concentration of 1.5×10^6 conidia/mL. For each accession, the 12 plantlets were divided into two blocks and then separately inoculated with the conidial suspension. After dipping, the two blocks of plants were transplanted in 54-hole trays and randomized in two different greenhouses until disease symptoms evaluation. For each line and progeny, 9 plants were

mock-inoculated with water and maintained in greenhouse as negative control.

Symptoms evaluation (Supplementary Data 21) was assessed on each plant at 30 Days After Inoculation (DAI) according to a scale ranging from 0 to 1, where 0 corresponds to 0 to “dead plant” and 1 to “fully resistant plant with complete absence of symptoms” (see Tasosone et al.³² for a detailed description, Fig. 8b). Two different resistance parameters to *F. oxysporum* f.sp. *melongenae* (*frws* and *frs*) were calculated for each block:

The *Fusarium* resistance ratio without score (*frws*), expressed as value ranging 0–100, was calculated as:

$$frws = \frac{\text{N}^\circ \text{ of plants with Score} > 0.5}{\text{Total N}^\circ \text{ of inoculated plants}} \times 100 \quad (1)$$

The *Fusarium* resistance parameter taking into account also resistance scores (*frs*), expressed as scale ranging 0–1, was calculated as:

$$frs = \frac{\sum(\text{Plant} \times \text{Score Assigned})}{\text{Total N}^\circ \text{ of inoculated plants}} \quad (2)$$

Polar extraction and quantification of CQA and its derivatives

Each eggplant accession of the CC was sown in the greenhouse, and plantlets were transplanted in an open field at the four- to five-leaf stage. Fruits of more than 300 accessions were subsequently harvested from two agronomical field trials carried out by CREA - Italy (from June to October 2019), and by UPV in Valencia - Spain (from June to October 2020). For each accession, samples of peel and flesh from at least 3–6 representative fruits at the commercial stage of ripening (~30 days after flowering) were collected, put in 50 mL Falcon tube, quick frozen in liquid nitrogen and stored at –80 °C. Two biological replicates from each environment were sampled: one for all the accessions of the CC, and, for a subset of 30 accessions, a second one, from two different randomized blocks. All samples were then lyophilized, ground and weighed. Metabolites from peel and flesh tissues were analyzed as reported by Sulli et al.¹²⁵ with the following modifications: five milligrams of powder from either flesh or peel were extracted using 1.5 mL of 75% methanol/0.05% v/v trifluoroacetic acid for peel and 75% methanol/0.1% v/v formic acid for flesh, spiked with 0.5 µg/mL formononetin (Sigma-Aldrich) as an internal standard. After vortexing for 30 s, shaking for 30 min at 25 °C and centrifugation (20 min at 20,000 × *g*, 15 °C), 0.5 mL of the supernatant was removed and transferred into PTFE filter vials for LC/MS analysis (Whatman). Five microliters of the filtered extract were injected into the liquid chromatography–heated electrospray ionization–mass spectrometry system (LC-HESI-MS), using a Q-Exactive mass spectrometer (Thermo Fisher Scientific). The ionization was performed using the heated electrospray ionization source, with nitrogen used as a sheath and auxiliary gas, and set to 35 and 10 units, respectively. The capillary temperature was 250 °C, the spray voltage was set to 3.5 kV, the probe heater temperature was 330 °C, and the S-lens RF level was set at 50^{125,126}. The acquisition was performed with a mass range of 110–1600 *m/z* both in positive and negative ion mode. Compounds, including chlorogenic acid (CQA), 3,5-DiCQA and 4,5-DiCQA; Supplementary Data 22), were identified by their Full MS and MS² spectra (as reported in Supplementary Data 23), and quantified relative to the internal standard (Fold-IS) by calculating the major adduct ion peak areas, and normalized to the ion peak area of the internal standard (Fold-Internal standard, Supplementary Data 22), using the Trace-Finder (v5.1) software (Thermo Fisher Scientific).

Pan-GWA studies

Adjusted means for each accession and for each resistance score were estimated using a best linear unbiased prediction (BLUP) model with the “inti” R package (v.0.6.6)¹²⁷ and were used as the phenotypic input for the GWAS. The BLUPs were calculated for single environments (correcting for block and genotype–block interaction effects when the number of measurements was sufficient) and across environments/trials (correcting for block and environment–block interaction effects). The broad-sense heritability was computed using the “inti” R package (Supplementary Data 24).

For the GWA studies, we used SNPs, short indels, and SVs derived from the genotyping of the eggplant CC based on the *PG-SM* pangenome graph.

The whole SNPs set was filtered using bcftools (v1.21)¹¹² to retain bi-allelic SNPs with a MAF > 0.05, a mean read depth > 15, < 10% missing data, and a heterozygosity rate < 20% of the accessions. Accessions with a heterozygosity level > 5% and/or a missing rate > 20% were removed. Comparable filtering criteria were applied to the SVs, selecting variants that were longer than 50 bp. SnpEff (v5.0)¹²⁸ was used to annotate and predict the effects of SNPs and SVs. SNPs and SVs were named according to their assigned chromosomes in the variation graph, followed by the base pair start position relative to the reference path in the *PG-SM* pangenome.

After excluding accessions that exhibited large phenotypic variation within the same accession, as well as those with > 20% missing data or a heterozygosity rate > 20%, we selected 308 accessions for SNP and 311 for short indels and SV-based Pan-GWA studies, which were carried out using 2,133,370 high-quality SNPs, 186,343 short indels and 4480 SVs.

GWA studies on single-environment and multi-environment BLUPs were carried out using the R package GAPIT-3 (v3)¹²⁹ and the Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) model¹³⁰. To reduce the number of false positives caused by population stratification, up to three principal components (PCs) were added as covariate variables to the BLINK model for each trait. The significance threshold for each trait was determined by estimating the FDR with the qvalue (v1.2) R package¹³¹. CMplot (v4.2) was used to produce rectangular Manhattan plots and QQ-plots¹³². The QTL confidence intervals were defined according to the LD decay distances estimated for each chromosome using SNPs from the *PG-SM* pangenome, extending the region around the significant marker in the upstream and downstream directions. Each QTL name was defined as the trait code, the chromosome number, and the QTL position. The gene models of the *S. melongena* genome of the reference line “67/3” (GPE001970) was then used to retrieve annotations for genes within the most promising QTL regions. Genes found within each QTL interval were investigated in literature to identify putative causal genes. To visualize specific regions of the pangenome graphs corresponding to variants involved in the resistance to fungal wilt, Sequence Tube Map (v0.8.1)¹³³ and odgi viz command were used.

Conserved and divergent regions in the *PG-SM* pangenome were identified using PanSel¹³⁴. A fixed sliding window of 100 kb was chosen to estimate the average number of edit distance between each haplotype per bin by computing the number of paths that are found in each window between two anchor nodes. Bins with low numbers of paths were designated as under selection genomic regions. The divergent regions in the *PG-SM* graph were then intersected with the whole set of QTLs identified by the Pan-GWA analyses using bedtools intersect (v2.31)¹³⁵.

Transcriptomic data for GPE001970 and GPE001490 were used to confirm the presence and potential transcriptional impact of mutations identified through Pan-GWA analyses. Total RNA was extracted from leaves of the selected accessions using the Qiagen RNeasy Plant Kit and sequenced on the DNBseq BGI platform at BGI

group. Raw RNA-seq reads were trimmed using fastp (v0.24.0)¹³⁶ and aligned to the eggplant reference line (GPE001970) genome sequence using HISAT2 (v2.2.1)¹³⁷. The visual inspection of loci harboring candidate mutations was carried out using the Integrative Genomics Viewer (IGV; v2.18.4)^{138,139}.

To validate the presence of rearrangements identified through both graphs and the Pan-GWA analyses, the ONT reads of the *S. incanum* accessions GPE008350, GPE008290, and GPE016100 were aligned against the Hi-C-guided genome sequences of *S. incanum* (GPE008290 and GPE016100) and the GPE008530 genome sequence using minimap2⁶⁴. Loci harboring candidate variants were visually inspected using the Integrative Genomics Viewer (IGV; v2.18.4)^{138,139}.

Reporting summary

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

Data availability

The sequencing raw data generated in this study have been deposited in the NCBI database under BioProject accession PRJNA1276259 and in the ENA database under BioProject accession PRJEB89803. The genome assemblies generated in this study are also available in the ENA database under BioProject accession PRJEB89803. Additionally, the genome and pangenome assemblies and annotations are available at Sol Genomics Network (<https://solgenomics.net/>). Source data are provided with this paper.

References

1. FAOSTAT 2023. <https://www.fao.org/faostat/en/#home> (accessed 30 March 2023).
2. Spooner, D. M., McLean, K., Ramsay, G., Waugh, R. & Bryan, G. J. A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proc. Natl. Acad. Sci. USA* **102**, 14694–14699 (2005).
3. Portis, E., Nervo, G., Cavallanti, F., Barchi, L. & Lanteri, S. Multivariate analysis of genetic relationships between Italian pepper landraces. *Crop Sci.* **46**, 2517–2525 (2006).
4. Bai, Y. & Lindhout, P. Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? *Ann. Bot.* **100**, 1085–1094 (2007).
5. Razifard, H. et al. Genomic evidence for complex domestication history of the cultivated tomato in Latin America. *Mol. Biol. Evol.* **37**, 1118–1132 (2020).
6. Tripodi, P. et al. Global range expansion history of pepper (*Capsicum* spp.) revealed by over 10,000 genebank accessions. *Proc. Natl. Acad. Sci. USA* **118**, e2104315118 (2021).
7. Vorontsova, M. S., Stern, S., Bohs, L. & Knapp, S. African spiny *Solanum* (subgenus *Leptostemonum*, Solanaceae): a thorny phylogenetic tangle. *Bot. J. Linn. Soc.* **173**, 176–193 (2013).
8. Barchi, L. et al. Analysis of >3400 worldwide eggplant accessions reveals two independent domestication events and multiple migration-diversification routes. *Plant J.* **116**, 1667–1680 (2023).
9. Barchi, L. et al. A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. *Sci. Rep.* **9**, 11769 (2019).
10. Omondi, E. et al. Association analyses reveal both anthropic and environmental selective events during eggplant domestication. *Plant J.* **121**, e17229 (2025).
11. Barchi, L. et al. Improved genome assembly and pan-genome provide key insights into eggplant domestication and breeding. *Plant J.* **107**, 579–596 (2021).
12. Golicz, A. A., Bayer, P. E., Bhalla, P. L., Batley, J. & Edwards, D. Pangenomics comes of age: from bacteria to plant and animal applications. *Trends Genet.* **36**, 132–145 (2020).

13. He, W., Li, X., Qian, Q. & Shang, L. The developments and prospects of plant super-pangenomes: Demands, approaches, and applications. *Plant Commun.* **6**, 101230 (2025).
14. Garrison, E. et al. Building pangenome graphs. *Nat. Methods* **21**, 2008–2012 (2024).
15. Gu, R. et al. Developments on core collections of plant genetic resources: do we know enough? *Forests* **14**, 926 (2023).
16. Barchi, L. et al. Single primer enrichment technology (SPET) for high-throughput genotyping in tomato and eggplant germplasm. *Front. Plant Sci.* **10**, 1005 (2019).
17. Simão, F. A., Waterhouse, R. M., Panagiotis, I., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
18. Ou, S., Chen, J. & Jiang, N. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* **46**, e126 (2018).
19. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
20. Zhou, Y. et al. Importance of incomplete lineage sorting and introgression in the origin of shared genetic variation between two closely related pines with overlapping distributions. *Heredity* **118**, 211–220 (2017).
21. Herbert, T. D. The mid-pleistocene climate transition. *Annu. Rev. Earth Planet. Sci.* **51**, 389–418 (2023).
22. Page, A., Gibson, J., Meyer, R. S. & Chapman, M. A. Eggplant domestication: pervasive gene flow, feralization, and transcriptomic divergence. *Mol. Biol. Evol.* **36**, 1359–1372 (2019).
23. Acquadro, A. et al. Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes. *PLoS ONE* **12**, e0180774 (2017).
24. Gramazio, P. et al. Development and genetic characterization of advanced backcross materials and an introgression line population of *Solanum incanum* in a *S. melongena* background. *Front. Plant Sci.* **8**, 1477 (2017).
25. Arnoux, S., Fraïsse, C. & Sauvage, C. Genomic inference of complex domestication histories in three Solanaceae species. *J. Evol. Biol.* **34**, 270–283 (2021).
26. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
27. Fitak, R. R. OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biol. Methods Protoc.* **6**, bpab017 (2021).
28. Hickey, G. et al. Pangenome graph construction from genome alignments with Minigraph-Cactus. *Nat. Biotechnol.* **42**, 663–673 (2024).
29. Ke, C. et al. Map-based cloning of LPD, a major gene positively regulates leaf prickles development in eggplant. *Theor. Appl. Genet.* **137**, 216 (2024).
30. Satterlee, J. W. et al. Convergent evolution of plant prickles by repeated gene co-option over deep time. *Science* **385**, eado1663 (2024).
31. Barchi, L. et al. QTL analysis reveals new eggplant loci involved in resistance to fungal wilts. *Euphytica* **214**, 20 (2018).
32. Tassone, M. R. et al. A genomic BSAseq approach for the characterization of QTLs underlying resistance to *Fusarium oxysporum* in eggplant. *Cells* **11**, 2548 (2022).
33. Thanyasiriwat, T. et al. Genetic loci associated with *Fusarium* wilt resistance in tomato (*Solanum lycopersicum* L.) discovered by genome-wide association study. *Plant Breed.* **142**, 788–797 (2023).
34. Miyatake, K. et al. Detailed mapping of a resistance locus against *Fusarium* wilt in cultivated eggplant (*Solanum melongena*). *Theor. Appl. Genet.* **129**, 357–367 (2016).
35. Bittner-Eddy, P. D., Crute, I. R., Holub, E. B. & Beynon, J. L. *RPP13* is a simple locus in *Arabidopsis thaliana* for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica*. *Plant J. Cell Mol. Biol.* **21**, 177–188 (2000).
36. Kim, S.-Y. et al. The chili pepper *CaATL1*: an *AT-hook motif-containing transcription factor* implicated in defence responses against pathogens. *Mol. Plant Pathol.* **8**, 761–771 (2007).
37. Martinčová, M. & Soukup, A. Twenty years of *AT-HOOK MOTIF NUCLEAR LOCALIZED (AHL)* gene family research – Their potential in crop improvement. *Curr. Plant Biol.* **42**, 100460 (2025).
38. Whitaker, B. D. & Stommel, J. R. Distribution of hydroxycinnamic acid conjugates in fruit of commercial eggplant (*Solanum melongena* L.) cultivars. *J. Agric. Food Chem.* **51**, 3448–3454 (2003).
39. Sharma, M. & Kaushik, P. Biochemical composition of eggplant fruits: a review. *Appl. Sci.* **11**, 7078 (2021).
40. Stommel, J. R., Whitaker, B. D., Haynes, K. G. & Prohens, J. Genotype × environment interactions in eggplant for fruit phenolic acid content. *Euphytica* **205**, 823–836 (2015).
41. Miguel, S. et al. A *GDSL lipase-like* from *Ipomoea batatas* catalyzes efficient production of 3,5-diCQA when expressed in *Pichia pastoris*. *Commun. Biol.* **3**, 1–13 (2020).
42. Li, N. et al. Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nat. Genet.* **55**, 852–860 (2023).
43. Zhou, Y. et al. Graph pangenome captures missing heritability and empowers tomato breeding. *Nature* **606**, 527–534 (2022).
44. Liu, F. et al. Genomes of cultivated and wild *Capsicum* species provide insights into pepper domestication and population differentiation. *Nat. Commun.* **14**, 5487 (2023).
45. Cheng, L. et al. Leveraging a phased pangenome for haplotype design of hybrid potato. *Nature* 1–10 <https://doi.org/10.1038/s41586-024-08476-9> (2025).
46. Guo, L. et al. Super pangenome of *Vitis* empowers identification of downy mildew resistance genes for grapevine improvement. *Nat. Genet.* **57**, 741–753 (2025).
47. Miyatake, K. et al. Fine mapping of a major locus representing the lack of prickles in eggplant revealed the availability of a 0.5-kb insertion/deletion for marker-assisted selection. *Breed. Sci.* **70**, 438–448 (2020).
48. Stravato, V. M., Cappelli, C. & Polverari, A. Attacks of *Fusarium oxysporum* f. sp. *melongenae*, causing vascular wilt of aubergine in central Italy. *Inf. Fitopatologico* **43**, 51–54 (1993). (in Italian).
49. Monma, S., Akazawa, S., Simosaka, K., Sakata, Y. & Matsunaga, H. ‘Diataro’, a bacterial wilt-and *Fusarium* wilt-resistant hybrid eggplant for rootstock. *Bull. Natl Inst. Veg. Orn. Plants Tea* **12**, 73–83 (1997). in Japanese with English summary.
50. Rizza, F. et al. Androgenic dihaploids from somatic hybrids between *Solanum melongena* and *S. aethiopicum* group gilo as a source of resistance to *Fusarium oxysporum* f. sp. *melongenae*. *Plant Cell Rep.* **20**, 1022–1032 (2002).
51. Toppino, L., Vale, G. & Rotino, G. L. Inheritance of *Fusarium* wilt resistance introgressed from *Solanum aethiopicum* Gilo and *Aculeatum* groups into cultivated eggplant (*S. melongena*) and development of associated PCR-based markers. *Mol. Breed.* **22**, 237–250 (2008).
52. Zhang, X. et al. A truncated *CC-NB-ARC* gene *TaRPP13L1-3D* positively regulates powdery mildew resistance in wheat via the RanGAP-WPP complex-mediated nucleocytoplasmic shuttle. *Planta* **255**, 60 (2022).
53. Chen, Y. et al. Grapevine VaRPP13 protein enhances oomycetes resistance by activating SA signal pathway. *Plant Cell Rep.* **41**, 2341–2350 (2022).
54. Yang, H. et al. A new adenylyl cyclase, putative disease-resistance RPP13-like protein 3, participates in abscisic acid-mediated resistance to heat stress in maize. *J. Exp. Bot.* **72**, 283–301 (2021).

55. Kwon, Y.-I., Apostolidis, E. & Shetty, K. In vitro studies of eggplant (*Solanum melongena*) phenolics as inhibitors of key enzymes relevant for type 2 diabetes and hypertension. *Bioresour. Technol.* **99**, 2981–2988 (2008).
56. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at <https://www.R-project.org/> (2024).
57. Brouwer, M. & de Blok, R. *coreCollection: Creating a Core Collection*. (Wageningen University and Research, Department Plant Breeding, Wageningen, The Netherlands, 2022).
58. Jansen, J. & van Hintum, Th. Genetic distance sampling: a novel sampling method for obtaining core collections using genetic distances with an application to cultivated lettuce. *Theor. Appl. Genet.* **114**, 421–428 (2007).
59. Vilanova, S. et al. SILEX: a fast and inexpensive high-quality DNA extraction method suitable for multiple sequencing platforms and recalcitrant plant species. *Plant Methods* **16**, 110 (2020).
60. Gramazio, P. et al. Whole-genome resequencing of seven eggplant (*Solanum melongena*) and one wild relative (*S. incanum*) accessions provides new insights and breeding tools for eggplant enhancement. *Front. Plant Sci.* **10**, 1220 (2019).
61. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **20**, 129 (2019).
62. Hu, J. et al. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol.* **25**, 107 (2024).
63. Medaka. *Oxford Nanopore Technologies*. <https://github.com/nanoporetech/medaka> (2022).
64. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
65. Mapping pipeline for data generated using Arima-HiC. *Arima Genomics, Inc.* https://github.com/ArimaGenomics/mapping_pipeline (2025).
66. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808 (2023).
67. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
68. Coombe, L., Kazemi, P., Wong, J., Birol, I. & Warren, R. L. Multi-genome synteny detection using minimizer graph mappings. Preprint at <https://doi.org/10.1101/2024.02.07.579356> (2024).
69. Alonge, M. et al. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* **23**, 258 (2022).
70. Ou, S. & Jiang, N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retro-transposons. *Plant Physiol.* **176**, 1410–1422 (2018).
71. Gremme, G., Steinbiss, S. & Kurtz, S. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEEACM Trans. Comput. Biol. Bioinforma.* **10**, 645–656 (2013).
72. Ou, S. et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275 (2019).
73. Tarailo-Graovac, M., & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics*. Chapter 4:4.10.1–4.10.14 (2009).
74. Stiehler, F. et al. Helixer: cross-species gene annotation of large eukaryotic genomes using deep learning. *Bioinformatics* **36**, 5291–5298 (2021).
75. Gabriel, L. et al. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* **34**, 769–777 (2024).
76. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
77. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
78. Brůna, T., Lomsadze, A. & Borodovsky, M. GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *Genome Res.* **34**, 757–768 (2024).
79. Dainat J. AGAT: another Gff analysis toolkit to handle annotations in any GTF/GFF format. (Version v0.7.0). Zenodo. <https://doi.org/10.5281/zenodo.3552717>.
80. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).
81. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinforma. Oxf. Engl.* **30**, 1236–1240 (2014).
82. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
83. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
84. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
85. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
86. Zhang, C., Scornavacca, C., Molloy, E. K. & Mirarab, S. ASTRAL-Pro: quartet-based species-tree inference despite paralogy. *Mol. Biol. Evol.* **37**, 3292–3307 (2020).
87. Zhang, C. & Mirarab, S. ASTRAL-Pro 2: ultrafast species tree reconstruction from multi-copy gene family trees. *Bioinformatics* **38**, 4949–4950 (2022).
88. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
89. Zharkikh, A. Estimation of evolutionary distances between nucleotide sequences. *J. Mol. Evol.* **39**, 315–329 (1994).
90. Yang, Z. Estimating the pattern of nucleotide substitution. *J. Mol. Evol.* **39**, 105–111 (1994).
91. Kumar, S. et al. TimeTree 5: an expanded resource for species divergence times. *Mol. Biol. Evol.* **39**, msac174 (2022).
92. Puttick, M. MCMCtreeR: a package to estimate distribution values for temporal node constraints and prepare input files for MCMCtree. <https://github.com/PuttickMacroevolution/MCMCtreeR>.
93. Han, M. V., Thomas, G. W. C., Lugo-Martinez, J. & Hahn, M. W. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* **30**, 1987–1997 (2013).
94. 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).
95. Coombe, L., Warren, R. L. & Birol, I. ntSynt-viz: visualizing synteny patterns across multiple genomes. Preprint at <https://doi.org/10.1101/2025.01.15.633221> (2025).
96. Almeida-Silva, F. & Van de Peer, Y. doubletrouble: an R/Bio-conductor package for the identification, classification, and analysis of gene and genome duplications. *Bioinformatics* **41**, btaf043 (2025).
97. Kille, B., Garrison, E., Treangen, T. J. & Phillippy, A. M. Minmers are a generalization of minimizers that enable unbiased local Jaccard estimation. *Bioinformatics* **39**, btad512 (2023).
98. Guarracino, A., Mwaniki, N., Marco-Sola, S. & Garrison, E. wfmash: whole-chromosome pairwise alignment using the hierarchical wavefront algorithm. <https://github.com/waveygang/wfmash>
99. Garrison, E. & Guarracino, A. Unbiased pangenome graphs. *Bioinformatics* **39**, btac743 (2023).

100. smoothxg: linearize and simplify variation graphs using blocked partial order alignment. <https://github.com/pangenome/smoothxg>.
101. Doerr, D. & Marijon, P. GFAffix: GFAffix identifies walk-preserving shared affixes in variation graphs and collapses them into a non-redundant graph structure. <https://github.com/codialab/GFAffix>.
102. Guarracino, A., Heumos, S., Nahnsen, S., Prins, P. & Garrison, E. ODGI: understanding pangenome graphs. *Bioinformatics* **38**, 3319–3326 (2022).
103. Ondov, B. D. et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).
104. Garrison, E. et al. Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nat. Biotechnol.* **36**, 875–879 (2018).
105. Novak, A. M. et al. Efficient indexing and querying of annotations in a pangenome graph. Preprint at <https://doi.org/10.1101/2024.10.12.618009> (2024).
106. Garrison, E., Kronenberg, Z. N., Dawson, E. T., Pedersen, B. S. & Prins, P. A spectrum of free software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2, hts-nim and slivar. *PLoS Comput. Biol.* **18**, e1009123 (2022).
107. Romain, S., Dubois, S., Legeai, F. & Lemaitre, C. Investigating the topological motifs of inversions in pangenome graphs. Preprint at <https://doi.org/10.1101/2025.03.14.643331> (2025).
108. Parmigiani, L., Garrison, E., Stoye, J., Marschall, T. & Doerr, D. Panacus: fast and exact pangenome growth and core size estimation. *Bioinformatics* **40**, btac720 (2024).
109. Vorbrugg, S., Bezrukov, I., Bao, Z. & Weigel, D. Gretl - Variation GRaph evaluation ToolKit. *Bioinformatics* **41**, btac755 (2024).
110. Zhao, Y. et al. PanGP: a tool for quickly analyzing bacterial pangenome profile. *Bioinforma. Oxf. Engl.* **30**, 1297–1299 (2014).
111. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
112. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
113. Zheng, X. et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
114. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
115. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
116. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).
117. Fricot, E. & François, O. LEA: an R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**, 925–929 (2015).
118. Francis, R. M. pophelper: an R package and web app to analyse and visualize population structure. *Mol. Ecol. Resour.* **17**, 27–32 (2017).
119. Cattell, R. B. The scree test for the number of factors. *Multivar. Behav. Res.* **1**, 245–276 (1966).
120. Maier, R. et al. On the limits of fitting complex models of population history to f-statistics. *Elife* **12**, e85492 (2023).
121. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population-structure. *Evolution* **38**, 1358–1370 (1984).
122. Milanesi, M. et al. BITE: an R package for biodiversity analyses. 181610 Preprint at <https://doi.org/10.1101/181610> (2017).
123. Terhorst, J., Kamm, J. A. & Song, Y. S. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).
124. Cappelli, C., Stravato, V. M., Rotino, G. L. & Buonaurio, R. Sources of resistance among *Solanum* spp. to an Italian isolate of *Fusarium oxysporum* f. sp. *melongenae*. In *EUCARPIA, Proceedings of the 9th Meeting of Genetics and Breeding of Capsicum and Eggplant*. (eds Andrásfalvi, A., Moòr, A. & Zatykò, L.) 221–224 (SINCOP, Budapest, Hungary, 1995).
125. Sulli, M. et al. An eggplant recombinant inbred population allows the discovery of metabolic QTLs controlling fruit nutritional quality. *Front. Plant Sci.* **12**, 638195 (2021).
126. Ranawaka, B. et al. A multi-omic *Nicotiana benthamiana* resource for fundamental research and biotechnology. *Nat. Plants* **9**, 1558–1571 (2023).
127. Lozano-Isla, F. inti: tools and statistical procedures in plant science. R package version 0.5.6. <https://CRAN.R-project.org/package=inti>.
128. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly. (Austin)* **6**, 80–92 (2012).
129. Wang, J. & Zhang, Z. GAPIT Version 3: boosting power and accuracy for genomic association and prediction. *Genomics Proteom. Bioinforma.* **19**, 629–640 (2021).
130. Huang, M., Liu, X., Zhou, Y., Summers, R. M. & Zhang, Z. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* **8**, giy154 (2019).
131. Storey, J., Bass, A., Dabney, A., & Robinson, D. Package ‘qvalue’. <https://bioconductor.org/packages/qvalue>.
132. Yin, L. et al. rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, Proteom. Bioinforma.* **19**, 619–628 (2021).
133. Beyer, W. et al. Sequence tube maps: making graph genomes intuitive to commuters. *Bioinformatics* **35**, 5318–5320 (2019).
134. Zytnecki, M. Assessing genome conservation on pangenome graphs with PanSel. *Bioinforma. Adv.* **5**, vbafo18 (2025).
135. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
136. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
137. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
138. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
139. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

Acknowledgements

This work has been funded by the European Commission, Horizon 2020 G2P-SOL project (grant no. 677379 to G.G.) and by the Horizon Europe “Promoting a Plant Genetic Resource Community for Europe (PRO-GRACE)” project (grant agreement no. 101094738 to G.G.). The overall work also partially fulfills some goals of the Agritech National Research Center and received funding from the European Union Next-Generation EU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR)—MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4—D.D. 1032 17/06/2022, CN00000022). In particular, this study covers some activities comprised in: Spoke 4 (Task 4.1.1.) “Next-generation genotyping and -omics technologies for the molecular prediction of multiple resilient traits in crop plants”; Spoke 1 (Task 1.2.1 Linking phenotype and genotype: discovery of loci/genes/alleles for traits of interest) and Spoke 2 (Task 2.2.1: “Improved genetic materials to reduce the use of agrochemicals”). The World Vegetable Center also acknowledges the long-term strategic donors, including Taiwan, the United Kingdom, the United States, Australia, Germany, Thailand, South Korea, the Philippines, and Japan. Furthermore, Worldveg acknowledges the funding from NSCT Taiwan: Promoting a Plant Genetic Resource Community (NSTC 112-2923-B-125-

002-MY2 to R.S.). B.U. received funding from the DFG under Germany's Excellence Strategy (CEPLAS Cluster, EXC 2048/1, Project ID: 390686111). We finally thank Manuela Costanzo (ENEA, Casaccia Res Ctr, 00123 Roma, Italy) for help with metabolite extractions.

Author contributions

B.U., L.B., and G.G. conceived the study. E.I., M. Schmidt, S.L., and E.P. produced sequencing data. R.F. and M. Brouwer constructed the eggplant CC. L.G. and M. Bolger performed the genome assemblies and annotations. L.G. and L.B. performed pangenome graphs construction and downstream analyses. L.G., L.B., and G.A. analyzed resequencing data and GWA analyses. L.T., G.L.R., D.A., V.L., R.S., A.B., J.P., and H.F. B. provided plant materials. L.T., M.R.T., G.L.R., and H.F.B. performed fungal tests. L. T., M.R.T., G.L.R., H.F.B., J.P., D.A., and P.F. provided field data. M. Sulli produced and analyzed metabolic data. L.T. and S.G. performed real time analyzes on *Fusarium* candidate genes. L.G., L.B., B.U., and G.G. wrote the manuscript. All authors critically revised and approved the manuscript.

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/s41467-025-64866-1>.

Correspondence and requests for materials should be addressed to Giovanni Giuliano, Björn Usadel or Lorenzo Barchi.

Peer review information *Nature Communications* thanks Kozo Nakamura, Junpeng Shi and the other, anonymous, reviewer(s) 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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2025