NATURE | LETTER OPEN

The tomato genome sequence provides insights into fleshy fruit evolution

The Tomato Genome Consortium

Nature **485**, 635–641 (31 May 2012) doi:10.1038/nature11119
Received 17 August 2011 Accepted 03 April 2012 Published online 30 May 2012

Tomato (*Solanum lycopersicum*) is a major crop plant and a model system for fruit development. *Solanum* is one of the largest angiosperm genera[1] and includes annual and perennial plants from diverse habitats. Here we present a high-quality genome sequence of domesticated tomato, a draft sequence of its closest wild relative, *Solanum pimpinellifolium*[2], and compare them to each other and to the potato genome (*Solanum tuberosum*). The two tomato genomes show only 0.6% nucleotide divergence and signs of recent admixture, but show more than 8% divergence from potato, with nine large and several smaller inversions. In contrast to *Arabidopsis*, but similar to soybean, tomato and potato small RNAs map predominantly to gene-rich chromosomal regions, including gene promoters. The *Solanum* lineage has experienced two consecutive genome triplications: one that is ancient and shared with rosids, and a more recent one. These triplications set the stage for the neofunctionalization of genes controlling fruit characteristics, such as colour and fleshiness.

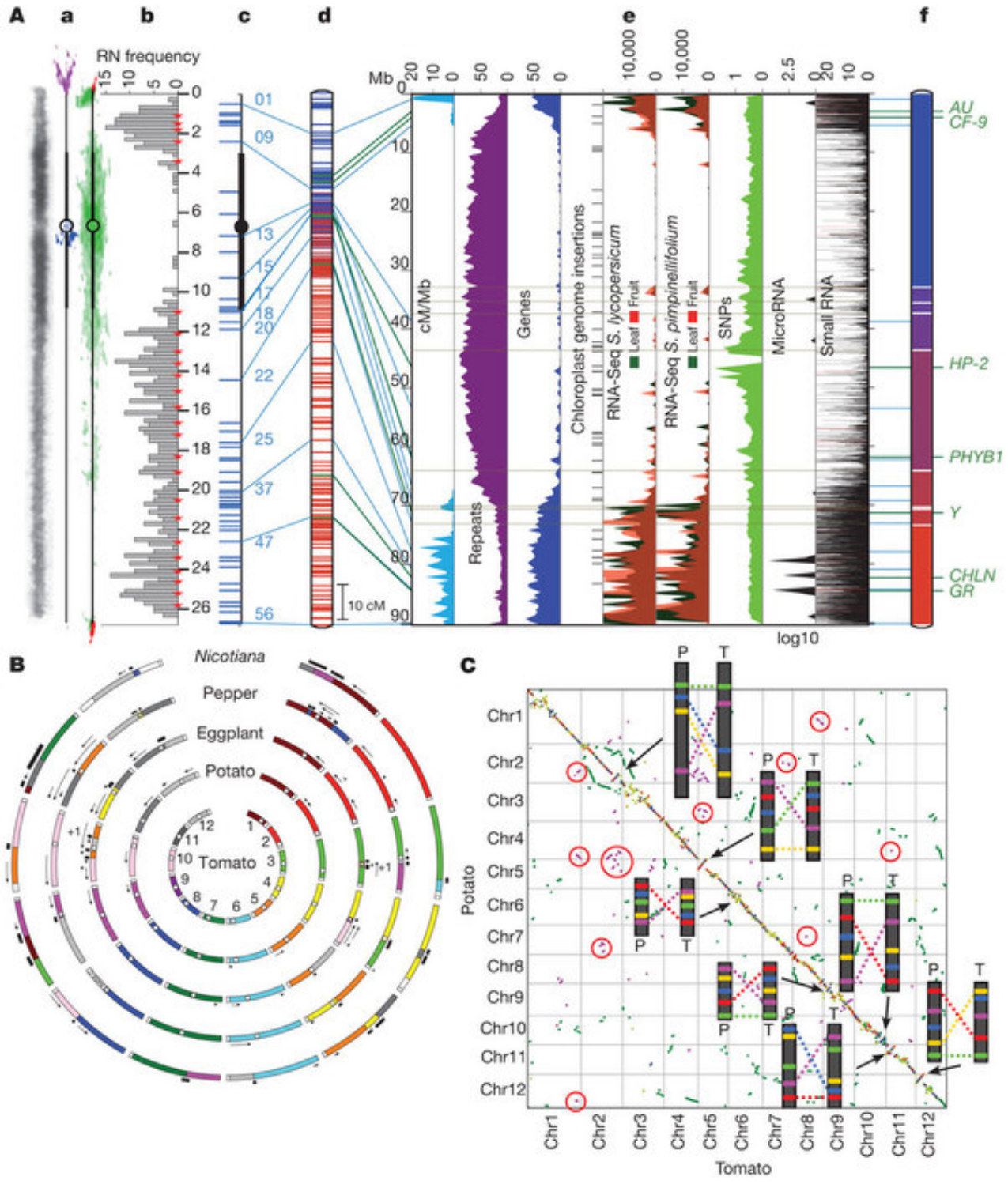
Subject terms: Genomics Plant sciences Evolution Genetics

Main

Main **Methods** **Accession Codes** **References** **Acknowledgements** **Author Information** **Supplementary Information**

The genome of the inbred tomato cultivar ‘Heinz 1706’ was sequenced and assembled using a combination of Sanger and ‘next generation’ technologies (Supplementary Information section 1). The predicted genome size is approximately 900 megabases (Mb), consistent with previous estimates[3], of which 760 Mb were assembled in 91 scaffolds aligned to the 12 tomato chromosomes, with most gaps restricted to pericentromeric regions (Fig. 1A and Supplementary Fig. 1). Base accuracy is approximately one substitution error per 29.4 kilobases (kb) and one indel error per 6.4 kb. The scaffolds were linked with two bacterial artificial chromosome (BAC)-based physical maps and anchored/oriented using a high-density genetic map, introgression line mapping and BAC fluorescence *in situ* hybridization (FISH).

Figure 1: Tomato genome topography and synteny.



A, Multi-dimensional topography of tomato chromosome 1 (chromosomes 2–12 are shown in Supplementary Fig. 1). **a**, Left: contrast-reversed, 4′,6-diamidino-2-phenylindole (DAPI)-stained pachytene chromosome; centre and right: FISH signals for repeat sequences on diagrammatic pachytene chromosomes (purple, TGR1; blue, TGR4; red, telomere repeat; green, Cot 100 DNA (including most repeats)). **b**, Frequency distribution of recombination nodules (RNs) representing crossovers on 249 chromosomes. Red stars mark 5 cM intervals starting from the end of the short arm (top). Scale is in micrometres. **c**, FISH-based locations of selected BACs (horizontal blue lines on left). **d**, Kazusa F2-2000 linkage map. Blue lines to the left connect linkage map markers on the BAC-FISH map (**c**), and to the right to heat maps (**e**) and the DNA pseudomolecule (**f**). **e**, From

left to right: linkage map distance (cM/Mb, turquoise), repeated sequences (% nucleotides per 500 kb, purple), genes (% nucleotides per 500 kb, blue), chloroplast insertions; RNA-Seq reads from leaves and breaker fruits of *S. lycopersicum* and *S. pimpinellifolium* (number of reads per 500 kb, green and red, respectively), microRNA genes (transcripts per million per 500 kb, black), small RNAs (thin horizontal black and red lines, sum of hits-normalized abundances). Horizontal grey lines represent gaps in the pseudomolecule (**f**). **f**, DNA pseudomolecule consisting of nine scaffolds. Unsequenced gaps (approximately 9.8 Mb, Supplementary Table 13) are indicated by white horizontal lines. Tomato genes identified by map-based cloning (Supplementary Table 14) are indicated on the right. For more details, see legend to Supplementary Fig. 1. **B**, Syntenic relationships in the *Solanaceae*. COSII-based comparative maps of potato, aubergine (eggplant), pepper and *Nicotiana* with respect to the tomato genome (Supplementary Information section 4.5 and Supplementary Fig. 14). Each tomato chromosome is assigned a different colour and orthologous chromosome segment(s) in other species are shown in the same colour. White dots indicate approximate centromere locations. Each black arrow indicates an inversion relative to tomato and '+1' indicates a minimum of one inversion. Each black bar beside a chromosome indicates translocation breakpoints relative to tomato. Chromosome lengths are not to scale, but segments within chromosomes are. **C**, Tomato–potato syntenic relationships dot plot of tomato (T) and potato (P) genomic sequences based on collinear blocks (Supplementary Information section 4.1). Red and blue dots represent gene pairs with statistically significant high and low ω (K_a/K_s) in collinear blocks, which average $K_s \leq 0.5$, respectively. Green and magenta dots represent genes in collinear blocks which average $0.5 < K_s \leq 1.5$ and $K_s > 1.5$, respectively. Yellow dots represent all other gene pairs. Blocks circled in red are examples of pan-eudicot triplication. Inserts represent schematic drawings of BAC-FISH patterns of cytologically demonstrated chromosome inversions (also in Supplementary Fig. 15).

The genome of *S. pimpinellifolium* LA1589 was sequenced and assembled *de novo* using Illumina short reads, yielding a 739 Mb draft genome (Supplementary Information section 3). Estimated divergence between the wild and domesticated genomes is 0.6% (5.4 million single nucleotide polymorphisms (SNPs) distributed along the chromosomes (Fig. 1A and Supplementary Fig. 1)).

Tomato chromosomes consist of pericentric heterochromatin and distal euchromatin, with repeats concentrated within and around centromeres, in chromomeres and at telomeres (Fig. 1A and Supplementary Fig. 1). Substantially higher densities of recombination, genes and transcripts are observed in euchromatin, whereas chloroplast insertions (Supplementary Information sections 1.22 and 1.23) and conserved microRNA (miRNA) genes (Supplementary Information section 2.9) are more evenly distributed throughout the genome. The genome is highly syntenic with those of other economically important Solanaceae (Fig. 1B). Compared to the genomes of *Arabidopsis*[4] and *Sorghum*[5], tomato has fewer high-copy, full-length long terminal repeat (LTR) retrotransposons with older average insertion ages (2.8 versus 0.8 million years (Myr) ago) and fewer high-frequency *k*-mers (Supplementary Information section 2.10). This supports previous findings that the tomato genome is unusual among angiosperms by being largely comprised of low-copy DNA[6, 7].

The pipeline used to annotate the tomato and potato[8] genomes is described in Supplementary Information section 2. It predicted 34,727 and 35,004 protein-coding genes, respectively. Of these, 30,855 and 32,988, respectively, are supported by RNA sequencing (RNA-Seq) data, and 31,741 and 32,056, respectively, show high similarity to *Arabidopsis* genes (Supplementary Information section 2.1). Chromosomal organization of genes, transcripts, repeats and small RNAs (sRNAs) is very similar in the two species (Supplementary Figs 2–4). The protein-coding genes of tomato, potato, *Arabidopsis*, rice and grape were clustered into 23,208 gene groups (≥2 members), of which 8,615 are common to all five genomes, 1,727 are confined to eudicots (tomato, potato, grape and *Arabidopsis*), and 727 are confined to plants with fleshy fruits (tomato, potato and grape) (Supplementary Information section 5.1 and

Supplementary Fig. 5). Relative expression of all tomato genes was determined by replicated strand-specific Illumina RNA-Seq of root, leaf, flower (two stages) and fruit (six stages) in addition to leaf and fruit (three stages) of *S. pimpinellifolium* (Supplementary Table 1).

sRNA sequencing data supported the prediction of 96 conserved miRNA genes in tomato and 120 in potato, a number consistent with other plant species (Fig. 1A, Supplementary Figs 1 and 3 and Supplementary Information section 2.9). Among the 34 miRNA families identified, 10 are highly conserved in plants and similarly represented in the two species, whereas other, less conserved families are more abundant in potato. Several miRNAs, predicted to target Toll interleukin receptor, nucleotide-binding site and leucine-rich repeat (TIR-NBS-LRR) genes, seemed to be preferentially or exclusively expressed in potato (Supplementary Information section 2.9).

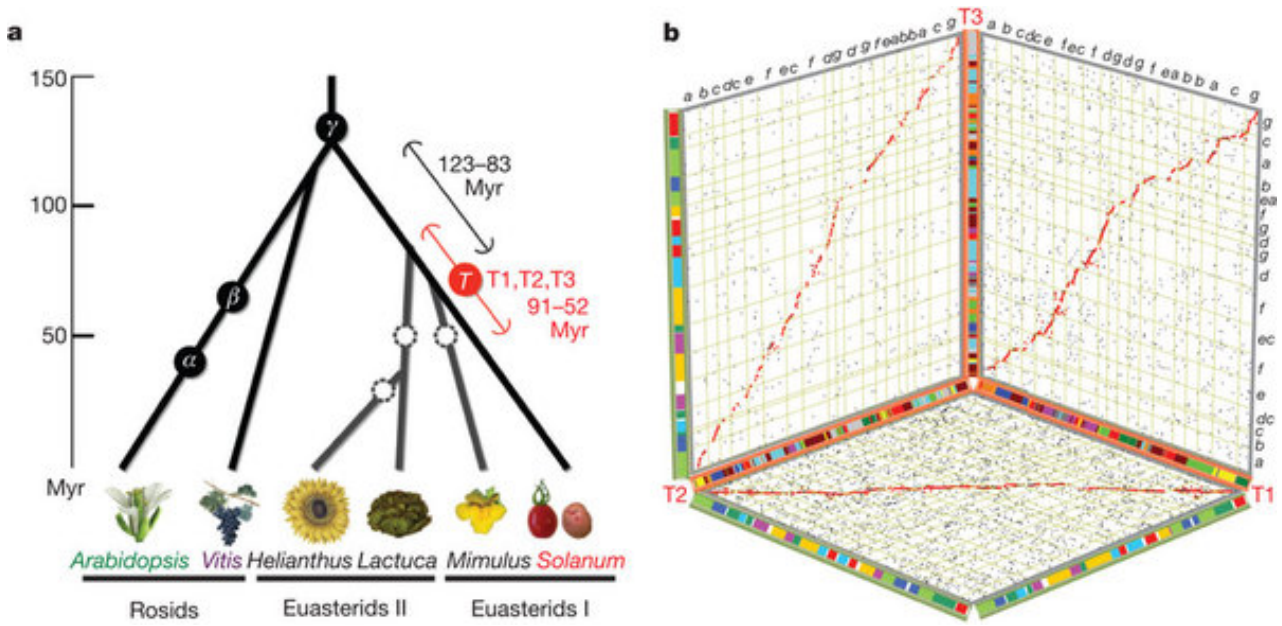
Comparative genomic studies are reported in Supplementary Information section 4. Sequence alignment of 71 Mb of euchromatic tomato genomic DNA to their potato[8] counterparts revealed 8.7% nucleotide divergence (Supplementary Information section 4.1). Intergenic and repeat-rich heterochromatic sequences showed more than 30% nucleotide divergence, consistent with the high sequence diversity in these regions among potato genotypes[8]. Alignment of tomato–potato orthologous regions confirmed nine large inversions known from cytological or genetic studies and several smaller ones (Fig. 1C). The exact number of small inversions is difficult to determine due to the lack of orientation of most potato scaffolds.

A total of 18,320 clearly orthologous tomato–potato gene pairs were identified. Of these, 138 (0.75%) had significantly higher than average non-synonymous (K_a) versus synonymous (K_s) nucleotide substitution rate ratios (ω), indicating diversifying selection, whereas 147 (0.80%) had significantly lower than average ω, indicating purifying selection (Supplementary Table 2). The proportions of high and low ω between sorghum and maize (*Zea mays*) are 0.70% and 1.19%, respectively, after 11.9 Myr of divergence[9], indicating that diversifying selection may have been stronger in tomato–potato. The highest densities of low-ω genes are found in collinear blocks with average $K_s > 1.5$, tracing to a genome triplication shared with grape (see below) (Fig. 1C, Supplementary Fig. 6 and Supplementary Table 3). These genes, which have been preserved in paleo-duplicated locations for more than 100 Myr[10, 11], are more constrained than 'average' genes and are enriched for transcription factors and genes otherwise related to gene regulation (Supplementary Tables 3 and 4).

Sequence comparison of 31,760 Heinz 1706 genes with >5× *S. pimpinellifolium* read coverage in over 90% of their coding regions revealed 7,378 identical genes and 11,753 with only synonymous changes. The remaining 12,629 genes had non-synonymous changes, including gains and losses of stop codons with potential consequences for gene function (Supplementary Tables 5–7). Several pericentric regions, predicted to contain genes, are absent or polymorphic in the broader *S. pimpinellifolium* germplasm (Supplementary Table 8 and Supplementary Fig. 7). Within cultivated germplasm, particularly among the small-fruited cherry tomatoes, several chromosomal segments are more closely related to *S. pimpinellifolium* than to Heinz 1706 (Supplementary Figs 8 and 9), supporting previous observations on recent admixture of these gene pools due to breeding[12]. Heinz 1706 itself has been reported to carry introgressions from *S. pimpinellifolium*[13], traces of which are detectable on chromosomes 4, 9, 11 and 12 (Supplementary Table 9).

Comparison of the tomato and grape genomes supports the hypothesis that a whole-genome triplication affecting the rosid lineage occurred in a common eudicot ancestor[11] (Fig. 2a). The distribution of K_s between corresponding gene pairs in duplicated blocks suggests that one polyploidization in the solanaceous lineage preceded the rosid–asterid (tomato–grape) divergence (Supplementary Fig. 10).

Figure 2: The *Solanum* whole genome triplication.

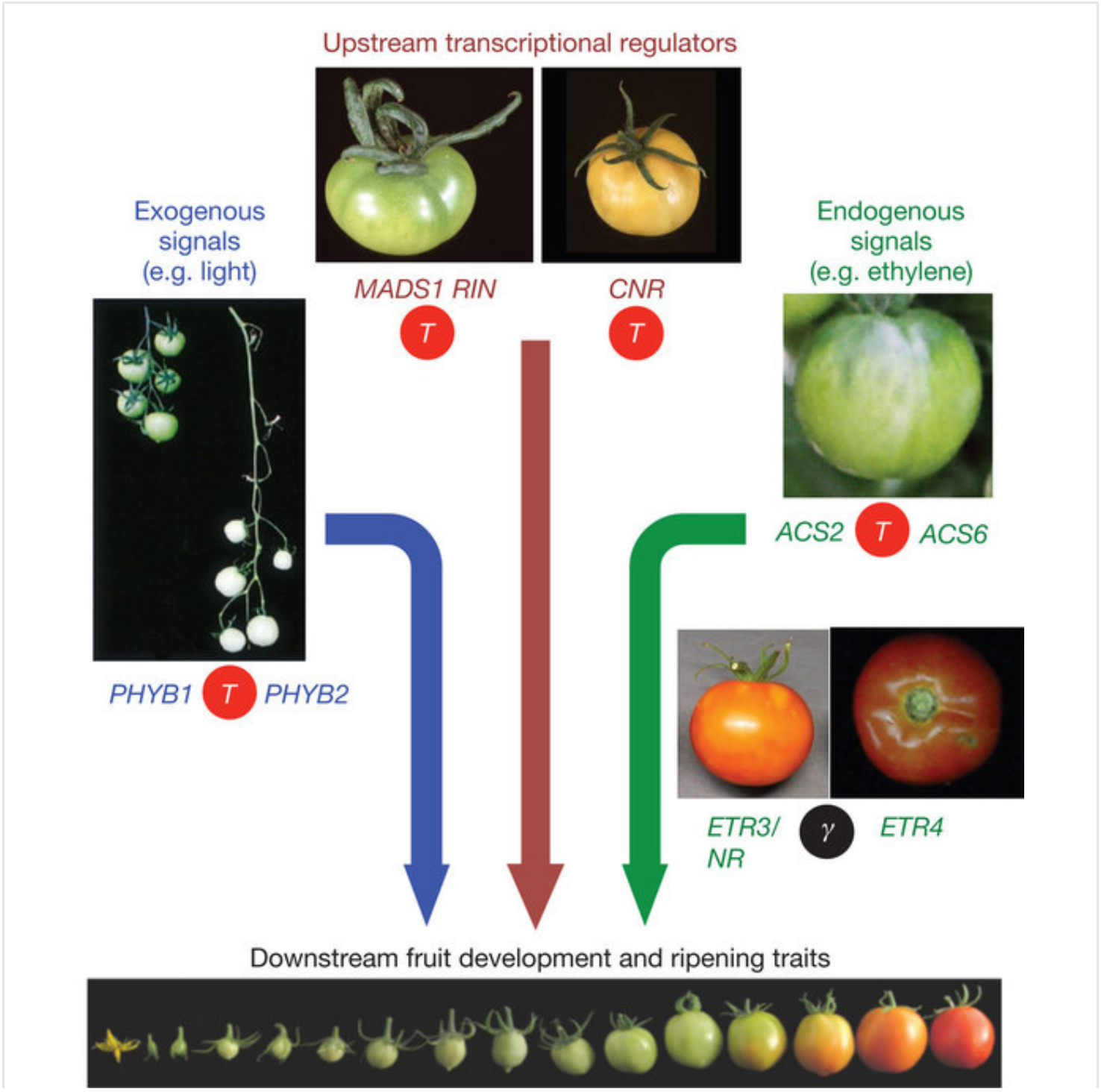


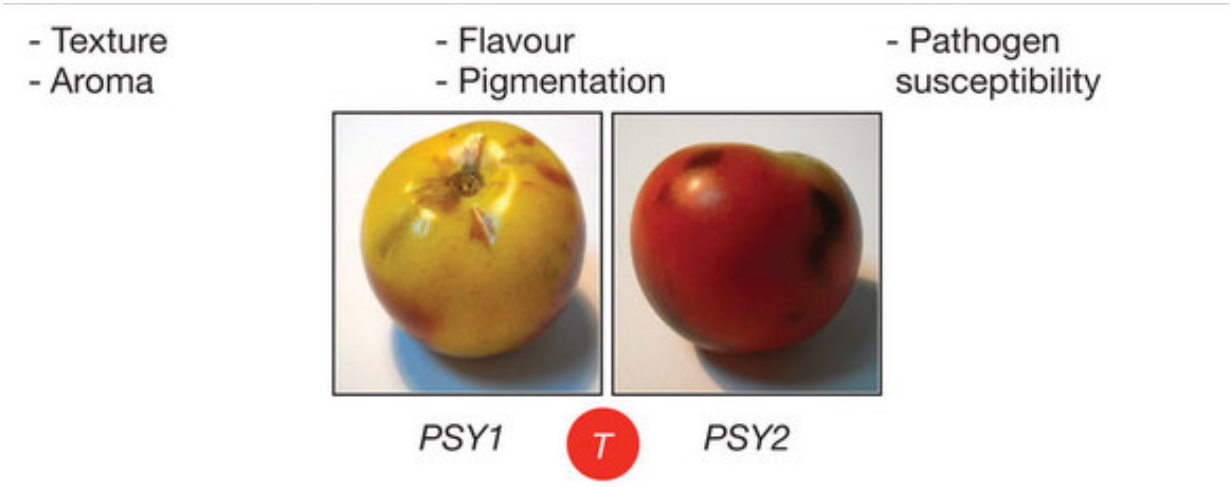
a, Speciation and polyploidization in eudicot lineages. Confirmed whole-genome duplications and triplications are shown with annotated circles, including '*T*' (this paper) and previously discovered events *α*, *β*, *γ*[10, 11, 14]. Dashed circles represent one or more suspected polyploidies reported in previous publications that need further support from genome assemblies[27, 28]. Grey branches indicate unpublished genomes. Black and red error bars bracket indicate the likely timings of divergence of major asterid lineages and of '*T*', respectively. The post-'*T*' subgenomes, designated T1, T2, and T3, are further detailed in Supplementary Fig. 10. **b**, On the basis of alignments of multiple tomato genome segments to single grape genome segments, the tomato genome is partitioned into three non-overlapping 'subgenomes' (T1, T2, T3), each represented by one axis in the three-dimensional plot. The ancestral gene order of each subgenome is inferred according to orthologous grape regions, with tomato chromosomal affinities shown by red (inner) bars. Segments tracing to pan-eudicot triplication (*γ*) are shown by green (outer) bars with colours representing the seven putative pre-*γ* eudicot ancestral chromosomes[10], also coded *a*–*g*.

Comparison with the grape genome also reveals a more recent triplication in tomato and potato. Whereas few individual tomato/potato genes remain triplicated (Supplementary Tables 10 and 11), 73% of tomato gene models are in blocks that are orthologous to one grape region, collectively covering 84% of the grape gene space. Among these grape genomic regions, 22.5% have one orthologous region in tomato, 39.9% have two, and 21.6% have three, indicating that a whole-genome triplication occurred in the *Solanum* lineage, followed by widespread gene loss. This triplication, also evident in potato (Supplementary Fig. 11), is estimated at 71 (±19.4) Myr on the basis of the K_S of paralogous genes (Supplementary Fig. 10), and therefore predates the ~7.3 Myr tomato–potato divergence. On the basis of alignments to single grape genome segments, the tomato genome can be partitioned into three non-overlapping 'subgenomes' (Fig. 2b). The number of euasterid lineages that have experienced the recent triplication remains unclear and awaits complete euasterid I and II genome sequences. K_S distributions show that euasterids I and II, and indeed the rosid–asterid lineages, all diverged from common ancestry at or near the pan-eudicot triplication (Fig. 2a), suggesting that this event may have contributed to the formation of major eudicot lineages in a short period of several million years[14], partially explaining the explosive radiation of angiosperm plants on Earth[15].

Fleshy fruits (Supplementary Fig. 12) are an important means of attracting vertebrate frugivores for seed dispersal[16]. Combined orthology and synteny analyses indicate that both genome triplications added new gene family members that mediate important fruit-specific functions (Fig. 3). These include transcription factors and enzymes necessary for ethylene biosynthesis (*RIN*, *CNR*, *ACS*) and perception (*ETR3/NR*, *ETR4*)[17], red light photoreceptors influencing fruit quality (*PHYB1/PHYB2*) and ethylene- and light-regulated genes mediating lycopene biosynthesis (*PSY1/PSY2*). Several cytochrome P450 subfamilies associated with toxic alkaloid biosynthesis show contraction or complete loss in tomato and the extant genes show negligible expression in ripe fruits (Supplementary Information section 5.4).

Figure 3: Whole-genome triplications set the stage for fruit-specific gene neofunctionalization.

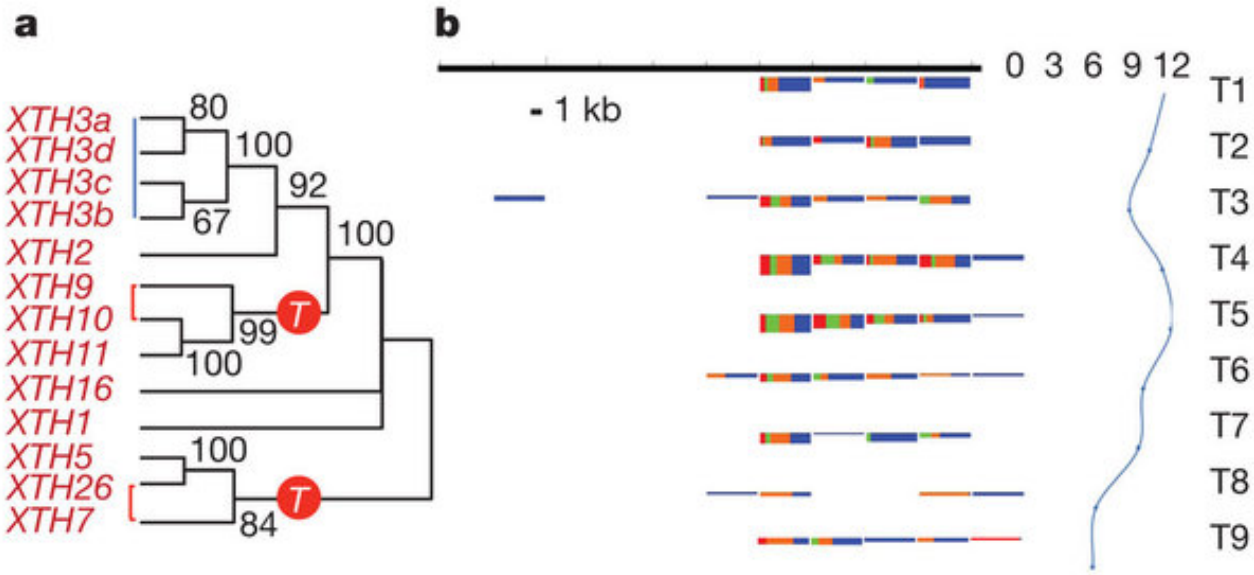




The genes shown represent a fruit ripening control network regulated by transcription factors (*MADS-RIN*, *CNR*) necessary for production of the ripening hormone ethylene, the production of which is regulated by ACC synthase (*ACS*). Ethylene interacts with ethylene receptors (*ETR*s) to drive expression changes in output genes, including phytoene synthase (*PSY*), the rate-limiting step in carotenoid biosynthesis. Light, acting through phytochromes, controls fruit pigmentation through an ethylene-independent pathway. Paralogous gene pairs with different physiological roles (*MADS1/RIN*, *PHYB1/PHYB2*, *ACS2/ACS6*, *ETR3/ETR4*, *PSY1/PSY2*), were generated during the eudicot (*γ*, black circle) or the more recent *Solanum* (*T*, red circle) triplications. Complete dendrograms of the respective protein families are shown in Supplementary Figs 16 and 17.

Fruit texture has profound agronomic and sensory importance and is controlled in part by cell wall structure and composition[18]. More than 50 genes showing differential expression during fruit development and ripening encode proteins involved in modification of cell wall architecture (Fig. 4a and Supplementary Information section 5.7). For example, a family of xyloglucan endotransglucosylase/hydrolases (XTHs) has expanded both in the recent whole-genome triplication and through tandem duplication. One of the triplicated members, *XTH10*, shows differential loss between tomato and potato (Fig. 4a and Supplementary Table 12), suggesting genetically driven specialization in the remodelling of fruit cell walls.

Figure 4: The tomato genome allows systems approaches to fruit biology.



a, Xyloglucan transglucosylase/hydrolases (XTHs) differentially expressed between mature green and ripe fruits (Supplementary Information section 5.7). These *XTH* genes and many others are expressed in ripening fruits and are linked with the *Solanum* triplication, marked with a red circle on the phylogenetic tree. Red lines on the tree denote paralogues derived from the *Solanum* triplication, and blue lines are tandem duplications. **b**, Developmentally regulated accumulation of sRNAs mapping to the promoter region of a fruit-regulated cell wall gene (pectin acetylesterase, Solyc08g005800). Variation of abundance of sRNAs (left) and messenger RNA expression levels from the corresponding gene (right) over a tomato fruit developmental series (T1, bud; T2, flower; T3, fruit 1–3 mm; T4, fruit 5–7 mm; T5, fruit 11–13 mm; T6, fruit mature green; T7, breaker; T8, breaker + 3 days; T9, breaker + 7 days). The promoter regions are grouped in 100-nucleotide windows. For each window the size class distribution of sRNAs is shown (red, 21; green, 22; orange, 23; blue, 24). The height of the box corresponding to the first time point shows the cumulative sRNA abundance in log scale. The height of the following boxes is proportional to the log offset fold change (offset = 20) relative to the first time point. The expression profile of the mRNA is shown in $\log_2$ scale. The horizontal black line represents 1 kb of the promoter region. 0 to 12 represent arbitrary units of gene expression.

Similar to soybean and potato and in contrast to *Arabidopsis*, tomato sRNAs map preferentially to euchromatin (Supplementary Fig. 2). sRNAs from tomato flowers and fruits[19] map to 8,416 gene promoters. Differential expression of sRNAs during fruit development is apparent for 2,687 promoters, including those of cell-wall-related genes (Fig. 4b) and occurs preferentially at key developmental transitions (for example, flower to fruit, fruit growth to fruit ripening, Supplementary Information section 2.8).

The genome sequences of tomato, *S. pimpinellifolium* and potato provide a starting point for comparing gene family evolution and sub-functionalization in the *Solanaceae*. A striking example is the *SELF PRUNING* (*SP*) gene family, which includes the homologue of *Arabidopsis FT*, encoding the mobile flowering signal florigen[20] and its antagonist *SP*, encoding the orthologue of *TFL1*. Nearly a century ago, a spontaneous mutation in *SP* spawned the 'determinate' varieties that now dominate the tomato mechanical harvesting industry[21]. The genome sequence has revealed that the

SP family has expanded in the *Solanum* lineage compared to *Arabidopsis*, driven by the *Solanum* triplication and tandem duplication (Supplementary Fig. 13). In potato, *SP3D* and *SP6A* control flowering and tuberization, respectively[22], whereas *SP3D* in tomato, known as *SINGLE FLOWER TRUSS*, similarly controls flowering, but also drives heterosis for fruit yield in an epistatic relationship with *SP*[23, 24, 25]. Interestingly, *SP6A* in *S. lycopersicum* is inactivated by a premature stop codon, but remains functionally intact in *S. pimpinellifolium*. Thus, allelic variation in a subset of *SP* family genes has played a major role in the generation of both shared and species-specific variation in solanaceous agricultural traits.

The genome sequences of tomato and *S. pimpinellifolium* also provide a basis for understanding the bottlenecks that have narrowed tomato genetic diversity: the domestication of *S. pimpinellifolium* in the Americas, the export of a small number of genotypes to Europe in the 16th century, and the intensive breeding that followed. Charles Rick pioneered the use of trait introgression from wild tomato relatives to increase genetic diversity of cultivated tomatoes[26]. Introgression lines exist for seven wild tomato species, including *S. pimpinellifolium*, in the background of cultivated tomato. The genome sequences presented here and the availability of millions of SNPs will allow breeders to revisit this rich trait reservoir and identify domestication genes, providing biological knowledge and empowering biodiversity-based breeding.

Methods

Main Methods **Accession Codes** **References** **Acknowledgements** **Author Information** **Supplementary Information**

A total of 21 gigabases (Gb) of Roche/454 Titanium shotgun and mate pair reads and 3.3 Gb of Sanger paired-end reads, including ~200,000 BAC and fosmid end sequence pairs, were generated from the 'Heinz 1706' inbred line (Supplementary Information sections 1.1–1.7), assembled using both Newbler and CABOG and integrated into a single assembly (Supplementary Information sections 1.17 and 1.18). The scaffolds were anchored using two BAC-based physical maps, one high density genetic map, overgo hybridization and genome-wide BAC FISH (Supplementary Information sections 1.8–1.16 and 1.19). Over 99.9% of BAC/fosmid end pairs mapped consistently on the assembly and over 98% of EST sequences could be aligned to the assembly (Supplementary Information section 1.20). Chloroplast genome insertions in the nuclear genome were validated using a mate pair method and the flanking regions were identified (Supplementary Information sections 1.22–1.24). Annotation was carried out using a pipeline based on EuGene that integrates *de novo* gene prediction, RNA-Seq alignment and rich function annotation (Supplementary Information section 2). To facilitate interspecies comparison, the potato genome was re-annotated using the same pipeline. LTR retrotransposons were detected *de novo* with the LTR-STRUC program and dated by the sequence divergence between left and right solo LTR (Supplementary Information section 2.10). The genome of *S. pimpinellifolium* was sequenced to ×40 depth using Illumina paired end reads and assembled using ABySS (Supplementary Information section 3). The tomato and potato genomes were aligned using LASTZ (Supplementary Information section 4.1). Identification of triplicated regions was done using BLASTP, in-house-generated scripts and three-way comparisons between tomato, potato and *S. pimpinellifolium* using MCSCAN (Supplementary Information sections 4.2–4.4). Specific gene families/groups (genes for ascorbate, carotenoid and jasmonate biosynthesis, cytochrome P450s, genes controlling cell wall architecture, hormonal and transcriptional regulators, resistance genes) were subjected to expert curation/analysis (Supplementary Information section 5). PHYML and MEGA were used to reconstruct phylogenetic trees and MCSCAN was used to infer gene collinearity (Supplementary Information section 5.2).

Accession codes

Main **Methods** Accession Codes **References** **Acknowledgements** **Author Information** **Supplementary Information**

Primary accessions

GenBank

SRA049915
GSE33507
SRA050797
SRA048144

References

Main **Methods** **Accession Codes** References **Acknowledgements** **Author Information** **Supplementary Information**

1. Frodin, D. G. History and concepts of big plant genera. *Taxon* **53**, 753–776 (2004)

   Show context Article

2. Peralta, I. E., Spooner, D. M. & Knapp, S. Taxonomy of tomatoes: a revision of wild tomatoes (*Solanum* section *Lycopersicon*) and their outgroup relatives in sections *Juglandifolia* and *Lycopersicoides*. *Syst. Bot. Monogr.* **84**, 1–186 (2008)

   Show context

3. Michaelson, M. J., Price, H. J., Ellison, J. R. & Johnston, J. S. Comparison of plant DNA contents determined by Feulgen microspectrophotometry and laser flow cytometry. *Am. J. Bot.* **78**, 183–188 (1991)

   Show context Article CAS

4. The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000)

   Show context Article PubMed ISI CAS

5. Paterson, A. H. *et al.* The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**, 551–556 (2009)

   Show context Article PubMed ISI CAS

6. Zamir, D. & Tanksley, S. D. Tomato genome is comprised largely of fast-evolving, low copy-number sequences. *Mol. Gen. Genet.* **213**, 254–261 (1988)

   Show context Article CAS

7. Peterson, D. G., Pearson, W. R. & Stack, S. M. Characterization of the tomato (*Lycopersicon esculentum*) genome using *in vitro* and *in situ* DNA reassociation. *Genome* **41**, 346–356 (1998)

   Show context CAS

8. Xu, X. *et al.* Genome sequence and analysis of the tuber crop potato. *Nature* **475**, 189–195 (2011)

   Show context Article PubMed CAS

9. Swigoňová, Z. *et al.* Close split of sorghum and maize genome progenitors. *Genome Res.* **14**, 1916–1923 (2004)

Show context Article PubMed ISI CAS

10. Jaillon, O. *et al.* The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463–467 (2007)

Show context Article PubMed ISI CAS

11. Tang, H. *et al.* Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008)

Show context Article PubMed ISI CAS

12. Ranc, N., Munos, S., Santoni, S. & Causse, M. A clarified position for *Solanum lycopersicum* var. *cerasiforme* in the evolutionary history of tomatoes (solanaceae). *BMC Plant Biol.* **8**, 130 (2008)

Show context Article PubMed

13. Ozminkowski, R. Pedigree of variety Heinz 1706. *Rep. Tomato Genet. Coop.* **54**, 26 (2004)

Show context

14. Moore, M. J., Soltis, P. S., Bell, C. D., Burleigh, J. G. & Soltis, D. E. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc. Natl Acad. Sci. USA* **107**, 4623–4628 (2010)

Show context Article PubMed

15. Stockey, R. A., Graham, S. W. & Crane, P. R. Introduction to the Darwin special issue: the abominable mystery. *Am. J. Bot.* **96**, 3–4 (2009)

Show context Article PubMed

16. Howe, H. F. & Smallwood, J. Ecology of seed dispersal. *Annu. Rev. Ecol. Syst.* **13**, 201–228 (1982)

Show context Article ISI

17. Klee, H. J. & Giovannoni, J. J. Genetics and control of tomato fruit ripening and quality attributes. *Annu. Rev. Genet.* **45**, 41–59 (2011)

Show context Article PubMed CAS

18. Vicente, A. R., Saladie, M., Rose, J. K. C. & Labavitch, J. M. The linkage between cell wall metabolism and fruit softening: looking to the future. *J. Sci. Food Agric.* **87**, 1435–1448 (2007)

Show context Article CAS

19. Mohorianu, I. *et al.* Profiling of short RNAs during fleshy fruit development reveals stage-specific sRNAome expression patterns. *Plant J.* **67**, 232–246 (2011)

Show context Article PubMed CAS

20. Corbesier, L. *et al.* FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science* **316**, 1030–1033 (2007)

Show context Article PubMed ISI CAS

21. Rick, C. M. The tomato. *Sci. Am.* **239**, 76–87 (1978)

Show context Article

22. Navarro, C. *et al.* Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. *Nature* **478**, 119–122 (2011)

Show context Article PubMed CAS

23. Lifschitz, E. *et al.* The tomato *FT* ortholog triggers systemic signals that regulate growth and flowering and substitute for diverse environmental stimuli. *Proc. Natl Acad. Sci. USA* **103**, 6398–6403 (2006)

Show context Article PubMed CAS

24. Krieger, U., Lippman, Z. B. & Zamir, D. The flowering gene *SINGLE FLOWER TRUSS* drives heterosis for yield in tomato. *Nature Genet.* **42**, 459–463 (2010)

Show context Article

25. Pnueli, L. *et al.* The *SELF-PRUNING* gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of *CEN* and *TFL1*. *Development* **125**, 1979–1989 (1998)

Show context PubMed ISI CAS

26. Rick, C. M. Hybridization between *Lycopersicon esculentum* and *Solanum pennellii*: phylogenetic and cytogenetic significance. *Proc. Natl Acad. Sci. USA* **46**, 78–82 (1960)

Show context Article PubMed CAS

27. Barker, M. S. *et al.* Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Mol. Biol. Evol.* **25**, 2445–2455 (2008)

Show context Article PubMed CAS

28. Aagaard, J. E., Willis, J. H. & Phillips, P. C. Relaxed selection among duplicate floral regulatory genes in Lamiales. *J. Mol. Evol.* **63**, 493–503 (2006)

Show context Article PubMed CAS

Download references

Acknowledgements

Main **Methods** **Accession Codes** **References** Acknowledgements **Author Information** **Supplementary Information**

This work was supported by: Argentina: INTA and CONICET. Belgium: Flemish Institute for Biotechnology and Ghent University. China: The State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences; Ministry of Science and Technology (2006AA10A116, 2004CB720405, 2006CB101907, 2007DFB30080) Ministry of Agriculture ('948' Program: 2007-Z5); National Natural Science Foundation (36171319); Postdoctoral Science Foundation (20070420446). European Union: FP6 Integrated Project EU-SOL PL 016214. France: Institute National de la Recherche Agronomique and Agence Nationale de la Recherche. Germany: the Max Planck Society. India: Department of Biotechnology, Government of India; Indian Council of Agricultural Research. Italy: Ministry of Research (FIRB-SOL, FIRB-Parallelomics, ItaLyco and GenoPOM projects); Ministry of Agriculture (Agronanotech and Biomassval projects); FILAS foundation; ENEA; CNR-ENEA project L. 191/2009. Japan: Kazusa DNA Research Institute Foundation and National Institute of Vegetable and Tea Science. Korea: KRIBB Basic Research Fund and Crop Functional Genomics Research Center (CFGC), MEST. Netherlands: Centre for BioSystems Genomics, Netherlands Organization for Scientific Research. Spain: Fundación Genoma España; Cajamar; FEPEX; Fundación Séneca; ICIA; IFAPA; Fundación Manrique de Lara; Instituto Nacional de Bioinformatica. UK: BBSRC grant BB/C509731/1; DEFRA; SEERAD. USA: NSF (DBI-0116076; DBI-0421634; DBI-0606595; IOS-0923312; DBI-0820612; DBI-0605659; DEB-0316614; DBI 0849896 and MCB 1021718); USDA (2007-02773 and 2007-35300-19739); USDA-ARS. We acknowledge the Potato Genome Sequencing Consortium for sharing data before publication; potato RNA-Seq data was provided by C. R. Buell from the NSF-funded Potato Genome Sequence and Annotation project; tomato RNA-Seq data by the USDA-funded SolCAP project, N. Sinha and J. Maloof; the Amplicon Express team for BAC pooling services; construction of the Whole Genome Profiling (WGP) physical map was supported by EnzaZaden, RijkZwaan, Vilmorin & Cie, and Takii & Co. Keygene N.V. owns patents and patent applications covering its AFLP and Whole Genome Profiling technologies; AFLP and Keygene are registered trademarks of Keygene N.V. The following individuals are also acknowledged for their contribution to the work described: J. Park, B. Wang, C. Niu, D. Liu, F. Cojutti, S. Pescarolo, A. Zambon, G. Xiao, J. Chen, J. Shi, L. Zhang, L. Zeng, M. Caccamo, D. Bolser, D. Martin, M. Gonzalez, P. A. Bedinger, P. A. Covey, P. Pachori, R. R. Pousada, S. Hakim, S. Sims, V. Cahais, W. Long, X. Zhou, Y. Lu, W. Haso, C. Lai, S. Lepp, C. Peluso, H. Teramu, H. De Jong, R. Lizarralde, E. R. May and Z. Li. M. Zabeau is thanked for his support and encouragement and S. van den Brink for her secretarial support. We dedicate this work to the late C. Rick who pioneered tomato genetics, collection of wild germplasm and the distribution of seed and knowledge.

Author information

Main **Methods** **Accession Codes** **References** **Acknowledgements** Author Information **Supplementary Information**

Affiliations

Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan.
Shusei Sato, Satoshi Tabata, Hideki Hirakawa, Erika Asamizu, Kenta Shirasawa, Sachiko Isobe, Takakazu Kaneko, Yasukazu Nakamura, Daisuke Shibata & Koh Aoki

454 Life Sciences, a Roche company, 15 Commercial Street, Branford, Connecticut 06405, USA.
Michael Egholm & James Knight

Amplicon Express Inc., 2345 Hopkins Court, Pullman, Washington 99163, USA.
Robert Bogden

Beijing Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China.
Changbao Li

National Center for Gene Research, Chinese Academy of Sciences, Shanghai 200233, China.
Changbao Li, Bin Han, Qi Feng & Danlin Fan

BGI-Shenzhen, Shenzhen 518083, China.
Yang Shuang, Xun Xu, Shengkai Pan, Shifeng Cheng, Xin Liu, Yuanyuan Ren & Jun Wang

BMR-Genomics SrL, via Redipuglia 21/A, 35131 Padova, Italy.
Alessandro Albiero, Francesca Dal Pero & Sara Todesco

Boyce Thompson Institute for Plant Research, Tower Road, Cornell University campus, Ithaca, New York 14853, USA.
Joyce Van Eck, Robert M. Buels, Aureliano Bombarely, Joseph R. Gosselin, Minyun Huang, Jonathan A. Leto, Naama Menda, Susan Strickler, Linyong Mao, Shan Gao, Isaak Y. Tecle, Thomas York, Yi Zheng, Julia T. Vrebalov, JeMin Lee, Silin Zhong, Lukas A. Mueller, Zhangjun Fei, Ruth A. White, Theodore W. Thannhauser & James J. Giovannoni

Centre for BioSystems Genomics, PO Box 98, 6700 AB Wageningen, The Netherlands.
Willem J. Stiekema

Centro Nacional de Análisis Genómico (CNAG), C/ Baldiri Reixac 4, Torre I, 08028 Barcelona, Spain.
Paolo Ribeca & Tyler Alioto

Department of Vegetable Science, College of Agronomy and Biotechnology, China Agricultural University, No. 2 Yuanmingyuan Xi Lu, Haidian District, Beijing 100193, China.
Wencai Yang

Key Laboratory of Horticultural Crops Genetic Improvement of Ministry of Agriculture, Sino-Dutch Joint Lab of Horticultural Genomics Technology, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China.
Sanwen Huang, Yongchen Du, Zhonghua Zhang, Jianchang Gao, Yanmei Guo, Xiaoxuan Wang, Ying Li & Jun He

State Key Laboratory of Plant Genomics and National Centre for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China.
Chuanyou Li, Zhukuan Cheng, Jianru Zuo, Jianfeng Ren, Jiuhai Zhao, Liuhua Yan, Hongling Jiang, Bao Wang, Hongshuang Li, Zhenjun Li, Fuyou Fu & Bingtang Chen

Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China.
Ying Wang

State Key Laboratory of Plant Cell and Chromosome Engineering and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China.
Hongqing Ling

Laboratory of Molecular and Developmental Biology and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100080, China.
Yongbiao Xue

Cold Spring Harbor Laboratory, One Bungtown Road, Cold Spring Harbor, New York 11724, USA.
Doreen Ware, W. Richard McCombie, Zachary B. Lippman, Jer-Ming Chia, Ke Jiang, Shiran Pasternak, Laura Gelley & Melissa Kramer

Department of Biology, Colorado State University, Fort Collins, Colorado 80523, USA.
Lorinda K. Anderson, Suzanne M. Royer, Lindsay A. Shearer & Stephen M. Stack

Department of Agronomy, National Taiwan University, Taipei 107, Taiwan.
Song-Bin Chang

Department of Plant Biology, Cornell University, Ithaca, New York 14853, USA.
Jocelyn K. C. Rose, Yimin Xu, Nancy Eannetta, Antonio J. Matas, Ryan McQuinn, Steven D. Tanksley & James J. Giovannoni

Genome Bioinformatics Laboratory. Center for Genomic Regulation (CRG), University Pompeu Fabra, Barcelona 08003, Spain.
Francisco Camara & Roderic Guigó

Department of Plant Systems Biology, VIB. Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, 9052 Gent, Belgium.
Stephane Rombauts, Jeffrey Fawcett & Yves Van de Peer

Faculty of Agriculture, The Hebrew University of Jerusalem, PO Box 12, Rehovot 76100, Israel.
Dani Zamir

Institute of Industrial Crops, Heilongjiang Academy of Agricultural Sciences, Harbin 150086, China.
Chunbo Liang

Institute for Bioinformatics and Systems Biology (MIPS), Helmholtz Center for Health and Environment, Ingolstädter Landstr. 1, D-85764 Neuherberg, Germany.
Manuel Spannagl, Heidrun Gundlach, Remy Bruggmann & Klaus Mayer

College of Horticulture, Henan Agricultural University, Zhengzhou 450002, China.
Zhiqi Jia

National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China.
Junhong Zhang & Zhibiao Ye

Department of Life Sciences, Imperial College London, London SW7 1AZ, UK.
Gerard J. Bishop, Sarah Butcher, Rosa Lopez-Cobollo, Daniel Buchan, Ioannis Filippis & James Abbott

NRC on Plant Biotechnology, Indian Agricultural Research Institute, New Delhi 110 012, India.
Rekha Dixit, Manju Singh, Archana Singh, Jitendra Kumar Pal, Awadhesh Pandit, Pradeep Kumar Singh, Ajay Kumar Mahato, Vivek Dogra, Kishor Gaikwad, Tilak Raj Sharma, Trilochan Mohapatra & Nagendra Kumar Singh

INRA, UR1052 Génétique et amélioration des fruits et légumes, BP 94, 84143 Monfavet Cedex, France.
Mathilde Causse

INRA, Biologie du Fruit et Pathologie, 71 rue E. Bourleaux, 33883 Villenave d'Ornon, France.
Christophe Rothan

Unité de Biométrie et d'Intelligence Artificielle UR 875, INRA, F-31320 Castanet-Tolosan, France.
Thomas Schiex & Céline Noirot

INRA-CNRGV BP52627, 31326 Castanet-Tolosan, France.
Arnaud Bellec, Hélène Berges & Sonia Vautrin

Plateforme bioinformatique Genotoul, UR875 Biométrie et Intelligence Artificielle, INRA, 31326 Castanet-Tolosan, France.
Christophe Klopp & Jérôme Mariette

Institut National Polytechnique de Toulouse - ENSAT, Université de Toulouse, Avenue de l'Agrobiopole BP 32607, 31326 Castanet-Tolosan, France.
Corinne Delalande, Pierre Frasse, Mohamed Zouine, Alain Latché, Christine Rousseau, Farid Regad, Jean-Claude Pech, Murielle Philippot & Mondher Bouzayen

Instituto de Biología Molecular y Celular de Plantas (CSIC-UPV), Ciudad Politecnica de la Innovación, escalera 8E, Ingeniero Fausto Elios s/n, 46022 Valencia, Spain.
Pierre Pericard, Sonia Osorio, Asunción Fernandez del Carmen, Antonio Monforte & Antonio Granell

Instituto de Hortofruticultura Subtropical y Mediterránea "La Mayora", Universidad de Malaga - Consejo Superior de Investigaciones Cientificas (IHSM-UMA-CSIC), 29750 Algarrobo-Costa (Málaga), Spain.
Rafael Fernandez-Muñoz

Instituto de Biotecnología, PO Box 25, B1712WAA Castelar, Argentina.
Mariana Conte, Gabriel Lichtenstein & Fernando Carrari

Institute for Biomedical Technologies, National Research Council of Italy, Via F. Cervi 93, 20090 Segrate (Milano), Italy.
Gianluca De Bellis, Fabio Fuligni & Clelia Peano

Institute of Plant Genetics, Research Division Portici, National Research Council of Italy, Via Università 133,

80055 Portici (Naples), Italy.
Silvana Grandillo & Pasquale Termolino

ENEA, Casaccia Research Center, Via Anguillarese 301, 00123 Roma, Italy.
Marco Pietrella, Elio Fantini, Giulia Falcone, Alessia Fiore & Giovanni Giuliano

Scuola Superiore Sant'Anna, Piazza Martiri della Libertà 33 - 56127 Pisa, Italy.
Marco Pietrella

ENEA, Trisaia Research Center, S.S. Ionica - Km 419.5, 75026 Rotondella (Matera), Italy.
Loredana Lopez, Paolo Facella, Gaetano Perrotta & Loretta Daddiego

James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK.
Glenn Bryan

Barcelona Supercomputing Center, Nexus II Building, c/ Jordi Girona, 29, 08034 Barcelona, Spain.
Modesto Orozco, Xavier Pastor & David Torrents

Institute of Research in Biomedicine, c/ Josep Samiter 1-5, 08028 Barcelona, Spain
Modesto Orozco

ICREA, Pg Lluís Companys, 23, 08010 Barcelona, Spain.
David Torrents

Keygene N.V., Agro Business Park 90, 6708 PW Wageningen, The Netherlands.
Marco G. M. van Schriek, Richard M.C. Feron, Jan van Oeveren, Peter de Heer, Lorena daPonte, Saskia Jacobs-Oomen, Mike Cariaso, Marcel Prins, Michiel J. T. van Eijk, Antoine Janssen, Mark J. J. van Haaren, Erwin Datema & Roeland C. H. J. van Ham

Plant Systems Engineering Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, 305-806, Republic of Korea.
Sung-Hwan Jo, Jungeun Kim, Suk-Yoon Kwon, Sangmi Kim, Dal-Hoe Koo, Sanghyeob Lee & Cheol-Goo Hur

Life Technologies, 500 Cummings Center, Beverly, Massachusetts 01915, USA.
Christopher Clouser

Life Technologies, 25 avenue de la Baltique, BP 96, 91943 Courtaboeuf Cedex 3, France.
Alain Rico

Max Planck Institute for Plant Breeding Research, Carl von Linné Weg 10, 50829 Cologne, Germany.
Asis Hallab, Christiane Gebhardt, Kathrin Klee, Anika Jöcker, Jens Warfsmann & Ulrike Göbel

School of Agriculture, Meiji University, 1-1-1 Higashi-Mita, Tama-ku, Kawasaki-shi, Kanagawa 214-8571, Japan.
Shingo Kawamura & Kentaro Yano

Department of Plant Science and Plant Pathology, Montana State University, Bozeman, Montana 59717, USA.
Jamie D. Sherman

NARO Institute of Vegetable and Tea Science, 360 Kusawa, Ano, Tsu, Mie 514-2392, Japan.
Hiroyuki Fukuoka & Satomi Negoro

National Institute of Plant Genome Research, New Delhi 110 067, India.
Sarita Bhutty, Parul Chowdhury & Debasis Chattopadhyay

Plant Research International, Business Unit Bioscience, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.
Erwin Datema, Sandra Smit, Elio G. W. M. Schijlen, Jose van de Belt, Jan C. van Haarst, Sander A. Peters, Marjo J. van Staveren, Marleen H. C. Henkens, Paul J. W. Mooyman, Thamara Hesselink & Roeland C. H. J. van Ham

Institute of Plant Genetic Engineering, Qingdao Agricultural University, Qingdao 266109, China.
Guoyong Jiang

Roche Applied Science, D-82377 Penzberg, Germany.
Marcus Droege

Seoul National University, Department of Plant Science and Plant Genomics and Breeding Institute, Seoul 151-921, Republic of Korea.
Doil Choi, Byung-Cheol Kang, Byung Dong Kim, Minkyu Park, Seungill Kim & Seon-In Yeom

Seoul National University, Department of Agricultural Biotechnology, Seoul 151-921, Republic of Korea.
Yong-Hwan Lee

Seoul National University, Crop Functional Genomics Center, College of Agriculture and Life Sciences, Seoul 151-921, Republic of Korea.
Yang-Do Choi

High-Tech Research Center, Shandong Academy of Agricultural Sciences, Jinan, 250000 Shandong, China.
Guangcun Li

Institute of Vegetables, Shandong Academy of Agricultural Sciences, Jinan, 250100 Shandong, China.
Jianwei Gao

School of Life Sciences, Sichuan University, Chengdu, Sichuan 610064, China.
Yongsheng Liu & Shengxiong Huang

Sistemas Genomicos, Parque Tecnológico de Valencia, Ronda G. Marconi, 6, 46980 Paterna (Valencia), Spain.
Victoria Fernandez-Pedrosa, Carmen Collado & Sheila Zuñiga

College of Horticulture, South China Agricultural University, 510642 Guangzhou, China.

Guoping Wang

Syngenta Biotechnology, Inc. 3054 East Cornwallis Road, Research Triangle Park, North Carolina 27709 Durham, USA.
Rebecca Cade & Robert A. Dietrich

Norwich Research Park, Norwich NR4 7UH, UK.
Jane Rogers

Department of Botany, The Natural History Museum, Cromwell Road, London SW7 5BD, UK.
Sandra Knapp

United States Department of Agriculture - Agricultural Research Service, Robert W. Holley Center, Tower Road, Cornell University campus, Ithaca, New York 14853, USA.
Zhangjun Fei, Ruth A. White, Theodore W. Thannhauser & James J. Giovannoni

Instituto de Hortofruticultura Subtropical y Mediterranea. Departamento de Biologia Molecular y Bioquimica, 29071 Málaga, Spain.
Miguel Angel Botella & Louise Gilbert

Centre de Regulacio Genomica, Universitat Pompeu Fabra, Dr Aiguader, 88, E-08003 Barcelona, Spain.
Ramon Gonzalez

Arizona Genomics Institute, BIO-5 Institute for Collaborative Research, School of Plant Sciences, Thomas W. Keating Building, 1657 E. Helen Street, Tucson, Arizona 85721, USA.
Jose Luis Goicoechea, Yeisoo Yu, David Kudrna, Kristi Collura, Marina Wissotski & Rod Wing

Crop Bioinformatics, Institute of Crop Science and Resource Conservation, University of Bonn, 53115 Bonn, Germany.
Heiko Schoof

Department of Plant and Soil Sciences, and Delaware Biotechnology Institute, University of Delaware, Newark, Delaware 19711, USA.
Blake C. Meyers, Aishwarya Bala Gurazada & Pamela J. Green

Interdisciplinary Centre for Plant Genomics and Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi 110 021, India.
Saloni Mathur, Shailendra Vyas, Amolkumar U. Solanke, Rahul Kumar, Vikrant Gupta, Arun K. Sharma, Paramjit Khurana, Jitendra P. Khurana & Akhilesh K. Tyagi

University of East Anglia, BIO, Norwich NR4 7TJ, UK.
Tamas Dalmay

University of East Anglia, CMP, Norwich NR4 7TJ, UK.

Irina Mohorianu

Department of Biology and the UF Genetics Institute, Cancer and Genetics Research Complex 2033 Mowry Road, PO Box 103610, Gainesville, Florida 32610, USA.
Brandon Walts, Srikar Chamala & W. Brad Barbazuk

Plant Genome Mapping Laboratory, 111 Riverbend Road, University of Georgia, Athens, Georgia 30602, USA.
Jingping Li, Hui Guo, Tae-Ho Lee, Yupeng Wang, Dong Zhang, Andrew H. Paterson, Xiyin Wang & Haibao Tang

Center for Genomics and Computational Biology, School of Life Sciences, and School of Sciences, Hebei United University, Tangshan, Hebei 063000, China.
Xiyin Wang

J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, Maryland 20850, USA.
Haibao Tang

University of Naples "Federico II" Department of Soil, Plant, Environmental and Animal Production Sciences, Via Universita', 100, 80055 Portici (Naples), Italy.
Amalia Barone, Maria Luisa Chiusano, Maria Raffaella Ercolano, Nunzio D'Agostino, Miriam Di Filippo, Alessandra Traini, Walter Sanseverino & Luigi Frusciante

Division of Plant and Crop Sciences, University of Nottingham, Sutton Bonington, Loughborough LE12 5RD, UK.
Graham B. Seymour

Department of Chemistry and Biochemistry, Stephenson Research and Technology Center, University of Oklahoma, Norman, Oklahoma 73019, USA.
Mounir Elharam, Ying Fu, Axin Hua, Steven Kenton, Jennifer Lewis, Shaoping Lin, Fares Najar, Hongshing Lai, Baifang Qin, Chunmei Qu, Ruihua Shi, Douglas White, James White, Yanbo Xing, Keqin Yang, Jing Yi, Ziyun Yao, Liping Zhou & Bruce A. Roe

CRIBI, University of Padua, via Ugo Bassi 58/B, 35131 Padova, Italy.
Alessandro Vezzi, Michela D'Angelo, Rosanna Zimbello, Riccardo Schiavon, Elisa Caniato, Chiara Rigobello, Davide Campagna, Nicola Vitulo & Giorgio Valle

Department of Microbiology, Immunology and Biochemistry, University of Tennessee Health Science Center, Memphis, Tennessee 38163, USA.
David R. Nelson

Department of Agriculture and Environmental Sciences, University of Udine, via delle Scienze 208, 33100, Udine, Italy.
Emanuele De Paoli

Wageningen University, Laboratory of Genetics, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.
Dora Szinay & Hans H. de Jong

Wageningen University, Laboratory of Plant Breeding, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.
Dora Szinay, Yuling Bai & Richard G. F. Visser

Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.
René M. Klein Lankhorst

Wellcome Trust Sanger Institute Hinxton, Cambridge CB10 1SA, UK.
Helen Beasley, Karen McLaren, Christine Nicholson & Claire Riddle

Ylichron SrL, Casaccia Research Center, Via Anguillarese 301, 00123 Roma, Italy.
Giulio Gianese

Present address: Plant Engineering Research Institute, Sejong University, Seoul, 143-747, Republic of Korea.
Sanghyeob Lee

Consortia
The Tomato Genome Consortium

Kazusa DNA Research Institute
Shusei Sato, Satoshi Tabata, Hideki Hirakawa, Erika Asamizu, Kenta Shirasawa, Sachiko Isobe, Takakazu Kaneko, Yasukazu Nakamura, Daisuke Shibata & Koh Aoki

454 Life Sciences, a Roche company
Michael Egholm & James Knight

Amplicon Express Inc.
Robert Bogden

Beijing Academy of Agriculture and Forestry Sciences
Changbao Li

BGI-Shenzhen
Yang Shuang, Xun Xu, Shengkai Pan, Shifeng Cheng, Xin Liu, Yuanyuan Ren & Jun Wang

BMR-Genomics SrL
Alessandro Albiero, Francesca Dal Pero & Sara Todesco

Boyce Thompson Institute for Plant Research
Joyce Van Eck, Robert M. Buels, Aureliano Bombarely, Joseph R. Gosselin, Minyun Huang, Jonathan A. Leto, Naama Menda, Susan Strickler, Linyong Mao, Shan Gao, Isaak Y. Tecle, Thomas York, Yi Zheng, Julia T. Vrebalov, JeMin Lee, Silin Zhong & Lukas A. Mueller

Centre for BioSystems Genomics
Willem J. Stiekema

Centro Nacional de Análisis Genómico (CNAG)
Paolo Ribeca & Tyler Alioto

China Agricultural University
Wencai Yang

Chinese Academy of Agricultural Sciences
Sanwen Huang, Yongchen Du, Zhonghua Zhang, Jianchang Gao, Yanmei Guo, Xiaoxuan Wang, Ying Li & Jun He

Chinese Academy of Sciences
Chuanyou Li, Zhukuan Cheng, Jianru Zuo, Jianfeng Ren, Jiuhai Zhao, Liuhua Yan, Hongling Jiang, Bao Wang, Hongshuang Li, Zhenjun Li, Fuyou Fu, Bingtang Chen, Bin Han, Qi Feng, Danlin Fan, Ying Wang, Hongqing Ling & Yongbiao Xue

Cold Spring Harbor Laboratory and United States Department of Agriculture – Agricultural Research Service
Doreen Ware, W. Richard McCombie, Zachary B. Lippman, Jer-Ming Chia, Ke Jiang, Shiran Pasternak, Laura Gelley & Melissa Kramer

Colorado State University
Lorinda K. Anderson, Song-Bin Chang, Suzanne M. Royer, Lindsay A. Shearer & Stephen M. Stack

Cornell University
Jocelyn K. C. Rose, Yimin Xu, Nancy Eannetta, Antonio J. Matas, Ryan McQuinn & Steven D. Tanksley

Genome Bioinformatics Laboratory GRIB–IMIM/UPF/CRG
Francisco Camara & Roderic Guigó

Ghent University-VIB
Stephane Rombauts, Jeffrey Fawcett & Yves Van de Peer

Hebrew University of Jerusalem
Dani Zamir

Heilongjiang Academy of Agricultural Sciences
Chunbo Liang

Helmholtz Center for Health and Environment
Manuel Spannagl, Heidrun Gundlach, Remy Bruggmann & Klaus Mayer

Henan Agricultural University
Zhiqi Jia

Huazhong Agricultural University
Junhong Zhang & Zhibiao Ye

Imperial College London
Gerard J. Bishop, Sarah Butcher, Rosa Lopez-Cobollo, Daniel Buchan, Ioannis Filippis & James Abbott

Indian Agricultural Research Institute
Rekha Dixit, Manju Singh, Archana Singh, Jitendra Kumar Pal, Awadhesh Pandit, Pradeep Kumar Singh, Ajay Kumar Mahato, Vivek Dogra, Kishor Gaikwad, Tilak Raj Sharma, Trilochan Mohapatra & Nagendra Kumar Singh

INRA Avignon
Mathilde Causse

INRA Bordeaux
Christophe Rothan

INRA Toulouse
Thomas Schiex, Céline Noirot, Arnaud Bellec, Christophe Klopp, Corinne Delalande, Hélène Berges, Jérôme Mariette, Pierre Frasse & Sonia Vautrin

Institut National Polytechnique de Toulouse
Mohamed Zouine, Alain Latché, Christine Rousseau, Farid Regad, Jean-Claude Pech, Murielle Philippot & Mondher Bouzayen

Instituto de Biología Molecular y Celular de Plantas (CSIC-UPV)
Pierre Pericard, Sonia Osorio, Asunción Fernandez del Carmen, Antonio Monforte & Antonio Granell

Instituto de Hortofruticultura Subtropical y Mediterránea (IHSM-UMA-CSIC)
Rafael Fernandez-Muñoz

Instituto Nacional de Tecnología Agropecuaría (IB-INTA) and Consejo Nacionalde Investigaciones Científicas y Técnicas (CONICET)
Mariana Conte, Gabriel Lichtenstein & Fernando Carrari

Italian National Res Council, Institute for Biomedical Technologies
Gianluca De Bellis, Fabio Fuligni & Clelia Peano

Italian National Research Council, Institute of Plant Genetics, Research Division Portici
Silvana Grandillo & Pasquale Termolino

Italian National Agency for New technologies, Energy and Sustainable Development
Marco Pietrella, Elio Fantini, Giulia Falcone, Alessia Fiore, Giovanni Giuliano, Loredana Lopez, Paolo Facella, Gaetano Perrotta & Loretta Daddiego

James Hutton Institute
Glenn Bryan

Joint IRB-BSC program on Computational Biology
Modesto Orozco, Xavier Pastor & David Torrents

Keygene N.V.
Marco G. M. van Schriek, Richard M.C. Feron, Jan van Oeveren, Peter de Heer, Lorena daPonte, Saskia Jacobs-Oomen, Mike Cariaso, Marcel Prins, Michiel J. T. van Eijk, Antoine Janssen & Mark J. J. van Haaren

Korea Research Institute of Bioscience and Biotechnology
Sung-Hwan Jo, Jungeun Kim, Suk-Yoon Kwon, Sangmi Kim, Dal-Hoe Koo, Sanghyeob Lee & Cheol-Goo Hur

Life Technologies
Christopher Clouser & Alain Rico

Max Planck Institute for Plant Breeding Research
Asis Hallab, Christiane Gebhardt, Kathrin Klee, Anika Jöcker, Jens Warfsmann & Ulrike Göbel

Meiji University
Shingo Kawamura & Kentaro Yano

Montana State University
Jamie D. Sherman

NARO Institute of Vegetable and Tea Science
Hiroyuki Fukuoka & Satomi Negoro

National Institute of Plant Genome Research
Sarita Bhutty, Parul Chowdhury & Debasis Chattopadhyay

Plant Research International
Erwin Datema, Sandra Smit, Elio G. W. M. Schijlen, Jose van de Belt, Jan C. van Haarst, Sander A. Peters, Marjo J. van Staveren, Marleen H. C. Henkens, Paul J. W. Mooyman, Thamara Hesselink & Roeland C. H. J. van Ham

Qingdao Agricultural University
Guoyong Jiang

Roche Applied Science
Marcus Droege

Seoul National University
Doil Choi, Byung-Cheol Kang, Byung Dong Kim, Minkyu Park, Seungill Kim, Seon-In Yeom, Yong-Hwan Lee & Yang-Do Choi

Shandong Academy of Agricultural Sciences
Guangcun Li & Jianwei Gao

Sichuan University
Yongsheng Liu & Shengxiong Huang

Sistemas Genomicos
Victoria Fernandez-Pedrosa, Carmen Collado & Sheila Zuñiga

South China Agricultural University
Guoping Wang

Syngenta Biotechnology
Rebecca Cade & Robert A. Dietrich

The Genome Analysis Centre
Jane Rogers

The Natural History Museum
Sandra Knapp

United States Department of Agriculture – Agricultural Research Service, Robert W. Holley Center and Boyce Thompson Institute for Plant Research
Zhangjun Fei, Ruth A. White, Theodore W. Thannhauser & James J. Giovannoni

Universidad de Malaga-Consejo Superior de Investigaciones Cientificas
Miguel Angel Botella & Louise Gilbert

Universitat Pompeu Fabra
Ramon Gonzalez

University of Arizona
Jose Luis Goicoechea, Yeisoo Yu, David Kudrna, Kristi Collura, Marina Wissotski & Rod Wing

University of Bonn
Heiko Schoof

University of Delaware
Blake C. Meyers, Aishwarya Bala Gurazada & Pamela J. Green

University of Delhi South Campus
Saloni Mathur, Shailendra Vyas, Amolkumar U. Solanke, Rahul Kumar, Vikrant Gupta, Arun K. Sharma, Paramjit Khurana, Jitendra P. Khurana & Akhilesh K. Tyagi

University of East Anglia, School of Biological Sciences
Tamas Dalmay

University of East Anglia, School of Computing Sciences
Irina Mohorianu

University of Florida
Brandon Walts, Srikar Chamala & W. Brad Barbazuk

University of Georgia
Jingping Li, Hui Guo, Tae-Ho Lee, Yupeng Wang, Dong Zhang, Andrew H. Paterson, Xiyin Wang & Haibao Tang

University of Naples "Federico II"
Amalia Barone, Maria Luisa Chiusano, Maria Raffaella Ercolano, Nunzio D'Agostino, Miriam Di Filippo, Alessandra Traini, Walter Sanseverino & Luigi Frusciante

University of Nottingham
Graham B. Seymour

University of Oklahoma
Mounir Elharam, Ying Fu, Axin Hua, Steven Kenton, Jennifer Lewis, Shaoping Lin, Fares Najar, Hongshing Lai, Baifang Qin, Chunmei Qu, Ruihua Shi, Douglas White, James White, Yanbo Xing, Keqin Yang, Jing Yi, Ziyun Yao, Liping Zhou & Bruce A. Roe

University of Padua
Alessandro Vezzi, Michela D'Angelo, Rosanna Zimbello, Riccardo Schiavon, Elisa Caniato, Chiara Rigobello, Davide Campagna, Nicola Vitulo & Giorgio Valle

University of Tennessee Health Science Center
David R. Nelson

University of Udine
Emanuele De Paoli

Wageningen University
Dora Szinay, Hans H. de Jong, Yuling Bai, Richard G. F. Visser & René M. Klein Lankhorst

Wellcome Trust Sanger Institute
Helen Beasley, Karen McLaren, Christine Nicholson & Claire Riddle

Ylichron SrL
Giulio Gianese

Principal Investigators

Shusei Sato, Satoshi Tabata, Lukas A. Mueller, Sanwen Huang, Yongchen Du, Chuanyou Li, Zhukuan Cheng, Jianru Zuo, Bin Han, Ying Wang, Hongqing Ling, Yongbiao Xue, Doreen Ware, W. Richard McCombie, Zachary B. Lippman, Stephen M. Stack, Steven D. Tanksley, Yves Van de Peer, Klaus Mayer, Gerard J. Bishop, Sarah Butcher, Nagendra Kumar Singh, Thomas Schiex, Mondher Bouzayen, Antonio Granell, Fernando Carrari, Gianluca De Bellis, Giovanni Giuliano, Glenn Bryan, Michiel J. T. van Eijk, Hiroyuki Fukuoka, Debasis Chattopadhyay, Roeland C. H. J. van Ham, Doil Choi, Jane Rogers, Zhangjun Fei, James J. Giovannoni, Rod Wing, Heiko Schoof, Blake C. Meyers, Jitendra P. Khurana, Akhilesh K. Tyagi, Tamas Dalmay, Andrew H. Paterson, Xiyin Wang, Luigi Frusciante, Graham B. Seymour, Bruce A. Roe, Giorgio Valle, Hans H. de Jong & René M. Klein Lankhorst

Contributions

For full details of author contributions, please see the Supplementary Information.

Competing financial interests

The authors declare no competing financial interests.

Corresponding authors

Correspondence to: Dani Zamir or Giovanni Giuliano

The genomic data generated by the whole project are available in GenBank as accession number AEKE00000000, and the individual chromosome sequences as numbersCM001064–CM001075. TheRNA-Seq data are available in the Sequence Read Archive under accession number SRA049915, GSE33507, SRA050797 and SRA048144. Further information on data access can be found in Supplementary Information section 2.2.

Supplementary information

Main **Methods** **Accession Codes** **References** **Acknowledgements** **Author Information** Supplementary Information

PDF files

1. Supplementary Information (16.4M)
   This file contains Supplementary Methods, Supplementary Results, Supplementary Figures 1-56 and additional references –see Contents for details.

Zip files

1. Supplementary Tables (28.7M)
   This zipped excel file contains Supplementary Tables 1-78.

2. Supplementary Tables (250K)
   This zipped archive contains Supplementary HTM Tables 1-269. The tables can be opened by any web browser and contain additional text that can be visualized by placing the mouse over the image.

Nature ISSN 0028-0836 EISSN 1476-4687