

The tomato genome sequence provides insights into fleshy fruit evolution

The Tomato Genome Consortium*

Tomato (*Solanum lycopersicum*) is a major crop plant and a model system for fruit development. *Solanum* is one of the largest angiosperm genera¹ 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.

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 estimates³, 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).

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 potato⁸ 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 potato⁸ 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 genotypes⁸. 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 divergence⁹, 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

*Lists of participants and their affiliations appear at the end of the paper.

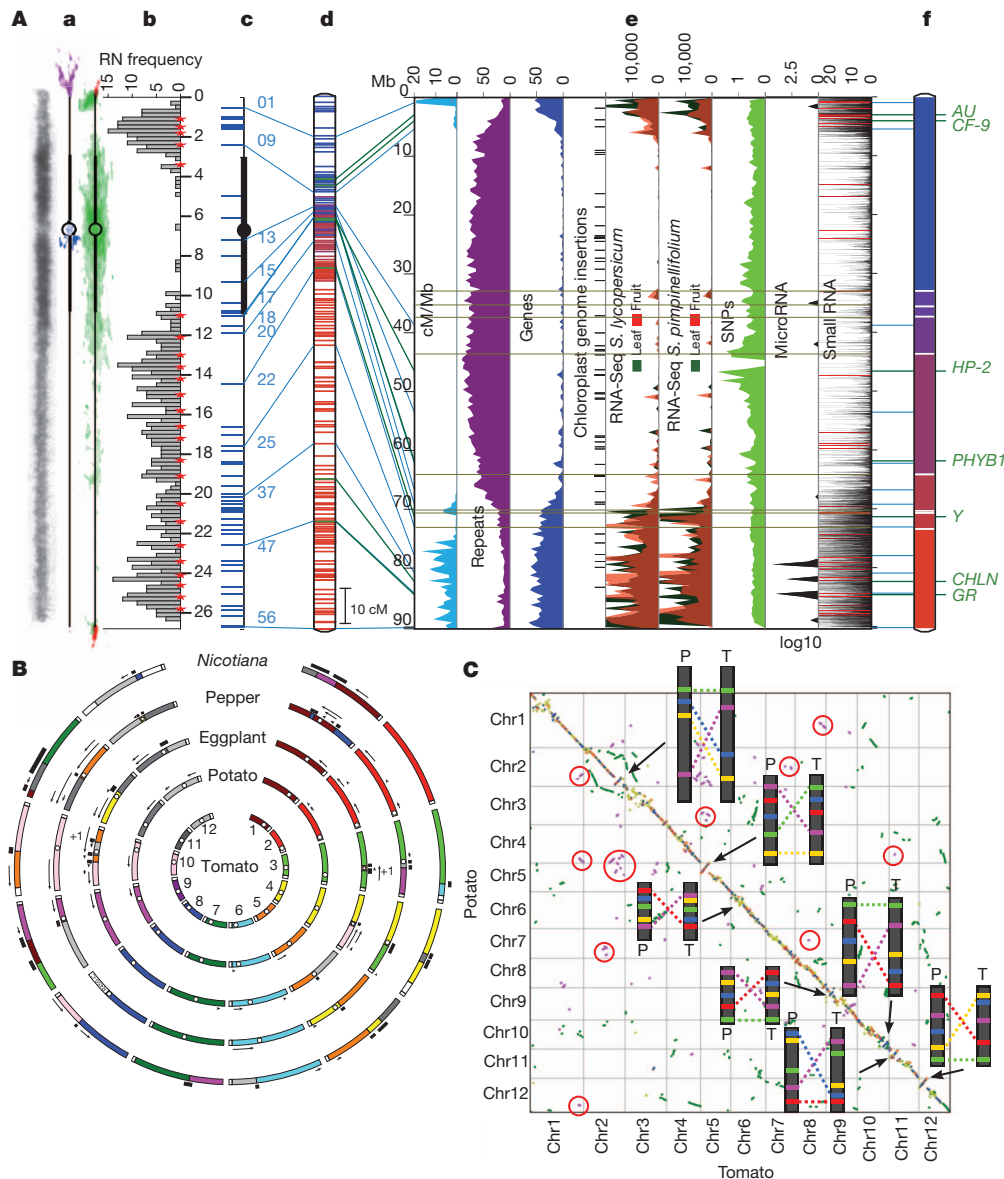


Figure 1 | Tomato genome topography and synteny. **A**, Multi-dimensional topography of tomato chromosome 1 (chromosomes 2–12 are shown in Supplementary Fig. 1). **a**, Left: contrast-reversed, 4',6-diamidino-2-phenylindole (DAPI)-stained pachytene chromosome; centre and right: FISH signals for repeat 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

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 $>5\times$ *S. pimpinellifolium* read coverage in over 90% of their coding regions

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 synteny 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 $\omega \leq 0.5$, 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).

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 breeding¹². 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 ancestor¹¹ (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).

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 years¹⁴, partially explaining the explosive radiation of angiosperm plants on Earth¹⁵.

Fleshy fruits (Supplementary Fig. 12) are an important means of attracting vertebrate frugivores for seed dispersal¹⁶. 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).

Fruit texture has profound agronomic and sensory importance and is controlled in part by cell wall structure and composition¹⁸. 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.

Similar to soybean and potato and in contrast to *Arabidopsis*, tomato sRNAs map preferentially to euchromatin (Supplementary Fig. 2). sRNAs from tomato flowers and fruits¹⁹ 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 florigen²⁰ and its antagonist *SP*, encoding the orthologue of *TFL1*. Nearly a century ago, a spontaneous mutation in *SP* spawned the 'determinate' varieties

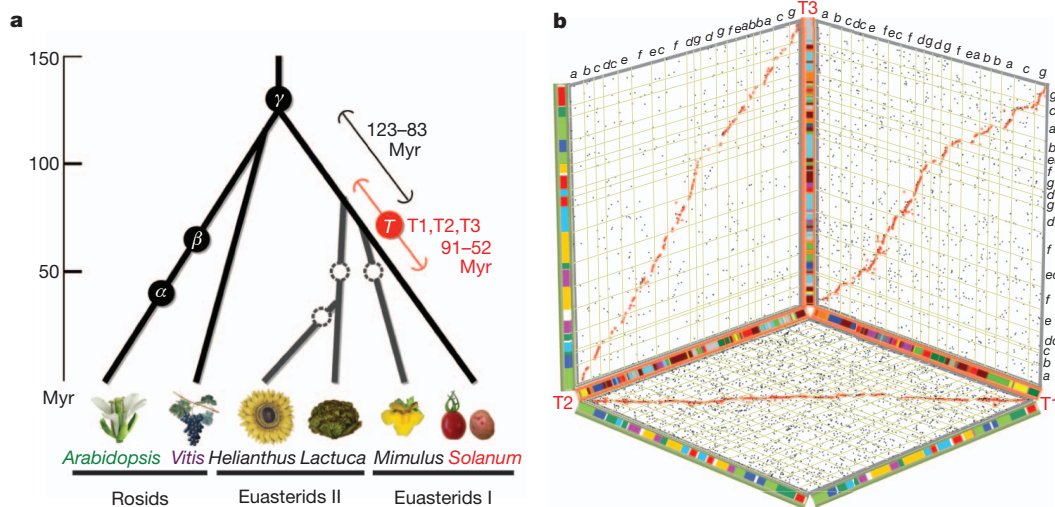


Figure 2 | The *Solanum* whole genome triplication. **a**, Speciation and polyploidization in eudicot lineages. Confirmed whole-genome duplications and triplications are shown with annotated circles, including 'T' (this paper) and previously discovered events α , β , γ ^{10,11,14}. Dashed circles represent 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 chromosomes¹⁰, also coded a–g.

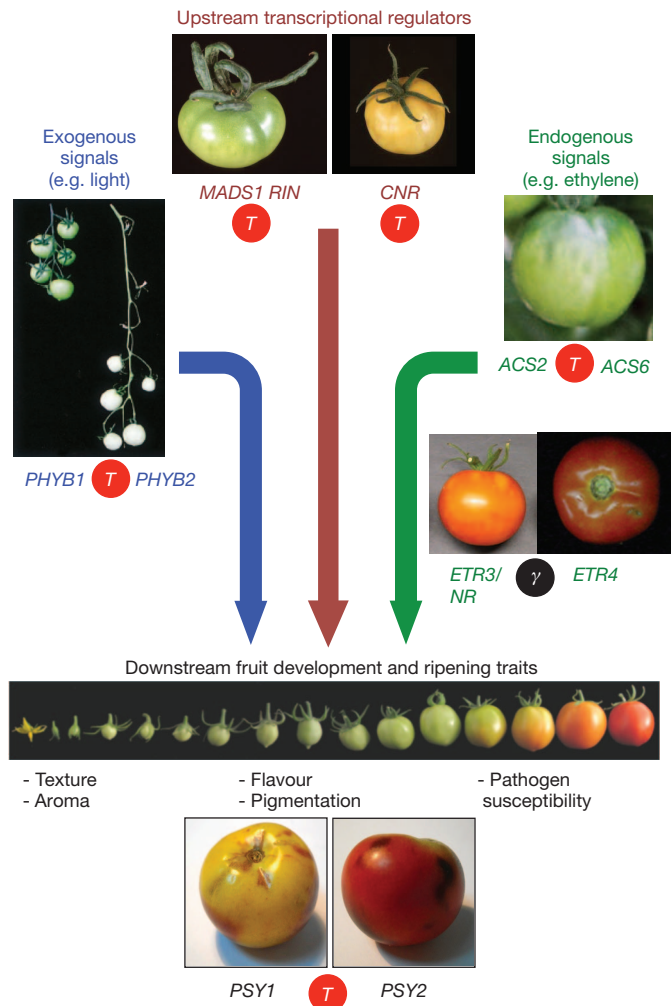


Figure 3 | Whole-genome triplications set the stage for fruit-specific gene neofunctionalization. The genes shown represent a fruit ripening control network regulated by transcription factors (*MADS-RIN*, *CNR*) necessary for production of the ripening hormone ethylene, the production of which is regulated by ACC synthase (*ACS*). Ethylene interacts with ethylene receptors (*ETRs*) 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.

that now dominate the tomato mechanical harvesting industry²¹. 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, respectively²², whereas *SP3D* in tomato, known as *SINGLE FLOWER TRUSS*, similarly controls flowering, but also drives heterosis for fruit yield in an epistatic relationship with *SP23-25*. 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

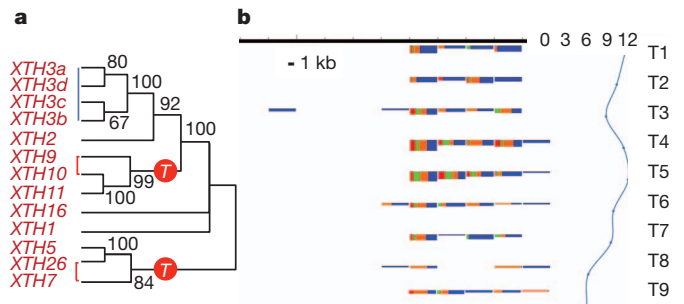


Figure 4 | The tomato genome allows systems approaches to fruit biology. a, Xyloglucan transglucosylase/hydrolases (*XTHs*) differentially expressed between mature green and ripe fruits (Supplementary Information section 5.7). These *XTH* genes and many others are expressed in ripening fruits and are 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 log₂ scale. The horizontal black line represents 1 kb of the promoter region. 0 to 12 represent arbitrary units of gene expression.

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 tomatoes²⁶. 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 SUMMARY

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 $\times 40$ 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 MCLUST was used to infer gene collinearity (Supplementary Information section 5.2).

Received 17 August 2011; accepted 3 April 2012.

1. Frodin, D. G. History and concepts of big plant genera. *Taxon* **53**, 753–776 (2004).
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).
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).
4. The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
5. Paterson, A. H. et al. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**, 551–556 (2009).
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).
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).
8. Xu, X. et al. Genome sequence and analysis of the tuber crop potato. *Nature* **475**, 189–195 (2011).
9. Swigonová, Z. et al. Close split of sorghum and maize genome progenitors. *Genome Res.* **14**, 1916–1923 (2004).
10. Jaillon, O. et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463–467 (2007).
11. Tang, H. et al. Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008).
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).
13. Ozminkowski, R. Pedigree of variety Heinz 1706. *Rep. Tomato Genet. Coop.* **54**, 26 (2004).
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).
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).
16. Howe, H. F. & Smallwood, J. Ecology of seed dispersal. *Annu. Rev. Ecol. Syst.* **13**, 201–228 (1982).
17. Klee, H. J. & Giovannoni, J. J. Genetics and control of tomato fruit ripening and quality attributes. *Annu. Rev. Genet.* **45**, 41–59 (2011).
18. Vicente, A. R., Saladié, 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).
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).
20. Corbesier, L. et al. FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science* **316**, 1030–1033 (2007).
21. Rick, C. M. The tomato. *Sci. Am.* **239**, 76–87 (1978).
22. Navarro, C. et al. Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. *Nature* **478**, 119–122 (2011).
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).
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).
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).
26. Rick, C. M. Hybridization between *Lycopersicon esculentum* and *Solanum pennellii*: phylogenetic and cytogenetic significance. *Proc. Natl Acad. Sci. USA* **46**, 78–82 (1960).
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).
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).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements 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, 2006CB8101907, 2007DFB30080) Ministry of Agriculture (‘948’ Program: 2002–25); National Natural Science Foundation (36171319); Postdoctoral Science Foundation (20070420446). European Union: FP6 Integrated Project EU-SOL PL 016214. France: Institut 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 (Agronotech 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 Bioinformática. 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 Contributions For full details of author contributions, please see the Supplementary Information.

Author Information The genomic data generated by the whole project are available in GenBank as accession number AEKE00000000, and the individual chromosome sequences as numbers CM001064–CM001075. The RNA-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. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.Z. (zamir@agri.huji.ac.il) or G.G. (giovanni.giuliano@enea.it).

The Tomato Genome Consortium

Kazusa DNA Research Institute Shusei Sato (Principal Investigator)¹, Satoshi Tabata (Principal Investigator)¹, Hideki Hirakawa¹, Erika Asamizu¹, Kenta Shirasawa¹, Sachiko Isobe¹, Takakazu Kaneko¹, Yasukazu Nakamura¹, Daisuke Shibata¹, Koh Aoki¹; **454 Life Sciences, a Roche company** Michael Egholm², James Knight²; **Amplicon Express Inc.** Robert Bogden³; **Beijing Academy of Agriculture and Forestry Sciences** Changbao Li^{4,5}; **BGI-Shenzhen** Yang Shuang⁶, Xun Xu⁶, Shengkai Pan⁶, Shifeng Cheng⁶, Xin Liu⁶, Yuanyuan Ren⁶, Jun Wang⁶; **BMR-Genomics SRL** Alessandro Albiero⁷, Francesca Dal Pero⁷, Sara Todisco⁷; **Boyce Thompson Institute for Plant Research** Joyce Van Eck⁸, Robert M. Buels⁸, Aureliano Bombarely⁸, Joseph R. Gosselin⁸, Minyun Huang⁸, Jonathan A. Leto⁸, Naama Menda⁸, Susan Strickler⁸, Linyong Mao⁸, Shan Gao⁸, Isaac Y. Tecle⁸, Thomas York⁸, Yi Zheng⁸, Julia T. Vrebalov⁸, JeMin Lee⁸, Silin Zhong⁸, Lukas A. Mueller (Principal Investigator)⁸; **Centre for BioSystems Genomics** Willem J. Stiekema⁹; **Centro Nacional de Análisis Genómico (CNAG)** Paolo Ribeca¹⁰, Tyler Alioto¹⁰; **China Agricultural University** Wencai Yang¹¹; **Chinese Academy of Agricultural Sciences** Sanwen Huang (Principal Investigator)¹², Yongchen Du (Principal Investigator)¹², Zhonghua Zhang¹², Jianchang Gao¹², Yanmei Guo¹², Xiaoxuan Wang¹², Ying Li¹², Jun He¹²; **Chinese Academy of Sciences** Chuanyou Li (Principal Investigator)¹³, Zhukuan Cheng (Principal Investigator)¹³, Jianru Zuo (Principal Investigator)¹³, Jianfeng Ren¹³, Jiahai Zhao¹³, Lihua Yan¹³, Hongling Jiang¹³, Bao Wang¹³, Hongshuang Li¹³, Zhenjun Li¹³, Fuyou Fu¹³, Bingtang Chen¹³, Bin Han (Principal Investigator)¹³, Qi Feng¹³, Danlin Fan¹³, Ying Wang (Principal Investigator)¹⁴, Hongqing Ling (Principal Investigator)¹⁵, Yongbiao Xue (Principal Investigator)¹⁶; **Cold Spring Harbor Laboratory and United States Department of Agriculture – Agricultural Research Service** Doreen Ware (Principal Investigator)¹⁷, W. Richard McCombie (Principal Investigator)¹⁷, Zachary B. Lippman (Principal Investigator)¹⁷, Jer-Ming Chia¹⁷, Ke Jiang¹⁷, Shiran Pasternak¹⁷, Laura Kelley¹⁷, Melissa Kramer¹⁷; **Colorado State University** Lorinda K. Anderson¹⁸, Song-Bin Chang¹⁹, Suzanne M. Royer¹⁸, Lindsay A. Shearer¹⁸, Stephen M. Stack (Principal Investigator)¹⁸; **Cornell University** Jocelyn K. C. Rose²⁰, Yimin Xu²⁰, Nancy Eannetta²⁰, Antonio J. Matas²⁰, Ryan McQuinn²⁰, Steven D. Tanksley (Principal Investigator)²⁰; **Genome Bioinformatics Laboratory GRIB-IMIM/UPF/CRG** Francisco Camara²¹, Roderic Guigó²¹; **Ghent University-VIB** Stephane Rombauts²², Jeffrey Fawcett²², Yves Van de Peer (Principal Investigator)²²; **Hebrew University of Jerusalem** Dani Zamir²³; **Heilongjiang Academy of Agricultural Sciences** Chunbo Liang²⁴; **Helmholtz Center for Health and Environment** Manuel Spannagl²⁵, Heidrun Gundlach²⁵, Remy Bruggmann²⁵, Klaus Mayer (Principal Investigator)²⁵; **Henan**

Agricultural University Zhiqi Jia²⁶, Huazhong Agricultural University **Junhong Zhang**²⁷, Zhibiao Ye²⁷; Imperial College London **Gerard J. Bishop (Principal Investigator)**²⁸, Sarah Butcher (Principal Investigator)²⁸, Rosa Lopez-Cobollo²⁸, Daniel Buchan²⁸, Ioannis Filippis²⁸, James Abbott²⁸; Indian Agricultural Research Institute **Rekha Dixit**²⁹, Manju Singh²⁹, Archana Singh²⁹, Jitendra Kumar Pal²⁹, Awadhesh Pandit²⁹, Pradeep Kumar Singh²⁹, Ajay Kumar Mahato²⁹, Vivek Dogra²⁹, Kishor Gaikwad²⁹, Tilak Raj Sharma²⁹, Trilochan Mohapatra²⁹, Nagendra Kumar Singh (Principal Investigator)²⁹; INRA Avignon **Mathilde Causse**³⁰; INRA Bordeaux **Christophe Rothan**³¹; INRA Toulouse **Thomas Schiex (Principal Investigator)**³², Céline Noirot³², Arnaud Bellec³³, Christophe Klopp³⁴, Corinne Delalande³⁵, Hélène Berges³³, Jérôme Mariette³⁴, Pierre Frasse³⁵, Sonia Vautrin³³; Institut National Polytechnique de Toulouse **Mohamed Zouine**³⁵, Alain Latché³⁵, Christine Rousseau³⁵, Farid Regad³⁵, Jean-Claude Pech³⁵, Murielle Philippot³⁵, Mondher Bouzayen (Principal Investigator)³⁵; Instituto de Biología Molecular y Celular de Plantas (CSIC-UPV) **Pierre Pericard**³⁶, Sonia Osorio³⁶, Asunción Fernandez del Carmen³⁶, Antonio Monforte³⁶, Antonio Granell (Principal Investigator)³⁶; Instituto de Hortofruticultura Subtropical y Mediterránea (IHSM-UMA-CSIC) **Rafael Fernandez-Muñoz**³⁷; Instituto Nacional de Tecnología Agropecuaria (IB-INTA) and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) **Mariana Conte**³⁸, Gabriel Lichtenstein³⁸, Fernando Carrari (Principal Investigator)³⁸; Italian National Res Council, Institute for Biomedical Technologies **Gianluca De Bellis (Principal Investigator)**³⁹, Fabio Fuligni³⁹, Clelia Peano³⁹; Italian National Research Council, Institute of Plant Genetics, Research Division Portici **Silvana Grandillo**⁴⁰, Pasquale Termolino⁴⁰; Italian National Agency for New technologies, Energy and Sustainable Development **Mario Pietrella**^{41,42}, Elio Fantini⁴¹, Giulia Falcone⁴¹, Alessia Fiore⁴¹, Giovanni Giuliano (Principal Investigator)⁴¹, Loredana Lopez⁴³, Paolo Facella⁴³, Gaetano Perrotta⁴³, Loretta Daddiego⁴³; James Hutton Institute **Glenn Bryan (Principal Investigator)**⁴⁴; Joint IRB-BSC program on Computational Biology **Modesto Orozco**^{45,46}, Xavier Pastor⁴⁵, David Torrents^{45,47}; Keygene N.V. **Marco G. M. van Schriek**⁴⁸, Richard M.C. Feron⁴⁸, Jan van Oeveren⁴⁸, Peter de Heer⁴⁸, Lorena daPonte⁴⁸, Saskia Jacobs-Oomen⁴⁸, Mike Carriaso⁴⁸, Marcel Prins⁴⁸, Michiel J. T. van Eijk (Principal Investigator)⁴⁸, Antoine Janssen⁴⁸, Mark J. J. van Haaren⁴⁸; Korea Research Institute of Bioscience and Biotechnology **Sung-Hwan Jo**⁴⁹, Jungeun Kim⁴⁹, Suk-Yoon Kwon⁴⁹, Sangmi Kim⁴⁹, Dal-Hoe Koo⁴⁹, Sanghyeob Lee⁴⁹, Cheol-Goo Hur⁴⁹; Life Technologies **Christopher Clouser**⁵⁰, Alain Rico⁵¹; Max Planck Institute for Plant Breeding Research **Asis Hallab**⁵², Christiane Gebhardt⁵², Kathrin Klee⁵², Anika Jöcker⁵², Jens Warfsmann⁵², Ulrike Göbel⁵²; Meiji University **Shingo Kawamura**⁵³, Kentaro Yano⁵³; Montana State University **Jamie D. Sherman**⁵⁴, NARO Institute of Vegetable and Tea Science **Hiroyuki Fukuoka (Principal Investigator)**⁵⁵, Satomi Negoro⁵⁵; National Institute of Plant Genome Research **Sarita Bhutty**⁵⁶, Parul Chowdhury⁵⁶, Debasis Chattopadhyay (Principal Investigator)⁵⁶; Plant Research International **Erwin Datema**^{48,57}, Sandra Smit⁵⁷, Elio G. W. M. Schiljen⁵⁷, Jose van de Belt⁵⁷, Jan C. van Haarst⁵⁷, Sander A. Peters⁵⁷, Marjo J. van Staveren⁵⁷, Marleen H. C. Henkens⁵⁷, Paul J. W. Mooyman⁵⁷, Tamara Hesselink⁵⁷, Roeland C. H. J. van Ham (Principal Investigator)^{48,57}; Qingdao Agricultural University **Guoyong Jiang**⁵⁸; Roche Applied Science **Marcus Droege**⁵⁹; Seoul National University **Doil Choi (Principal Investigator)**⁶⁰, Byung-Cheol Kang⁶⁰, Byung Dong Kim⁶⁰, Minkyu Park⁶⁰, Seungill Kim⁶⁰, Seon-In Yeom⁶⁰, Yong-Hwan Lee⁶¹, Yang-Do Choi⁶²; Shandong Academy of Agricultural Sciences **Guangcun Li**⁶³, Jianwei Gao⁶⁴; Sichuan University **Yongsheng Liu**⁶⁵, Shengxiong Huang⁶⁵; Sistemas Genómicos **Victoria Fernandez-Pedrosa**⁶⁶, Carmen Collado⁶⁶, Sheila Zuñiga⁶⁶; South China Agricultural University **Guoping Wang**⁶⁷; Syngenta Biotechnology **Rebecca Cade**⁶⁸, Robert A. Dietrich⁶⁸; The Genome Analysis Centre **Jane Rogers (Principal Investigator)**⁶⁹; The Natural History Museum **Sandra Knapp**⁷⁰; United States Department of Agriculture – Agricultural Research Service, Robert W. Holley Center and Boyce Thompson Institute for Plant Research **Zhangjun Fei (Principal Investigator)**⁷¹, Ruth A. White^{8,71}, Theodore W. Thannhauser^{8,71}, James J. Giovannoni (Principal Investigator)^{8,20,71}; Universidad de Malaga-Consejo Superior de Investigaciones Científicas **Miguel Angel Botella**⁷², Louise Gilbert⁷²; Universitat Pompeu Fabra **Ramon Gonzalez**⁷³; University of Arizona **Jose Luis Goicoechea**⁷⁴, Yeisoo Yu⁷⁴, David Kudrna⁷⁴, Kristi Collura⁷⁴, Marina Wissotski⁷⁴, Rod Wing (Principal Investigator)⁷⁴; University of Bonn **Heiko Schoof (Principal Investigator)**⁷⁵; University of Delaware **Blake C. Meyers (Principal Investigator)**⁷⁶, Aishwarya Bala Gurazada⁷⁶, Pamela J. Green⁷⁶; University of Delhi South Campus **Saloni Mathur**⁷⁷, Shailendra Vyas⁷⁷, Amolkumar A. Solanke⁷⁷, Rahul Kumar⁷⁷, Vikrant Gupta⁷⁷, Arun K. Sharma⁷⁷, Paramjit Khurana⁷⁷, Jitendra P. Khurana (Principal Investigator)⁷⁷, Akhilesh K. Tyagi (Principal Investigator)⁷⁷; University of East Anglia, School of Biological Sciences **Tamas Dalmay (Principal Investigator)**⁷⁸; University of East Anglia, School of Computing Sciences **Irina Mohorianu**⁷⁹; University of Florida **Brandon Watts**⁸⁰, Srikanth Chamala⁸⁰, W. Brad Barbazuk⁸⁰; University of Georgia **Jingping Li**⁸¹, Hui Guo⁸¹, Tae-Ho Lee⁸¹, Yupeng Wang⁸¹, Dong Zhang⁸¹, Andrew H. Paterson (Principal Investigator)⁸¹, Xiyan Wang (Principal Investigator)^{81,82}, Haibao Tang^{81,83}; University of Naples “Federico II” **Amalia Barone**⁸⁴, Maria Luisa Chiusano⁸⁴, Maria Raffaella Ercolano⁸⁴, Nunzio D’Agostino⁸⁴, Miriam Di Filippo⁸⁴, Alessandra Traini⁸⁴, Walter Sanseverino⁸⁴, Luigi Frusciant (Principal Investigator)⁸⁴; University of Nottingham **Graham B. Seymour (Principal Investigator)**⁸⁵; University of Oklahoma **Mounir Elharam**⁸⁶, Ying Fu⁸⁶, Axin Hua⁸⁶, Steven Kenton⁸⁶, Jennifer Lewis⁸⁶, Shaoping Lin⁸⁶, Fares Najja⁸⁶, Hongshing Lai⁸⁶, Baifang Qin⁸⁶, Chunmei Qu⁸⁶, Ruihua Shi⁸⁶, Douglas White⁸⁶, James White⁸⁶, Yanbo Xing⁸⁶, Keqin Yang⁸⁶, Jing Yi⁸⁶, Ziyun Yao⁸⁶, Liping Zhou⁸⁶, Bruce A. Roe (Principal Investigator)⁸⁶; University of Padua **Alessandro Vezzi**⁸⁷, Michela D’Angelo⁸⁷, Rosanna Zimbello⁸⁷, Riccardo Schiavon⁸⁷, Elisa Caniato⁸⁷, Chiara Rigobello⁸⁷, Davide Campagna⁸⁷, Nicola Vitolo⁸⁷, Giorgio Valle (Principal Investigator)⁸⁷; University of Tennessee Health Science Center **David R. Nelson**⁸⁸; University of Udine **Emanuele De Paoli**⁸⁹; Wageningen University **Dora Szinay**^{90,91}, Hans H. de Jong (Principal Investigator)⁹⁰, Yuling Bai⁹¹, Richard G. F. Visser⁹¹, René M. Klein Lankhorst (Principal Investigator)⁹²; Wellcome Trust Sanger Institute **Helen**

Beasley⁹³, Karen McLaren⁹³, Christine Nicholson⁹³, Claire Riddle⁹³; Ylichron Srl **Giulio Gianese**⁹⁴

¹Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan. ²454 Life Sciences, a Roche company, 15 Commercial Street, Branford, Connecticut 06405, USA. ³Amplicon Express Inc, 2345 Hopkins Court, Pullman, Washington 99163, USA. ⁴Beijing Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China. ⁵National Center for Gene Research, Chinese Academy of Sciences, Shanghai 200233, China. ⁶BGI-Shenzhen, Shenzhen 518083, China. ⁷BMR-Genomics Srl, via Redipuglia 21/A, 35131 Padova, Italy. ⁸Boyce Thompson Institute for Plant Research, Tower Road, Cornell University campus, Ithaca, New York 14853, USA. ⁹Centre for BioSystems Genomics, PO Box 98, 6700 AB Wageningen, The Netherlands. ¹⁰Centro Nacional de Análisis Genómico (CNAG), C/ Baldori Reixac 4, Torre I, 08028 Barcelona, Spain. ¹¹Department of Vegetable Science, College of Agronomy and Biotechnology, China Agricultural University, No. 2 Yuanmingyuan Xi Lu, Haidian District, Beijing 100193, China. ¹²Key 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. ¹³State 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. ¹⁴Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China. ¹⁵State 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 100080, China. ¹⁶Laboratory 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. ¹⁷Cold Spring Harbor Laboratory, One Bungtown Road, Cold Spring Harbor, New York 11724, USA. ¹⁸Department of Biology, Colorado State University, Fort Collins, Colorado 80523, USA. ¹⁹Department of Agronomy, National Taiwan University, Taipei 107, Taiwan. ²⁰Department of Plant Biology, Cornell University, Ithaca, New York 14853, USA. ²¹Genome Bioinformatics Laboratory, Center for Genomic Regulation (CRG), University Pompeu Fabra, Barcelona 08003, Spain. ²²Department of Plant Systems Biology, VIB, Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, 9052 Gent, Belgium. ²³Faculty of Agriculture, The Hebrew University of Jerusalem, PO Box 12, Rehovot 76100, Israel. ²⁴Institute of Industrial Crops, Heilongjiang Academy of Agricultural Sciences, Harbin 150086, China. ²⁵Institute for Bioinformatics and Systems Biology (MIPS), Helmholtz Center for Health and Environment, Ingolstädter Landstr. 1, D-85764 Neuherberg, Germany. ²⁶College of Horticulture, Henan Agricultural University, Zhengzhou 450002, China. ²⁷National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China. ²⁸Department of Life Sciences, Imperial College London, London SW7 1AZ, UK. ²⁹NRC on Plant Biotechnology, Indian Agricultural Research Institute, New Delhi 110 012, India. ³⁰INRA, UR1052 Génétique et amélioration des fruits et légumes, BP 94, 84143 Monfavet Cedex, France. ³¹INRA, Biologie du Fruit et Pathologie, 71 rue E. Bourleaux, 33883 Villenave d’Ornon, France. ³²Unité de Biométrie et d’Intelligence Artificielle UR 875, INRA, F-31320 Castanet-Tolosan, France. ³³INRA-CNRGV BP52627, 31326 Castanet-Tolosan, France. ³⁴Plateforme bioinformatique Genotoul, UR875 Biométrie et Intelligence Artificielle, INRA, 31326 Castanet-Tolosan, France. ³⁵Institut National Polytechnique de Toulouse – ENSAT, Université de Toulouse, Avenue de l’Agrobiopole BP 32607, 31326 Castanet-Tolosan, France. ³⁶Instituto de Biología Molecular y Celular de Plantas (CSIC-UPV), Ciudad Politécnica de la Innovación, escalera 8E, Ingeniero Fausto Elio s/n, 46022 Valencia, Spain. ³⁷Instituto de Hortofruticultura Subtropical y Mediterránea “La Mayora”, Universidad de Malaga – Consejo Superior de Investigaciones Científicas (IHSM-UMA-CSIC), 29750 Algarrobo-Costa (Málaga), Spain. ³⁸Instituto de Biotecnología, PO Box 25, B1712WAA Castelar, Argentina. ³⁹Institute for Biomedical Technologies, National Research Council of Italy, Via F. Cervi 93, 20090 Segrate (Milano), Italy. ⁴⁰Institute of Plant Genetics, Research Division Portici, National Research Council of Italy, Via Università 133, 80055 Portici (Naples), Italy. ⁴¹ENEA, Casaccia Research Center, Via Anguillarese 301, 00123 Roma, Italy. ⁴²Scuola Superiore Sant’Anna, Piazza Martiri della Libertà 33 - 56127 Pisa, Italy. ⁴³ENEA, Trisaia Research Center, S.S. Ionica - Km 419.5, 75026 Rotondella (Matera), Italy. ⁴⁴James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK. ⁴⁵Barcelona Supercomputing Center, Nexus II Building, c/ Jordi Girona, 29, 08034 Barcelona, Spain. ⁴⁶Institute of Research in Biomedicine, c/ Josep Samitier 1-5, 08028 Barcelona, Spain. ⁴⁷ICREA, Pg Lluís Companys, 23, 08010 Barcelona, Spain. ⁴⁸Keygene N.V., Agro Business Park 90, 6708 PW Wageningen, The Netherlands. ⁴⁹Plant Systems Engineering Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, 305-806, Republic of Korea. ⁵⁰Life Technologies, 500 Cummings Center, Beverly, Massachusetts 01915, USA. ⁵¹Life Technologies, 25 avenue de la Baltique, BP 96, 91943 Courtabouet Cedex 3, France. ⁵²Max Planck Institute for Plant Breeding Research, Carl von Linné Weg 10, 50829 Cologne, Germany. ⁵³School of Agriculture, Meiji University, 1-1-1 Higashi-Mita, Tama-ku, Kawasaki-shi, Kanagawa 214-8571, Japan. ⁵⁴Department of Plant Science and Plant Pathology, Montana State University, Bozeman, Montana 59717, USA. ⁵⁵NARO Institute of Vegetable and Tea Science, 360 Kusawa, Ano, Tsu, Mie 514-2392, Japan. ⁵⁶National Institute of Plant Genome Research, New Delhi 110 067, India. ⁵⁷Plant Research International, Business Unit Bioscience, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands. ⁵⁸Institute of Plant Genetic Engineering, Qingdao Agricultural University, Qingdao 266109, China. ⁵⁹Roche Applied Science, D-82377 Penzberg, Germany. ⁶⁰Seoul National University, Department of Plant Science and Plant Genomics and Breeding Institute, Seoul 151-921, Republic of Korea. ⁶¹Seoul National University, Department of Agricultural Biotechnology, Seoul 151-921, Republic of Korea. ⁶²Seoul National University, Crop Functional Genomics Center, College of Agriculture and Life Sciences, Seoul 151-921, Republic of Korea. ⁶³High-Tech Research Center, Shandong Academy of Agricultural Sciences, Jinan, 250000 Shandong, China. ⁶⁴Institute of Vegetables, Shandong Academy of Agricultural Sciences, Jinan, 250100 Shandong,

China. ⁶⁵School of Life Sciences, Sichuan University, Chengdu, Sichuan 610064, China. ⁶⁶Sistemas Genómicos, Parque Tecnológico de Valencia, Ronda G. Marconi, 6, 46980 Paterna (Valencia), Spain. ⁶⁷College of Horticulture, South China Agricultural University, 510642 Guangzhou, China. ⁶⁸Syngenta Biotechnology, Inc. 3054 East Cornwallis Road, Research Triangle Park, North Carolina 27709 Durham, USA. ⁶⁹Norwich Research Park, Norwich NR4 7UH, UK. ⁷⁰Department of Botany, The Natural History Museum, Cromwell Road, London SW7 5BD, UK. ⁷¹United States Department of Agriculture - Agricultural Research Service, Robert W. Holley Center, Tower Road, Cornell University campus, Ithaca, New York 14853, USA. ⁷²Instituto de Hortofruticultura Subtropical y Mediterránea. Departamento de Biología Molecular y Bioquímica, 29071 Málaga, Spain. ⁷³Centre de Regulació Genòmica, Universitat Pompeu Fabra, Dr Aiguader, 88, E-08003 Barcelona, Spain. ⁷⁴Arizona Genomics Institute, BIO-5 Institute for Collaborative Research, School of Plant Sciences, Thomas W. Keating Building, 1657 E. Helen Street, Tucson, Arizona 85721, USA. ⁷⁵Crop Bioinformatics, Institute of Crop Science and Resource Conservation, University of Bonn, 53115 Bonn, Germany. ⁷⁶Department of Plant and Soil Sciences, and Delaware Biotechnology Institute, University of Delaware, Newark, Delaware 19711, USA. ⁷⁷Interdisciplinary Centre for Plant Genomics and Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi 110 021, India. ⁷⁸University of East Anglia, BIO, Norwich NR4 7TJ, UK. ⁷⁹University of East Anglia, CMP, Norwich NR4 7TJ, UK. ⁸⁰Department of Biology and the UF Genetics Institute, Cancer and Genetics Research Complex 2033 Mowry Road, PO Box 103610, Gainesville, Florida 32610, USA. ⁸¹Plant

Genome Mapping Laboratory, 111 Riverbend Road, University of Georgia, Athens, Georgia 30602, USA. ⁸²Center for Genomics and Computational Biology, School of Life Sciences, and School of Sciences, Hebei United University, Tangshan, Hebei 063000, China. ⁸³J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, Maryland 20850, USA. ⁸⁴University of Naples "Federico II" Department of Soil, Plant, Environmental and Animal Production Sciences, Via Università, 100, 80055 Portici (Naples), Italy. ⁸⁵Division of Plant and Crop Sciences, University of Nottingham, Sutton Bonington, Loughborough LE12 5RD, UK. ⁸⁶Department of Chemistry and Biochemistry, Stephenson Research and Technology Center, University of Oklahoma, Norman, Oklahoma 73019, USA. ⁸⁷CRIBI, University of Padua, via Ugo Bassi 58/B, 35131 Padova, Italy. ⁸⁸Department of Microbiology, Immunology and Biochemistry, University of Tennessee Health Science Center, Memphis, Tennessee 38163, USA. ⁸⁹Department of Agriculture and Environmental Sciences, University of Udine, via delle Scienze 208, 33100, Udine, Italy. ⁹⁰Wageningen University, Laboratory of Genetics, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands. ⁹¹Wageningen University, Laboratory of Plant Breeding, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands. ⁹²Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands. ⁹³Wellcome Trust Sanger Institute Hinxton, Cambridge CB10 1SA, UK. ⁹⁴Ylichron Srl, Casaccia Research Center, Via Anguillarese 301, 00123 Roma, Italy. †Present address: Plant Engineering Research Institute, Sejong University, Seoul, 143-747, Republic of Korea.