Document downloaded from:

http://hdl.handle.net/10251/80510

This paper must be cited as:

Sato, S.; Tabata, S.; Hirakawa, H.; Asamizu, E.; Shirasawa, K.; Isobe, S.; Kaneko, T.... (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature. 485(7400):635-641. doi:10.1038/nature11119

The final publication is available at

http://dx.doi.org/10.1038/nature11119

Copyright MOST/2006AA10A116

Additional Information

Europe PMC Funders Group
Author Manuscript
Nature. Author manuscript; available in PMC 2012 November 30.

Published in final edited form as:
Nature. ; 485(7400): 635–641. doi:10.1038/nature11119.

The tomato genome sequence provides insights into fleshy fruit evolution

The Tomato Genome Consortium (TGC)

[1] Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan [2] 454 Life Sciences, a Roche company, 15 Commercial Street, Branford, CT 06405, USA [3] Amplicon Express Inc., 2345 Hopkins Court, Pullman, WA 99163, USA [4] Beijing Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China [5] BGI-Shenzhen, Shenzhen 518083, China [6] BMR-Genomics SrL, via Redipuglia 21/A, 35131 Padova, Italy [7] Boyce Thompson Institute for Plant Research, Tower Road, Cornell University campus, Ithaca, NY 14853, USA [8] Centre for BioSystems Genomics, PO Box 98, 6700 AB Wageningen, The Netherlands [9] Centro Nacional de Análisis Genómico (CNAG) and National Bioinformatics Institute, C/ Baldiri Reixac 4, Torre I, 08028 Barcelona, Spain [10] Genome Bioinformatics Laboratory Center for Genomic Regulation, Dr Aiguader, 88, E-08003 Barcelona, Spain [11] Department of Vegetable Science, College of Agronomy and Biotechnology, China Agricultural University, No. 2 Yuanmingyuan Xi Lu, Haidian District, Beijing 100193, China [12] 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 [13] 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 [14] National Center for Gene Research, Chinese Academy of Sciences, Shanghai 200233, China [15] Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China [16] State Key Laboratory of Plant

Correspondence should be addressed to: Dani Zamir (zamir@agri.huji.ac.il) and Giovanni Giuliano (giovanni.giuliano@enea.it)..

AUTHOR CONTRIBUTIONS

AAL, ABA, ABE, ABG, ABO, AFI, AGR, AHA, AHU, AJA, AJM, AJO, AKM, AKS, AKT, AMO, APA, ARI, ASI, ATR, AUS, AVE, BAR, BAW, BCK, BCM, BDK, BFQ, BMW, BTC, CBL, CCL, CCO, CDE, CGE, CGH, CHR, CKL, CLI, CLR, CMQ, CNI, CNO, CPE, CRI, CYL, DCA, DCH, DDW, DHK, DIC, DKU, DLF, DOZ, DRN, DSH, DSZ, DTO, DWB, DZA, EAS, ECA, EDA, EDP, EFA, EGS, FCA, FDP, FEA, FFU, FRE, FYF, FYU, FZN, GBS, GDB, GEB, GFA, GGI, GHE, GIG, GJB, GLI, GPE, GUL, GVA, GWA, HBE, HBT, HDJ, HEB, HFU, HGU, HHI, HLA, HLJ, HSC, HSL, IFI, IMO, IYT, JAL, JCA, JDS, JDW, JEK, JFA, JFR, JGA, JHE, JHZ, JJG, JKN, JKP, JKR, JLE, JLG, JLI, JLL, JMA, JMC, JPK, JRG, JRO, JTV, JUW, JVB, JVE, JVH, JVO, JWA, JYI, JZH, KAO, KCO, KGA, KJI, KKL, KMA, KML, KQW, KSH, KYA, LAM, LAS, LDA, LDP, LGE, LHY, LKA, LLO, LMA, LOG, LPX, MAB, MAF, MAV, MBO, MCC, MCO, MDA, MDB, MDF, MDR, MEG, MEL, MHM, MHU, MKP, MLC, MOR, MPH, MPI, MRE, MSI, MSP, MVS, MWI, MZO, NDA, NEA, NKS, NME, NVI, PCH, PDH, PFA, PFR, PJM, PKH, PKS, PPE, PRI, PTE, QFE, RAD, RAW, RBO, RBR, RCR, RDI, RFE, RGO, RGU, RHS, RKU, RLC, RMB, RMC, RMF, RMK, RSC, RVH, RWI, RZI, SAP, SBC, SBH, SCH, SDT, SGA, SGR, SHH, SHJ, SHL, SHU, SIK, SIS, SIY, SJO, SKA, SKE, SKP, SMA, SMK, SMR, SMS, SNE, SOS, SPA, SPL, SSA, SSM, SST, STO, SUR, SVA, SVY, SXH, SYK, SZH, SZU, TAL, TDA, THE, THL, TKA, TMO, TRS, TSC, TWT, TYO, UGO, VDO, VGU, VYF, WBB, WRM, WSA, WYA, XLI, XPA, XWA, XXU, XXW, YBA, YBX, YDU, YGU, YHL, YLI, YNA, YPW, YRE, YSH, YXU, YYU, YZH, ZBL, ZFE, ZJI, ZJL, ZYA, ZYE, ZZH were involved in data generation and/or analysis. AAL, AGR, AHP, AKT, AVE, BAR, CCO, CYL, DCH, DIC, DRN, DSZ, DWA, DZA, ECA, EDA, EDP, EGS, FEA, GBS, GEB, GHE, GIG, GJB, GVA, HDJ, HHI, HSC, IFI, JFR, JJG, JKR, JLE, JLG, JMC, JPK, JTV, JVE, KJI, KMA, LAM, LKA, MAB, MBO, MCO, MKP, MLC, MPI, MRE, MSP, MVE, MZO, NKS, RAW, RMK, RVH, RWI, SAP, SDT, SGR, SIK, SIY, SKN, SMR, SMS, SPA, SSA, SSM, TDA, TMO, TRS, TSC, WBB, YVP, YXU, ZBL, ZFE wrote the manuscript. AGR, AHU, AJA, AKS, AKT, ALA, AVE, BAR, BCM, BHA, CGH, CRO, CYL, DCH, DIC, DTO, DWA, DZA, FEA, FZN, GBS, GDB, GEB, GIG, GJB, GVA, GYJ, HDJ, HFU, HQL, HSC, JCP, JDW, JJG, JPK, JRO, JRZ, JVE, JWG, KMA, LAM, LFR, MAB, MBO, MCA, MPR, MSP, MVE, MVH, MVS, PJG, PKH, RAD, RGU, RGV, RHS, RMK, RVH, RWI, SAB, SDT, SHU, SMR, SMS, SPL, SSA, STA, TDA, TSC, VYF, WJS, WRM, XPA, YBX, YDC, YDU, YOX, YSL, YVP, YWA, ZBL, ZKC designed experiments, supervised data generation/analysis and managed subprojects/tasks.

The authors declare no competing financial interests.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Cell and Chromosome Engineering and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China [17] 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 [18] Cold Spring Harbor Laboratory, One Bungtown Road, Cold Spring Harbor, NY 11724, USA [19] Department of Biology, Colorado State University, Fort Collins, CO 80523, USA [20] Department of Agronomy, National Taiwan University, Taipei, Taiwan [21] Department of Plant Biology, Cornell University, Ithaca, NY 14853, USA [22] Genome Bioinformatics Laboratory; Center for Genomic Regulation (CRG), University Pompeu Fabra, Barcelona, Catalonia, 08003, Spain [23] Department of Plant Systems Biology, VIB; Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, 9052 Gent, Belgium [24] Faculty of Agriculture, The Hebrew University of Jerusalem, PO Box 12, Rehovot 76100, Israel [25] Institute of Industrial Crops, Heilongjiang Academy of Agricultural Sciences, Harbin 150086, China [26] Institute for Bioinformatics and Systems Biology (MIPS), Helmholtz Center for Health and Environment, Ingolstädter Landstr. 1, D-85764 Neuherberg, Germany [27] College of Horticulture, Henan Agricultural University, Zhengzhou 450002, China [28] National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China. [29] Department of Life Sciences, Imperial College London, London SW7 1AZ, UK [30] NRC on Plant Biotechnology, Indian Agricultural Research Institute, New Delhi, 110 012, India [31] INRA, UR1052 Génétique et amélioration des fruits et légumes, BP 94, 84143 Monfavet CEDEX, France [32] INRA, Biologie du Fruit et Pathologie, 71 rue E. Bourleaux, 33883 Villenave d'Ornon, France [33] Unité de Biométrie et d'Intelligence Artificielle UR 875, INRA, F-31320, Castanet Tolosan, France [34] INRA-CNRGV BP52627 31326 Castanet-Tolosan, France [35] Plateforme bioinformatique Genotoul, UR875 Biométrie et Intelligence Artificielle, INRA, 31326 Castanet-Tolosan, France [36] ENSAT, Avenue de l'Agrobiopole BP 32607 31326 Castanet-Tolosan, France [37] 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 [38] Instituto de Hortofruticultura Subtropical y Mediterránea "La Mayora", Universidad de Malaga - Consejo Superior de Investigaciones Cientificas (IHSM-UMA-CSIC), 29750 Algarrobo-Costa (Málaga), Spain [39] Instituto Nacional de Tecnología Agropecuaría (IB-INTA) and Consejo Nacionalde Investigaciones Científicas y Técnicas (CONICET):; Instituto de Biotecnología, PO Box 25, B1712WAA Castelar, Argentina [40] Institute for Biomedical Technologies, National Research Council of Italy, Via F. Cervi 93, 20090 Segrate (Milano), Italy [41] Institute of Plant Genetics, Research Division Portici, National Research Council of Italy, Via Università 133, 80055 Portici, Italy [42] Italian National Agency for New technologies, Energy and Sustainable Development:; ENEA, Casaccia Research Center, Via Anguillarese 301, 00123 Roma, Italy [43] Scuola Superiore Sant'Anna, Piazza Martiri della Libertà 33 - 56127 Pisa, Italy [44] ENEA, Trisaia Research Center, S.S. Ionica - Km 419.5, 75026 Rotondella (Matera), Italy [45] James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK [46] Barcelona Supercomputing Center, Nexus II Building, c/ Jordi Girona, 29, 08034 Barcelona, Spain [47] ICREA, Pg Lluís Companys, 23, 08010, Barcelona, Spain [48] Keygene N.V., Agro Business Park 90, 6708 PW Wageningen, The Netherlands [49] Plant Systems Engineering Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, 305-806, Republic of Korea [50] Life Technologies, 500 Cummings Center, Beverly, MA 01915, U.S.A [51] Life Technologies, 25 avenue de la Baltique, BP 96, 91943 Courtaboeuf Cedex 3, France [52] Max Planck Institute for Plant Breeding Research, Carl von Linné Weg 10, 50829 Cologne, Germany [53] School of Agriculture, Meiji University, 1-1-1 Higashi-Mita, Tama-ku, Kawasaki-shi, Kanagawa 214-8571, Japan [54] Department of Plant Science and Plant Pathology, Montana State University, Bozeman, MT 59717, USA [55] NARO Institute of Vegetable and Tea Science, 360 Kusawa, Ano, Tsu, Mie 514-2392, Japan [56] National Institute of Plant Genome Research, New Delhi, 110 067, India [57] Plant Research International, Business Unit Bioscience, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands [58] Institute of Plant Genetic

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Engineering, Qingdao Agricultural University, Qingdao 266109, China [59] Roche Applied Science, D-82377 Penzberg, Germany [60] Seoul National University, Department of Plant Science and Plant Genomics and Breeding Institute, Seoul, 151-921, Republic of Korea [61] Seoul National University, Department of Agricultural Biotechnology, Seoul, 151-921, Republic of Korea [62] Seoul National University, Crop Functional Genomics Center, College of Agriculture and Life Sciences, Seoul, 151-921, Republic of Korea [63] High-tech Research center, Shandong Academy of Agricultural Sciences, Jinan 250000, China [64] Institute of Vegetables, Shandong Academy of Agricultural Sciences, Jinan, Shandong, 250100, China [65] School of life sciences, Sichuan University, Chengdu, Sichuan, 610064, China [66] Sistemas Genomicos, Parque Tecnológico de Valencia, Ronda G. Marconi, 6,46980 Paterna (Valencia), Spain [67] College of Horticulture, South China Agricultural University, 510642 Guangzhou, China. [68] Syngenta Biotechnology, Inc. 3054 East Cornwallis Rd, Research Triangle Park, NC 27709 Durham, USA [69] Norwich Research Park, Norwich NR4 7UH, UK [70] Department of Botany, The Natural History Museum, Cromwell Road, London SW7 5BD, United Kingdom [71] Robert W. Holley Center and Boyce Thompson Institute for Plant Research:; United States Department of Agriculture - Agricultural Research Service, Robert W. Holley Center, Tower Road, Cornell University campus, Ithaca NY 14853, USA [72] Universidad de Malaga-Consejo Superior de Investigaciones Cientificas:; Instituto de Hortofruticultura Subtropical y Mediterranea. Departamento de Biologia Molecular y Bioquimica, 29071 Málaga, Spain [73] Centre de Regulacio Genomica, Universitat Pompeu Fabra, Dr Aiguader, 88, E-08003 Barcelona, Spain [74] Arizona Genomics Institute, BIO-5 Institute for Collaborative Research, School of Plant Sciences, Thomas W. Keating Building, 1657 E. Helen Street, Tucson AZ 85721, USA [75] Crop Bioinformatics, Institute of Crop Science and Resource Conservation, University of Bonn, 53115 Bonn, Germany [76] Department of Plant & Soil Sciences, and Delaware Biotechnology Institute, University of Delaware, Newark, Delaware 19711, USA [77] Interdisciplinary Centre for Plant Genomics and Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi, 110 021, India [78] University of East Anglia, School of Biological Sciences:; University of East Anglia, BIO, Norwich NR4 7TJ, UK [79] University of East Anglia, School of Computing Sciences:; University of East Anglia, CMP, Norwich NR4 7TJ, UK [80] Department of Biology and the UF Genetics Institute, Cancer & Genetics Research Complex 2033 Mowry Road, PO Box 103610, Gainesville FL, USA [81] Plant Genome Mapping Laboratory, 111 Riverbend Road, University of Georgia, Athens, GA 30602, USA [82] Center for Genomics and Computational Biology, School of Life Sciences, and School of Sciences, Hebei United University, Tangshan, Hebei 063000, China [83] J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA [84] University of Naples “Federico II” Department of Soil, Plant, Environmental and Animal Production Sciences, Via Universita’, 100, 80055 Portici (Naples), Italy [85] Division of Plant and Crop Sciences, University of Nottingham, Sutton Bonington, Loughborough LE12 5RD, UK [86] Department of Chemistry and Biochemistry, Stephenson Research and Technology Center, University of Oklahoma, Norman, OK 73019, USA [87] CRIBI, University of Padua, via Ugo Bassi 58/B, 35131 Padova, Italy [88] Department of Microbiology, Immunology and Biochemistry, University of Tennessee Health Science Center, Memphis, USA [89] Department of Agriculture and Environmental Sciences, University of Udine, via delle Scienze 208, 33100, Udine, Italy [90] Wageningen University, Laboratory of Genetics, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands [91] Wageningen University, Laboratory of Plant Breeding, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands [92] Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands [93] Wellcome Trust Sanger Institute Hinxton, Cambridge CB10 1SA, UK [94] Ylichron SrL, Casaccia Research Center, Via Anguillarese 301, 00123 Roma, Italy

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Introductory Paragraph

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. We present a high quality genome sequence of domesticated tomato, a draft sequence of its closest wild relative, *S. pimpinellifolium*[2], and compare them to each other and to potato (*S. tuberosum*). The two tomato genomes show only 0.6% nucleotide divergence and signs of recent admixture, but show >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.

Main Text

The genome of the inbred tomato cultivar 'Heinz 1706' was sequenced and assembled using a combination of Sanger and "next generation" technologies (Supplementary Section 1). The predicted genome size is ~900 Mb, consistent with prior 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; Supplementary Fig. 1). Base accuracy is approximately one substitution error per 29.4 kb and one indel error per 6.4 kb. The scaffolds were linked with two BAC-based physical maps and anchored/oriented using a high-density genetic map, introgression line mapping and BAC fluorescence *in situ* hybridisation (FISH).

The genome of *S. pimpinellifolium* (accession LA1589) was sequenced and assembled *de novo* using Illumina short reads, yielding a 739 Mb draft genome (Supplementary Section 3). Estimated divergence between the wild and domesticated genomes is 0.6% (5.4M SNPs distributed along the chromosomes (Fig. 1A, 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, Supplementary Fig. 1). Substantially higher densities of recombination, genes and transcripts are observed in euchromatin, while chloroplast insertions (Supplementary Sections 1.22-1.23) and conserved miRNA genes (Supplementary 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 LTR retrotransposons with older average insertion ages (2.8 versus 0.8 mya) and fewer high-frequency k-mers (Supplementary 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 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-Seq data, and 31,741 and 32,056, respectively, show high similarity to *Arabidopsis* genes (Supplementary section 2.1). Chromosomal organisation of genes, transcripts, repeats and sRNAs is very similar in the two species (Supplementary Figures 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 Section 5.1, Supplementary Fig. 5). Relative expression of all tomato genes was determined by replicated strand-specific Illumina RNA-Seq of root, leaf, flower

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

(2 stages) and fruit (6 stages) in addition to leaf and fruit (3 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 Figures 1 and 3, Supplementary 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 TIR-NBS-LRR genes, appeared to be preferentially or exclusively expressed in potato (Supplementary Section 2.9).

Supplementary section 4 deals with comparative genomic studies. Sequence alignment of 71 Mb of euchromatic tomato genomic DNA to their potato[8] counterparts revealed 8.7% nucleotide divergence (Supplementary 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 9 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. 18,320 clearly orthologous tomato-potato gene pairs were identified. Of these, 138 (0.75%) had significantly higher than average non-synonymous (*Ka*) versus synonymous (*Ks*) nucleotide substitution rate ratios (ω), suggesting diversifying selection, whereas 147 (0.80%) had significantly lower than average ω, suggesting 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], suggesting that diversifying selection may have been stronger in tomato-potato. The highest densities of low-ω genes are found in collinear blocks with average *Ks* >1.5, tracing to a genome triplication shared with grape (see below) (Fig. 1C, Supplementary Fig. 6, 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 3-4).

Sequence comparison of 32,955 annotated genes in tomato and *S. pimpinellifolium* revealed 6,659 identical genes and 3,730 with only synonymous changes. A total of 22,888 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, 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 Figures 8-9), 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. 2B). The distribution of *Ks* between corresponding gene pairs in duplicated blocks suggests that one polyploidisation in the solanaceous lineage preceded the rosid-asterid (tomato-grape) divergence (Supplementary Fig. 10).

Comparison to the grape genome also reveals a more recent triplication in tomato and potato. While few individual tomato/potato genes remain triplicated (Supplementary Tables

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

10-11), 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) mya based on *Ks* of paralogous genes (Supplementary Fig. 10), and therefore predates the ~7.3 mya tomato-potato divergence. Based on alignments to single grape genome segments, the tomato genome can be partitioned into three non-overlapping 'subgenomes' (Fig. 2A). The number of euasterid lineages that have experienced the recent triplication remains unclear and awaits complete euasterid I and II genome sequences. Ks 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. 2B), suggesting that this event may have contributed to 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].

Supplementary section 5 reports on the analysis of specific gene families. Fleshy fruits (Supplementary Fig. 12) are an important means of attracting vertebrate frugivores for seed dispersal[16]. Combined orthology and synteny analyses suggest 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 (*LeETR3/NR, LeETR4*)[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 Section 5.4).

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 wall architecture (Fig. 4A and Supplementary 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, *SlXTH10,* shows differential loss between tomato and potato (Fig. 4A, Supplementary Table 12), suggesting genetically driven specialisation 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 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 (e.g. flower to fruit, fruit growth to fruit ripening, Supplementary 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 homolog of *Arabidopsis FT*, encoding the mobile flowering hormone florigen[20] and its antagonist *SP*, encoding the ortholog 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 tuberisation, respectively[22], whereas *SP3D* in tomato, known as *SINGLE FLOWER*

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

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 accessions 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 Summary

A total of 21 Gb of Roche/454 Titanium shotgun and matepair 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 Sections 1.1-1.7), assembled using both Newbler and CABOG and integrated into a single assembly (Supplementary Sections 1.17-1.18). The scaffolds were anchored using two BAC-based physical maps, one high density genetic map, overgo hybridization and genome-wide BAC FISH (Supplementary 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 Section 1.20). Chloroplast genome insertions in the nuclear genome were validated using a matepair method and the flanking regions were identified (Supplementary 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 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 Section 2.10). The genome of *S. pimpinellifolium* was sequenced to 40x depth using Illumina paired end reads and assembled using ABySS (Supplementary Section 3). The tomato and potato genomes were aligned using LASTZ (Supplementary 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 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 Section 5). PHYML and MEGA were used to reconstruct phylogenetic trees and MCSCAN was used to infer gene collinearity (Supplementary Section 5.2).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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 prior to publication; potato RNA-Seq data 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 (UC Davis); the Amplicon Express team (Pullman, WA, USA) 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: Jongsun Park, Biao Wang, Chengfeng Niu, Di Liu, Francesco Cojutti, Silvia Pescarolo, Alessandro Zambon, Gong Xiao, Jianjun Chen, Jinfeng Shi, Lei Zhang, Liping Zeng, Mario Caccamo, Dan Bolser, David Martin, Mireia Gonzalez, Patricia A. Bedinger, Paul A. Covey, Purnima Pachori, Renato Rodriguez Pousada, Sana Hakim, Sarah Sims, Vincent Cahais, Wenbo Long, Xincheng Zhou, Yiqi Lu, Waleed Haso, Cathy Lai, Stephanie Lepp, Cass Peluso, Homa Teramu, Hannah De Jong, Raphael Lizarralde, Eliel Ruiz May, and Zi Li. Marc Zabeau is thanked for his support and encouragement and Sandra van den Brink for her secretarial support. We dedicate this work to the late Professor Charles Rick (U. C. Davis) who pioneered tomato genetics, collection of wild germplasm and the distribution of seed and knowledge.

LIST OF TOMATO GENOME CONSORTIUM AUTHORS

Kazusa DNA Research Institute: Shusei Sato (Principal Investigator)[1], Satoshi Tabata (Principal Investigator)[1], Hideki Hirakawa[1], Erika Asamizu[1], Kenta Shirasawa[1], Sachiko Isobe[1], Takakazu Kaneko[1], Yasukazu Nakamura[1], Daisuke Shibata[1], Koh Aoki[1]; **454 Life Sciences, a Roche company:** Michael Egholm[2], James Knight[2]; **Amplicon Express Inc.:** Robert Bogden[3]; **Beijing Academy of Agriculture and Forestry Sciences:** Changbao Li[4,13]; **BGI-Shenzhen:** Yang Shuang[5], Xun Xu[5], Shengkai Pan[5], Shifeng Cheng[5], Xin Liu[5], Yuanyuan Ren[5], Jun Wang[5]; **BMR-Genomics SrL:** Alessandro Albiero[6], Francesca Dal Pero[6], Sara Todesco[6]; **Boyce Thompson Institute for Plant Research:** Joyce Van Eck[7], Robert M. Buels[7], Aureliano Bombarely[7], Joseph R. Gosselin[7], Minyun Huang[7], Jonathan A. Leto[7], Naama Menda[7], Susan Strickler[7], Linyong Mao[7], Shan Gao[7], Isaak Y. Tecle[7], Thomas York[7], Yi Zheng[7], Julia T. Vrebalov[7], JeMin Lee[7], Silin Zhong[7], Lukas A. Mueller (Principal Investigator)[7]; **Centre for BioSystems Genomics:** Willem J. Stiekema[8]; **Centro Nacional de Análisis Genómico (CNAG):** Paolo Ribeca[9], Tyler Alioto[9,10]; **China Agricultural University:** Wencai Yang[11]; **Chinese Academy of Agricultural Sciences:** Sanwen Huang (Principal Investigator)[12], Yongchen Du (Principal Investigator)[12], Zhonghua Zhang[12], Jianchang Gao[12], Yanmei Guo[12], Xiaoxuan Wang[12], Ying Li[12], Jun He[12]; **Chinese Academy of Agricultural Sciences:** Chuanyou Li (Principal Investigator)[13], Zhukuan Cheng (Principal Investigator)[13], Jianru Zuo (Principal Investigator)[13], Jianfeng Ren[13], Jiuhai Zhao[13], Liuhua Yan[13], Hongling Jiang[13], Bao Wang[13], Hongshuang Li[13], Zhenjun Li[13], Fuyou Fu[13], Bingtang Chen[13], Bin Han (Principal Investigator)[14], Qi Feng[14], Danlin Fan[14], Ying Wang (Principal Investigator)[15], Hongqing Ling (Principal Investigator)[16], Yongbiao Xue (Principal Investigator)[17]; **Cold Spring Harbor Laboratory and United States Department of Agriculture - Agricultural Research Service:** Doreen Ware (Principal Investigator)[18], W. Richard McCombie (Principal Investigator)[18], Zachary

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

B. Lippman (Principal Investigator)[18], Jer-Ming Chia[18], Ke Jiang[18], Shiran Pasternak[18], Laura Gelley[18], Melissa Kramer[18]; **Colorado State University:** Lorinda K. Anderson[19], Song-Bin Chang[20], Suzanne M. Royer[19], Lindsay A. Shearer[19], Stephen M. Stack (Principal Investigator)[19]; **Cornell University:** Jocelyn K. C. Rose[21], Yimin Xu[21], Nancy Eannetta[21], Antonio J. Matas[21], Ryan McQuinn[21], Steven D. Tanksley (Principal Investigator)[21]; **Genome Bioinformatics Laboratory GRIB -- IMIM/UPF/CRG:** Francisco Camara[22], Roderic Guigó[22], Stephane Rombauts[23], Jeffrey Fawcett[23], Yves Van de Peer (Principal Investigator)[23]; **Hebrew University of Jerusalem:** Dani Zamir[24]; **Heilongjiang Academy of Agricultural Sciences:** Chunbo Liang[25]; **Helmholtz Center for Health and Environment:** Manuel Spannagl[26], Heidrun Gundlach[26], Remy Bruggmann[26], Klaus Mayer (Principal Investigator)[26]; **Henan Agricultural University:** Zhiqi Jia[27]; **Huazhong Agricultural University:** Junhong Zhang[28], Zhibiao Ye[28]; **Imperial College London:** Gerard J Bishop (Principal Investigator)[29], Sarah Butcher (Principal Investigator)[29], Rosa Lopez-Cobollo[29], Daniel Buchan[29], Ioannis Filippis[29], James Abbott[29]; **Indian Agricultural Research Institute:** Rekha Dixit[30], Manju Singh[30], Archana Singh[30], Jitendra Kumar Pal[30], Awadhesh Pandit[30], Pradeep Kumar Singh[30], Ajay Kumar Mahato[30], Vivek Dogra[30], Kishor Gaikwad[30], Tilak Raj Sharma[30], Trilochan Mohapatra[30], Nagendra Kumar Singh (Principal Investigator)[30]; **INRA Avignon:** Mathilde Causse[31]; **INRA Bordeaux:** Christophe Rothan[32]; **INRA Toulouse:** Thomas Schiex (Principal Investigator)[33], Céline Noirot[33], Arnaud Bellec[34], Christophe Klopp[35], Corinne Delalande[36], Hélène Berges[34], Jérôme Mariette[35], Pierre Frasse[36], Sonia Vautrin[34]; **Institut National Polytechnique de Toulouse:** Mohamed Zouine[36], Alain Latché[36], Christine Rousseau[36], Farid Regad[36], Jean-Claude Pech[36], Murielle Philippot[36], Mondher Bouzayen (Principal Investigator)[36]; **Instituto de Biología Molecular y Celular de Plantas (CSIC-UPV):** Pierre Pericard[37], Sonia Osorio[37], Asunción Fernandez del Carmen[37], Antonio Monforte[37], Antonio Granell (Principal Investigator)[37]; **Instituto de Hortofruticultura Subtropical y Mediterránea (IHSM-UMA-CSIC):** Rafael Fernandez-Muñoz[38]; **Instituto Nacional de Tecnología Agropecuaría (IB-INTA) and Consejo Nacionalde Investigaciones Científicas y Técnicas (CONICET):** Mariana Conte[39], Gabriel Lichtenstein[39], Fernando Carrari (Principal Investigator)[39]; **Italian National Res Council, Institute for Biomedical Technologies:** Gianluca De Bellis (Principal Investigator)[40], Fabio Fuligni[40], Clelia Peano[40]; **Italian National Res Council, Institute of Plant Genetics, Research Division Portici:** Silvana Grandillo[41], Pasquale Termolino[41]; **Italian National Agency for New technologies, Energy and Sustainable Development:** Marco Pietrella[42,43], Elio Fantini[42], Giulia Falcone[42], Alessia Fiore[42], Giovanni Giuliano (Principal Investigator)[42], Loredana Lopez[44], Paolo Facella[44], Gaetano Perrotta[44], Loretta Daddiego[44]; **James Hutton Institute:** Glenn Bryan (Principal Investigator)[45]; **Joint IRB-BSC program on Computational Biology:** Modesto Orozco[46], Xavier Pastor[46], David Torrents[46,47]; **Keygene N.V.:** Marco G. M. van Schriek[48], Richard M.C. Feron[48], Jan van Oeveren[48], Peter de Heer[48], Lorena daPonte[48], Saskia Jacobs-Oomen[48], Mike Cariaso[48], Marcel Prins[48], Michiel J.T. van Eijk (Principal Investigator)[48], Antoine Janssen[48], Mark J.J. van Haaren[48]; **Korea Research Institute of Bioscience and Biotechnology:** Sung-Hwan Jo[49], Jungeun Kim[49], Suk-Yoon Kwon[49], Sangmi Kim[49], Dal-Hoe Koo[49], Sanghyeob Lee[49], Cheol-Goo Hur[49]; **Life Technologies:** Christopher Clouser[50], Alain Rico[51]; **Max Planck Institute for Plant Breeding Research:** Asis Hallab[52], Christiane Gebhardt[52], Kathrin Klee[52], Anika Jöcker[52], Jens Warfsmann[52], Ulrike Göbel[52]; **Meiji University:** Shingo Kawamura[53], Kentaro Yano[53]; **Montana State University:** Jamie D. Sherman[54]; **NARO Institute of Vegetable and Tea Science:** Hiroyuki Fukuoka (Principal Investigator)[55], Satomi Negoro[55]; **National Institute of Plant Genome Research:** Sarita Bhutty[56], Parul Chowdhury[56], Debasis Chattopadhyay (Principal Investigator)[56]; **Plant Research International:** Erwin Datema[57], Sandra Smit[57], Elio G.W.M. Schijlen[57], Jose van de Belt[57], Jan C. van Haarst[57], Sander A. Peters[57], Marjo J. van Staveren[57], Marleen H.C. Henkens[57], Paul J.W. Mooyman[57],

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Thamara Hesselink[57], 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 Kang[60], Byung Dong Kim[60], Minkyu Park[60], Seungill Kim[60], Seon-In Yeom[60], Yong-Hwan Lee[61], Yang-Do Choi[62]; **Shandong Academy of Agricultural Sciences:** Guangcun Li[63], Jianwei Gao[64]; **Sichuan University:** Yongsheng Liu[65], Shengxiong Huang[65]; **Sistemas Genomicos:** Victoria Fernandez-Pedrosa[66], Carmen Collado[66], Sheila Zuñiga[66]; **South China Agricultural University:** Guoping Wang[67]; **Syngenta Biotechnology:** Rebecca Cade[68], Robert A. Dietrich[68]; **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)[7,71], Ruth A. White[7,71], Theodore W. Thannhauser[71], James J. Giovannoni (Principal Investigator)[7,21,71]; **Universidad de Malaga-Consejo Superior de Investigaciones Cientificas:** Miguel Angel Botella[72], Louise Gilbert[72]; **Universitat Pompeu Fabra:** Ramon Gonzalez[73]; **University of Arizona:** Jose Luis Goicoechea[74], Yeisoo Yu[74], David Kudrna[74], Kristi Collura[74], Marina Wissotski[74], Rod Wing (Principal Investigator)[74]; **University of Bonn:** Heiko Schoof (Principal Investigator)[75]; **University of Delaware:**Blake C. Meyers (Principal Investigator)[76], Aishwarya Bala Gurazada[76], Pamela J. Green[76]; **University of Delhi South Campus:** Saloni Mathur[77], Shailendra Vyas[77], Amolkumar U. Solanke[77], Rahul Kumar[77], Vikrant Gupta[77], Arun K. Sharma[77], Paramjit Khurana[77], 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 Walts[80], Srikar Chamala[80], W. Brad Barbazuk[80]; **University of Georgia:** Jingping Li[81], Hui Guo[81], Tae-Ho Lee[81], Yupeng Wang[81], Dong Zhang[81], Andrew H. Paterson (Principal Investigator)[81], Xiyin Wang (Principal Investigator)[81,82], Haibao Tang[81,83]; **University of Naples "Federico II":** Amalia Barone[84], Maria Luisa Chiusano[84], Maria Raffaella Ercolano[84], Nunzio D'Agostino[84], Miriam Di Filippo[84], Alessandra Traini[84], Walter Sanseverino[84], Luigi Frusciante (Principal Investigator)[84]; **University of Nottingham:** Graham B. Seymour (Principal Investigator)[85]; **University of Oklahoma:** Mounir Elharam[86], Ying Fu[86], Axin Hua[86], Steven Kenton[86], Jennifer Lewis[86], Shaoping Lin[86], Fares Najar[86], Hongshing Lai[86], Baifang Qin[86], Chunmei Qu[86], Ruihua Shi[86], Douglas White[86], James White[86], Yanbo Xing[86], Keqin Yang[86], Jing Yi[86], Ziyun Yao[86], Liping Zhou[86], Bruce A. Roe (Principal Investigator)[86]; **University of Padua:** Alessandro Vezzi[87], Michela D'Angelo[87], Rosanna Zimbello[87], Riccardo Schiavon[87], Elisa Caniato[87], Chiara Rigobello[87], Davide Campagna[87], Nicola Vitulo[87], 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 Bai[91], Richard G.F. Visser[91], René M. Klein Lankhorst (Principal Investigator)[92]; **Wellcome Trust Sanger Institute:** Helen Beasley[93], Karen McLaren[93], Christine Nicholson[93], Claire Riddle[93]; **Ylichron SrL:** Giulio Gianese[94]

REFERENCES

1. Frodin DG. History and concepts of big plant genera. Taxon. 2004; 53:753–776.
2. Peralta IE, Knapp S, Spooner DM. Taxonomy of tomatoes: A revision of wild tomatoes (Solanum section Lycopersicon) and their outgroup relatives in sections Juglandifolia and Lycopersicoides. Systematic Botany Monographs. 2008; 84:1–186.

Europe PMC Funders Author Manuscripts

3. Michaelson MJ, Price HJ, Ellison JR, Johnston JS. Comparison of plant DNA contents determined by Feulgen microspectrophotometry and laser flow cytometry. American journal of botany. 1991:183–188.
4. The_Arabidopsis_Genome_Initiative. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature. 2000; 408:796–815. [PubMed: 11130711]
5. Paterson AH, et al. The Sorghum bicolor genome and the diversification of grasses. Nature. 2009; 457:551–556. [PubMed: 19189423]
6. Zamir D, Tanksley SD. Tomato genome is comprised largely of fast- evolving, low copy-number sequences. Molecular and General Genetics MGG. 1988; 213:254–261.
7. Peterson DG, Pearson WR, Stack SM. Characterization of the tomato (Lycopersicon esculentum) genome using in vitro and in situ DNA reassociation. Genome. 1998; 41:346–356.
8. Xu X, et al. Genome sequence and analysis of the tuber crop potato. Nature. 2011; 475:189–195. [PubMed: 21743474]
9. Swigonova Z, et al. Close split of sorghum and maize genome progenitors. Genome Res. 2004; 14:1916–1923. [PubMed: 15466289]
10. Jaillon O, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature. 2007; 449:463–467. [PubMed: 17721507]
11. Tang H, et al. Synteny and collinearity in plant genomes. Science. 2008; 320:486–488. [PubMed: 18436778]
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. 2008; 8:130. [PubMed: 19099601]
13. Ozminkowski R. Pedigree of variety Heinz 1706. Rep Tomato Genet Coop. 2004; 54:26.
14. Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc Natl Acad Sci U S A. 2010; 107:4623–4628. [PubMed: 20176954]
15. Stockey RA, Graham SW, Crane PR. Introduction to the Darwin special issue: The abominable mystery. Am J Bot. 2009; 96:3–4. [PubMed: 21628173]
16. Howe HF, Smallwood J. Ecology of seed dispersal. Annual review of ecology and systematics. 1982:201–228.
17. Klee HJ, Giovannoni JJ. Genetics and control of tomato fruit ripening and quality attributes. Annu Rev Genet. 2011; 45:41–59. [PubMed: 22060040]
18. Vicente AR, Saladie M, Rose JKC, Labavitch JM. The linkage between cell wall metabolism and fruit softening: looking to the future. Journal of the Science of Food and Agriculture. 2007; 87:1435–1448.
19. Mohorianu I, et al. Profiling of short RNAs during fleshy fruit development reveals stage-specific sRNAome expression patterns. Plant J. 2011; 67:232–246. [PubMed: 21443685]
20. Corbesier L, et al. FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. Science. 2007; 316:1030–1033. [PubMed: 17446353]
21. Rick CM. The tomato. Scientific American. 1978; 239:67–76.
22. Navarro C, et al. Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. Nature. 2011; 478:119–122. [PubMed: 21947007]
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 U S A. 2006; 103:6398–6403. [PubMed: 16606827]
24. Krieger U, Lippman ZB, Zamir D. The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. Nat Genet. 2010; 42:459–463. [PubMed: 20348958]
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. 1998; 125:1979–1989. [PubMed: 9570763]
26. Rick CM. Hybridization between Lycopersicon esculentum and Solanum pennellii: Phylogenetic and Cytogenetic Significance. Proc Natl Acad Sci U S A. 1960; 46:78–82. [PubMed: 16590600]

27. Barker MS, et al. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. Mol Biol Evol. 2008; 25:2445–2455. [PubMed: 18728074]

28. Aagaard JE, Willis JH, Phillips PC. Relaxed selection among duplicate floral regulatory genes in Lamiales. J Mol Evol. 2006; 63:493–503. [PubMed: 17021928]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

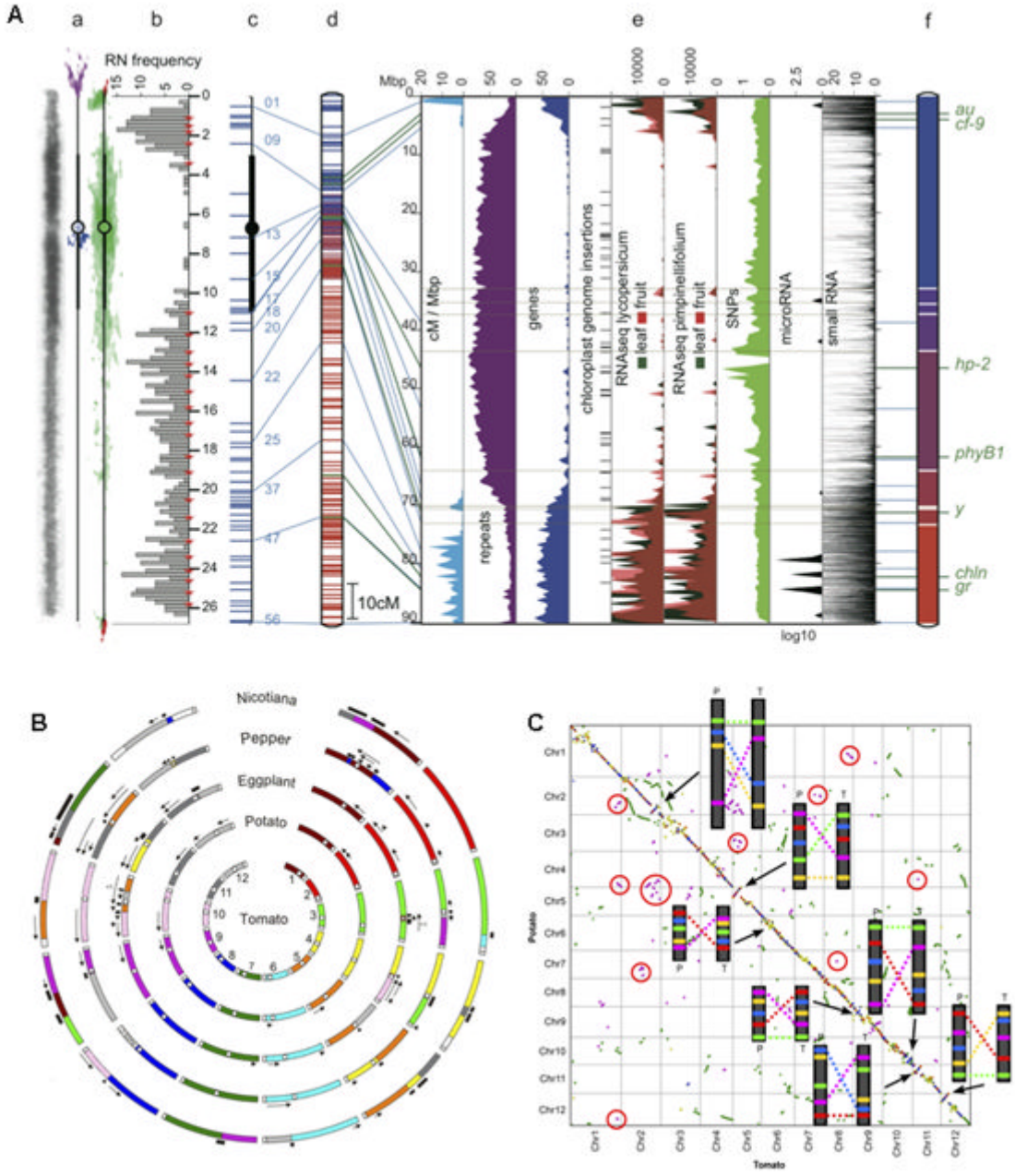

Figure 1.

A. Multi-dimensional topography of tomato chromosome *1* (chromosomes 2-12 are shown in Supplementary Figure 1).

(a) Left: contrast-reversed, DAPI-stained pachytene chromosome; centre and right: FISH signals for repeat sequences on diagrammatic pachytene chromosomes: TGR1 purple, TGR4 blue, telomere repeat red, Cot 100 DNA (including most repeats) green. (b) Frequency distribution of recombination nodules representing crossovers on 249 chromosomes. Red stars mark 5 cM intervals starting from the end of the short arm (top). Scale is in micrometers. (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 (c) BAC-FISH map, (e) heat maps and (f) DNA pseudomolecule. (e) From left to right: linkage

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

map distance (cM/Mb, turquoise); repeated sequences (% nucleotides/500 kb, purple); genes (% nucleotides/500 kb, blue); chloroplast insertions; RNA-Seq reads from leaves and breaker fruits of *S. lycopersicum* and *S. pimpinellifolium* (number of reads/500 kb, green and red, respectively); microRNA genes (transcripts per million/500 kb, black); small RNAs (thin horizontal black 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 Figure 1.

B. Syntenic relationships in the *Solanaceae*.

COSII-based comparative maps of potato, eggplant, pepper and *Nicotiana* with respect to the tomato genome (Supplementary section 4.5, 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 and potato genomic sequences based on collinear blocks Supplementary Section 4.1). Red and blue dots represent gene pairs with statistically significant high and low ω (*Ka/Ks*) in collinear blocks, which average $Ks \leq 0.5$, respectively. Green and magenta dots represent genes in collinear blocks which average $0.5 < Ks \leq 1.5$ and $Ks > 1.5$, 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).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

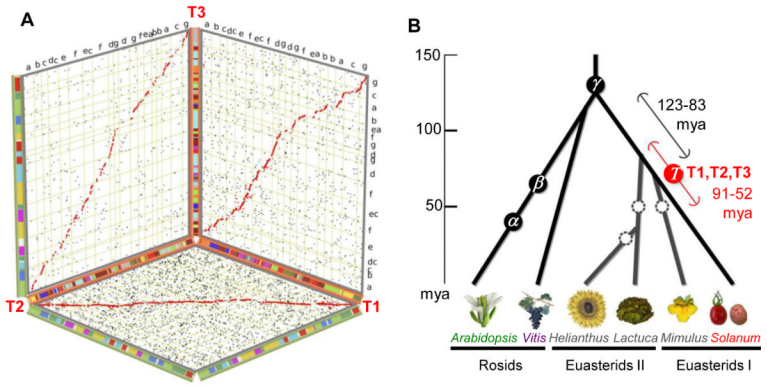

Figure 2. The *Solanum* whole genome triplication

A. Based on 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 3D plot. The ancestral gene order of each subgenome is inferred according to orthologous grape regions, with tomato chromosomal affinities shown by red-shaded (inner) bars. Segments tracing to pan-eudicot triplication (γ) are shown by green-shaded (outer) bars with colours representing the seven putative pre-γ eudicot ancestral chromosomes[10], also coded a-g.

B. Speciation and polyploidisation 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, respectively, the likely timings of divergence of major asterid lineages and of "T". The post-"T" subgenomes, designated T1, T2, and T3, are further detailed in Supplementary Fig. 10.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

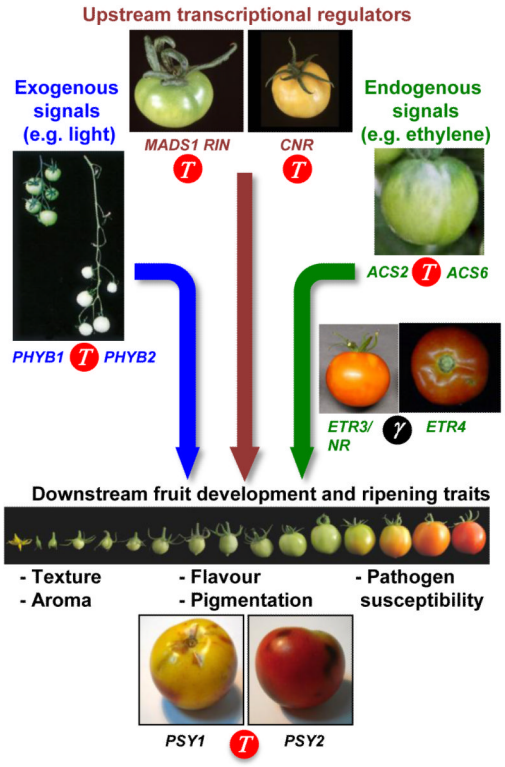

Figure 3. Whole genome triplications set the stage for fruit-specific gene neofunctionalisation

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 Figures 16 and 17.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

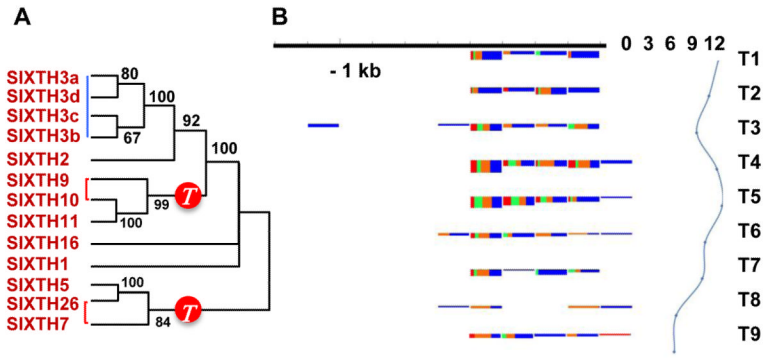

Figure 4. The tomato genome allows systems approaches to fruit biology

A. Xyloglucan transglucosylase-hydrolases (XTHs) differentially expressed between mature green and ripe fruits (Supplementary 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 paralogs 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 mRNA expression levels from the corresponding gene (right) over a tomato fruit developmental series (T1 – bud, T2 – flower, T3 – fruit 1- 3mm, T4 – fruit 5-7mm, T5 – fruit 11-13mm, T6 – fruit mature green, T7 – breaker, T8 – breaker +3days, T9 – breaker+7days). The promoter regions are grouped in 100nt windows. For each window the size class distribution of sRNAs is shown (21 – red, 22 – green, 23 – orange, 24 – blue). 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.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts