Patterns of gene-tree variation leveraged from herbarium records reveal a complex evolutionary ancestry of early European tomatoes

Thomas Grubinger[1], Gülfirde Akgül[2], Alessia Guggisberg[3], Reto Nyffeler[4], Jurriaan M. de Vos[5], Verena J. Schünemann[2] and Simon Aeschbacher[1]

[1]Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland; [2]Institute of Evolutionary Medicine, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland; [3]Department of Environmental Systems Science, ETH Zurich, Universitätstrasse 16, 8092 Zurich, Switzerland; [4]Department of Systematic and Evolutionary Botany, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland; [5]Department of Environmental Sciences – Botany, University of Basel, Schönbeinstrasse 6, 4056, Basel, Switzerland

Introduction

The cultivated tomato, *Solanum lycopersicum* L., was likely domesticated in Andean South America (Peru, Ecuador) and Mexico in a multi-step process [1–4]. One hypothesis suggests that the wild ancestor *S. pimpinellifolium* L. (SP) was domesticated to intermediate, cherry-sized tomatoes, *S. lycopersicum* L. var. *cerasiforme* (SLC), in Peru, and then further improved to large-fruited cultivated tomatoes, *S. lycopersicum* L. var. *lycopersicum* (SLL), in Mexico [2, 3]. A more recent study proposed that SLC evolved as a natural taxon from SP. Then, an initial domestication in Peru from wild to larger-fruited SLC was followed by independent (re)domestication in Mexico from small-fruited, undomesticated SLC of Peruvian origin [4].

In the 16th century, the cultivated tomato was introduced to Europe [5]. However, historical records disagree about whether from Peru or Mexico [2, 4, 5]. Here, we aim at unravelling the ancestry of European tomato specimens collected between around 1600 and 1915 both in terms of geographic and taxonomic origin.

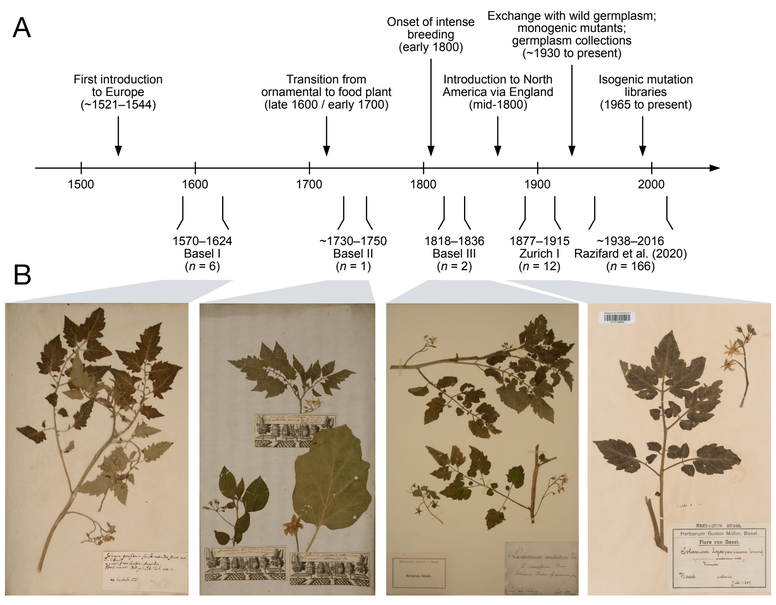

Fig. 1: History of tomato breeding in Europe. A) Timeline with key events on top, and ages of sampled herbarium specimens and the contemporary dataset at the bottom. B) Digital scans of herbarium specimens from the Herbaria Basel (BAS, left-most) and the United Herbaria of the University and ETH Zurich (Z+ZT, remaining ones). Photographs courtesy of BAS and Z+ZT.

Materials & Methods

- 21 herbarium specimens dated between ~1600 and 1915 (**Fig. 1**), denoted as SL HIST
- DNA extraction and library preparation in clean laboratory dedicated to historical DNA processing
- Whole-genome sequencing: Illumina NovaSeq 6000
- DNA resequencing data from 166 modern tomato accessions from South America, Central America, and Mexico [4] for comparison with historical specimens
- Bioinformatics and data filtering for quality and missingness
- Principal Component analysis (PCA): smartPCA [6]
- Inference of genetic ancestry components under a model of admixture: ADMIXTURE [7]
- Gene-tree topology weighting along the genome: RAxML + Twisst [8, 9]
- Statistical test for gene flow versus incomplete lineage sorting: Admixtools2 [10]

Objectives

1. Establishing the genomic and geographic ancestry of European tomato cultivars
2. Identifying the taxonomic groups represented among early European tomatoes
3. Exploring the extent of variation in gene-tree topologies along the genome

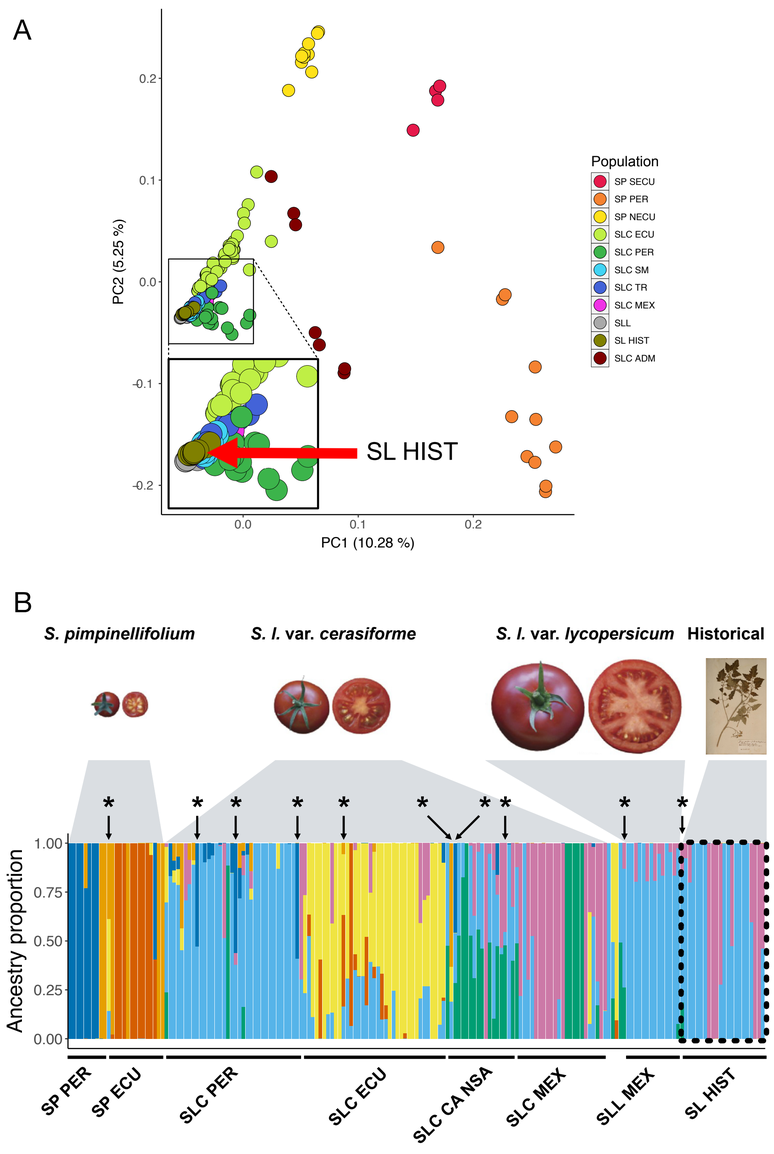

Fig. 2: Genomic and geographic ancestry of tomato cultivars at synonymous SNPs. A) First two principal components (PC) from a PCA. Historical specimens projected onto PCs inferred from accessions in reference [4]. PC1 and PC2 reflect the extent of domestication and sub-structure in SP, respectively. B) Ancestry proportions assuming $K = 7$ ancestral components (chosen by cross-validation). Bars and labels at the bottom show taxon and geographic origin according to passport data. Asterisks indicate accessions for which passport and genetic data disagree. Photographs of fruits modified from reference [3].

Results

Alignment of reads from the 21 herbarium specimens to the SL4.0 tomato reference genome [11] resulted in alignment depths ranging from 8.7 to 21.4-fold at filtered single-nucleotide polymorphisms (SNPs). Phylogenomic and population genomic analyses at 103,010 synonymous SNPs revealed:

- All historical specimens draw their genomic ancestry from two ancestral components (**Fig. 2**; pink and light blue).
- Additional analyses beyond synonymous SNPs show that historical specimens derive from two taxonomic groups (data not shown): large-fruited (SLL) cultivars (14 specimens, 3 dated around 1600) and cherry (SLC) accessions (7 specimens, 1 dated around 1600).
- Gene-tree topologies vary strongly along the genome. The most abundant gene-tree topology places the three ancient-most SLL (SL BAS) specimens from around 1600 as sister to the large-fruited Mexican landraces (SLL MEX) (**Fig. 3**). This finding implies that early European tomatoes originate from Mexico.
- The second-most abundant gene-tree topology suggests gene flow between semi-domesticated Peruvian cherries (SLC SM) and SLL MEX (**Fig. 3**), as confirmed by Patterson's *D*-statistic ($D = 0.00285$, $|z| = 11.723$, $p = <0.001$).

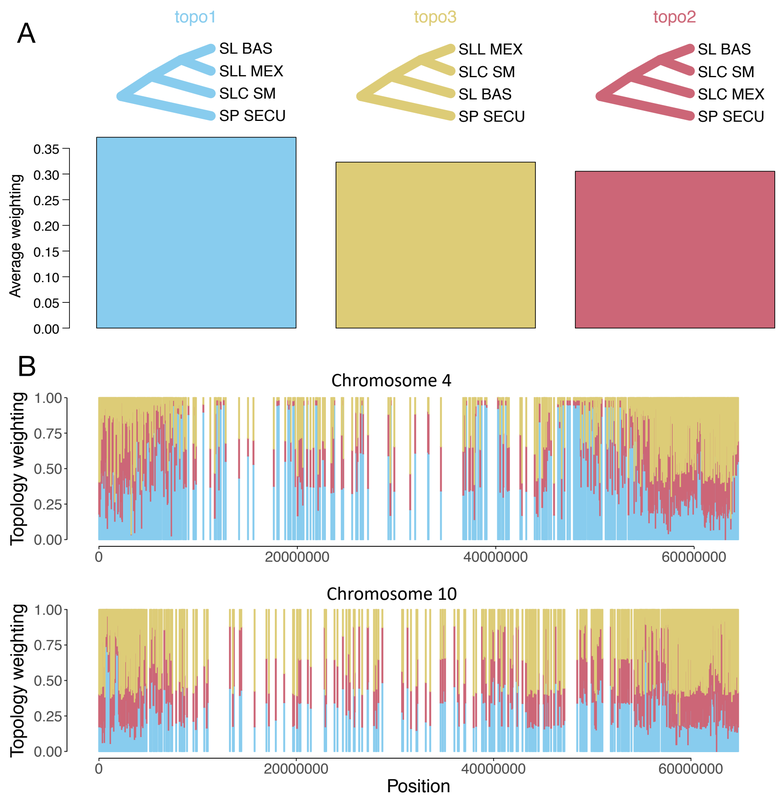

Fig. 3: Weighting of gene-tree topologies across 50-kilobase pair windows at synonymous SNPs. A) Average weight of topologies, with topo1 (blue) representing the likely species-tree. Topologies topo3 (yellow) and topo2 (red) represent 'incompatible' topologies arising from incomplete lineage sorting and/or historical gene flow. B) Topology weights along two of the twelve chromosomes. Chromosome 4 and 10 predominantly feature topo1 and topo3, respectively. Only windows with more than two SNPs are shown.

Conclusions

1. The genomic ancestry of European tomato specimens from around 1600 to 1915 traces back to a geographic origin in Mexico, but a Peruvian component is evident in large parts of the genome (**Fig. 2 & 3**).
2. Plant material from two different taxonomic groups was available to early European tomato breeders: large-fruited (SLL) cultivars and small-fruited cherry (SLC) cultivars.
3. The evolution of large fruit size in Mexico might not have been independent, but fuelled by adaptive gene flow from semi-domesticated Peruvian cherry tomatoes (**Fig. 3**).

We acknowledge funding from the Basler Stiftung für biologische Forschung and the Research Priority Program "Evolution in Action" of the University of Zurich .

References and Acknowledgments

[1] Peralta IE, Spooner DM, Knapp S (2008) Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoides*, sect. *Juglandifolia*, sect. *Lycopersicon*; Solanaceae). Syst. Bot. Monogr. 84.

[2] Blanca J, Montero-Pau J, Sauvage C, Bauchet G, Illa E, Díez MJ, Francis D, Causse M, van der Knaap E, Cañizares J (2015). Genomic variation in tomato, from wild ancestors to breeding accessions. BMC Genomics 16, 1–19

[3] Lin T, Zhu G, Zhang J, Xu X, Yu Q, Zheng Z, Zhang Z, Lun Y, Li S, Wang X, et al. (2014). Genomic analyses provide insights into the history of tomato breeding. Nat. Genet. 46, 1220–1226.

[4] Razifard H, Ramos A, Della Valle AL, Bodary C, Goetz E, Manser EJ, Li X, Zhang L, Visa S, Tieman D, et al. (2020). Genomic Evidence for Complex Domestication History of the Cultivated Tomato in Latin America. Mol. Biol. Evol. 37, 1118–1132.

[5] Bergougnoux V (2014) The history of tomato: From domestication to biopharming. Biotechnology Advances 32(1):170–189.

[6] Patterson N, Price AL, and Reich D (2006). Population Structure and Eigenanalysis. PLoS Genet. 2, 1–20

[7] Alexander DH, Novembre J, Lange K (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664

[8] Stamatakis A (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

[9] Martin SH, Van Belleghem SM (2017) Exploring Evolutionary Relationships Across the Genome Using Topology Weighting. Genetics 206(1), 429–438.

[10] Maier R, Patterson N (2021) Admixtools2. https://uqrmaie1.github.io/admixtools/index.html

[11] The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485(7400):635–641.