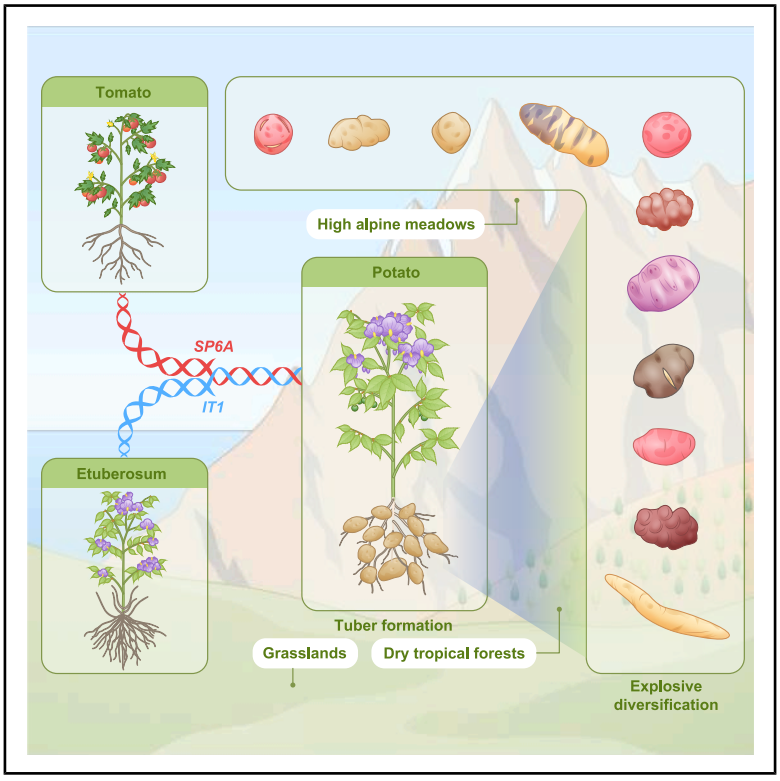
Cell

Ancient hybridization underlies tuberization and radiation of the potato lineage

Graphical abstract



Highlights

- The potato lineage is of ancient homoploid hybrid origin
- Alternate inheritance of highly divergent parental genes contributed to tuberization
- Hybridization and tuberization triggered species radiation and niche expansion

Authors

Zhiyang Zhang, Pingxian Zhang, Yiyuan Ding, ..., Loren H. Rieseberg, Jianquan Liu, Sanwen Huang

Correspondence

s.knapp@nhm.ac.uk (S.K.), loren.rieseberg@botany.ubc.ca (L.H.R.), liujq@nwipb.cas.cn (J.L.), huangsanwen@caas.cn (S.H.)

In brief

Genomic and functional analyses reveal that the potato lineage originated from a homoploid interspecific hybridization event between the Tomato and Etuberosum lineages 8–9 million years ago. The alternate inheritance of highly divergent parental alleles drove the origin of tubers and explosive species diversification and niche expansion into the high Andes.

Zhang et al., 2025, Cell *188*, 1–17
September 18, 2025
https://doi.org/10.1016/j.cell.2025.06.034

CellPress

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

Cell

CellPress
OPEN ACCESS

Article

Ancient hybridization underlies tuberization and radiation of the potato lineage

Zhiyang Zhang,[1,17] Pingxian Zhang,[1,17] Yiyuan Ding,[2,17] Zefu Wang,[3,17] Zhaoxu Ma,[1,2,18] Edeline Gagnon,[4,18] Yuxin Jia,[5] Lin Cheng,[1] Zhigui Bao,[1,6] Zinan Liu,[1] Yaoyao Wu,[1,7] Yong Hu,[1] Qun Lian,[1] Weichao Lin,[1] Nan Wang,[1] Keyi Ye,[1] Hongru Wang,[1] Jinzhe Zhang,[8] Yongfeng Zhou,[1] Liang Liu,[9] Suhua Li,[1] William J. Lucas,[1,10] Tiina Särkinen,[11] Sandra Knapp,[12,*] Loren H. Rieseberg,[13,*] Jianquan Liu,[14,15,*] and Sanwen Huang[1,16,19,*]

[1]National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, Guangdong 518120, China
[2]Huazhong Agricultural University, Wuhan, Hubei 430070, China
[3]State Key Laboratory of Tree Genetics and Breeding, Co-Innovation Center for Sustainable Forestry in Southern China, College of Ecology and Environment, Nanjing Forestry University, Nanjing, Jiangsu 210037, China
[4]Department of Integrative Biology, College of Biological Science, University of Guelph, Guelph, Ontario N1G 2W1, Canada
[5]The AGISCAAS-YNNU Joint Academy of Potato Sciences, Yunnan Normal University, Kunming, Yunnan 650500, China
[6]Department of Molecular Biology, Max Planck Institute for Biology Tübingen, Tübingen 72076, Germany
[7]State Key Laboratory of Crop Genetics and Germplasm Enhancement and Utilization, College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China
[8]State Key Laboratory of Vegetable Biobreeding, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China
[9]Department of Statistics, Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA
[10]Department of Plant Biology, College of Biological Sciences, University of California, Davis, Davis, CA 95616, USA
[11]Royal Botanic Garden Edinburgh, Edinburgh EH3 5LR, UK
[12]Natural History Museum, London SW7 5BD, UK
[13]Department of Botany and Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada
[14]Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, Sichuan 610065, China
[15]State Key Laboratory of Herbage Improvement and Grassland Agro-Ecosystem, College of Ecology, Lanzhou University, Lanzhou, Gansu 730000, China
[16]National Key Laboratory of Tropical Crop Breeding, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan 571101, China
[17]These authors contributed equally
[18]These authors contributed equally
[19]Lead contact
*Correspondence: s.knapp@nhm.ac.uk (S.K.), loren.rieseberg@botany.ubc.ca (L.H.R.), liujq@nwipb.cas.cn (J.L.), huangsanwen@caas.cn (S.H.)
https://doi.org/10.1016/j.cell.2025.06.034

SUMMARY

Interspecific hybridization may trigger species radiation by creating allele combinations and traits. Cultivated potato and its 107 wild relatives from the Petota lineage all share the distinctive trait of underground tubers, but the underlying mechanisms for tuberization and its relationship to extensive species diversification remain unclear. Through analyses of 128 genomes, including 88 haplotype-resolved genomes, we revealed that Petota is of ancient hybrid origin, with all members exhibiting stable mixed genomic ancestry, derived from the Etuberosum and Tomato lineages ca. 8–9 million years ago. Our functional experiments further validated the crucial roles of parental genes in tuberization, indicating that interspecific hybridization is a key driver of this innovative trait. This trait, along with the sorting and recombination of hybridization-derived polymorphisms, likely triggered the explosive species diversification of Petota by enabling occupation of broader ecological niches. These findings highlight how ancient hybridization fosters key innovation and drives subsequent species radiation.

INTRODUCTION

Hybrid speciation, the recombination of different alleles and genes from distinct species, is a significant evolutionary force.[1–6] Since standing genetic variation can be reassembled into combinations, via hybridization, it generates phenotypes that enable adaptation to environments and may further lead to the accelerated origin of many additional species.[7–10] An evolutionary

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

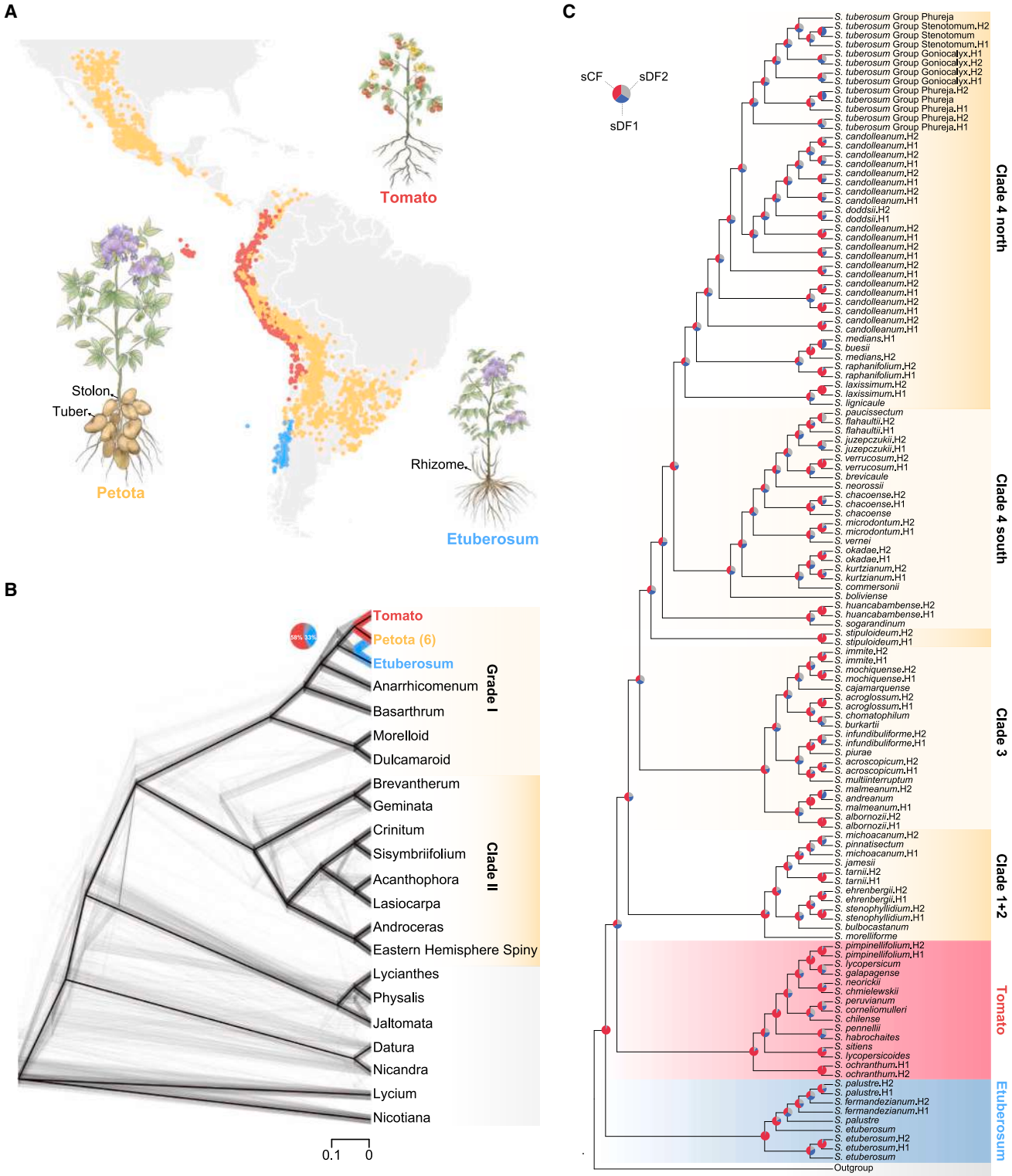


Figure 1. Geographical distribution and phylogenetic conflicts

(A) The geographic distribution of Petota (golden dots), Tomato (red dots), and Etuberosum (blue dots) (excluding cultivated species) is shown based on verified georeferenced collection records.

(B) The phylogeny is depicted using 500 trees randomly sampled from the 1-Mb windows with 200-kb step size (gray lines), and a coalescent-based species tree (black lines) with all nodes resolved.

(legend continued on next page)

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

radiation in one lineage is likely to occur when ancient hybrid speciation had taken place.[11,12] While ancient hybrid speciation has been detected previously,[13–15] adaptive species radiation triggered by it, especially through the key innovation, remains unexplored. Here, we not only show that the cultivated potato and its 107 wild relatives are derived from an ancient hybrid speciation event, but also that tuber formation itself, a key innovative trait, has a hybrid ancestry.

The cultivated potato (*Solanum tuberosum* L.) is currently the world's third most important staple crop, and with wheat, rice, and maize, is responsible for 80% of human caloric intake.[16] Within the genus *Solanum*, the cultivated potato and its closely related 107 wild species comprise the monophyletic Petota lineage (hereafter referred to as Petota), which is ecologically and morphologically distinct from the closely related Etuberosum and Tomato lineages (hereafter referred to as Etuberosum and Tomato, respectively), the former being a small monophyletic lineage of three species from southern South America, and the similarly monophyletic latter consisting of 17 cultivated and wild tomato species.[17–23] In contrast with Tomato, both Petota and Etuberosum are geophyte species (plants with underground re-sprouting organs), which enable vegetative propagation through underground stems, termed stolons in Petota and rhizomes in Etuberosum (Figure 1A).[24] Additionally, Petota has specialized tuber organs, the swollen portions of stolons, which store water and carbohydrates and serve as important nutritional and reproductive organs.

Although somatic hybridizations between Petota and Etuberosum, and between Petota and Tomato, with the same chromosome number, have been successfully produced,[25,26] strong postzygotic reproductive isolations (RIs) exist between the three lineages, with artificial inter-lineage crosses producing only small aborted seeds.[27] All species of Etuberosum and 35% of Tomato species are self-compatible, in contrast to mostly self-incompatible Petota species.[28,29] Tomato and Etuberosum species are mainly diploids, whereas Petota species include both diploids and polyploids. Diploids in all three lineages have the same chromosome number (2n = 24) and similar chromosome morphology with extensive synteny.[27,30,31] Recent studies have placed Etuberosum or Tomato as sister to Petota and have shown widespread phylogenetic conflict among the three lineages, possibly due to hybridization and/or incomplete lineage sorting (ILS).[32,33] However, it is still unclear how the underlying genetic mechanisms behind morphological divergence and phylogenetic incongruence, particularly the evolutionary origin of the tuber as a pivotal human food source, are linked to this controversial relationship.

Here, we resolve this discordance by showing that Petota originated through an ancient hybridization between Tomato and Etuberosum. The consistent genomic admixture, derived from parental lineages, was observed across 101 high-quality *de novo* genomes and 349 re-sequenced diploid potato genomes from 57 species within Petota. In addition, we show that this ancient hybrid speciation drove both tuberization and subsequent explosive adaptive species diversification of Petota, via complementary alleles and genes from its two parental lineages. The ability to form tubers appears to represent a key innovation, enabling the Petota ancestor to reproduce asexually and expand into seasonal high-elevation habitats. Collectively, we provide a deep revolutionary understanding of the potato genomes, the origins of the tubers, and the radiation of wild potato species.

RESULTS

Phylogenetic conflicts between Petota and its closest two lineages

To assess phylogenetic discordance, at the whole-genome scale, and identify potential relatives of Petota, we reconstructed a phylogeny of the family Solanaceae, using fully sequenced genomes across the phylogenetic range of Petota (six species covering Clade 1 + 2, Clade 3, and Clade 4), and Etuberosum, Tomato, and 19 additional Solanaceae lineages (21 species, each representing a single monophyletic lineage) (Table S1). A genome-scale phylogenetic analysis of these 27 Solanaceae species was performed using a maximum likelihood (ML) approach, based on 1-Mb sliding windows with a 200-kb step size. For 3,451 window trees constructed, the topology of Solanaceae species was mainly consistent with that reported in previous studies.[32–35] Extensive phylogenetic discordance was, however, observed among Petota, Tomato, and Etuberosum across windows with only 58.13% (2,006) of window trees supporting a sister relationship between Petota and Tomato, as found in the species tree (Figure 1B). Fbranch statistics also indicated that the gene flow was greater between Petota and Etuberosum (f4-ratio = 0.09) than between Petota and the other Solanaceae lineages (Figure S1D). This suggests that the closest relatives of Petota are Tomato and Etuberosum but that neither is strongly supported as sister.

We then collected a total of 128 high-quality diploid genomes (88 haplotype-resolved and 40 collapsed haplotype genomes) to address controversial relationships among the three lineages, including 101 from Petota (including 44 out of 107 diploid wild species), 15 from Tomato (including 12 out of 16 diploid wild species), 9 from Etuberosum (3 out of 3 diploid species), and 3 outgroups from the *Solanum* Clade II (Table S1), representing the most comprehensive diversity in wild potato genomic datasets developed to date. Of the 128 genomes, 58 haplotype-resolved genomes were assembled, covering 24 sequenced wild species from Petota, 2 species from Tomato, and all 3 species from Etubersosum. We generated an average of 29.54 Gb (~38-fold coverage compared with the estimated haploid potato genome size of approximately 800 Mb[36,37]) PacBio high-fidelity (HiFi) reads. Our initial *de novo* assemblies ranged from 590 to 905 Mb, with a mean contig N50 size of 19.84 Mb (Table S1).

(C) A rooted concatenated tree based on four-fold degenerate sites with site concordance factor (sCF) and site discordance factor (sDF) values shown by the pie chart on each node. In the pie charts: red represents sCF proportions averaged over 100 sites, blue indicates the sDF proportions for the first alternative topology, and gray represents sDF proportions for the remaining alternative topologies. The suffixes H1 and H2 denote haplotype 1 and haplotype 2 of the haplotype-resolved genomes, respectively.
See also Figure S1 and Table S1.

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

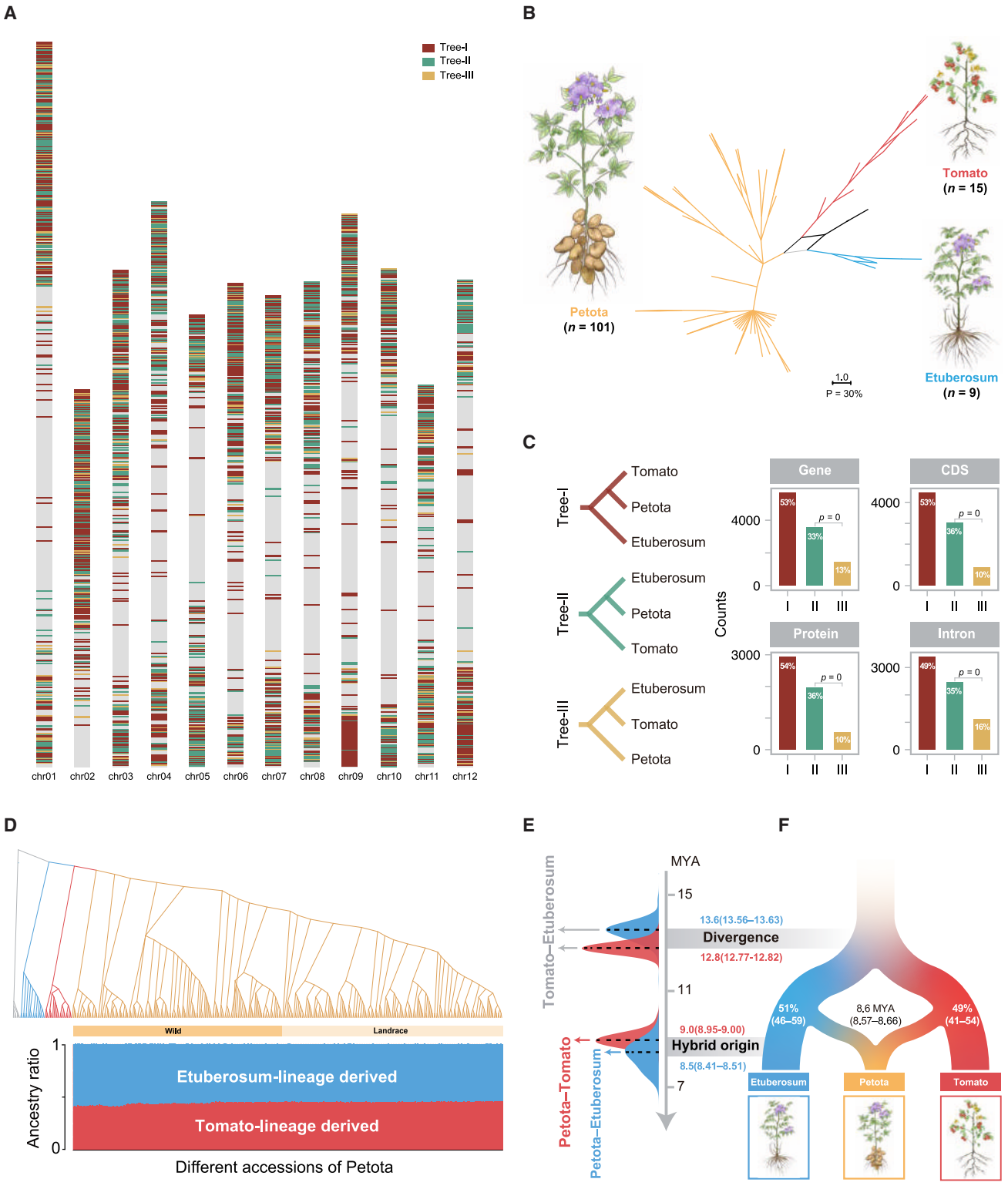


Figure 2. Phylogenomic analysis reveals the hybrid origin scenario

(A) The topology types are displayed for 100-kb non-overlapping windows. Each gray bar represents a chromosome, with DM v6.1 as the reference. Colored bands indicate the corresponding topologies of each 100-kb window.

(legend continued on next page)

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

These assemblies were then scaffolded and manually corrected in each diploid based on high-throughput chromatin conformation capture (Hi-C) contact information (~123-fold coverage), resulting in 58 haploid pseudo-chromosome assemblies, with an average anchor rate of 93.87% (Figure S1A; Table S1). The completeness of these 58 haplotypes was supported by k-mer spectra and benchmarking universal single-copy orthologs (BUSCO) analysis, with an average score of 97.76%, indicative of the high quality of these haplotype-resolved genomes (Figure S1B; Table S1). By integrating transcript alignment, *ab initio* prediction, and evidence of protein homology, the number of predicted genes in the 58 haplotypes ranged from 32,530 to 41,997 with an average BUSCO score of 95.44%.

We next performed whole-genome alignments among the 128 genomes by comparing to the potato reference genome (accession DM1-3 516 R44 v6.1 from *S. tuberosum* Group Phureja; hereafter referred to as DM v6.1)[36,37] The alignment rates within Petota ranged from 39.55% to 86.03%, indicating high levels of Petota genomic diversity. The Etuberosum accessions showed an average alignment rate of 32.66% to DM v6.1, which is higher than the Tomato accessions average of 23.58%, whereas the outgroups had an average alignment rate of only 11.98% (Figure S1C). Then, 3.7 Mb of 4-fold degenerate sites were extracted from the alignments to infer the comprehensive species tree using a concatenated approach. Petota, Tomato, and Etuberosum form deep splits, with Petota resolved as sister to Tomato, but with low site concordance factor (sCF, 34.89%) and high site discordance factors (sDF), including the 1st alternative quartet topology ratio of 38.83%, and the 2nd alternative quartet topology ratio of 26.28%, indicating that a minority of decisive alignment sites support this relationship (Figure 1C).

Our phylogenomic analyses partitioned Petota into three main nuclear clades: Clade 1 + 2, Clade 3, and Clade 4, as previously defined.[19] Although low sCFs were frequent within Petota with an average of 53.50%, species from Clade 1 + 2 comprised a monophyletic sublineage representing the basic branch of Petota, whereas Clade 3 and Clade 4 mutually formed a sister relationship, with mixing of a few species (Figure 1C). From this deep phylogeny of Petota, we assigned previously unclassified or uncertain species, such as *S. flahaultii* and *S. commersonii* to Clade 4, and *S. malmeanum* to Clade 3. Moreover, Clade 4 is divided into northern and southern subclades, with all cultivated accessions nested within the northern group. Most haplotype genomes clustered with each other, indicating that the sampled species typically represent "good" species.

Observable genetic mixture throughout the genomes of Petota

Hybridization is expected to result in an unequal distribution of genealogies, whereas ILS should produce a similar proportion and distribution of genealogies.[38] Phylogenetic reconstruction utilizing concatenated sequences assumes a singular evolutionary history, limiting the ability to estimate genealogical variation across the genomes.[39] We therefore reconstructed phylogenies for 100-kb non-overlapping sliding windows, using the above whole-genome alignments of 128 genomes, to explore the distribution of discordance throughout the genome. Of the 3,436 window trees constructed, 3,121 showed the three lineages are each monophyletic. The most common tree, Tree-I (1,588, 50.88%), was consistent with the species tree and the plastome phylogeny, placing Tomato as a sister to Petota (Figures 1C and S2A). The second-most common tree, Tree-II (1,162, 37.23%), placed Etuberosum sister to Petota. The third tree, Tree-III (371, 11.89%), supported a sister relationship between Tomato and Etuberosum and was significantly less frequent than Tree-I and Tree-II (two-tailed one-sample binomial test, $p = 0$). Tree-I and Tree-II topologies were evenly distributed across the genome of the potato, except in centromere regions, which exhibited a high degree of variation and were difficult to align, resulting in insufficient genetic variation for analysis (Figures 2A and S2C). These window trees were imported into Splitstree using the ConsensusNetwork method (threshold of 30%); results indicated a reticulate evolutionary history between the three lineages (Figure 2B).[40] A distance-based method, quantifying introgression via branch lengths (QuIBL), showed that the Delta Bayesian information criterion (BIC) value of Tree-III is lower than Tree-I and Tree-II, indicating that the hybridization was more likely a cause of the relationships shown within Tree-I and Tree-II than ILS (Figure S2B).[41] The Patterson's *D* (ABBA-BABA statistics) also detected hybridization-derived gene flow between Petota and Etuberosum ($D = 17.8\%$, $Z = 33.2$; Figure S2D).

Although window-based analysis provides a broad overview of evolutionary history, it does not represent actual evolutionary relationships as accurately as phylogenetic analyses based on gene trees. We therefore built a comprehensive gene repertoire to resolve the phylogenetic relationships of these three lineages, at the gene level, by clustering the 4,675,781 predicted gene models from the 128 genomes into 54,816 orthologous gene clusters, using the Markov clustering algorithm. The extensive synteny observed among Tomato, Etuberosum, and Petota, based on genome-wide synteny-constrained orthologous gene clusters, likely facilitated the formation of a hybrid lineage with

(B) Phylogenetic network analysis, based on ConsensusNetwork in SplitsTree (threshold level of 30%), indicates a reticulate evolutionary history of Tomato, Petota, and Etuberosum.

(C) Phylogenetic topologies and bar plots reflect that all counts of Tree-II were significantly higher than Tree-III in the gene, protein, CDS, and intron trees, by two-tailed one-sample binomial tests.

(D) Population genomic analysis of Petota. The phylogenetic tree on the top includes 349 Petota (golden), 22 Tomato (red), 21 Etuberosum (blue), and 6 outgroup accessions (gray). Ancestry component analysis of the corresponding Petota individuals is displayed at the bottom, with the blue and red stacked blocks representing the proportions deriving from Etuberosum and Tomato, respectively.

(E) The divergence time estimated from the concatenated CDSs based on the Tree-I and Tree-II genealogy scenarios, using the MCMCTree program.

(F) A simplified schematic illustrating the hybrid origin of Petota. The introgression probability and the hybridization time are estimated using the BPP program, with 95% confidence intervals shown below.

See also Figure S2 and Table S2.

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

ancestry dispersed across chromosomes from both parental lineages (Figure S2E). We next extracted 47,556 out of 54,816 clusters and classified them into four categories based on their frequency of occurrence within Petota: core clusters (present in all 101 genomes; 7,631; 16.05%), near core clusters (present in 70–100 individuals; 11,216; 23.58%), shell clusters (found in 2–69 individuals; 25,950; 54.57%), and accession-specific clusters (2,759; 5.80%) (Figure S2F).

By employing core and near core clusters that span most species in the three lineages, we identified 17,872 homologous gene groups and used them for phylogeny reconstruction using the ML method. For each gene group, phylogenies were reconstructed using their coding sequences (CDSs), protein sequences, and genic and intronic sequences, respectively. The generated trees exhibited three main topologies as before. Tree-I was most frequent (53.26% of gene trees, 53.70% of protein trees, 53.49% of CDS trees, and 48.61% of intron trees showed this topology); Tree-II was the next most common (33.33% of gene trees, 36.14% of protein trees, 36.18% of CDS trees, and 35.38% of intron trees); whereas Tree-III was the least frequent (13.41% of gene trees, 10.16% of protein trees, 10.32% of CDS trees, and 16.00% of intron trees). Tree-I and -II were significantly more frequent (two-tailed one-sample binomial test, all $p = 0$) than Tree-III (Figure 2C).

To exclude ILS as the source of these patterns, we simulated ILS under the species tree with branch lengths in coalescent units, based on the multispecies coalescent model. For each sequence type, the ratios of Tree-III to Tree-II were all significantly different from those simulated under the solely ILS scenario (labeled by gray curves), leading to the rejection of the solely ILS scenario (Figure S2G). Moreover, the gene flows quantified by ABBA-BABA statistics between Petota and different Tomato/Etuberosum accessions showed consistent signals (Figure S2H). Taken together, our results show that the phylogenetic conflicts detected among the three lineages are widespread across genomes in Petota and were caused by hybridization events from the ancestors of Tomato and Etuberosum.

Stability of genomic signals of a hybrid origin of Petota

Differentiation between a stable hybrid lineage, with a single ancient hybrid origin, and recent introgressions into its few species, requires observing consistent hybrid signals and stable parental ancestry tracts across all offspring, in the former, whereas the latter would likely result in only a subset of derived offspring displaying hybridization signals and wide variation in genome ancestry among individuals.[6,42] In order to discern between these possibilities, we conducted an extensive genomic assessment by collecting whole-genome resequencing data of 349 accessions (162 wild and 187 landrace accessions) from 55 diploid species belonging to Petota (Table S2). Petota exhibited higher nucleotide diversity ($\pi = 0.069$) compared with its two putative parental lineages (Tomato: $\pi = 0.060$; Etuberosum: $\pi = 0.029$) (Figure S2K). We further compared population genetic differentiation (F_{ST}) and genetic divergence (D_{XY}) between Petota members and species from its two putative parental lineages. In each case, F_{ST} and D_{XY} values were lower than those between the two parental lineage members ($F_{ST} = 0.53$ and $D_{XY} = 0.18$ between Petota and Tomato; $F_{ST} = 0.52$ and $D_{XY} = 0.15$ between Petota and Etuberosum; $F_{ST} = 0.69$ and $D_{XY} = 0.22$ between Tomato and Etuberosum) (Figures S2I and S2J). These patterns are consistent with the differentiation of hybrid lineages being approximately intermediate between the two parental lineages.

For each Petota accession, Loter and HyDe analyses were performed to quantify the genomic contributions of the two parental lineages (Etuberosum and Tomato).[43,44] We estimated the Euberosum ancestry proportion to be ~60%, and the Tomato ancestry proportion to be ~40% in all Petota species analyzed. The parental ancestry tract segments are consistent in Clade 1 + 2, Clade 3, and Clade 4 of Petota (Figures 2D, S2L, and S2M). The further f_d genomic scans with a 100-kb window show that the 330.4 Mb region of the potato genome exhibits consistent excess allele sharing with the Etuberosum members. These genome-wide signatures of admixture are highly overlapping among the three main clades of Petota with an average correlation coefficient of 0.75, suggesting that introgression events from the ancestor of Etuberosum likely occurred at the origin of Petota (Figures S2N and S2O).

Dating the hybrid origin of Petota

In the hybrid origin scenario, the estimated divergence times of the hybrid lineage, with respect to each of the putative parental lineages, should be nearly equal, and obviously younger than the divergence time of the parental lineages.[45] Nearly equal divergence times would be expected between all three lineages if ILS was solely responsible for the observed phylogenetic conflicts. To test whether Petota is of ancient hybrid origin or if its species received recent introgression from Etuberosum, we evaluated the divergence time among the three lineages by partitioning the coding region of Petota into three topology categories. The concatenated CDSs of each topology were estimated by the MCMCTree. The estimated divergence times of Petota/Tomato (~9.0 million years ago [MYA], 95% confidence interval 8.95–9.00 MYA) and Petota/Etuberosum (~8.5 MYA, 95% confidence interval 8.41–8.51 MYA) were relatively close and much younger than that of the two parental lineages (Tomato/Etuberosum: ~12.8–13.6 MYA) (Figure 2E).

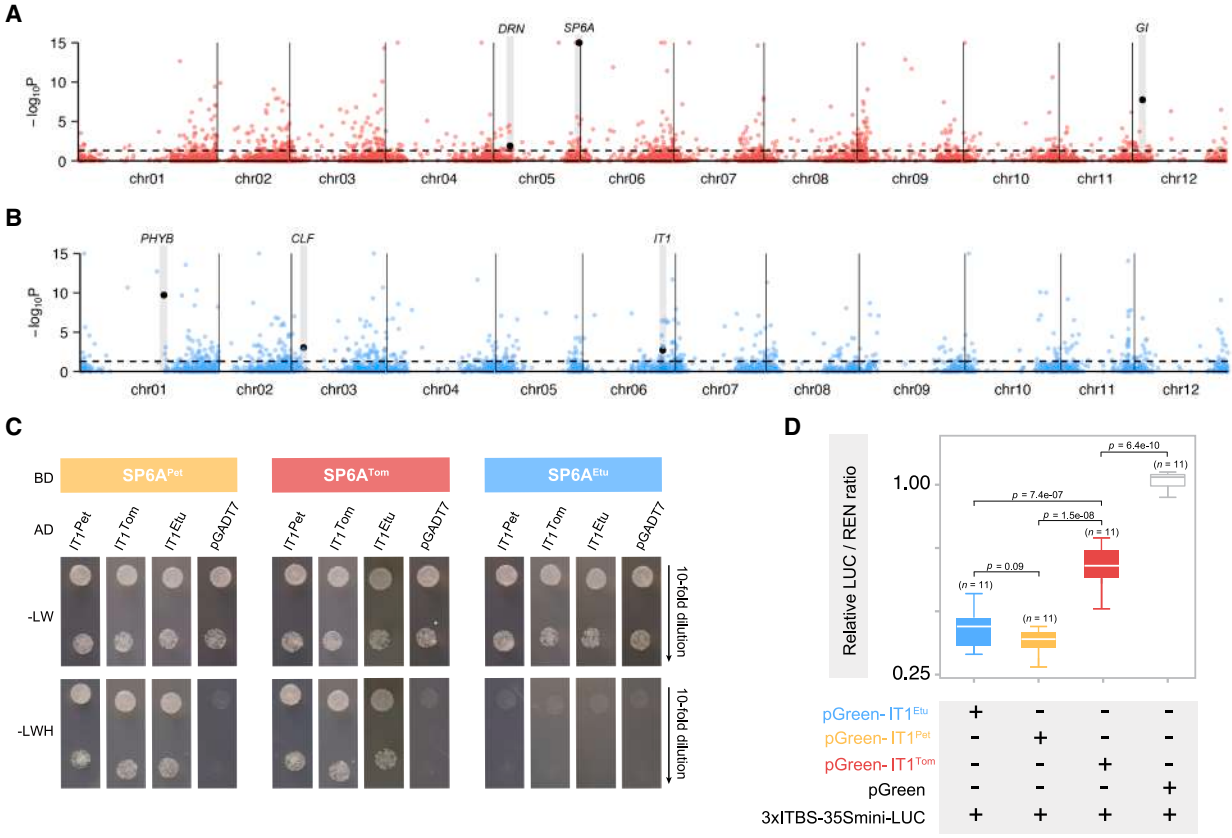
We further applied the multispecies coalescent-with-introgression (MSCi) model to test the hybrid origin scenario for Petota, using the haplotype-resolved genomes from outgroup, Tomato, Etuberosum, and Petota. The estimated hybridization time (τ) of Petota, approximately 8.6 MYA with a 95% high posterior density (HPD) interval of 8.57–8.66 MYA, coincides closely with the above divergence time calculated by MCMCTree, with a 51.26% (95% HPD: 45.80%–58.69%) contribution from Tomato and 48.74% (95% HPD: 41.31%–54.20%) contribution from Etuberosum (Figure 2F). These coalescent-based estimates are consistent with Petota being a hybrid lineage formed through admixture between Tomato and Etuberosum.

Genetic basis of tuber formation in Petota

Hybrid speciation refers to the genetic mixture, between distinct species or the ancestors of current lineages with multiple species, that may lead to prezygotic RIs through the production of morphological traits.[7–10] Both Petota and Etuberosum species produce underground stems, referred to as stolons or rhizomes,

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

CellPress
OPEN ACCESS



(legend on next page)

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

but only the stolons of Petota species swell at their tips to form tubers (Figure 1A). We hypothesize that the formation of the tuber, as an innovative organ in Petota, represented a key innovation that may have arisen via hybridization and subsequently triggered adaptive species diversification. In such a scenario, genes involved in tuberization are expected to undergo rapid evolution, due to natural selection, with diverged alleles potentially becoming alternately fixed in the hybrid offspring lineage.[10] We therefore searched for positively selected genes (PSGs) from the 17,872 homologous gene groups in our dataset of 128 genomes (see above), leading to the detection of 229 PSGs derived from Etuberosum and 269 PSGs from Tomato (Table S3), which were under natural selection in the two parental lineages prior to interspecific hybridization and were alternately inherited by Petota during its origin.

Some of these PSGs had been previously reported to be involved in tuberization, with induction by various exogenous and endogenous signals.[46,47] We observed a mosaic inheritance pattern in tuberization-related genes (Figures 3A, 3B, S3A, and S3B). For example, two light signal response genes, *GIGANTEA* (*GI*) and *SELF-PRUNING 6A* (*SP6A*), both derived from Tomato, play essential roles in tuberization.[48,49] Notably, *SP6A*, an ortholog of the *Arabidopsis* vascular-mobile florigen, *FLOWERING LOCUS T* (*FT*), acts as a key regulator of tuber formation, whereas another light signal gene, *PHYTOCHROME B* (*PHYB*), derived from Etuberosum, functions upstream of *SP6A* to control tuberization.[49,50] In addition to light signaling, transcription factors (TFs) are also critical for tuberization. For instance, *Identity of Tuber 1* (*IT1*)/*BRANCHED1b* (*BRC1b*), a TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR (TCP) family TF derived from Etuberosum, and *DORNRONSCHEN* (*DRN*)/*LEAFLESS* (*LFS*), an *ERF*/*AP2* family TF derived from Tomato, regulate axillary meristem initiation and specialized stolon balancing to control potato tuberization.[51,52] Furthermore, the Polycomb group (PcG) gene *CURLY LEAF* (*CLF*), derived from Etuberosum, modulates tuberization by balancing the deposition of H3K27me3 and H3K4me3.[47,53] The mosaic nature of the regulatory network implies that natural selection recruited alternately inherited parental genes (brought together by hybridization) for specific pathways that contributed to tuberization.

Next, we confirmed the crucial functions of two PSGs associated with tuberization, *SP6A* and *IT1*, which were derived from Tomato and Etuberosum, respectively. *SP6A* is not expressed in any tissues in the Etuberosum species, whereas *IT1* is mainly expressed in rhizome tissues of the Etuberosum species and stolon tissues of the Petota species (Figures S3C and S3D). To explore their hypothesized role in tuberization, we conducted yeast two-hybrid assays and split luciferase assays using orthologous IT1 and SP6A proteins from Petota (*S. tuberosum*), Tomato (*S. pimpinellifolium*), and Etuberosum (*S. etuberosum*) (Figures 3C and S3G). These results collectively support that $SP6A^{Pet}$ (Petota-derived SP6A protein) originated from the $SP6A^{Tom}$ (Tomato-derived SP6A protein) ortholog, which can form an SP6A-IT1 complex, whereas a truncated phosphatidylethanolamine-binding protein (PEBP) domain in $SP6A^{Etu}$ (Etuberosum-derived SP6A protein) would eliminate its interaction with IT1 (Figure S3E). Notably, there are 14 non-synonymous mutations fixed in Petota, 13 of which are derived from Etuberosum, with 4 located in the TCP domain region (Figure S3F). Next, we investigated whether this sequence difference in IT1 affects its TF activity by conducting luciferase reporter assays (Figures 3D and S3H). Our results indicated that the transcriptional activity of $IT1^{Pet}$ (Petota-derived IT1 protein) is not significantly different (two-tailed Student's *t* test, $p = 0.09$) from that of $IT1^{Etu}$ (Etuberosum-derived IT1 protein), suggesting their activities are comparable. However, both $IT1^{Pet}$ and $IT1^{Etu}$ exhibit activities that are significantly different from that of $IT1^{Tom}$ (Tomato-derived IT1 protein) (two-tailed Student's *t* test, $p = 1.5e{-}08$ and $p = 7.4e{-}07$, respectively), supporting the hypothesis that $IT1^{Pet}$ originated from Etuberosum and contributed to tuberization. These findings indicate that a combination of tuberization-associated PSGs from Tomato and Etuberosum played a crucial role in the origin of tuber formation in Petota.

Additionally, we selected another two genes, *DRN* and *CLF* with hybrid recombination to further examine the function of PSGs in tuberization. In the *DRN* gene of Petota, a total of 23 fixed mutations were identified, with 9 originating from Etuberosum and 14 from Tomato (Figure S3I). Similarly, in the *CLF* gene of Petota, 23 fixed mutations were observed, of which 17 were derived from Etuberosum and 6 from Tomato (Figure S3K). We then generated knockout mutants using CRISPR-Cas9-based genome editing in the wild potato *S. chacoense* (Figures S3J and S3L). In these *drn* mutants, we observed that aerial axillary branches were converted into ectopic stolon branches (but no tubers formed in aerial stolons) (Figure 3E); furthermore, below-ground stolons were produced in *drn* mutants, which formed smaller tubers than those in wild-type plants, limiting their function as organs for vegetative reproduction. In contrast, *clf* mutants were dwarf and unable to initiate underground stolons, consequently losing the ability to generate tubers (Figure 3F). Together, these two mutants—one causing ectopic organ formation and the other leading to organ loss—suggest

Figure 3. PSGs and experimental validation of positively selected tuberization-associated genes

(A) The gene variant comparisons between Tomato/Petota and Etuberosum by the Hudson-Kreitman-Aguadé (HKA) test.[10] The results of genes are displayed (red dots) in the plot when Petota is grouped with Tomato in the CDS topology. A gene highlighted by a gray bar represents its likely association with tuberization.

(B) The gene variant comparisons between Etuberosum/Petota and Tomato by the HKA test. The results of genes are displayed (blue dots) in the plot when Petota is grouped with Etuberosum in the CDS topology. A gene highlighted by a gray bar represents its likely association with tuberization.

(C) Validation in a yeast two-hybrid assay. Yeast cells were grown on the selective medium of SD3 (−LWH) without Leu (L), Trp (W), and His (H), and cells on SD2 (−LW) served as growth control.

(D) Validation in a dual-luciferase reporter assay. Boxplots show the median, box edges denote the 25th and 75th percentiles, and whiskers denote the maximum and minimum data points within 1.5× the interquartile range outside the box edges. $n = 11$. Statistical significance was determined by a two-tailed Student's *t* test.

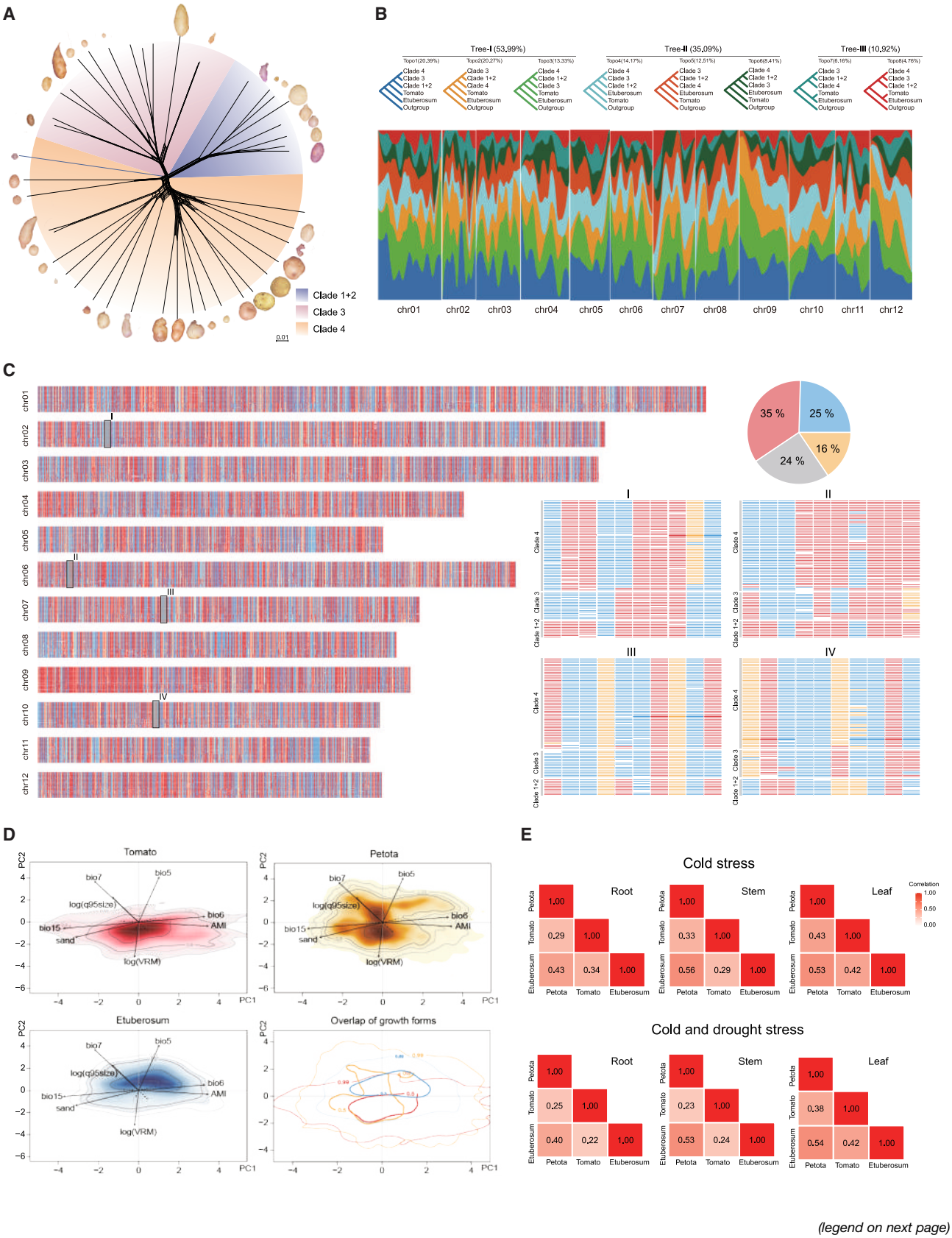
(E) Phenotypes of the *DRN* gene knockout mutant.

(F) Phenotypes of the *CLF* gene knockout mutant.

See also Figure S3 and Table S3.

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

CellPress
OPEN ACCESS



(legend on next page)

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

that the alternate inheritance of highly divergent alleles of key loci from Tomato and Etuberosum contributed to tuber formation through hybrid combinations and recombinations. This critical innovation, which facilitates vegetative reproduction, likely allowed for the stabilization of beneficial hybrid gene combinations and recombinations, potentially including those that confer hybrid vigor.

Rapid species radiation and environmental adaptation of Petota

Ancient hybridization may enable rapid species radiation by introducing innovative traits that help adapt to various niches, along with recombination and sorting of hybridization-derived polymorphisms.[1–4] To investigate this possibility in Petota, we constructed phylogenetic networks using NeighbourNet methods, based on a genomic distances matrix derived from the variant call format (VCF) file containing 44 wild potato species (Figure 4A). This network also identified Clade 1 + 2, Clade 3, and Clade 4, but the central knot and basal position exhibited limited structural differentiation, which is consistent with the low support observed in these nodes in coalescence reconstruction, indicating a basal polytomy within Petota.

The species within Clade 4 displayed a more complex network structure than those in Clade 1 + 2 and Clade 3, consistent with the strong signature of gene flow, especially among the species of the southern group (Figure S4A). Significant Patterson's *D* values ($p < 0.01$) indicative of extensive introgression were inferred in 12,912 of the 17,296 possible trios. Species within Clade 4 also exhibit higher gene flows with significant *D*-statistic values (mean value: 0.083) than the species in Clade 1 + 2 (mean value: 0.062) and Clade 3 (mean value: 0.048). The admixture analysis also grouped the species of Petota into three main clades when $K = 4$, consistent with the genomic principal-components analysis (PCA) (Figures S4B and S4E). *S. sogarandinum*, *S. huancabambense*, and *S. stipuloideum* formed separate sublineages from Clade 4, with admixture components from Clade 1 + 2, Clade 3, and Clade 4.

To further quantify the distribution of three main clades across the genome, the topology weighting by the iterative sampling of subtrees (TWISST) analysis was conducted, which also supported Tree-I (Topo1, 2, and 3) and Tree-II (Topo4, 5, and 6) comprising 89.08% of the eight most common trees (Figure 4B). Additionally, pervasive conflict signals were detected among clades of Petota: Topo1 (20.39%), Topo4 (14.17%), and Topo8 (4.76%) grouped Clade 3 with Clade 4, consistent with the species tree; Topo2 (20.27%), Topo5 (12.51%), and Topo7 (6.16%) grouped Clade 1 + 2 with Clade 3; and Topo3 (13.33%), Topo6 (8.41%) grouped Clade 1 + 2 with Clade 4.

These admixture genealogies and mosaic-like ancestry patterns were further validated in the 17,872 homologous gene groups with widespread discordance. The multispecies coalescent method utilized CDSs, protein sequences, genic, and intronic sequences from the 17,872 homologous groups. In the Astral tree reconstructed by CDSs, protein, and genic sequences, Clade 3 is sister to Clade 4, which is consistent with the species tree, whereas Clade 3 is sister to Clade 1 + 2 based on intronic sequences (Figure S4C). We determined that 35% of homologous gene groups were derived from Tomato (all genomes are labeled red in a gene group), whereas 25% were derived from Etuberosum (all genomes are labeled blue in a gene group) (Figure 4C). Notably, 40% of the homologous gene groups showed extreme differentiation from other derivatives among species within Petota. Of these, 16% of homologous gene groups were associated with the deeply divergent haplotype groups derived from the donor parental lineage (all genomes are labeled golden in a gene group), and 24% were polymorphic within Petota, with the Clade 1 + 2, Clade 3, and Clade 4 derived from different parental species. The PCA analysis based on these gene topologies also clustered the species of Petota into three main clades (Figure S4F). In the exceptionally differentiated bi-allelic SNPs between Tomato and Etuberosum ($F_{ST} = 1$), we show that ~41.44% (406,041 sites) of differentiated parental sites from Etuberosum are fixed in Petota, which is slightly higher than ~35.93% (352,049 sites) from Tomato (Figure S4G). Approximately 22.63% (221,816 sites) of differentiated parental sites remain polymorphic in different clades within Petota. These results imply that the recombination and subsequent sorting of differentiated parental sites likely represent an important contribution to the divergence of these three clades.

We further examined whether accelerated species diversification occurred in Petota as a result of the production of the tuber organ and sorting and recombination of hybridization-derived polymorphisms. Petota comprises 107 wild species, all characterized by possession of tubers, and exhibits the fastest species-level radiation in *Solanum* Grade I, with an estimated per-lineage

Figure 4. Hybrid recombination contributes to the radiation and adaptation of Petota

(A) Genomic neighbor network constructed from the distance matrix of the SNP sites from the whole-genome alignment.

(B) The eight most common topologies were identified from 100-kb non-overlapping window trees. The bottom plot shows the topology weighting across the 12 chromosomes with loess smoothing (span = 200 bp).

(C) Parental-derived inference of the homologous gene groups across the 101 Petota genomes, based on their topological relationships. Four highlighted regions are zoomed in at the bottom right corner, labeled Clade 1 + 2, Clade 3, and Clade 4. Each column represents a homologous gene group, and each row represents each Petota genome. The pie chart illustrates the proportions derived from Tomato (red), Etuberosum (blue), Petota (golden), and the mixed region within Petota (gray).

(D) PCA kernel density estimate of species within Tomato, Etuberosum, and Petota, and the overlap of kernel density estimates from all three growth forms in PCA space. The interpretation of the selected environmental variables: seasonal precipitation (bio15), temperature annual range (bio7), minimum temperature of coldest month (bio6), maximum temperature seasonality (bio5), annual moisture index (AMI), log of 95th quantile for fire size (log[q95size]), log of vector ruggedness metric, log of topographic heterogeneity (log[VRM]), and proportion of sand particles (>0.05 mm) in the fine earth fraction at a depth of between 0–5 cm (sand).

(E) Correlation analysis of RNA sequencing (RNA-seq) data from three species and tissue types under cold stress and combined cold-drought stress conditions. See also Figure S4.

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

CellPress
OPEN ACCESS

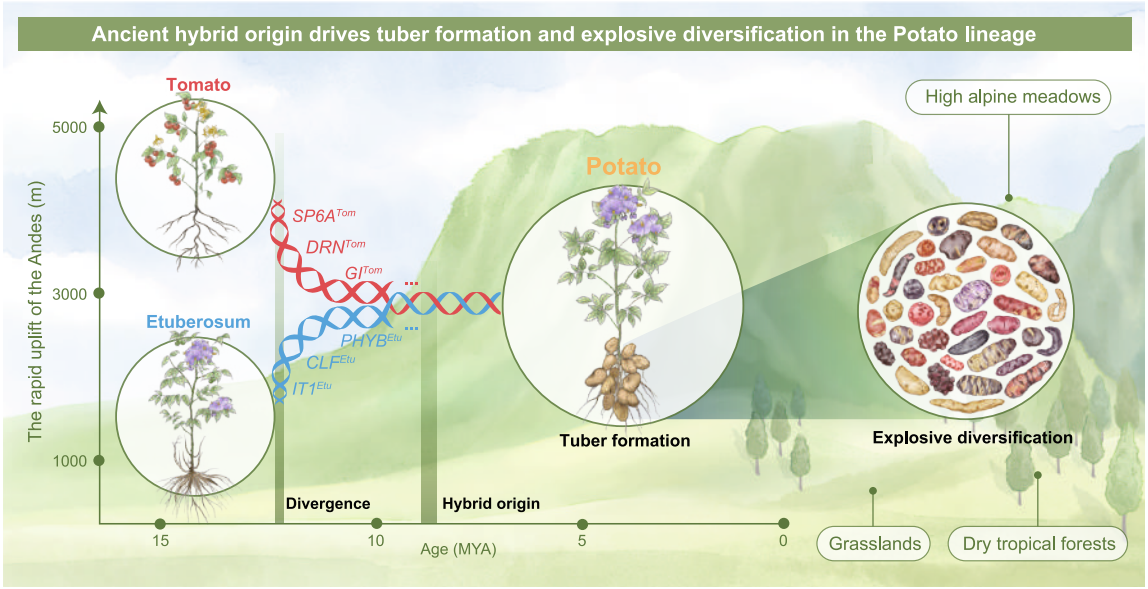


Figure 5. Hybrid origin evolutionary scenario for the potato lineage

Petota originated from an ancient hybridization event between Tomato and Etuberosum around 8–9 MYA, coinciding with the rapid uplift of the Andes (6–10 MYA). Members of Petota alternately inherited parental genes, which contributed to the formation of tubers and triggered explosive species diversification, enabling the species from Petota to adapt to and colonize various environments.

rate of ca. 0.53 lineages per million years (Myr^{-1}). This rate is higher than that of Etuberosum (0.39 lineages Myr^{-1}), Tomato (0.44 lineages Myr^{-1}), and the global *Solanum* mean (0.38 lineages Myr^{-1}) (Figure S4D).

Hybrid lineages often also occupy broader or different niches.[54] A previous study indicated that geophytes of the *Solanum* occupy distinct ecological niches compared with the non-geophyte species, influenced by eight bioclimatic environmental variables.[24] We extracted these eight variables for 4,703 occurrence records of the three species in Etuberosum, all 16 wild species in Tomato, and the 107 non-cultivated species in Petota. The highest diversity of wild species of Petota is found in high-elevation, cold-adapted montane habitats in central Mexico and the central Andes. Species of Petota have a much wider geographical distribution (Figure 1A) and tend to occupy a broader range of environmental conditions than species of either Tomato or Etuberosum.

Habitats occupied by Petota species are characterized by strong precipitation seasonality, and range from different types of grasslands to seasonally dry tropical forests to high alpine meadows.[30,55,56] The median niche breadth of Petota (0.18) is broader than that of its parental lineages, Tomato (0.09), and Etuberosum (0.05), according to ecological niche model results. The suitable climatic conditions of Petota are drier than species of Etuberosum but on average cooler and with greater precipitation (higher annual moisture index) than species of Tomato (Figures 4D and S4H). To compare the cold and dry adaptation mechanism between the hybrid and parental lineages, we conducted tolerance experiments on *S. tuberosum* (Petota), *S. lycopersicum* (Tomato), and *S. palustre* (Etuberosum). Correlation analyses of transcriptomic data revealed that the expression patterns of *S. tuberosum* exhibit a stronger correlation in expression patterns with *S. palustre* than with *S. lycopersicum* in stem, leaf, and root tissues, under both cold and cold-drought stresses (Figures 4E, S4I, and S4J). Differential gene expression analysis showed that the number of up- and downregulated genes is greater in *S. lycopersicum* than in *S. palustre* under cold conditions, and the expression pattern of *S. tuberosum* is similar to that of *S. palustre*, which is consistent with their ecological niche characteristics. Taken together, the innovative tuber organ, along with the sorting and recombination of hybridization-derived polymorphisms, likely facilitated the initial survival of hybrid offspring, including those with low fertility, and triggered the explosive species diversification of Petota by enabling the occupation of broader ecological niches as well as their establishment in distinct habitats, such as higher elevations with colder climates in the Andes.

DISCUSSION

Hybrid speciation is increasingly recognized as an "evolutionary catalyst," fostering the emergence of evolutionary innovation and the expansion of ecological niches.[1,54] However, a mechanistic understanding of how ancient hybrid speciation promotes phenotypic innovation and adaptive species radiation, within species-rich lineages, is only beginning to emerge. Here, our study provides genomic and functional evidence that Petota is of hybrid origin, resulting in the innovative trait of tuber formation and subsequent colonization of a broad ecological niche with further extensive species diversification (Figure 5). Phylogenetic analyses reveal that the closest relatives of Petota are Etuberosum and Tomato with substantial discordance among gene trees or genomic windows. All members of Petota exhibit a stable mixed genomic ancestry, derived from Etuberosum and Tomato.

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

We estimated that Tomato and Etuberosum diverged ca. 13–14 MYA, and that hybridization between them resulted in the origin of Petota around ca. 8–9 MYA (Figure 2F), coinciding with the rapid uplift of the Andes (6–10 MYA).[57] Therefore, this ancient homoploid hybrid speciation event between ancestors of Tomato and Etuberosum, which gave rise to the ancestor of Petota, may have taken place under geological oscillations and appearance of distinct niches, generating allelic combinations/recombinations and innovative phenotypes for further natural selection. This is consistent with the mosaic-like genetic patterns throughout the entire genomes of all sampled species in Petota.

Most importantly, we obtained evidence that the alternate inheritance of highly divergent alleles of certain key genes from Tomato and Etuberosum, respectively, may have established an innovative interaction network that contributed to tuber formation during ancient hybrid speciation. This critical innovation, which favors vegetative reproduction, likely aided the initial survival of hybrid offspring with low fertility due to postzygotic RIs between the divergent ancestors of Tomato and Etuberosum. It also facilitated their establishment in distinct habitats, such as higher elevations with colder climates in the Andes.[30,55] These distinct niches may have further enhanced prezygotic RIs of hybrid offspring by reducing backcrossing with either parent, allowing them to form an independently evolved Petota by recovering fertility during subsequent evolution.

Additionally, the emergence of more effective asexual reproduction through tuber formation, along with the sorting and recombination of hybridization-derived polymorphisms, may have triggered subsequent rapid species diversification of Petota, which now comprises 107 wild species in a broad range of environments.[56] More generally, our findings suggest that hybrid speciation can involve both the inheritance of alternate parental genes and the development of key trait innovations through allele combination and recombination. In the case of Petota, this process ultimately led to the ability to form tubers and, consequently, to the origin of the cultivated potato, one of humankind's most important foodstuffs.

Limitations of the study

While ancient population structure within a single species can mimic hybridization signals,[58] our coalescent analyses suggest that Tomato and Etuberosum ancestors had diverged ~5 MYA (Figure 2F), accumulating highly diverged alleles prior to hybridization. This seems to support hybridization between independently evolving species (lineages) over ILS, although further evidence is needed. Furthermore, the combination of highly divergent alleles at different loci, inherited alternately from the ancestors of Etuberosum and Tomato, may have contributed to postzygotic RIs in the Petota ancestor during ancient hybrid speciation. This inference aligns with the Bateson-Dobzhansky-Muller incompatibility model, which suggests that divergent alleles, at two loci, can lead to postzygotic RIs.[1,10,14] However, the specific genes responsible for these postzygotic RIs remain to be identified. It also remains unclear how the Petota ancestor regained the ability for normal sexual reproduction, after experiencing probable hybrid breakdown due to postzygotic RIs, following ancient hybridization, although the temporary survival of hybrids may have been supported by formed tubers. Finally, the specific pathways established through hybridization-triggered recombination for tuberization may alter the original functions of related genes, resulting in pleiotropic roles, as seen with *DRN* and *CLF* (Figures 3E and 3F), where their mutants affect both tuber formation and other phenotypes. Therefore, confirming the origin of tuber formation through artificial hybridization between Etuberosum and Tomato species, identifying the key genes contributing to postzygotic RIs, and clarifying multiple effects of tuber-related genes, remains a critical area for future research.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to, and will be fulfilled by, the lead contact, Sanwen Huang (huangsanwen@caas.cn).

Materials availability

Materials generated in this study are available from the lead contact.

Data and code availability

- All PacBio sequence data, Hi-C data, and RNA data generated in this study have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject accession number PRJNA1173794. Genome sequences and gene annotations were uploaded to the Solomics Database Potato Repositories website (http://solomics.agis.org.cn/potato/) and Figshare (https://doi.org/10.6084/m9.figshare.28942299).[59] The whole-genome resequencing data were downloaded from NCBI under BioProject accession PRJNA839598. Publicly available sequencing data used in this study were downloaded from NCBI with BioProject accession numbers PRJNA839598, PRJNA754534, PRJNA733299, and PRJNA809001. Additional information required to reanalyze the data reported in this study is available from the lead contact upon request.
- All codes were deposited at https://github.com/zhangzhiyangcs/hybrid_origin_of_the_potato.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We dedicate this paper to the memory of Richard J. Abbott, whose pioneering work in homoploid hybrid speciation laid the cornerstone for this research. We thank Chunzhi Zhang, Jianjian Qi, Zhuangwei Hou, Jialiang Li, Meiyan Shi, Yafei Mao, Liming Cai, Pei Wang, Dié Tang, Yao Zhou, Susy Echeverría-Londoño, and Nils Stein for their discussions and comments on the project. This work was supported by Guangdong Major Project of Basic and Applied Basic Research (2021B0301030004), the National Natural Science Foundation of China (32488302), the Science and Technology Major Project of Xizang (XZ202402ZD0005), the National Natural Science Foundation of China (32422053 and 31590821), the National Natural Science Foundation of China (32301896 and 32422079), Occurrence data collection for the lineages used in this study was funded by the US National Science Foundation Planetary Biodiversity Inventory project "PBI Solanum: a worldwide treatment" (DEB-0316614), Basic Research Center for Agricultural Frontiers and Interdisciplinary Sciences (BRC-AFIS), Innovation Program of Chinese Academy of Agricultural Sciences (CAAS-BRC-AFIS-2025-02), and Agricultural Science and Technology Innovation Program (CAAS-ZDRW202404).

AUTHOR CONTRIBUTIONS

S.H., Z.Z., and J.L. conceived and designed the research. S.H., J.L., L.H. R., and S.K. headed and managed the project. Z.Z., Z.M., and Y.D.

contributed to genome assembly, genome annotation, and synteny analyses. P.Z. and Z.L. contributed to the molecular experiments. Z.Z. and Y.D. contributed to the phylogenetic analysis, the population genetic analysis, and the network analyses. Z.Z., Z.W., and Y.D. contributed to the identification of positively selected genes. Z.Z., Y.D., and Z.W. contributed to the divergence time estimates. Z.Z., Z.B., and L.C., contributed to whole-genome alignment. Z.Z., E.G., and T.S. performed the ecological analysis. S.L. and W.L. participated in the material preparation. N.W., L. L., Y.W., Y.H., and Q.L. assisted in bioinformatics analyses. J.Z., Y.J., H. W., Y.Z., and K.Y. coordinated the project. Z.Z., Y.D., P.Z., and Z.W. wrote the manuscript. S.H., J.L., L.H.R., S.K., T.S., and W.J.L. revised the manuscript. S.H. supervised the research. All authors read, edited, and approved the manuscript.

DECLARATION OF INTERESTS

The Agricultural Genomics Institute at Shenzhen has one pending patent application based on this study. S.H. is a member of the Cell advisory board.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Plant materials and growth conditions
- METHOD DETAILS
  - Plant materials and sequencing
  - Genome *de novo* assembly and evaluation
  - Annotation of repetitive elements and protein-coding genes
  - Whole-genome alignments and variant calling
  - Identification of homologous gene groups
  - Phylogenetic analyses
  - Network analyses
  - Divergence time estimates
  - Population genetic analysis of 349 Petota re-sequenced genomes
  - Identification of positively selected genes
  - Yeast two-hybrid
  - Firefly luciferase complementation imaging assay
  - Dual-luciferase reporter assay
  - CRISPR/Cas9-mediated knockout of *StCLF* and *StDRN* genes
  - Maps and ecological niche identification
  - Gene expression under the cold treatment
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2025.06.034.

Received: September 25, 2023
Revised: January 3, 2025
Accepted: June 24, 2025

REFERENCES

1. Mallet, J. (2007). Hybrid speciation. Nature *446*, 279–283. https://doi.org/10.1038/nature05706.
2. Taylor, S.A., and Larson, E.L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. Nat. Ecol. Evol. *3*, 170–177. https://doi.org/10.1038/s41559-018-0777-y.
3. Marques, D.A., Meier, J.I., and Seehausen, O. (2019). A combinatorial view on speciation and adaptive radiation. Trends Ecol. Evol. *34*, 531–544. https://doi.org/10.1016/j.tree.2019.02.008.
4. Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C.A., Buggs, R., et al. (2013). Hybridization and speciation. J. Evol. Biol. *26*, 229–246. https://doi.org/10.1111/J.1420-9101.2012.02599.x.
5. Rosser, N., Seixas, F., Queste, L.M., Cama, B., Mori-Pezo, R., Kryvokhyzha, D., Nelson, M., Waite-Hudson, R., Goringe, M., Costa, M., et al. (2024). Hybrid speciation driven by multilocus introgression of ecological traits. Nature *628*, 811–817. https://doi.org/10.1038/s41586-024-07263-w.
6. Wang, Z., Kang, M., Li, J., Zhang, Z., Wang, Y., Chen, C., Yang, Y., and Liu, J. (2022). Genomic evidence for homoploid hybrid speciation between ancestors of two different genera. Nat. Commun. *13*, 1987. https://doi.org/10.1038/s41467-022-29643-4.
7. Zhang, B.-L., Chen, W., Wang, Z., Pang, W., Luo, M.-T., Wang, S., Shao, Y., He, W.-Q., Deng, Y., Zhou, L., et al. (2023). Comparative genomics reveals the hybrid origin of a macaque group. Sci. Adv. *9*, eadd3580. https://doi.org/10.1126/sciadv.add3580.
8. Meier, J.I., Marques, D.A., Mwaiko, S., Wagner, C.E., Excoffier, L., and Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nat. Commun. *8*, 14363. https://doi.org/10.1038/ncomms14363.
9. Meier, J.I., McGee, M.D., Marques, D.A., Mwaiko, S., Kishe, M., Wandera, S., Neumann, D., Mrosso, H., Chapman, L.J., Chapman, C.A., et al. (2023). Cycles of fusion and fission enabled rapid parallel adaptive radiations in African cichlids. Science *381*, eade2833. https://doi.org/10.1126/science.ade2833.
10. Wang, Z., Jiang, Y., Bi, H., Lu, Z., Ma, Y., Yang, X., Chen, N., Tian, B., Liu, B., Mao, X., et al. (2021). Hybrid speciation via inheritance of alternate alleles of parental isolating genes. Mol. Plant *14*, 208–222. https://doi.org/10.1016/j.molp.2020.11.008.
11. Seehausen, O. (2004). Hybridization and adaptive radiation. Trends Ecol. Evol. *19*, 198–207. https://doi.org/10.1016/j.tree.2004.01.003.
12. Brown, M.R., Hollingsworth, P.M., Forrest, L.L., Hart, M.L., Leitch, I.J., Jones, L., Ford, C., de Vere, N., and Twyford, A.D. (2023). Genetic factors predict hybrid formation in the British flora. Proc. Natl. Acad. Sci. USA *120*, e2220261120. https://doi.org/10.1073/pnas.2220261120.
13. Stull, G.W., Pham, K.K., Soltis, P.S., and Soltis, D.E. (2023). Deep reticulation: the long legacy of hybridization in vascular plant evolution. Plant J. *114*, 743–766. https://doi.org/10.1111/tpj.16142.
14. Wu, H., Wang, Z., Zhang, Y., Frantz, L., Roos, C., Irwin, D.M., Zhang, C., Liu, X., Wu, D., Huang, S., et al. (2023). Hybrid origin of a primate, the gray snub-nosed monkey. Science *380*, eabl4997. https://doi.org/10.1126/science.abl4997.
15. Long, Z., and Rieseberg, L.H. (2024). Documenting homoploid hybrid speciation. Mol. Ecol. e17412. https://doi.org/10.1111/mec.17412.
16. Burrell, M.M. (2003). Starch: the need for improved quality or quantity—an overview. J. Exp. Bot. *54*, 451–456. https://doi.org/10.1093/jxb/erg049.
17. Spooner, D.M., Anderson, G.J., and Jansen, R.K. (1993). Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes, and pepinos (Solanaceae). Am. J. Bot. *80*, 676–688. https://doi.org/10.1002/j.1537-2197.1993.tb15238.x.
18. Rodriguez, F., Wu, F., Ané, C., Tanksley, S., and Spooner, D.M. (2009). Do potatoes and tomatoes have a single evolutionary history, and what proportion of the genome supports this history? BMC Evol. Biol. *9*, 191. https://doi.org/10.1186/1471-2148-9-191.
19. Tang, D., Jia, Y., Zhang, J., Li, H., Cheng, L., Wang, P., Bao, Z., Liu, Z., Feng, S., Zhu, X., et al. (2022). Genome evolution and diversity of wild and cultivated potatoes. Nature *606*, 535–541. https://doi.org/10.1038/s41586-022-04822-x.
20. Gagnon, E., Hilgenhof, R., Orejuela, A., McDonnell, A., Sablok, G., Aubriot, X., Giacomin, L., Gouvêa, Y., Bragionis, T., Stehmann, J.R., et al. (2022). Phylogenomic discordance suggests polytomies along the

backbone of the large genus *Solanum*. Am. J. Bot. *109*, 580–601. https://doi.org/10.1002/ajb2.1827.

21. Sarkinen, T., Bohs, L., Olmstead, R.G., and Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): A dated 1000-tip tree. BMC Evol. Biol. *13*, 214. https://doi.org/10.1186/1471-2148-13-214.
22. Peralta, I.E., Spooner, D.M., and Knapp, S. (2008). Taxonomy of wild tomatoes and their relatives (Solanum sect. Lycopersicoides, sect. Juglandifolia, sect. Lycopersicon; Solanaceae). Syst. Bot. Monogr. *84*, 1–186.
23. Spooner, D.M., Ghislain, M., Simon, R., Jansky, S.H., and Gavrilenko, T. (2014). Systematics, diversity, genetics, and evolution of wild and cultivated potatoes. Bot. Rev. *80*, 283–383. https://doi.org/10.1007/s12229-014-9146-y.
24. Gagnon, E., Baldaszti, L., Moonlight, P., Knapp, S., Lehmann, C.E.R., and Särkinen, T. (2023). Functional and ecological diversification of underground organs in *Solanum*. Front. Genet. *14*, 1231413. https://doi.org/10.3389/fgene.2023.1231413.
25. Novy, R.G., Gillen, A.M., and Whitworth, J.L. (2007). Characterization of the expression and inheritance of potato leafroll virus (PLRV) and potato virus Y (PVY) resistance in three generations of germplasm derived from *Solanum etuberosum*. Theor. Appl. Genet. *114*, 1161–1172. https://doi.org/10.1007/s00122-007-0508-2.
26. Melchers, G., Sacristán, M.D., and Holder, A.A. (1978). Somatic hybrid plants of potato and tomato regenerated from fused protoplasts. Carlsberg Res. Commun. *43*, 203–218. https://doi.org/10.1007/BF02906548.
27. Nee, M., Symon, D.D.E., and Lester, R.N. (1999). Solanaceae IV: Advances in Biology and Utilization (Royal Botanic Gardens, Kew).
28. Peralta, I., and Spooner, D. (2000). Classification of wild tomatoes: a review. Kurtziana *28*, 45–54.
29. Hawkes, J.G. (1990). The potato: evolution, biodiversity and genetic resources. Am. J. Potato Res. *67*, 733–735. https://doi.org/10.1007/BF03044023.
30. Hijmans, R.J., Gavrilenko, T., Stephenson, S., Bamberg, J., Salas, A., and Spooner, D.M. (2007). Geographical and environmental range expansion through polyploidy in wild potatoes (*Solanum* section *Petota*). Glob. Ecol. Biogeogr. *16*, 485–495. https://doi.org/10.1111/j.1466-8238.2007.00308.x.
31. Gaiero, P., Vaio, M., Peters, S.A., Schranz, M.E., de Jong, H., and Speranza, P.R. (2019). Comparative analysis of repetitive sequences among species from the potato and the tomato clades. Ann. Bot. *123*, 521–532. https://doi.org/10.1093/aob/mcy186.
32. Hilgenhof, R., Gagnon, E., Knapp, S., Aubriot, X., Tepe, E.J., Bohs, L., Giacomin, L.L., Gouvêa, Y.F., Martine, C.T., Orejuela, A., et al. (2023). Morphological trait evolution in *Solanum* (Solanaceae): evolutionary lability of key taxonomic characters. Taxon *72*, 811–847. https://doi.org/10.1002/tax.12990.
33. Messeder, J.V.S., Carlo, T.A., Zhang, G., Tovar, J.D., Arana, C., Huang, J., Huang, C.H., and Ma, H. (2024). A highly resolved nuclear phylogeny uncovers strong phylogenetic conservatism and correlated evolution of fruit color and size in *Solanum* L. New Phytol. *243*, 765–780. https://doi.org/10.1111/nph.19849.
34. Wu, Y., Li, D., Hu, Y., Li, H., Ramstein, G.P., Zhou, S., Zhang, X., Bao, Z., Zhang, Y., Song, B., et al. (2023). Phylogenomic discovery of deleterious mutations facilitates hybrid potato breeding. Cell *186*, 2313–2328.e15. https://doi.org/10.1016/j.cell.2023.04.008.
35. Cheng, L., Wang, N., Bao, Z., Zhou, Q., Guarracino, A., Yang, Y., Wang, P., Zhang, Z., Tang, D., Zhang, P., et al. (2025). Leveraging a phased pangenome for haplotype design of hybrid potato. Nature *640*, 408–417. https://doi.org/10.1038/s41586-024-08476-9.
36. Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., Zhang, G., Yang, S., Li, R., et al.; Potato; Genome; Sequencing Consortium (2011). Genome sequence and analysis of the tuber crop potato. Nature *475*, 189–195. https://doi.org/10.1038/nature10158.
37. Pham, G.M., Hamilton, J.P., Wood, J.C., Burke, J.T., Zhao, H., Vaillancourt, B., Ou, S., Jiang, J., and Buell, C.R. (2020). Construction of a chromosome-scale long-read reference genome assembly for potato. GigaScience *9*, giaa100. https://doi.org/10.1093/gigascience/giaa100.
38. Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. Genetics *105*, 437–460. https://doi.org/10.1093/genetics/105.2.437.
39. Kubatko, L.S., and Degnan, J.H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. *56*, 17–24. https://doi.org/10.1080/10635150601146041.
40. Huson, D.H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. *23*, 254–267. https://doi.org/10.1093/molbev/msj030.
41. Edelman, N.B., Frandsen, P.B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R.B., García-Accinelli, G., Van Belleghem, S.M., Patterson, N., Neafsey, D.E., et al. (2019). Genomic architecture and introgression shape a butterfly radiation. Science *366*, 594–599. https://doi.org/10.1126/science.aaw2090.
42. Kong, S., and Kubatko, L.S. (2021). Comparative performance of popular methods for hybrid detection using genomic data. Syst. Biol. *70*, 891–907. https://doi.org/10.1093/sysbio/syaa092.
43. Dias-Alves, T., Mairal, J., and Blum, M.G.B. (2018). Loter: a software package to infer local ancestry for a wide range of species. Mol. Biol. Evol. *35*, 2318–2326. https://doi.org/10.1093/molbev/msy126.
44. Blischak, P.D., Chifman, J., Wolfe, A.D., and Kubatko, L.S. (2018). HyDe: a python package for genome-scale hybridization detection. Syst. Biol. *67*, 821–829. https://doi.org/10.1093/sysbio/syy023.
45. Hibbins, M.S., and Hahn, M.W. (2019). The timing and direction of introgression under the multispecies network coalescent. Genetics *211*, 1059–1073. https://doi.org/10.1534/genetics.118.301831.
46. Zierer, W., Rüscher, D., Sonnewald, U., and Sonnewald, S. (2021). Tuber and tuberous root development. Annu. Rev. Plant Biol. *72*, 551–580. https://doi.org/10.1146/annurev-arplant-080720-084456.
47. Kondhare, K.R., Kumar, A., Patil, N.S., Malankar, N.N., Saha, K., and Banerjee, A.K. (2021). Development of aerial and belowground tubers in potato is governed by photoperiod and epigenetic mechanism. Plant Physiol. *187*, 1071–1086. https://doi.org/10.1093/plphys/kiab409.
48. Kloosterman, B., Abelenda, J.A., Gomez, M., Oortwijn, M., de Boer, J.M., Kowitwanich, K., Horvath, B.M., van Eck, H.J., Smaczniak, C., Prat, S., et al. (2013). Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature *495*, 246–250. https://doi.org/10.1038/nature11912.
49. Navarro, C., Abelenda, J.A., Cruz-Oró, E., Cuéllar, C.A., Tamaki, S., Silva, J., Shimamoto, K., and Prat, S. (2011). Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. Nature *478*, 119–122. https://doi.org/10.1038/nature10431.
50. Rodríguez-Falcón, M., Bou, J., and Prat, S. (2006). Seasonal control of tuberization in potato: conserved elements with the flowering response. Annu. Rev. Plant Biol. *57*, 151–180. https://doi.org/10.1146/annurev.arplant.57.032905.105224.
51. Nicolas, M., Torres-Pérez, R., Wahl, V., Cruz-Oró, E., Rodríguez-Buey, M.L., Zamarreño, A.M., Martín-Jouve, B., García-Mina, J.M., Oliveros, J.C., Prat, S., et al. (2022). Spatial control of potato tuberization by the TCP transcription factor *BRANCHED1b*. Nat. Plants *8*, 281–294. https://doi.org/10.1038/s41477-022-01112-2.
52. Capua, Y., and Eshed, Y. (2017). Coordination of auxin-triggered leaf initiation by tomato *LEAFLESS*. Proc. Natl. Acad. Sci. USA *114*, 3246–3251. https://doi.org/10.1073/pnas.1617146114.
53. Kumar, A., Kondhare, K.R., Malankar, N.N., and Banerjee, A.K. (2021). The Polycomb group methyltransferase StE(z)2 and deposition of

H3K27me3 and H3K4me3 regulate the expression of tuberization genes in potato. J. Exp. Bot. *72*, 426–444. https://doi.org/10.1093/jxb/eraa468.

54. Wang, D., Xu, X., Zhang, H., Xi, Z., Abbott, R.J., Fu, J., and Liu, J. (2022). Abiotic niche divergence of hybrid species from their progenitors. Am. Nat. *200*, 634–645. https://doi.org/10.1086/721372.

55. Spooner, D.M., Berg, R.G., Rodríguez, A., Bamberg, J., Hijmans, R.J., and Cabrera, S.I.L. (2004). Wild potatoes (*Solanum* Section Petota; Solanaceae) of north and central America. Syst. Bot. Monogr. *68*, 1–209. https://doi.org/10.2307/25027915.

56. Hawkes, J.G. (1954). The ecology of wild potato species and its bearing on the origin of potato cultivation. J. Agric. Tradit. Bot. Appl. *1*, 356–358. https://doi.org/10.3406/jatba.1954.2166.

57. Pérez-Escobar, O.A., Zizka, A., Bermúdez, M.A., Meseguer, A.S., Condamine, F.L., Hoorn, C., Hooghiemstra, H., Pu, Y., Bogarín, D., Boschman, L.M., et al. (2022). The Andes through time: evolution and distribution of Andean floras. Trends Plant Sci. *27*, 364–378. https://doi.org/10.1016/j.tplants.2021.09.010.

58. Tournebize, R., and Chikhi, L. (2025). Ignoring population structure in hominin evolutionary models can lead to the inference of spurious admixture events. Nat. Ecol. Evol. *9*, 225–236. https://doi.org/10.1038/s41559-024-02591-6.

59. Zhang, Z. (2025). Ancient hybridization underlies tuberization and radiation of the potato lineage. figshare. https://figshare.com/articles/dataset/Ancient_hybridization_underlies_tuberization_and_radiation_of_the_potato_lineage/28942299.

60. Xing, H.L., Dong, L., Wang, Z.P., Zhang, H.Y., Han, C.Y., Liu, B., Wang, X.C., and Chen, Q.J. (2014). A CRISPR/Cas9 toolkit for multiplex genome editing in plants. BMC Plant Biol. *14*, 327. https://doi.org/10.1186/s12870-014-0327-y.

61. Hellens, R.P., Allan, A.C., Friel, E.N., Bolitho, K., Grafton, K., Templeton, M.D., Karunairetnam, S., Gleave, A.P., and Laing, W.A. (2005). Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. Plant Methods *1*, 13. https://doi.org/10.1186/1746-4811-1-13.

62. Ranallo-Benavidez, T.R., Jaron, K.S., and Schatz, M.C. (2020). GenomeScope 2.0 and smudgeplot for reference-free profiling of polyploid genomes. Nat. Commun. *11*, 1432. https://doi.org/10.1038/s41467-020-14998-3.

63. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods *18*, 170–175. https://doi.org/10.1038/s41592-020-01056-5.

64. Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., Wang, X., Lippman, Z.B., Schatz, M.C., and Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. *23*, 258. https://doi.org/10.1186/s13059-022-02823-7.

65. Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., et al. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science *356*, 92–95. https://doi.org/10.1126/science.aal3327.

66. Robinson, J.T., Turner, D., Durand, N.C., Thorvaldsdóttir, H., Mesirov, J.P., and Aiden, E.L. (2018). Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. Cell Syst. *6*, 256–258.e1. https://doi.org/10.1016/j.cels.2018.01.001.

67. Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

68. Mapleson, D., Garcia Accinelli, G., Kettleborough, G., Wright, J., and Clavijo, B.J. (2017). KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics *33*, 574–576. https://doi.org/10.1093/bioinformatics/btw663.

69. Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. *20*, 275. https://doi.org/10.1186/s13059-019-1905-y.

70. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360. https://doi.org/10.1038/nmeth.3317.

71. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295. https://doi.org/10.1038/nbt.3122.

72. Brůna, T., Hoff, K.J., Lomsadze, A., Stanke, M., and Borodovsky, M. (2021). BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom. Bioinform. *3*, lqaa108. https://doi.org/10.1093/nargab/lqaa108.

73. Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics *24*, 637–644. https://doi.org/10.1093/bioinformatics/btn013.

74. Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O., and Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. *33*, 6494–6506. https://doi.org/10.1093/nar/gki937.

75. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158.

76. Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. *9*, R7. https://doi.org/10.1186/gb-2008-9-1-r7.

77. Venturini, L., Caim, S., Kaithakottil, G.G., Mapleson, D.L., and Swarbreck, D. (2018). Leveraging multiple transcriptome assembly methods for improved gene structure annotation. GigaScience *7*, giy093. https://doi.org/10.1093/gigascience/giy093.

78. Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Jr., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. *31*, 5654–5666. https://doi.org/10.1093/nar/gkg770.

79. Wei, W., Gui, S., Yang, J., Garrison, E., Yan, J., and Liu, H.-J. (2024). Wgatools: an ultrafast toolkit for manipulating whole genome alignments. Preprint at arXiv. https://doi.org/10.48550/arXiv.2409.08569.

80. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

81. Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. *16*, 157. https://doi.org/10.1186/s13059-015-0721-2.

82. Zhao, Y., Jia, X., Yang, J., Ling, Y., Zhang, Z., Yu, J., Wu, J., and Xiao, J. (2014). PanGP: A tool for quickly analyzing bacterial pan-genome profile. Bioinformatics *30*, 1297–1299. https://doi.org/10.1093/bioinformatics/btu017.

83. Lovell, J.T., Sreedasyam, A., Schranz, M.E., Wilson, M., Carlson, J.W., Harkess, A., Emms, D., Goodstein, D.M., and Schmutz, J. (2022).

GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. Elife *11*, e78526. https://doi.org/10.7554/eLife.78526.

84. Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I.T., Novak, A.M., Deran, A., Fang, Q., Xie, D., Feng, S., Stiller, J., et al. (2020). Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature *587*, 246–251. https://doi.org/10.1038/s41586-020-2871-y.

85. Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274. https://doi.org/10.1093/molbev/msu300.

86. Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics *19*, 153. https://doi.org/10.1186/s12859-018-2129-y.

87. Malinsky, M., Matschiner, M., and Svardal, H. (2021). Dsuite - Fast *D*-statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. *21*, 584–595. https://doi.org/10.1111/1755-0998.13265.

88. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

89. Martin, S.H., and Van Belleghem, S.M. (2017). Exploring evolutionary relationships across the genome using topology weighting. Genetics *206*, 429–438. https://doi.org/10.1534/genetics.116.194720.

90. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

91. Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. *34*, W609–W612. https://doi.org/10.1093/nar/gkl315.

92. Liu, L., Yu, L., and Edwards, S.V. (2010). A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol. Biol. *10*, 302. https://doi.org/10.1186/1471-2148-10-302.

93. Sukumaran, J., and Holder, M.T. (2010). DendroPy: a Python library for phylogenetic computing. Bioinformatics *26*, 1569–1571. https://doi.org/10.1093/bioinformatics/btq228.

94. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

95. Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience *4*, 7. https://doi.org/10.1186/s13742-015-0047-8.

96. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088.

97. Flouri, T., Jiao, X., Rannala, B., and Yang, Z. (2020). A bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. Mol. Biol. Evol. *37*, 1211–1223. https://doi.org/10.1093/molbev/msz296.

98. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

99. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

100. Browning, B.L., Tian, X., Zhou, Y., and Browning, S.R. (2021). Fast two-stage phasing of large-scale sequence data. Am. J. Hum. Genet. *108*, 1880–1890. https://doi.org/10.1016/j.ajhg.2021.08.005.

101. Korunes, K.L., and Samuk, K. (2021). PIXY: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. Mol. Ecol. Resour. *21*, 1359–1368. https://doi.org/10.1111/1755-0998.13326.

102. Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J. A., and Harris, S.R. (2016). SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb. Genom. *2*, e000056. https://doi.org/10.1099/mgen.0.000056.

103. Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., et al. (2019). CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases. Methods Ecol. Evol. *10*, 744–751. https://doi.org/10.1111/2041-210X.13152.

104. Rabosky, D.L., Grundler, M., Anderson, C., Title, P., Shi, J.J., Brown, J.W., Huang, H., and Larson, J.G. (2014). BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods Ecol. Evol. *5*, 701–707. https://doi.org/10.1111/2041-210X.12199.

105. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

106. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527. https://doi.org/10.1038/nbt.3519.

107. Liao, Y., Smyth, G.K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

108. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

109. Belton, J.M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. Methods *58*, 268–276. https://doi.org/10.1016/j.ymeth.2012.05.001.

110. Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. *3*, 95–98. https://doi.org/10.1016/j.cels.2016.07.002.

111. Zhou, Y., Zhang, Z., Bao, Z., Li, H., Lyu, Y., Zan, Y., Wu, Y., Cheng, L., Fang, Y., Wu, K., et al. (2022). Graph pangenome captures missing heritability and empowers tomato breeding. Nature *606*, 527–534. https://doi.org/10.1038/s41586-022-04808-9.

112. Li, N., He, Q., Wang, J., Wang, B., Zhao, J., Huang, S., Yang, T., Tang, Y., Yang, S., Aisimutuola, P., et al. (2023). Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. Nat. Genet. *55*, 852–860. https://doi.org/10.1038/s41588-023-01340-y.

113. Minh, B.Q., Hahn, M.W., and Lanfear, R. (2020). New methods to calculate concordance factors for phylogenomic datasets. Mol. Biol. Evol. *37*, 2727–2733. https://doi.org/10.1093/molbev/msaa106.

114. Wilf, P., Carvalho, M.R., Gandolfo, M.A., and Cúneo, N.R. (2017). Eocene lantern fruits from Gondwanan Patagonia and the early origins of Solanaceae. Science *355*, 71–75. https://doi.org/10.1126/science.aag2737.

115. Lisa De-Silva, D., Mota, L.L., Chazot, N., Mallarino, R., Silva-Brandão, K.L., Piñerez, L.M.G., Freitas, A.V.L., Lamas, G., Joron, M., Mallet, J., et al. (2017). North Andean origin and diversification of the largest ithomiine butterfly genus. Sci. Rep. *7*, 45966. https://doi.org/10.1038/srep45966.

116. Thawornwattana, Y., Seixas, F.A., Yang, Z., and Mallet, J. (2022). Full-likelihood genomic analysis clarifies a complex history of species divergence and introgression: the example of the erato-sara group of Heliconius butterflies. Syst. Biol. *71*, 1159–1177. https://doi.org/10.1093/sysbio/syac009.

117. Hudson, R.R., Kreitman, M., and Aguadé, M. (1987). A test of neutral molecular evolution based on nucleotide data. Genetics *116*, 153–159. https://doi.org/10.1093/genetics/116.1.153.

118. Zhang, J., Lyu, H., Chen, J., Cao, X., Du, R., Ma, L., Wang, N., Zhu, Z., Rao, J., Wang, J., et al. (2024). Releasing a sugar brake generates sweeter tomato without yield penalty. Nature *635*, 647–656. https://doi.org/10.1038/s41586-024-08186-2.

119. Díaz, S., Kattge, J., Cornelissen, J.H.C., Wright, I.J., Lavorel, S., Dray, S., Reu, B., Kleyer, M., Wirth, C., Prentice, I.C., et al. (2016). The global spectrum of plant form and function. Nature *529*, 167–171. https://doi.org/10.1038/nature16489.

120. Chang, J., Rabosky, D.L., and Alfaro, M.E. (2020). Estimating diversification rates on incompletely sampled phylogenies: theoretical concerns and practical solutions. Syst. Biol. *69*, 602–611. https://doi.org/10.1093/sysbio/syz081.

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Agrobacterium tumefaciens GV3101	This study	N/A
Escherichia coli DH5a	This study	N/A
Saccharomyces cerevisiae strain AH109	This study	N/A
Chemicals, peptides, and recombinant proteins		
DO Supplement -Leu/-Trp with Agar	Coolaber	Cat#PM2222
DO Supplement -His/-Leu/-Trp with Agar	Coolaber	Cat#PM2152
Minimal SD Base	Coolaber	Cat#PM2040
Murashige and Skoog medium	BioRun	Cat#RFA01
Phytagel™	BioRun	Cat#RFR06
Kinetin (KT)	BioRun	Cat#RJR00
2,4-Dichlorophenoxyacetic acid (2,4-D)	BioRun	Cat#RJO01
Kanamycin	BioRun	Cat#RJG00
Rifampicin	BioRun	Cat#RJF00
Critical commercial assays		
Dual-Luciferase® Reporter Assay System	Promega	Cat#E1910
Firefly Luciferase Assay Kit	UElandy	Cat#F6024L
AH109-GAL4 Yeast Two-Hybrid interaction proving kit	Coolaber	Cat#WLYH2021
ClonExpress Ultra One Step Cloning Kit V3	Vazyme	Cat# C117
Deposited data		
PacBio CCS, Hi-C, and RNA sequencing data	This study	BioProject: PRJNA1173794
Genome assemblies	This study	http://solomics.agis.org.cn/potato/; https://doi.org/10.6084/m9.figshare.28942299
Public Solanaceae genome assemblies	Tang et al.[19]; Wu et al.[34]; Cheng et al.[35]	N/A
Public Solanaceae whole-genome sequencing data	Tang et al.[19]; Wu et al.[34]	N/A
Species occurrence data	Gagnon et al.[24]	N/A
Experimental models: Organisms/strains		
Nicotiana benthamiana	This study	N/A
Solanum chacoense	This study	N/A
Oligonucleotides		
Primers used for Y2H and dual LUC assays (see Table S3)	This study	N/A
Recombinant DNA		
pGBKT7-SP6A^{Pet}	This study	N/A
pGBKT7-SP6A^{Tom}	This study	N/A
pGBKT7-SP6A^{Etu}	This study	N/A
pGADT7-IT1^{Pet}	This study	N/A
pGADT7-IT1^{Tom}	This study	N/A
pGADT7-IT1^{Etu}	This study	N/A
pGreen-IT1^{Pet}	This study	N/A
pGreen-IT1^{Tom}	This study	N/A
pGreen-IT1^{Etu}	This study	N/A
3xITBS-35Smini-LUC	This study	N/A

(*Continued on next page*)

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pJW712-SP6A^{Pet}	This study	N/A
pJW712-SP6A^{Tom}	This study	N/A
pJW712-SP6A^{Etu}	This study	N/A
pJW711-IT1^{Pet}	This study	N/A
pJW711-IT1^{Tom}	This study	N/A
pJW711-IT1^{Etu}	This study	N/A
pKSE401-CLF^{CR}	This study	N/A
pKSE401-DRN^{CR}	This study	N/A
pGADT7	TaKaRa	Cat#630442
pGBKT7	TaKaRa	Cat#630489
pKSE401	Xing et al.[60]	N/A
pGreenII0800-LUC	Hellens et al.[61]	N/A
pGreenII 62 SK	Hellens et al.[61]	N/A
Software and algorithms		
SplitsTree v4.18.3	Huson and Bryant[40]	https://software-ab.cs.uni-tuebingen.de/download/splitstree4/welcome.html
QuIBL	Edelman et al.[41]	https://github.com/michaelmiyagi/QuIBL
LOTER	Dias-Alves et al.[43]	http://faculty.washington.edu/browning/beagle/beagle.html
HyDe	Blischak et al.[44]	https://github.com/pblischak/HyDe
GenomeScope v2.0	Ranallo-Benavidez et al.[62]	https://github.com/tbenavi1/genomescope2.0
hifiasm v0.19.8-r603	Cheng et al.[63]	https://github.com/chhylp123/hifiasm
Ragtag v.2.0.1	Alonge et al.[64]	https://github.com/malonge/RagTag
3D-DNA pipeline v201008	Dudchenko.[65]	https://github.com/aidenlab/3d-dna
Juicebox v.1.11.08	Robinson et al.[66]	https://github.com/aidenlab/Juicebox
BUSCO v5.7.0	Simao et al.[67]	https://busco.ezlab.org/
KAT v2.4.2	Mapleson et al.[68]	https://kat.readthedocs.io/en/latest/
EDTA v1.9.4	Ou et al.[69]	https://github.com/oushujun/EDTA
Hisat2 v2.2.1	Kim et al.[70]	https://github.com/DaehwanKimLab/hisat2
StringTie v2.2.1	Pertea et al.[71]	https://ccb.jhu.edu/software/stringtie/
BRAKER2 v2.1.5	Brůna et al.[72]	https://github.com/Gaius-Augustus/BRAKER
AUGUSTUS v3.4.0	Stanke et al.[73]	https://github.com/Gaius-Augustus/Augustus
GeneMark-ET v3.67	Lomsadze et al.[74]	https://exon.gatech.edu/GeneMark/
cd-hit-est v4.6.8	Li and Godzik.[75]	https://github.com/weizhongli/cdhit/
EVidenceModeler v1.1.1	Haas et al.[76]	https://github.com/EVidenceModeler/EVidenceModeler/releases
Mikado v2.3.4	Venturini et al.[77]	https://github.com/EI-CoreBioinformatics/mikado
PASA pipeline v2.5.1	Haas et al.[78]	https://github.com/PASApipeline/PASApipeline
wgatools v0.1.0	Wei et al.[79]	https://github.com/wjwei-handsome/wgatools
GATK v4.1.2	McKenna et al.[80]	https://gatk.broadinstitute.org
OrthoFinder v2.5.4	Emms and Kelly[81]	https://github.com/davidemms/OrthoFinder
PanGP v1.0.1	Zhao et al.[82]	https://pangp.zhaopage.com/download.html
GENESPACE v0.9.4	Lovell et al.[83]	https://github.com/jtlovell/GENESPACE
Progressive Cactus v2.0.3	Armstrong et al.[84]	https://github.com/ComparativeGenomicsToolkit/cactus
IQTREE v2.0.3	Nguyen et al.[85]	http://www.iqtree.org
ASTRAL v5.15.5	Zhang et al.[86]	https://github.com/smirarab/ASTRAL
Dsuite v0.5 r16	Malinsky et al.[87]	https://github.com/millanek/Dsuite

(*Continued on next page*)

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
trimAl v1.4	Capella-Gutiérrez et al.[88]	https://github.com/inab/trimal
TWISST pipeline	Martin and Van.[89]	https://github.com/simonhmartin/twisst
MAFFT v7.505	Katoh and Standley[90]	https://mafft.cbrc.jp/alignment/software/
PAL2NAL v.14	Suyama et al.[91]	http://www.bork.embl.de/pal2nal/
MP-EST v2.0	Liu et al.[92]	https://phylolab.franklinresearch.uga.edu/mp-est
DendroPy v4.1.0	Sukumaran and Holder[93]	https://dendropy.org/
VCFtools	Danecek et al.[94]	https://vcftools.github.io/index.html
Plink v1.90	Chang et al.[95]	https://zzz.bwh.harvard.edu/plink/fanal.shtml
MCMCTree program in PAML package v4.9	Yang[96]	http://abacus.gene.ucl.ac.uk/software/
BPP v4.7.0	Flouri et al.[97]	https://github.com/bpp/bpp
BWA	Li and Durbin[98]	https://bio-bwa.sourceforge.net/
SAMtools	Li et al.[99]	http://www.htslib.org/
BEAGLE v5.3	Browning et al.[100]	https://faculty.washington.edu/browning/beagle/
Pixy v1.2.11	Korunes and Samuk[101]	https://github.com/ksamuk/pixy
Snp-sites	Page et al.[102]	https://github.com/sanger-pathogens/snp-sites
CoordinateCleaner	Zizka et al.[103]	https://cran.r-project.org/web/packages/CoordinateCleaner/index.html
TACT v0.5.0	Chang et al.[103]	https://github.com/jonchang/tact
BAMM v2.5.0	Rabosky et al.[104]	https://cran.r-project.org/web/packages/BAMMtools/index.html
fastp v2.22.0	Chen et al.[105]	https://github.com/OpenGene/fastp
Kallisto v0.46.2	Bray et al.[106]	https://github.com/pachterlab/kallisto/releases
featureCounts v2.0.3	Liao et al.[107]	https://subread.sourceforge.net/featureCounts.html
DESeq2 v1.38.3	Love et al.[108]	https://github.com/thelovelab/DESeq2

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Plant materials and growth conditions

The 27 accessions for *de novo* assembly and RNA-seq, comprising Petota and Etuberosum, were planted in the greenhouses of the Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen (22°36′N and 114°30′E), Guangdong province, China in the spring of 2023. Transgenic plants were also cultivated in the greenhouses of the Agricultural Genomics Institute during spring 2024. The seedlings of three species for ecological adaption analysis were grown in an artificial climate room under various stress conditions.

METHOD DETAILS

Plant materials and sequencing

We collected a total of 144 high-quality diploid genomes, comprising 101 Petota, 15 Tomato, 9 Etuberosum, and 19 other Solanaceae genomes (Table S1). Of these, 54 haplotype-resolved genomes from Petota and Etuberosum are sequenced and assembled. The genomic DNA of genomes was extracted from *in vitro* seedlings. High-fidelity (HiFi) reads were sequenced using the Pacific Biosciences (PacBio) Sequel II platform in circular consensus sequencing (CCS) mode. HiFi reads ranging from 23.47 to 36.30 Gb were generated using the CCS program (https://github.com/PacificBiosciences/ccs). For the construction of the Hi-C libraries, the extracted DNA was digested with the restriction enzyme MboI following the previously described Hi-C library preparation protocol.[109] A total of 2.71 Tb of Hi-C data was generated with approximately 123-fold coverage for each accession on the DNBSEQ-T7 platform. To annotate the genome, we sequenced a total of 650.01 Gb RNA data for 27 accessions from roots, stems, leaves, and flowers. Additionally, 4 haplotype-resolved genomes from Tomato were assembled using the public data and Hi-C data in this study.[34]

Genome *de novo* assembly and evaluation

The genome size and heterozygosity were assessed utilizing a k-mer-based method through GenomeScope (v2.0).[62] We then generated the preliminary haplotype-resolved assemblies using hifiasm (v0.19.8-r603) with default parameters based on HiFi and Hi-C reads.[63] The two initial haplotypes of each sample were further ordered using Ragtag (v.2.0.1), with the DM v6.1 or *S. etuberosum* genome as a reference, resulting in haploid pseudo-chromosome assemblies.[64] These assemblies were aligned

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

and scaffolded using Juicer pipeline with Hi-C reads, and Hi-C contact maps were generated based on the 3D-DNA pipeline (v201008).[65,110] Finally, the misaligned regions and assembly errors were manually adjusted based on the chromatin interaction patterns of the Hi-C contact maps visualized by Juicebox (v.1.11.08).[66] Additionally, the Benchmarking Universal Single-Copy Orthologs (BUSCO v5.7.0) with the embryophyta_odb10 database was used to evaluate the quality of assemblies.[67] The completeness of the haplotype-resolved assemblies was further assessed by the K-mer Analysis Toolkit (KAT v2.4.2).[68]

Annotation of repetitive elements and protein-coding genes

Transposable elements (TEs) were identified utilizing the Extensive *de novo* TE Annotator (EDTA v1.9.4), including Helitron-like DNA transposons, long terminal repeat retrotransposons, DNA transposons with terminal inverted repeat sequences, and short/long interspersed nuclear element repeats.[69] RepeatMasker (v4.10) (https://www.repeatmasker.org) was employed to mask annotated repetitive elements with TE library as input.

Protein coding genes were predicted for each haplotype by integrating transcript expression evidence, *ab initio* prediction, and homologous sequences. Firstly, RNA-seq reads were aligned to the assembled genome using Hisat2 (v2.2.1) with the "–dta" parameter, and the alignments were subsequently assembled to potential transcripts using StringTie (v2.2.1) with the "–rf" parameter.[70,71] Secondly, for *ab initio* prediction of genes, we applied BRAKER2 (v2.1.5), combined with AUGUSTUS (v3.4.0) and hidden Markov model (HMM) of GeneMark-ET (v3.67), with the transcript assemblies as evidence and "–nocleanup –softmasking" as parameters.[72–74] Thirdly, to enhance prediction accuracy, we collected the high-confidence plant homologous protein sequences from the UniProt Swiss-Prot database (https://www.uniprot.org), combined them with the previously published amino acid sequences of Petota, Tomato, and Etuberosum, and filtered redundancy using the cd-hit-est (v4.6.8).[19,75,111,112] The three types of results above were integrated using EVidenceModeler (v1.1.1) to produce a highly confident gene set, with the "weights" parameters set as "ABINITIO_PREDICTION 1, PROTEIN 4, TRANSCRIPT 8".[76] Finally, the Mikado (v2.3.4) was utilized to identify the "best" set of transcripts from multiple transcript assemblies, which were then provided to the PASA pipeline (v2.5.1) for comparing and updating gene structures.[77,78]

Whole-genome alignments and variant calling

The wfmash (v0.13.0) (https://github.com/waveygang/wfmash) software was used to align the 127 genomes to the DM v6.1 potato reference genome by all-vs-one alignments. Given the high divergence among genomes, we set the segment length parameter (–segment-length) to 5000, the mapping percent identity (–map-pct-id) to 80, the kmer size (–kmer) to 19, and the mapping filter parameter (–hg-filter-ani-diff) to 30. The PAF-format alignments generated by wfmash were converted to MAF format and gVCF files for subsequent analysis via wgatools (v0.1.0).[79] The MAF-format files were merged into 128-genome multiple sequence alignments (including the reference genome DM v6.1). The msa_view (v0.1.2) (http://compgen.cshl.edu/phast) software was used to extract four-fold degenerate sites based on the 128-genome multiple alignments and CDS annotation of the DM v6.1 reference genome. The variants were called using Genome Analysis Toolkit (GATK v4.1.2)[80] from the gVCF files, retaining only bi-allelic SNPs, and excluding those with more than 20% missing data in any of the four lineages (Petota, Etuberosum, Tomato, and outgroup).

Identification of homologous gene groups

For the protein sequences from the 128-genome dataset, 54,816 orthologous pan-gene clusters were identified using OrthoFinder (v2.5.4).[81] Among these, 47,556 with Petota involved were classified into four categories: 1) core clusters conserved in all 101 Petota genomes; 2) near core clusters present in 70–100 genomes; 3) shell clusters found in 2–69 genomes; 4) accession-specific clusters unique to a single genome. The pan-genome profile, based on the number of gene clusters, was simulated by PanGP (v1.0.1) with the sample algorithm as "Totally Random", a sample size of 400, and the repeats of 30.[82] Subsequently, 17,872 homologous gene groups were extracted from core and near core clusters based on the synteny-anchored pan-genome annotation generated by GENESPACE (v0.9.4).[83]

Phylogenetic analyses

Phylogenetic inferences were conducted at three levels: 1) phylogeny of the 27 Solanaceae species; 2) phylogeny based on the whole-genome alignments of 128 genomes; and 3) phylogeny of the 17,872 homologous gene groups.

To reconstruct a comprehensive phylogeny of the 27 Solanaceae species, including 6 Petota species spanning Clade 1+2, Clade 3, and Clade 4, as well as 21 additional distinct Solanaceae species (14 *Solanum* species and 7 other Solanaceae species) (Table S1), we extracted those species from the reference-free whole-genome alignments generated by Progressive Cactus (v2.0.3), and partitioned them into 1-Mb sliding windows with a 200-Kb step size. Window trees were then inferred by IQ-TREE (v2.0.3) with the options "-m GTR -B 1000".[84,85] A coalescent-based species tree was reconstructed from all the window trees using ASTRAL (v5.15.5).[86] Then we combined the species tree and the 500 randomly selected 1-Mb window trees to visualize the topological discordance with a Densitree plot. To investigate the genomic patterns of introgression between Petota and the other Solanaceae lineages, we calculated fbranch statistics using Dsuite (v.0.5 r16).[87] A VCF dataset converted from Cactus whole-genome alignments and the above coalescent-based species tree were used. The resulting Z-score matrix was visualized as a heatmap with the dtools.py script.

To infer the phylogeny based on the multiple alignments of 128 genomes and assess phylogenetic discordance at the whole-genome scale, the concatenated sequence of four-fold degenerate sites was used to construct the phylogeny of the 128-genome

dataset using IQ-TREE (v2.0.3), with the parameter "–scf" included to estimate site concordance factors.[113] We then partitioned the 128-genome multiple alignments into 100-Kb non-overlapping windows. The partitioned alignments were cleaned using trimAl (v1.4) with a gap threshold of 70% (-gt 0.7), and the cleaned alignments were analyzed by IQ-TREE (v2.0.3) with the parameters "-m GTR -B 1000".[85,88] The topology of each window tree was summarized using a custom R script. To distinguish between ILS and hybridization in topology conflicts, we performed QuIBL analysis on the window trees.[41] From these, 1,000 window trees were randomly selected and pruned to include only four species from distinct lineages. This process was repeated 100 times, generating 100 QuIBL results. Delta Bayesian information criterion (Delta.BIC) values, representing the significance of introgression levels, were collected from all 100 results, and their distributions for Tree-I, Tree-II, and Tree-III topologies were compared. To further estimate the discordance of three main clades within Petota at the whole-genome scale, the 100-Kb non-overlapping window trees were analyzed using TWISST pipeline. The scripts twisst.py and plot_twisst.R (available at: https://github.com/simonhmartin/twisst) were used to calculate and visualize the topology weighting, respectively.[89] To explore the introgression among the species of Petota, we selected all 44 distinct species from the 128-genome VCF dataset and calculated Patterson's *D*.[87] The resulting heatmap of the Patterson's *D* values was plotted using the Ruby script plot_d.rb.

To estimate the phylogeny of the 17,872 homologous gene groups, we extracted and aligned (MAFFT v7.505) the CDSs, protein sequences, and genic and intronic sequences of these homologous groups.[90] CDSs were further aligned according to the protein alignments using PAL2NAL (v.14).[91] The ML inferences of the syntenic orthogroups were performed by IQ-TREE with the options "-m MFP -B 1000" for phylogenetic construction.[85] The tree topology from each homologous group was summarized using a custom R script. We then performed coalescent-based methods such as ASTRAL (v5.15.5) and MP-EST (v2.0) to infer the species tree based on the homologous group's tree of CDSs, protein sequences, and genic and intronic sequences, respectively.[86,92] The gene concordance factors were estimated through the species tree and CDSs/intron trees by IQ-TREE (v2.0.3) with the parameter "–gcf".[85] To simulate the solely ILS scenario, species trees with branch lengths in coalescent units were used as input for DendroPy (v4.1.0) to generate 2,000 simulation trees.[93] We then compared the observed frequency of topologies with the simulated frequency using the two-tailed one-sample Student's *t*-test to evaluate whether the phylogenetic discordance was caused by the solely ILS scenario.

Network analyses

Network analyses were conducted using SplitsTree (v4.18.3) based on the trees of the 128-genome 100-Kb windows with the "ConsensusNewtwork" model with a threshold of 30%.[40] The network was visualized by "EqualAngel" to illustrate the phylogenetic conflicts among Tomato, Petota, and Etuberosum.

To investigate the explosive species diversification within Petota, we carried out a genomic neighbor network analysis using the 128-genome VCF dataset. The dataset was further filtered to include only Petota by VCFtools (v0.1.16), and the linkage disequilibrium-pruned (LD-pruned) set was generated using Plink (v1.90) with the parameter "–indep-pairwise" set to "1000kb 1000 0.5".[94,95] The distance matrices were then calculated by the Python script distMat.py (https://github.com/simonhmartin/genomics_general) based on the LD-pruned VCF dataset. We used the distance matrices to construct the phylogenetic neighbor networks with NeighborNet in SplitsTree (v4.18.3).[40]

Divergence time estimates

To estimate the timing of the hybrid origin of Petota, we calibrated the Solanaceae phylogeny using two calibration points: (1) seed fossils placed in the stem node of the Berry lineage (i.e., the divergence between *Nicotiana longiflora* Cav. and *Solanum*) dating back to ~ 51.2–53.2 MYA,[114,115] and (2) a secondary calibration point placed at the divergence between *Solanum* Clade II and Etuberosum (~ 22.0–29.0 MYA).[34,115] We randomly selected 500 combinations of orthologues from 5 species, representing Petota, Tomato, and Etuberosum, alongside *Solanum* Clade II and *N. longiflora*. For each combination, the orthologues were aligned by MAFFT (v7.505) and used to obtain the ML trees by IQ-TREE (v2.0.3).[85,90] To reduce the effect of the gene tree estimation error on divergence time calculation, we selected two main topologies (Petota-Etuberosum and Petota-Tomato) of the trees that were built based on the codon alignments with the bootstrap value larger than 95. The dating analysis was run in MCMCtree program in PAML (v4.9).[96]

To estimate when the divergence between and the hybridization resulting in Petota would have occurred, we used the multispecies coalescent-with-introgression (MSCi) model implemented in BPP (v4.7.0) (A00 analysis).[97,116] We selected the phased genomes of one species from Petota, Tomato, and Etuberosum, alongside *Solanum* Clade II. We first cut the CDSs with a gap of 2 kb and with a length from 400 bp to 800 bp in a BED file using the script DefineLocusToSelect.R (https://github.com/FernandoSeixas/bpp-pipeline). Second, we used the prepared BED file to extract the sequence from the genomes with the repeat sequences and missing sites removed, and parse them into PHYLIP format as the input sequence file of A00 analysis through a custom script. Third, we performed the A00 analysis with the *thetaprior* and *tauprior* of Gamma (a = 2, b = 100) and *phiprior* of Beta (a = 1, b = 1). The MCMC was run for 5,000 iterations of sample frequencies, 2,000 iterations of burn-in, and 5 iterations of the number of samples.

Population genetic analysis of 349 Petota re-sequenced genomes

To obtain the population genomic variant information, we mapped the sequences of the 398 diploid accessions (including 349 Petota, 22 Tomato, 21 Etuberosum, and 6 outgroup accessions) to the DM v6.1 reference genome using the "BWA-GATK-SAMtools" pipeline.[80,98,99] SNPs were filtered based on quality metrics (QUAL $\geq$ 20, QD $\leq$ 2.0, and GQ_MEAN $\geq$ 20),[94] keeping only bi-allelic SNPs reserved, and those with a maximum of 20% of missing data in any of the four lineages (Petota, Etuberosum, Tomato, or outgroup).

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

SNPs were phased by BEAGLE (v5.3).[100] Loter software was used to infer parental ancestry,[43] where genomes of Etuberosum and Tomato were used as the parental panel and genomes from Petota as the hybrid panel. Custom R and Perl scripts were used to calculate the genomic distributions and proportions of the parental ancestry tract. The HyDe software was applied to quantify the level of inter-species hybridization.[44] Ancient hybridization events between the three lineages (Petota, Tomato, and Etuberosum) were analyzed using "run_hyde.py", with eggplant accessions (*Solanum* Clade II) used as the outgroup. Hybridization events at a shallower evolutionary level were analyzed using individual genomes from within the three lineages, using "indidual_hyde.py". Pairwise nucleotide diversity (π), population genetic differentiation (F_{ST}), and genetic divergence (D_{XY}) for Petota, Tomato, and Etuberosum were calculated using the Pixy (v1.2.11) using the variant information file of the 398 diploid accessions.[101] Genomic f_d scans were performed for each 100-Kb non-overlapping window based on the variant information files from Clade 1+2, Clade 3, and Clade 4 of Petota respectively, using the Python scripts (https://github.com/simonhmartin/genomics_general/blob/master/ABBABABAwindows.py).

Identification of positively selected genes

We employed a recently proposed method to identify positively selected genes (PSGs) in Petota, which were derived from either the Tomato or Etuberosum parental lineages.[10] First, all individuals were divided into two groups for Hudson-Kreitman-Aguadé (HKA) test.[117] To identify PSGs derived from Tomato, Petota and Tomato were treated as Lineage 1 and Etuberosum as Lineage 2. For each gene, we counted the number of polymorphic sites (SNPs) in Lineage 1 (termed as A) and the number of fixed differences (the number of the SNPs with $F_{ST} \geq 0.95$) between Lineage 1 and Lineage 2 (termed as B) for each gene, based on the variant information file converted from the 128-genome codon alignments using snp-sites software.[94,102] Next, the Pearson's Chi-square test with Bonferroni correction (corrected p-value ≤ 0.05) was applied to reject the null hypothesis A(gene)/B(gene)=A(genome-wide)/B(genome-wide), which denotes the gene in Lineage 1 is not positively selected compared to the whole genome sequence. The same process was repeated to identify PSGs from Etuberosum, where Petota and Etuberosum were treated as as Lineage 1, and Tomato as Lineage 2. Among the identified PSGs, we focused on genes meeting the following criteria: (1) with a significant p-value (≤ 0.05) in the HKA test, (2) with the top 5% for fixed nonsynonymous mutations, and (3) consistent lineage origin when compared to the CDS sequence topology.

Yeast two-hybrid

Yeast two-hybrid assays were generated using the Matchmaker Gold Yeast Two-Hybrid System (Coolaber) vectors pGBKT7 and pGADT7. The full-length CDS of *IT1* and *SP6A* from Petota (*S. tuberosum*), Tomato (*S. pimpinellifolium*), and Etuberosum (*S. etuberosum*), were cloned into the pGADT7 and pGBKT7 vectors and subsequently introduced into the yeast strain AH109. The yeast cells were plated on SD/−Leu/−Trp medium and SD/−Leu/−Trp/−His medium, followed by cultivation at 30 °C for 5 days. Primer sequences used in these experiments are provided in Table S2.

Firefly luciferase complementation imaging assay

A firefly luciferase complementation assay was conducted followed by a previous report.[118] The full-length CDS of *IT1* or *SP6A* from the three lineages above, were cloned into pJW771 or pJW772 and transiently expressed in *Nicotiana benthamiana* Domin leaves. After dark treatment with 24 h, plants grow in a growth chamber under long-day conditions (16 h light and 8 h dark). Two days later, 1 mM luciferin was sprayed on the infiltrated *N. benthamiana* leaves, and then the images were captured by the Lumazone imaging system equipped with a 2048B CCD camera. Primers used in these analyses are detailed in Table S3.

Dual-luciferase reporter assay

Luciferase reporter assay was conducted as previously reported to investigate the difference in transcription factor activity of IT1 among Petota, Etuberosum, and Tomato.[61] We synthesized a 3× IT1 binding site (ITBS) followed by a mini *35S* promoter (3x-ITBS-35Smini). This 3×-ITBS-35Smini was subsequently inserted into a pGreenII 0800-luciferase reporter vector, which was introduced into *Agrobacterium tumefaciens* GV3101 (pSoup-p19) competent cells. Effector vectors, including $IT1^{Pet}$, $IT1^{Tom}$, and $IT1^{Etu}$ overexpression (vector: pGreenII 62 SK) constructs, along with a pGreenII 62 SK empty vector as a control. These constructs were also introduced into GV3101 (pSoup-p19), and cells containing effectors and reporters were mixed in a proportion of 4:1 and injected into leaves of *N. benthamiana*. Luciferase activities were measured at 2.5 days post-injection. Primers used in these analyses are detailed in Table S3.

CRISPR/Cas9-mediated knockout of *StCLF* and *StDRN* genes

The diploid wild potato *S. chacoense* was used as the model system for targeted gene editing. Two 20-nucleotide single-guide RNA (sgRNA) were designed to target the *StCLF* gene (sgRNA1: TGAACGTGCTGATTATATTAAGG and sgRNA2: CCCTCCATATACT TCATGGATA) and the *StDRN* gene (sgRNA1: TGCGCTGCTAGAGCAATGCGTGG and sgRNA2: GGTTGGGTAGGAGTAGG ACAAGG), both derived from the *S. chacoense*. These sgRNAs were then integrated into the CRISPR-Cas9 vector pKSE401, following the protocol established by Xing et al.[60] Subsequently, Agrobacterium-mediated transformation was conducted on potato internodes, adhering to protocols established in our previous research.[19] To rigorously identify successful gene edits, positive transformants were screened for their resilience in a kanamycin-supplemented medium at a concentration of 50 mg/L. For the validated T0

Please cite this article in press as: Zhang et al., Ancient hybridization underlies tuberization and radiation of the potato lineage, Cell (2025), https://doi.org/10.1016/j.cell.2025.06.034

plants, targeted amplification of the *StCLF* and *StDRN* gene fragments was achieved using primers designed to flank the sgRNA targets. This was further substantiated through Sanger sequencing. Positive transformants harboring the desired mutations were then carefully transplanted into a greenhouse environment for comprehensive phenotype observation.

Maps and ecological niche identification

To evaluate the ecological differences between the three lineages, we extracted occurrence data for each lineage from the taxonomically verified Solanaceae Source (https://solanaceaesource.myspecies.info/). Only occurrence records identified to species were kept, removing all cultivated records as well as the species only known from cultivation. Occurrences without coordinates or with erroneous coordinates (e.g., falling into sea or outside their specified country based on label data) were removed using custom R scripts and R package CoordinateCleaner.[103] Next, we extracted environmental variables for each occurrence point from bioclimatic variables from WorldClim (http://www.worldclim.org/), annual moisture index (AMI), topographic heterogeneity (log[VRM]), fire size (log[q95size]), and sand particle proportion.[24] A previous study demonstrated that geophytes and non-geophytes typically occupy distinct ecological niches, with geophytes being significantly influenced by eight bioclimatic environmental variables.[24] We used these eight environmental variables associated with the filtered occurrence of Tomato, Petota, and Etuberosum, respectively, to perform principal coordinate analysis (PCA) with kernel density estimates.[119] PCA was executed through the "princomp" function, the kernel density was estimated by the "kde" function, and the results were visualized for each of the lineages using contour plots.

The *Solanum* phylogeny from Gagnon et al. was utilized as the backbone tree for diversification analysis.[20] Taxonomic ranks were assigned to all *Solanum* species listed in the online dataset.[32] Using this taxonomy and their available branch lengths, we employed the TACT (v0.5.0) to generate a pseudo-phylogenetic tree supplemented with the unsampled branch lengths.[120] Diversification rates were then estimated based on the pseudo tree with the BAMM (v2.5.0) software, running for 1 million generations with priors obtained through the BAMMtools R package.[104] After independently processing each of the 100 trees generated by TACT with BAMM, we combined the results by calculating the median of the estimated tip diversification rates.

Gene expression under the cold treatment

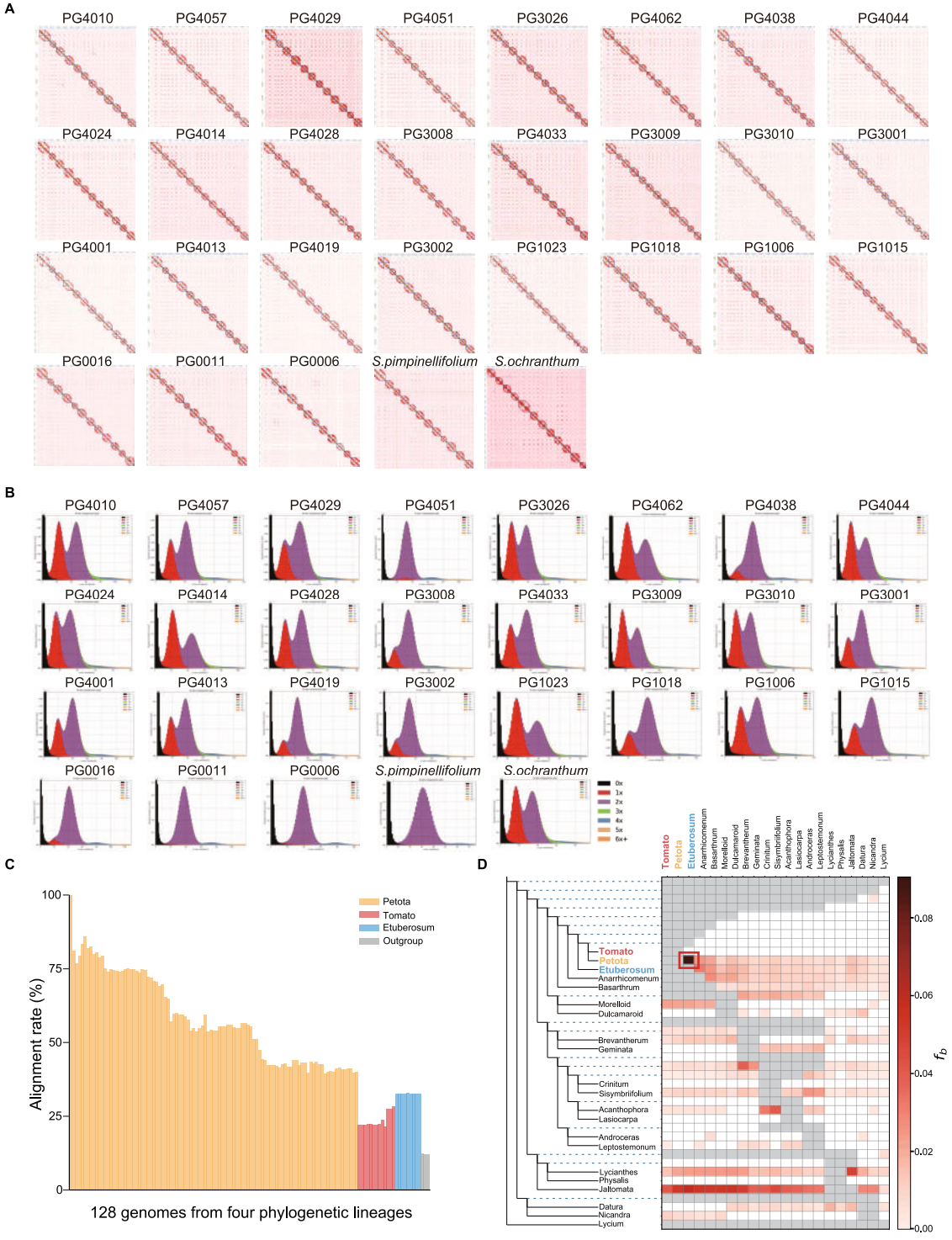
We performed RNA-seq experiments on three lineage species, one from each lineage (*S. tuberosum* [Petota], *S. lycopersicum* [-Tomato], and *S. palustre* [Etuberosum]) under 20 °C (control), 4 °C (cold treatment), and 4 °C with combined cold and drought treatment. At the seedling stage, we collected root, stem, and leaf tissues from each species, with six biological repeats, yielding 162 transcriptomes (3 species × 3 conditions × 3 tissues × 6 replicates). RNA was extracted and used for library preparation, followed by sequencing on the DNBSEQ-T7 platform at Annoroad Gene Technology, generating approximately 6 Gb of data per tissue sample, of which two samples failed. All raw reads were filtered using fastp (v2.22.0) with default parameters.[105] Kallisto (v0.46.2) was used to quantify the RNA-seq data as transcripts per million (TPM) per genes.[106] Gene-expression levels were quantified utilizing featureCounts (v2.0.3), and differential gene expression analysis was conducted with R package DESeq2 (v1.38.3) with adjusted p value < 0.05 and $|\log_2(\text{FoldChange}) \geq 1|$.[107,108]

QUANTIFICATION AND STATISTICAL ANALYSIS

Sample numbers (n) and statistical analyses are detailed in the results section and figure legends above. Statistical analyses were performed in R (v4.3.2). Two-tailed one-sample binomial test, two-tailed student's t-test, and the Pearson's Chi-square test with Bonferroni correction were employed to obtain p-values, with p-values < 0.05 considered statistically significant. p-values are indicated on each graph.

CellPress
OPEN ACCESS

Supplemental figures



(legend on next page)

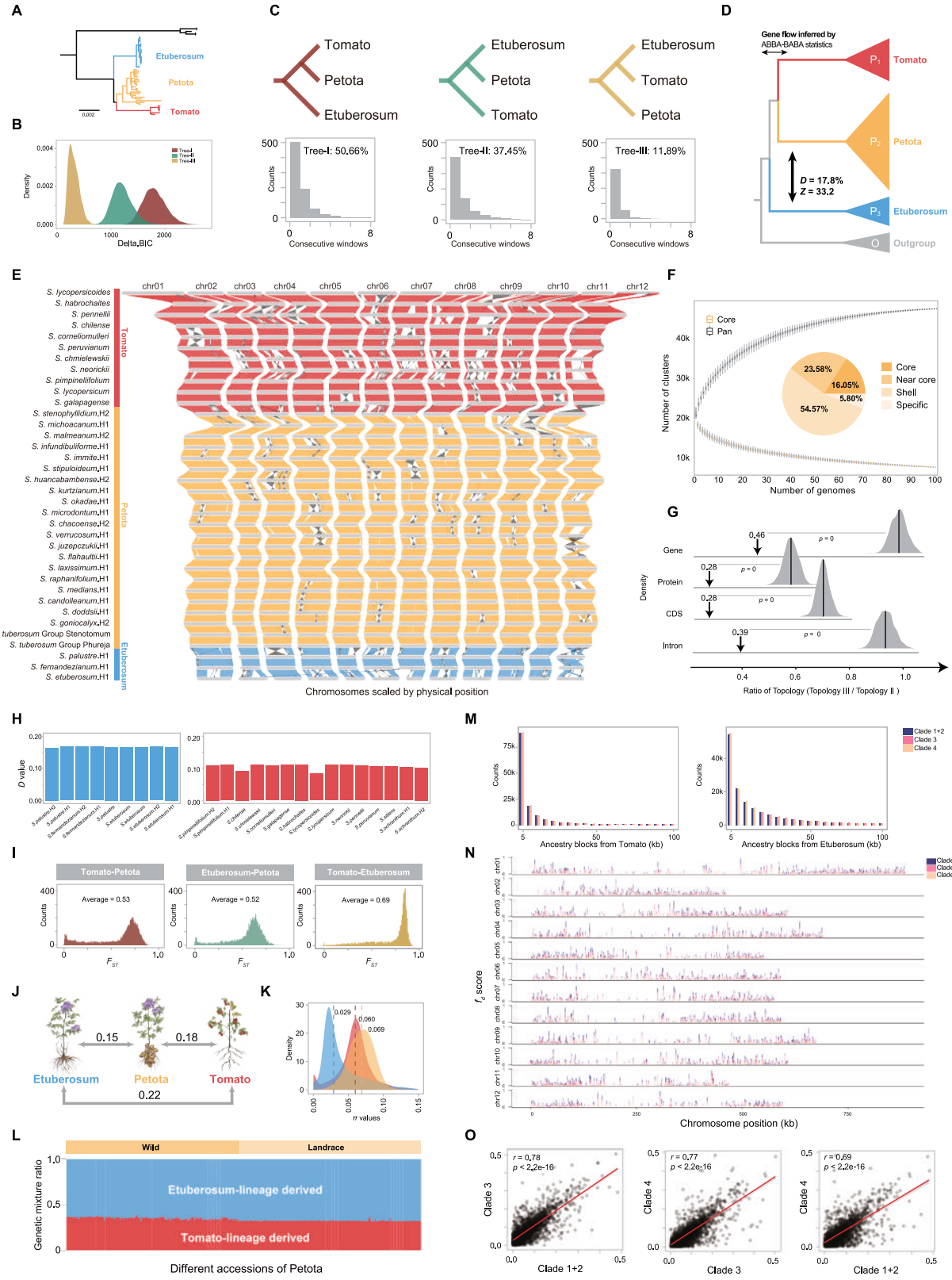
Figure S1. Assembly assessment and introgression events, related to Figure 1

(A) Hi-C contact maps for the 58 haplotype-resolved genomes.

(B) K-mer spectra assessment for the 58 haplotype-resolved genomes.

(C) Alignment rates of the 128 genomes with DM v6.1 as a reference, ordered according to Figure 1C.

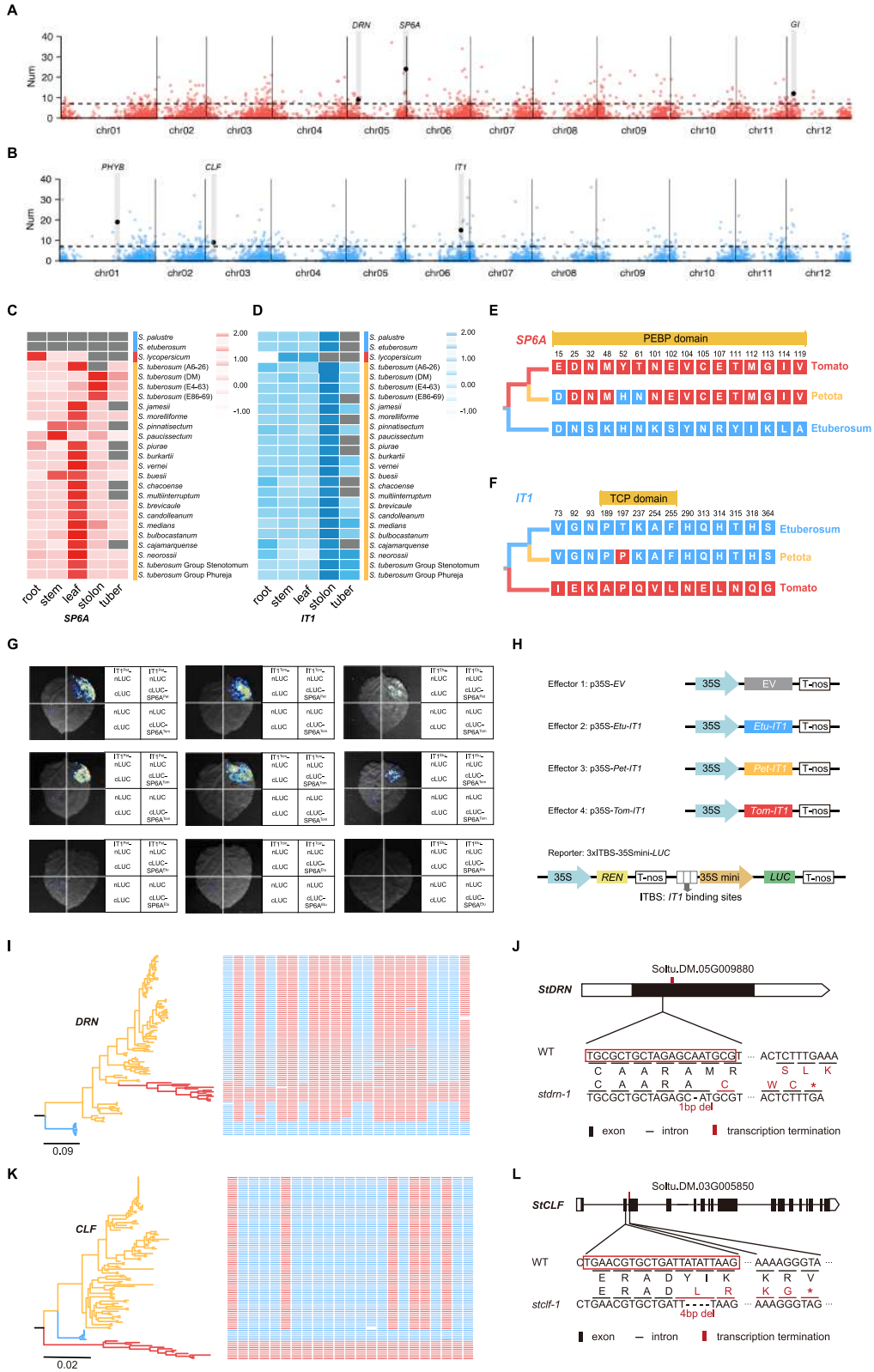
(D) Heatmap showing the introgression between the "expanded" tree branch (P2) on the *y* axis (relative to its sister branch P1) and the Solanaceae species (P3) on the *x* axis, as determined by Fbranch statistics. The gray cells are insufficient for introgression calculation. The red box highlights the high introgression value between Petota and Etuberosum. The "expanded" tree (*y* axis) includes species tips (*x* axis) and the internal branches of the species tree.



(legend on next page)

Figure S2. Phylogenetic and population genetic analysis reveals the extent and stability of genomic hybrid signals, related to Figure 2

(A) Phylogeny of chloroplast genomes constructed using the ML method with 1,000 bootstrap replicates.
(B) Distribution of QuIBL delta BIC values for the three topologies.
(C) The three major phylogenetic tree topologies and their distributions across consecutive 100-kb sliding windows.
(D) ABBA-BABA analysis reveals significant gene flow between Petota and Etuberosum.
(E) Synteny plots of the homologous genes from 36 genomes. The gray lines indicate inverted regions.
(F) Simulation of pan- and core-genome sizes, in terms of gene cluster numbers and pan-genome composition. For each specified number of genomes, 500 combinations were generated with 30 replicates each.
(G) Comparison between observed topology ratios and the ratios in a simulated ILS-only scenario. Black arrows point to the observed ratios (Tree-III/Tree-II) derived from gene, protein, CDS, and intron trees, respectively. Gray curves represent the ratios in the ILS-only scenario. Statistical significances were calculated by a two-tailed one-sample binomial test.
(H) ABBA-BABA analysis to estimate the gene flow between Petota and different Etuberosum (blue)/Tomato (red) accessions.
(I) Pairwise genetic differentiation (F_{ST}) among Tomato, Petota, and Etuberosum.
(J) Pairwise genetic divergence (D_{XY}) among Tomato, Petota, and Etuberosum.
(K) Distribution of nucleotide diversity (π) for Tomato, Petota, and Etuberosum.
(L) Individual ancestry component analysis for 349 Petota accessions using the HyDe program.
(M) Distributions of ancestry blocks for different clades of Petota inferred by Loter analyses.
(N) Genomic f_d windowed scans of the genetic mixture between Etuberosum and each of the three clades of Petota with 100-kb non-overlapping windows.
(O) The genomic distribution of introgressed regions from Etuberosum is highly correlated across Clade 1 + 2, Clade 3, and Clade 4 within Petota, suggesting a single ancestral hybridization event shared by all Petota species.



(legend on next page)

Figure S3. Genetic basis of tuberization with hybrid signals, related to Figure 3

(A) Counts of the non-synonymous sites for the genes with a sister relationship between Petota and Tomato.
(B) Counts of the non-synonymous sites for the genes that Petota are grouped with Etuberosum in gene topology.
(C) Heatmap displaying the expression pattern of the *SP6A* gene.
(D) Heatmap displaying the expression pattern of the *IT1* gene.
(E) All inter-lineage fixed amino acids in the *SP6A* gene.
(F) All inter-lineage fixed amino acids in the *IT1* gene.
(G) Images of the dual-luciferase assay.
(H) Effector and reporter structures in dual-luciferase assay.
(I) The *DRN* gene tree is displayed on the left. The right part shows the positive-selected signals for all individuals in each fixed SNP site in the CDS region of *DRN*.
(J) The *DRN* CRISPR-Cas9 knockout mutant.
(K) The *CLF* gene tree is displayed on the left. The right part shows the positive-selected signals for all individuals in each fixed SNP site in the CDS region of *CLF*.
(L) The *CLF* CRISPR-Cas9 knockout mutant.

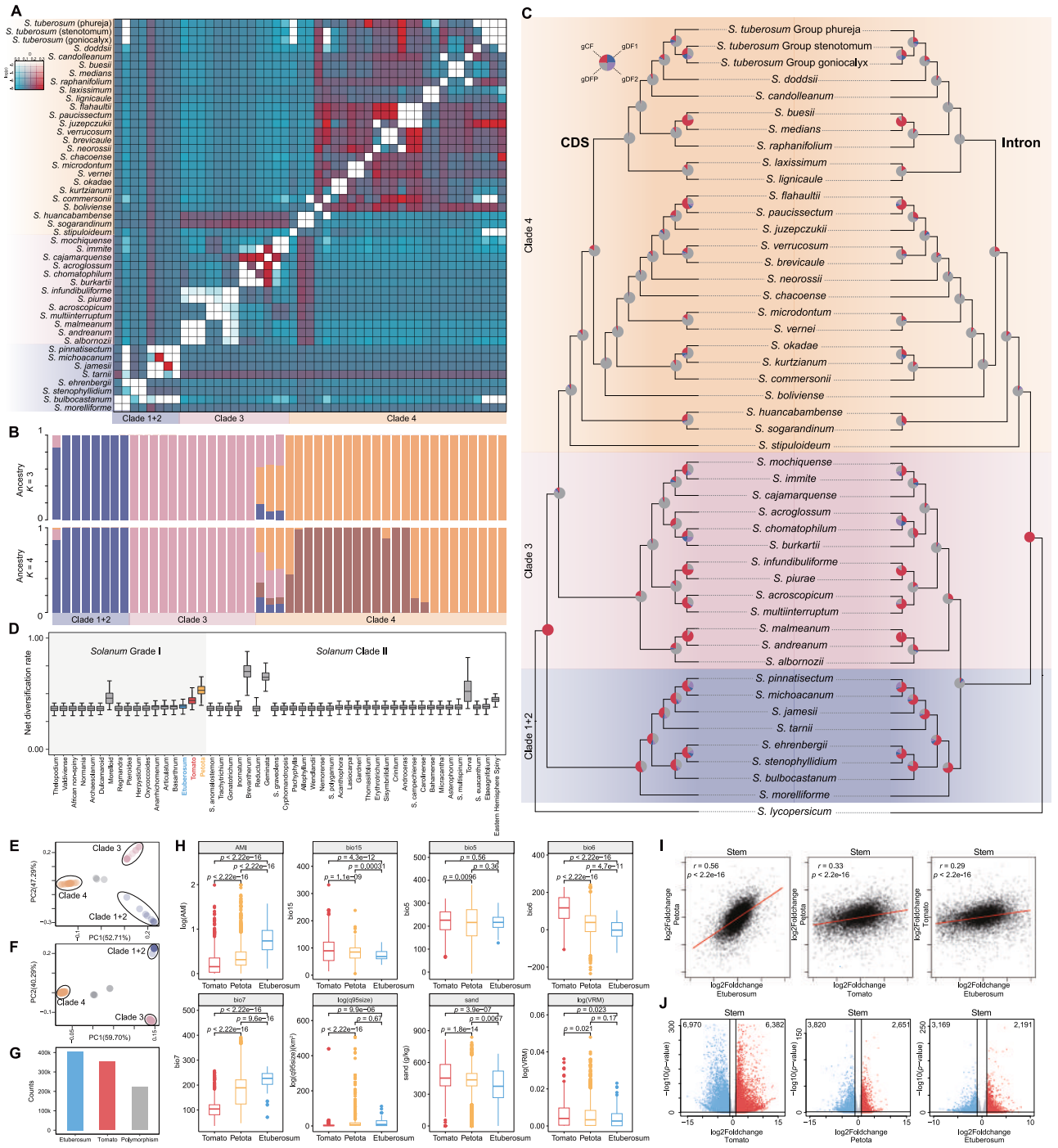


Figure S4. Rapid species radiation and ecological diversification of Petota, related to Figure 4

(A) Pairwise *D*-statistics between the Petota species. The red blocks in the heatmap indicate stronger introgression signals, with the blue blocks representing weaker signals. The light-colored blocks refer to larger (i.e., insignificant) *p* values.

(B) Admixture cluster analysis when $K = 3$ and $K = 4$.

(C) The CDS tree (left) and intron tree (right) with gene concordance factor (gCF) values shown as a pie chart on each node. In the gCF pie charts: red represents the proportion of CDS/Intron trees concordant with that branch (gCF), blue is the proportion of CDS/Intron trees concordant with the first alternative topology (gDF1), purple represents the ratio of the second alternative topology (gDF2), and gray refers to the remaining alternative topologies (gDFP).

(D) Median net diversification rates in *Solanum* lineages.

(E) PCA analysis based on genome-wide SNP data.

(F) PCA analysis based on topology relationships of Petota accessions.

(legend continued on next page)

(G) Hybridization-derived fixed and polymorphic sites in Petota.
(H) Boxplots showing variation across the selected environmental variables for the different lineages: seasonal precipitation (bio15), temperature annual range (bio7), minimum temperature of coldest month (bio6), maximum temperature seasonality (bio5), annual moisture index (AMI), log of 95th quantile for fire size (log [q95size]), log of vector ruggedness metric, log of topographic heterogeneity (log[VRM]), and proportion of sand particles (>0.05 mm) in the fine earth fraction at a depth of between 0–5 cm (sand).
(I) Correlation analysis of log2 fold-change of RNA expression, between the control and cold stress conditions, in stem tissue among Petota, Tomato, and Etuberosum.
(J) Volcano plot shows the significance of the log2 fold-change of RNA expression, between the control and cold stress conditions, in stem tissue for Petota, Tomato, and Etuberosum.