Preprints.org

Article

Comparative Identification of COSII Orthologs in Tomato, Potato, and Pepper Genomes by In Silico Analysis

Sibel Aziz [*] and Nasya Tomlekova

Article

Comparative Identification of COSII Orthologs in Tomato, Potato, and Pepper Genomes by In Silico Analysis

Sibel Aziz * and Nasya Tomlekova

Agricultural Academy, Maritsa Vegetable Crops Research Institute, Breeding Department, Laboratory of Molecular Biology

* Correspondence: sibeldaziz@gmail.com; Tel.: +359329952296

Abstract

Conserved Ortholog Set II (COSII) markers represent a well-established resource for comparative genomics and phylogenetic analyses in the Solanaceae family. In this study, we conducted a comprehensive in silico assessment of COSII orthologs in *Solanum lycopersicum* L., *Solanum tuberosum* L., and *Capsicum annuum* L. using an integrated workflow that combined OrthoFinder-based orthogroup inference, hierarchical orthogroup (HOG) reconstruction, synteny mapping, and evaluation of copy number. We identified 2,853 COSII-associated orthogroups, of which 2,359 (82.7%) were shared among all the three species, forming a deeply conserved solanaceous core. Among the three species, 1,839 orthogroups represented strict single-copy loci, reflecting their high evolutionary stability. Across these loci tomato and potato retained nearly complete single-copy status, whereas *C. annuum* L. displayed moderate copy-number variation (mean 1.35 genes per orthogroup; 22% multicopy), with duplicated clusters enriched on chromosomes 1-3, as well as on unplaced scaffolds (CA00). Hierarchical orthogroup analysis revealed substantial gene family expansion at the ancestral Solanaceae node, followed by lineage-specific diversification within Solanum and Capsicum. Synteny mapping showed extensive collinearity among genomes, combined with localized breaks and rearrangements in pepper. Together, these findings highlight a dual evolutionary pattern in Solanaceae: a highly conserved COSII genomic backbone, alongside lineage-specific structural innovations in *C. annuum*. COSII remains a reliable marker system for phylogenetics, comparative genomics, and marker-assisted breeding. The observed Capsicum-specific multicopy expansions overlap genomic regions enriched for stress-response gene families, suggesting links between structural variation and abiotic stress adaptation.

Keywords: COSII; solanaceae; tomato; potato; pepper; comparative genomics; orthologs; In silico analysis

1. Introduction

The Solanaceae family comprises more than 3,000 species distributed globally, including tomato (*Solanum lycopersicum* L.), potato (*Solanum tuberosum* L.), and pepper (*Capsicum annuum* L.), which are of major agronomic and economic importance. Their extensive morphological, physiological, and ecological diversity has made Solanaceae model systems for studies of genome evolution, domestication, chromosomal rearrangements, and trait diversification [1]. The availability of high-quality reference genomes has greatly advanced comparative genomics in Solanaceae. The chromosome-anchored tomato genome [2] laid the foundation for detailed structural and functional analyses, later refined by pangenome-level resources and improved annotations [3,4]. For potato, phased diploid genomes and pangenome analyses have revealed extensive structural variation and dynamic gene-family evolution [5,6]. Likewise, multiple high-continuity pepper genomes have elucidated lineage-specific chromosomal rearrangements, repeat-content variation, and copy-

number shifts [7–9]. Comparative studies consistently report strong macro-synteny among tomato, potato, and pepper, coupled with local inversions, translocations, and Copy Number Variation (CNV) hotspots [10–12]. The Conserved Ortholog Set II (COSII) consists of low-copy nuclear genes originally defined across euasterid species [13]. COSII markers have been widely used to construct comparative genetic maps, infer phylogenetic relationships, and reconstruct macro- and micro-synteny within Solanaceae [14,15]. Early COSII mapping studies highlighted extensive structural conservation among tomato, pepper, and potato [16,17], as well as conservation across more distantly related taxa such as coffee [18]. However, most classical COSII studies relied on EST sequences or PCR-based marker recovery and did not leverage modern graph-based orthology frameworks. Contemporary approaches, such as OrthoFinder and hierarchical orthogroup (HOG) reconstruction, enable genome-wide orthology inference, explicit resolution of duplication events, and deeper evolutionary interpretation across multiple species simultaneously [19]. Integrating these tools with today's high-continuity genomes provides an opportunity to re-evaluate COSII markers within a unified comparative framework. To date, no study has jointly integrated OrthoFinder-based orthology, HOG analysis, synteny mapping, and copy-number variation analysis for tomato, potato, and pepper. Therefore, the objective of the present work is to provide a comprehensive in silico assessment of COSII loci across these three species. Specifically, this study aims to: (i) identify and validate COSII orthologs using modern orthology and HOG-based inference; (ii) assess their sequence conservation, chromosomal positioning, and syntenic relationships; and (iii) characterize patterns of copy-number variation and lineage-specific structural divergence. This study provides a modern comparative-genomics framework for COSII loci and refines our understanding of Solanaceae genome evolution and COSII utility in phylogenetics and breeding.

2. Materials and Methods

Genome Resources

High-quality reference genomes for *S. lycopersicum*, *S. tuberosum*, and *C. annuum* were obtained from publicly available repositories and used as the basis for all comparative analyses. The tomato reference genome (Acc. No. GCA_000188115.5) and corresponding annotation files were retrieved from the Sol Genomics Network [20]. Potato genome assemblies and annotations were obtained from the Potato Genome Sequencing Consortium resources and validated through recent updates. Pepper genome sequences representing *C. annuum* L. were downloaded from the PGSC/CM334 reference dataset. Genome assemblies and gene annotations for *S. lycopersicum*, *S. tuberosum*, and *C. annuum* were obtained from Ensembl Plants and NCBI. The following assembly versions were used: Tomato SL4.0 [20], Potato DM v6.1 [6], and Pepper UCD10X v1.0/CA59/CM334 [8,9]. When multiple assemblies were available, the most complete reference-quality assembly for each species was selected. The *Arabidopsis thaliana* (TAIR10) gene models were included as an outgroup for orthology inference. All genome assemblies included chromosome-level scaffolds and curated gene models to ensure consistency in ortholog detection and downstream analyses.

Dataset Preparation and Gene Model Extraction

Gene sequences predicted protein-coding regions, and corresponding annotation metadata were extracted from each genome assembly using standard parsing tools. For each species, only primary transcripts were retained to avoid redundancy. Gene IDs, protein sequences, and coding sequences (CDS) were standardized across datasets, and genes lacking complete CDS information or showing annotation inconsistencies were removed [21].

COSII Marker Dataset Construction

The original COSII marker list and corresponding reference sequences were collected from previously published datasets and supplemented with curated markers from foundational COSII

studies. Reference sequences were aligned to each genome using BLAST-based similarity searches to identify candidate orthologs [22]. Hits with incomplete gene structures or ambiguous alignments were excluded. The final COSII dataset consisted of validated gene models across the three species, each mapped to genomic loci.

Orthology Inference Using OrthoFinder

Orthologous gene groups across the three species were identified using OrthoFinder (version consistent with recent releases). Protein sequences for all predicted genes were used as input to generate orthogroups, gene trees, and species trees. OrthoFinder's hierarchical orthogroup (HOG) framework was used to reconstruct gene-family evolution across internal phylogenetic nodes representing the Solanaceae lineage. Orthogroups corresponding to COSII markers were extracted, and their evolutionary consistency was assessed based on single-copy status and duplication patterns.

Phylogenetic Reconstruction of Single-Copy COSII Orthologs

Strict single-copy orthologs corresponding to COSII genes were aligned at the protein level using high-accuracy multiple sequence alignment tools. Codon-aware back-translation was performed to generate CDS alignments. Poorly aligned regions were removed to improve phylogenetic accuracy. Concatenated alignments were used for species-tree inference based on maximum-likelihood approaches with appropriate substitution models. Branch support was assessed using standard bootstrap procedures.

Copy Number Variation (CNV) Analysis

To evaluate gene copy-number variation among COSII loci, paralogous expansions were identified within orthogroups containing more than one gene per species. Copy number per orthogroup and per species was quantified, and paralog distribution across genomic regions was mapped. Genomic intervals showing enrichment for expanded COSII orthogroups were identified and compared across the three species. Special attention was given to patterns of CNV in *C. annuum* L., given its larger genome and known gene family expansions.

Chromosomal Mapping and Synteny Analyses

Genomic positions of COSII orthologs were extracted from GFF3 annotation files and mapped onto chromosomes. Synteny relationships among tomato, potato, and pepper were visualized using established genome-alignment tools based on reciprocal best-hit criteria and collinearity blocks. Syntenic clusters and rearranged segments were identified by comparing chromosome-level collinearity and breakpoint patterns. Chromosomal regions showing conserved COSII positions were distinguished from those exhibiting translocations, inversions, or local rearrangements.

Hierarchical Orthogroup (HOG) Analysis

To understand the evolutionary context of COSII loci, hierarchical orthogroup assignments were examined across ancestral Solanaceae nodes inferred by OrthoFinder. For each COSII orthogroup, expansion or contraction patterns were mapped across the Solanaceae phylogeny. Gene-family size changes were interpreted in the context of lineage-specific innovations and evolutionary divergence between *Solanum* and *Capsicum*.

Multiple alignment and tree inference

The protein sequences from the 1,839 three-way single-copy orthogroups (OGs) were aligned using MAFFT v7 with the L-INS-i algorithm [23]. The list of orthogroups used for concatenated alignment and species tree reconstruction is provided in Supplementary File S1 (*Orthogroups_for_concatenated_alignment.txt*). Poorly aligned regions were trimmed with trimAl v1.4

[24]. Maximum likelihood phylogenetic trees were generated using IQ-TREE v2 [25] with the best-fit substitution model determined by ModelFinder [26], and branch support was evaluated with 1,000 ultrafast bootstrap replicates [27].

Synteny assessment and annotation

Syntenic relationships among COSII loci were inferred based on OrthoFinder orthogroups and previously reported Solanaceae genome alignments [15,28]. Collinear relationships and potential rearrangements were visualized using Circos plots [29] based on orthogroup chromosomal mapping. Functional context of COSII loci was inferred from orthogroup assignments and cross-references to annotated Solanaceae gene models. Phylogenetic trees were visualized with iTOL v6 [30].

3. Results

Identification of COSII Orthogroups Across Tomato, Potato and Pepper

A total of 2,853 COSII-associated orthogroups were identified across *S. lycopersicum*, *S. tuberosum*, and *C. annuum*. Among them, 82.7% were shared across all the three species, forming a deeply conserved Solanaceae core. The remaining sets consisted of 244 orthogroups shared only between tomato and potato, 169 shared between potato and pepper, 66 shared between tomato and pepper, and a small number of lineage-specific orthogroups unique to each genome. These distributions reflect the expected phylogenetic relationships, with tomato and potato showing the highest level of shared orthology. The results of Orthogroups count is showing in file - *Supplementary_ S2_Orthogroups.GeneCount.tsv*
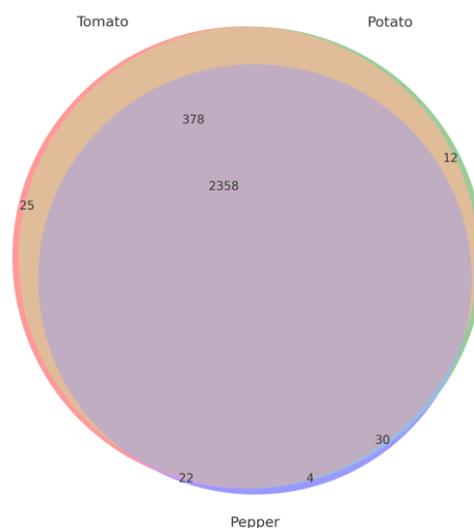
Single-Copy and Multicopy COSII Orthologs

A total of 1,839 COSII orthogroups (64.5%) were strictly single-copy across all three species. Across all COSII orthogroups, pepper exhibited a mean of 1.35 genes per orthogroup, with 22% of orthogroups showing multicopy expansions. These duplications were not randomly distributed but were concentrated on chromosomes 1, 2, and 3, as well as within unplaced scaffolds (CA00).

Phylogenetic Tree Reconstruction

Maximum likelihood (ML) phylogenetic analyses based on concatenated alignments of 1,839 single-copy COSII orthogroups yielded a well-resolved and strongly supported species tree for the three Solanaceae crops: *C. annuum*, *S. lycopersicum*, and *S. tuberosum* (UFboot ≥ 98; ***Figure 1***). *A. thaliana* was included as an outgroup to root the tree and to validate the orthology of the COSII loci employed. The resulting topology was fully consistent with established Solanaceae phylogenies: *C. annuum* diverged basally within the clade, while *S. lycopersicum* and *S. tuberosum* formed a strongly supported sister group. Branch length estimates indicated a deep divergence between *Arabidopsis* and the Solanaceae lineage (0.2496 substitutions per site), whereas interspecific distances within Solanaceae were substantially shorter, reflecting their more recent evolutionary separation. Individual gene trees largely recapitulated this topology; only a small fraction exhibited well-supported discordances, likely reflecting incomplete lineage sorting, ancient duplications, or lineage-specific gene losses. The uniformly high bootstrap support across the concatenated phylogeny underscores the suitability of COSII single-copy orthologues for robust species tree reconstruction within Solanaceae and across the broader euasterid clade. Comparative analysis of 2,852 COSII orthogroups revealed extensive conservation across the three species, with 2,358 (82.7%) shared among *S. lycopersicum*, *S. tuberosum*, and *C. annuum*, while an additional 378 orthogroups were shared exclusively between tomato and potato. Species-specific orthogroups were rare, underscoring the strong conservation of this low-copy nuclear gene set within Solanaceae. The concatenated alignment and orthogroup identifiers used for this analysis are provided in Supplementary Files S1 and S3

(Suplementary_S1_*Orthogroups_for_concatenated_alignment.txt;   Supplementary_S3_Orthogroups.txt*).
The results for phylogenetic HOGs are available in *Supplementary S4__HOG_lists.xlsx.*

**Figure 1.** Venn Diagram of COSII Orthogroups among Tomato, Potato, and Pepper Genomes.

Evolutionary Dynamics of Hierarchical Orthogroups (HOGs)

The distribution of HOGs (high-order gene groups) at evolutionary nodes provides a detailed picture of the diversification and retention of gene families within Solanaceae (*Supplementary S4__HOG_lists.xlsx).* Based on the obtained data from the partitioning of 40,499 phylogenetic hierarchical orthogroups, they were assigned to hierarchical levels N0–N3, corresponding to major ancestral and lineage-specific nodes. The largest expansion of gene families was observed at the ancestral node N1, preceding the divergence of the *Capsicum* and *Solanum* lineages (***Figure 2***). This early burst of gene family diversification likely reflects large-scale duplication events that provided a rich substrate for functional innovation. Many of these expanded orthogroups include receptor-like kinases (*RLKs*), receptor-like proteins (*RLPs*), and nucleotide-binding site leucine-rich repeat (NBS-LRR) genes families known to mediate pathogen perception and stress response [31–34]. The high number of orthogroups at N1 (40,499 HOGs) indicates extensive ancestral gene retention, consistent with the hypothesis of a Solanaceae-wide adaptive radiation supported by early genomic innovations [35,36]. In contrast, nodes N2 and N3 showed fewer but more specialized orthogroups, reflecting lineage-specific adaptation within *Capsicum* and *Solanum*. Node N2 encompasses orthogroups shared among tomato, potato, and pepper, likely corresponding to conserved Solanaceae traits including fleshy fruit development, secondary metabolism, and stress tolerance [2,36]. Node N3 encompasses orthogroups unique to the tomato–potato clade, highlighting Solanum-specific features such as tuber formation, fruit pigmentation, and growth habit diversification [37,38]. The conserved genomic core represented by node N0 (32,104 HOGs) remains substantial and corresponds to ancient orthogroups shared between Solanaceae and *A. thaliana*. These orthogroups encode essential cellular and metabolic functions, forming the evolutionary foundation upon which Solanaceae-specific gene expansions have occurred [35,39]. Together, these results support a two-phase model of genome evolution in Solanaceae: (i) an early phase of widespread gene family expansion (N0–N1) associated with the emergence of the ancestral Solanaceae lineage, and (ii) subsequent lineage-specific diversification (N2–N3) fine-tuning gene repertoires that contributed to phenotypic and metabolic differentiation among tomato, potato, and pepper. This model aligns with observations in other angiosperm clades, including Brassicaceae and Poaceae, suggesting convergent patterns of adaptive gene family evolution during diversification [40]. A large majority (82.7%) were shared among *S.*

lycopersicum, *S. tuberosum*, and *C. annuum*, while tomato and potato shared an additional 378 orthogroups absent from pepper. Species-specific orthogroups were rare, underscoring the strong conservation of this low-copy nuclear gene set within Solanaceae. Notably, several of the expanded HOGs at node N1 include gene families previously implicated in abiotic stress responses in Solanaceae, such as RLKs, RLPs, HSPs, LEA proteins, and secondary metabolism–related enzymes. The presence of these expansions at the ancestral Solanaceae node suggests that early gene family diversification may have contributed to environmental adaptation, particularly to drought, heat, and osmotic stress.

A

Node	HOG groups	Major taxa	Gene type
N0	32,104	Arabidopsis + Solanaceae	Conserved
N1	40,499	Arabidopsis	Receptor-like/signaling
N2	24,636	Tomato + Potato + Pepper	Solanum-specific
N3	24,416	Tomato + Potato	Internal Solanaceae
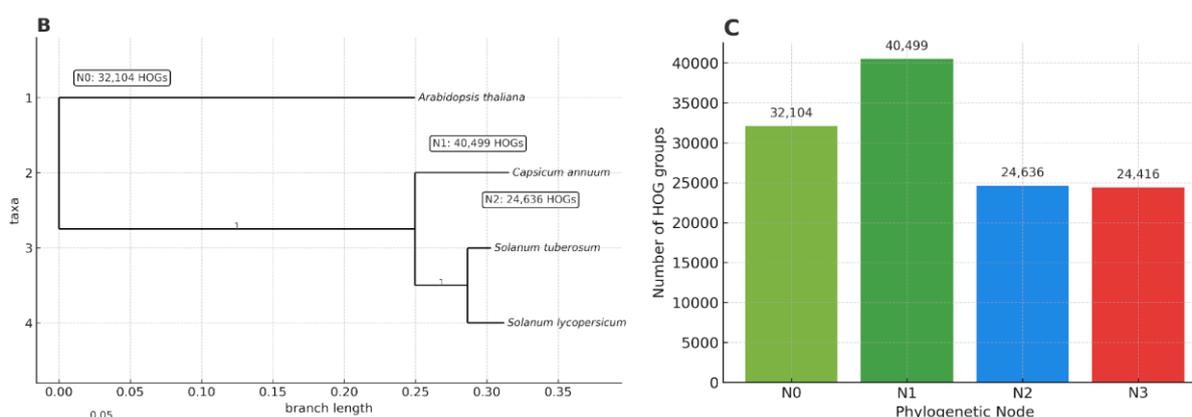
Figure 2. Distribution of HOG Counts across Phylogenetic Nodes in *Arabidopsis* and Solanaceae Species (A) Summary table of HOG counts, major taxa, and representative gene types for each node. (B) Rooted species tree with annotated HOG counts per internal node (N0–N3). (C) Bar plot showing the number of HOG groups at each phylogenetic node. The dominant contribution of N1 reflects major gene family expansions at the base of Solanaceae evolution.

Overall Distribution of HOGs Across Phylogenetic Nodes

The distribution of HOGs across internal nodes of the Solanaceae species phylogeny is summarized in ***Figure 2***. At the root of the tree (N0), representing the divergence between *A. thaliana* and Solanaceae, 32,104 HOGs were identified, corresponding to the conserved genomic core shared across the euasterid clade. The largest number of orthogroups (40,499) was observed at node N1, which represents the ancestral Solanaceae lineage prior to the divergence of *Capsicum* and *Solanum*. This substantial increase suggests extensive gene family expansion early in Solanaceae evolution, particularly within receptor-like and signaling-related gene families. Nodes N2 (24,636 HOGs) and N3 (24,416 HOGs) correspond to more recent diversification events within Solanaceae. The N2 comprises orthogroups shared among tomato (*S. lycopersicum* L.), potato (*S. tuberosum* L.), and pepper (*C. annuum* L.), reflecting Solanaceae-wide gene families retained across these lineages. In contrast, N3 encompasses orthogroups shared exclusively between tomato and potato, representing lineage-specific diversification within *Solanum*.

Species-Specific Patterns of HOG Representation

To investigate species-specific contributions to gene families across evolutionary nodes, we quantified both the number of HOGs containing genes from each species and the average gene copy number per HOG (*Figure 3*). Three evolutionary levels were considered: N0 — shared with *A.thaliana* (euasterid core); N1 — ancestral Solanaceae node; and N2 — *Solanum*-specific node. In the left panel of *Figure 3*, species-level HOG representation exhibits a clear phylogenetic signal. *A. thaliana* is present exclusively at N0, with approximately 15,000 HOGs, reflecting its role as the outgroup. The *Arabidopsis* genes are represented at N1 or N2. In *C. annuum, there* a marked increase in HOG representation from N0 to N1 (up to ~20,000 HOGs), followed by a sharp decline at N2, consistent with its early divergence relative to *Solanum*. The *S. lycopersicum* and *S. tuberosum* display similar patterns: representation is observed at N0, increases at N1, and peaks at N2 (>20,000 HOGs), indicating lineage-specific emergence and/or expansion of gene families within *Solanum*.

The right panel of *Figure 3* shows the average gene copy number per HOG, providing insights into duplication and retention dynamics. At N0, *A. thaliana* exhibits the highest copy numbers (>2.5), reflecting ancient duplication events predating the Solanaceae lineage. Pepper, tomato, and potato show moderately elevated values (~2.0) at N0, which gradually decrease through N1 and N2. *Solanum* species maintain relatively stable copy numbers (>1.0) across nodes, whereas pepper shows a pronounced reduction at N2, consistent with its exclusion from *Solanum*-specific orthogroups.
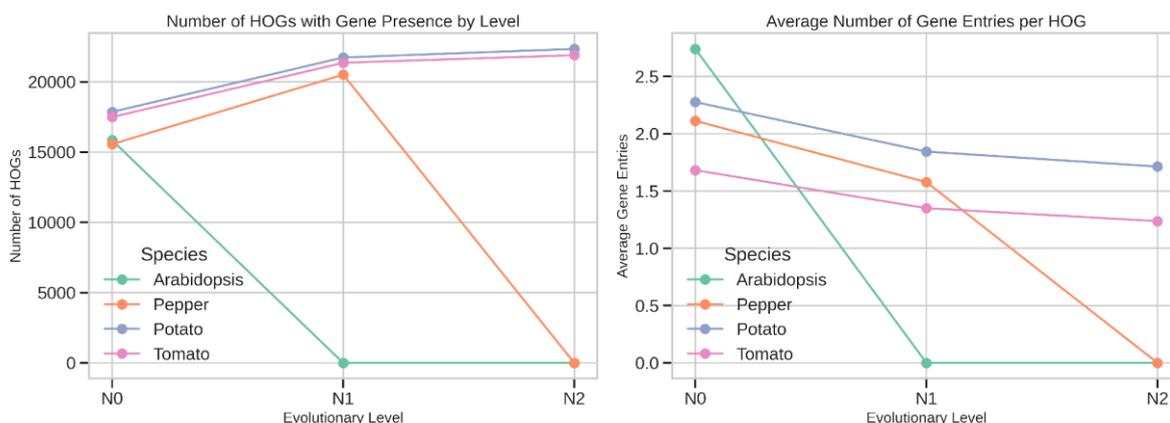
Figure 3. Species-Specific Distribution of Orthogroup Presence and Gene Copy Number across Evolutionary Levels (N0–N2) Left: Number of HOGs containing gene entries per species. **Right:** Average number of gene copies per HOG.Pepper exhibits a strong presence at N1 but not N2, whereas tomato and potato continue to increase through N2, reflecting Solanum-specific expansions.

Pepper-Specific HOG Expansions

Lineage-specific gene family expansions in pepper were examined by quantifying *C. annuum*-specific HOGs across evolutionary nodes (*Figure 4*). Approximately 900 pepper-specific HOGs were identified at N0, indicating that a subset of lineage-specific copies is associated with ancestral gene families conserved across the euasterid clade. A pronounced increase was observed at N1, with nearly 2,900 pepper-specific HOGs—more than threefold higher than at N0—suggesting substantial gene family diversification at the ancestral Solanaceae node. No pepper-specific HOGs were detected at N2, which corresponds to the *Solanum* lineage. This pattern indicates that most pepper-specific gene family innovations arose after the divergence from *A. thaliana* but prior to the diversification of *Solanum*. These expansions likely involve receptor-like kinases, defence-related genes, and metabolic pathways, reflecting lineage-specific ecological and physiological adaptations in *Capsicum*.
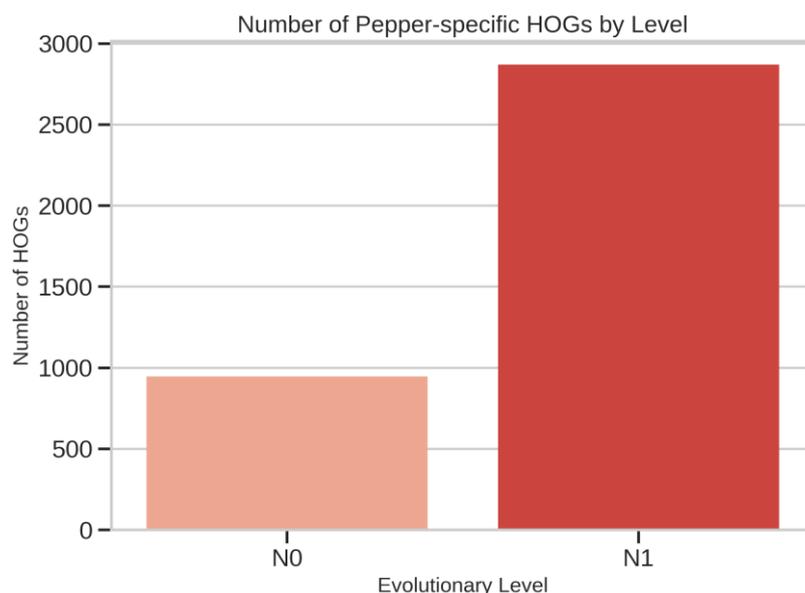
Figure 4. Pepper-Specific HOG Counts across Evolutionary Levels The highest number of pepper-specific groups is found at N1, suggesting a burst of lineage-specific expansions during early Solanaceae evolution.

Tomato–Pepper Lineage-Specific Differences

To assess lineage-specific patterns of gene family retention and loss, we compared the number of unique HOGs between tomato (*S.lycopersicum*) and pepper (*C. annuum*) across evolutionary nodes (***Figure 5***). Two categories were considered: (1) HOGs present in tomato but absent in pepper (blue), and (2) HOGs present in pepper but absent in tomato (orange). At N0, tomato-specific HOGs outnumbered pepper-specific ones (~3,000 vs. <1,000). A similar pattern was observed at N1, with approximately 3,000 tomato-specific HOGs compared to ~1,500 in pepper. The most pronounced difference occurred at N2, corresponding to the *Solanum*-specific node, where over 20,000 HOGs were present in tomato but absent in pepper. As expected, no pepper-specific HOGs were detected at N2, reflecting the absence of *Capsicum*-specific lineages in the *Solanum* clade. Overall, this analysis indicates that tomato has retained or acquired substantially more lineage-specific orthogroups than pepper, particularly within *Solanum*-specific lineages. This pronounced asymmetry likely reflects extensive functional diversification following the tomato–pepper split.
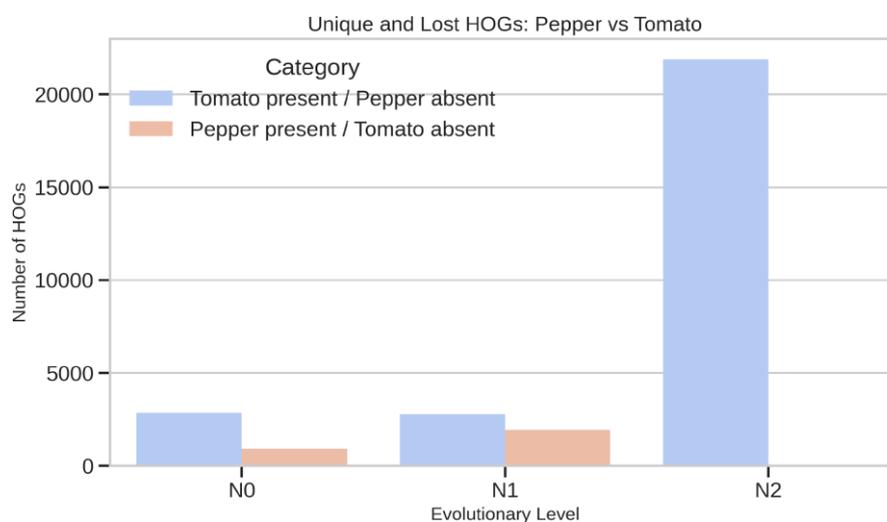
Figure 5. Unique and Lost HOGs in Tomato and Pepper across Evolutionary Levels (N0–N2) Tomato exhibits consistently higher numbers of lineage-specific orthogroups, especially at N2, reflecting major gene family innovations in Solanum relative to pepper.

Presence/Absence Variation and Heatmap Visualization

This visualization provides an intuitive overview of shared versus lineage-specific orthogroups and highlights phylogenetic structure within the dataset. Each column of the heatmap corresponds to a single HOG, while each row represents a species (***Figure 6***). Green cells indicate presence (≥1 gene per species per HOG), as example PEBP, CA2, CaAAT, CA00g00710, CA00g06820 and white cells indicate absence (*Supplementary S5_missing_pepper.txt; Supplementary S6_missing_potato.txt*). The resulting pattern reveals a pronounced phylogenetic signal. A. thaliana exhibits the sparsest distribution, with most HOGs absent (white), consistent with its role as an outgroup and its divergence prior to Solanaceae diversification (*Supplementary S7_unique_arabidopsis_ genes.txt*). In contrast, tomato and potato display nearly complete representation across the sampled HOGs, reflecting a high degree of overlap and shared orthogroups within Solanum. Pepper exhibits an intermediate pattern, with many shared orthogroups (*Supplementary S8_Unique_pepper_genes.txt*) but also a notable number of absences (*Supplementary S5_missing_pepper.txt*), particularly relative to tomato and potato and potato relative to pepper and Arabidopsis (*Supplementary S6_missing_potato.txt*). This presence/absence structure mirrors the species phylogeny: N0 (ancestral node) shows widespread presence across all species, corresponding to ancient, conserved gene families; N1 (ancestral Solanaceae) displays a mixed pattern, with pepper sharing many orthogroups but also exhibiting some lineage-specific gains and losses; N2 (Solanum-specific) shows extensive overlap between tomato and potato, while pepper exhibits frequent absences, reflecting post-divergence gene family innovations in Solanum.
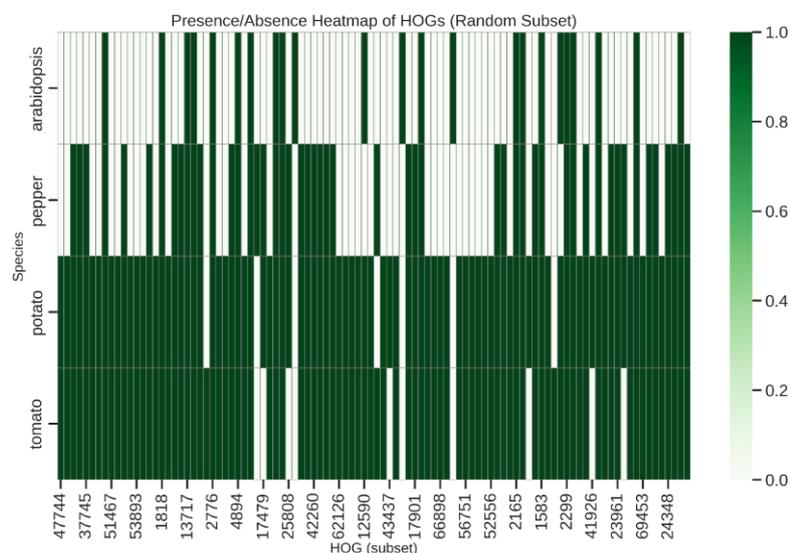
Figure 6. Presence/Absence Heatmap of a Random Representative Subset of 100 HOGs across *Arabidopsis*, Pepper, Potato, and Tomato Green indicates gene presence; white indicates absence. The heatmap highlights conserved ancestral gene families at N0, mixed patterns at N1, and Solanum-specific gene family expansions and retentions at N2.
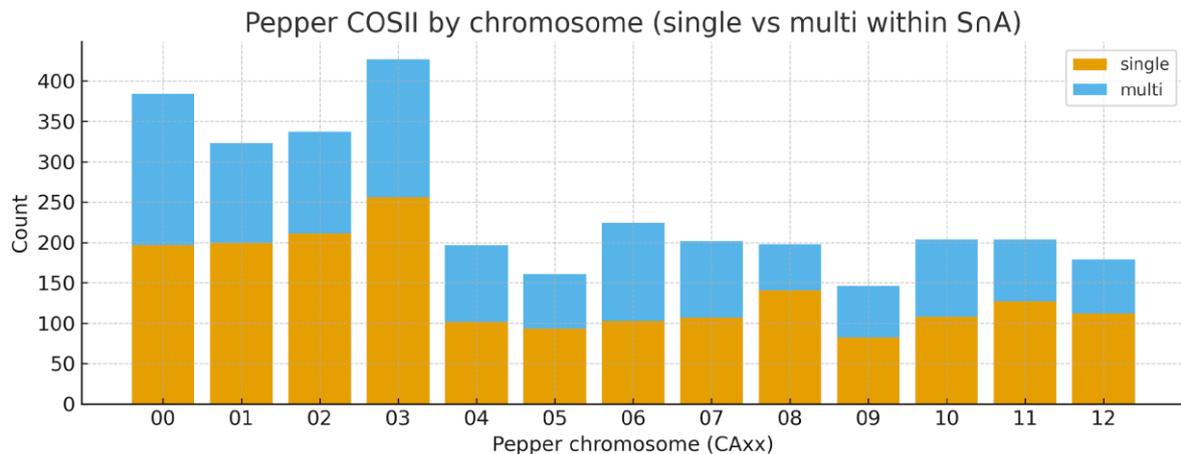
Figure 7. Chromosomal Distribution of COSII Loci in Pepper.

COSII Orthogroup Distribution and Synteny

A large proportion of COSII loci were conserved between tomato and potato (2,358 loci; 82.7%), reflecting their close evolutionary relationship within the Solanum lineage [15,28]. In contrast, fewer loci were recovered in pepper (2,414 loci; 84.7% shared, but with 438 missing), consistent with its greater phylogenetic distance and more fragmented genome assembly [7,43]. The chromosomal distribution and synteny of COSII loci were analysed to evaluate the conservation of genome structure and the occurrence of lineage-specific rearrangements among S. lycopersicum, S. tuberosum, and C. annuum. Mapping COSII orthogroups onto their respective reference genomes revealed broad coverage across all chromosomes, with pronounced clustering observed in specific genomic regions. Collectively, these results indicate that COSII orthogroups are broadly distributed across Solanaceae genomes but exhibit species-specific clustering and local disruptions in synteny, particularly in C. annuum. To further resolve the structural organization and conservation of COSII loci, we next analysed their chromosomal distribution and large-scale collinearity among the three Solanaceae genomes.

Chromosomal Distribution and Synteny of COSII Loci

Mapping COSII loci to chromosomes revealed broadly conserved positional patterns across tomato, potato, and pepper. Most single-copy COSII genes occupied syntenic regions corresponding to ancestral Solanaceae chromosome blocks. Synteny analysis confirmed large-scale collinearity among the three genomes, especially between tomato and potato. In tomato and potato, COSII loci exhibited a relatively even chromosomal distribution, with only modest clustering on chromosomes 1 and 9. In contrast, *C. annuum* displayed localized disruptions, including inversions, translocations, and small-scale rearrangements affecting COSII intervals. COSII loci in pepper were distributed more unevenly, with pronounced concentrations on chromosomes 3 (n = 427), 2 (n = 337), and 1 (n = 323), as well as notable enrichments on chromosomes 6, 10, and 11 (***Figure 11***). This skewed distribution in pepper corresponds to previously reported structural complexities and assembly fragmentation within the Capsicum genome. Several COSII loci that were single copy in tomato and potato but multicopy in pepper were located near syntheses breakpoints or within regions of known structural variation. These patterns indicate that gene duplication events in pepper are associated with genomic rearrangements rather than random expansion. Comparative synteny analyses revealed multiple conserved collinear blocks across the three genomes, with the greatest number shared between tomato and potato, consistent with their close evolutionary relationship. Conserved syntenic regions were also detected between tomato and pepper, although these frequently exhibited local structural rearrangements.

Conservation of the COSII Core

Intersecting COSII orthogroups across *S.lycopersicum*, *S. tuberosum*, *C. annuum*, and *A. thaliana* confirmed that nearly all Solanaceae COSII loci are embedded within deeply conserved euasterid-wide single-copy orthogroups. Integration of OrthoFinder identified 2,359 orthogroups shared between the three Solanaceae genomes and Arabidopsis (S∩A), of which 1,839 (78.0%) are strictly single copy across all three species. Tomato and potato displayed complete 1:1 orthology within this set, whereas *C.annuum* exhibited a modest average of 1,351 genes per orthogroup, corresponding to 520 multi-copy loci (22.0%). These duplications were predominantly localized to chromosomes 3, 2, and 1, consistent with previously reported structural dynamics in the Capsicum genome. Importantly, no overlap was detected between COSII loci and the top 20 expanded or contracted gene families in Solanaceae, underscoring that COSII genes occupy a functionally constrained, low-copy genomic space. The full embedding of the Solanaceae COSII set within Arabidopsis orthogroups (S ⊂ A; S\A = 0) further confirms their ancient origin and stability throughout euasterid evolution. Together, these findings establish COSII loci as a robust genomic backbone for comparative mapping, phylogenetic reconstruction, and marker transfer across Solanaceae species, while distinguishing them from the dynamic gene families driving lineage-specific innovations.

Copy-Number Variation and Lineage-Specific COSII Dynamics

Copy-number distribution among COSII orthogroups revealed strong stability in tomato and potato, with nearly all loci preserved as single-copy. Pepper, however, showed a non-uniform pattern of expansion, with clear hotspots corresponding to functionally enriched genomic intervals. To investigate patterns of gene duplication and retention in *C. annuum*, we analysed the distribution of gene copy numbers per orthogroup within the COSII dataset (***Figures 8–10***). This analysis provides insights into lineage-specific gene family expansions, gene loss, and the overall copy-number architecture of the pepper genome in comparison with *S. lycopersicum* and *S. tuberosum*. The COSII-based OrthoFinder analysis revealed that *C. annuum* harbors an average of 1,351 genes per orthogroup, corresponding to 520 multi-copy orthogroups (22.0%) within the conserved COSII subset (S∩A = 2,359) shared among Solanaceae species. In contrast, both tomato and potato maintained a strict single-copy status (1:1:1 orthology across all 1,839 loci). These duplications in pepper were predominantly localised to chromosomes 3 (n = 427), 2 (n = 337), and 1 (n = 323), with additional enrichment on CA00 (n = 384) a set of unplaced scaffolds consistent with assembly fragmentation. This uneven chromosomal clustering supports the presence of segmental duplications and local structural rearrangements previously reported in the Capsicum genome [7,41]. Importantly, the multi-copy COSII orthogroups in pepper did not overlap with the top expanded gene families in Solanaceae, indicating that these duplications represent structural, rather than functional, genomic events.
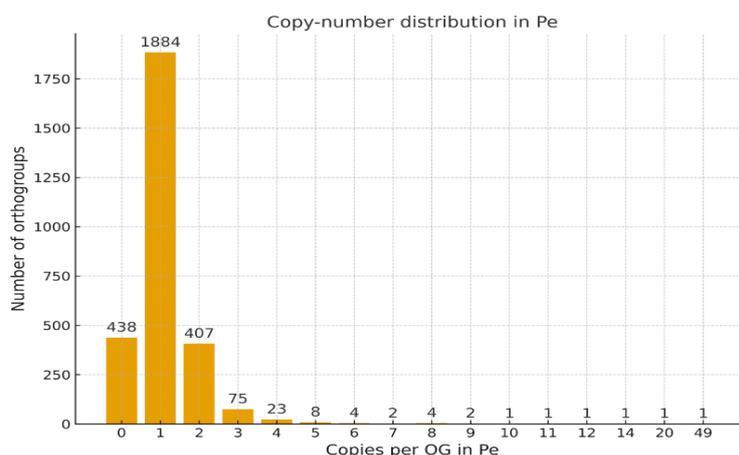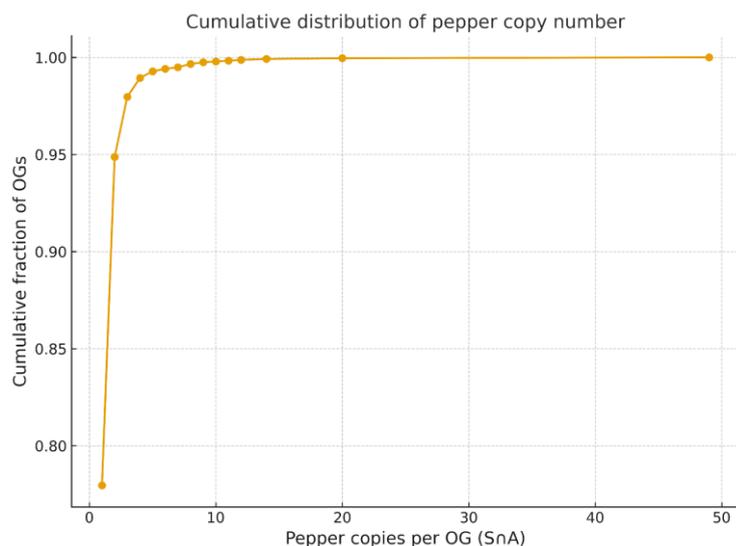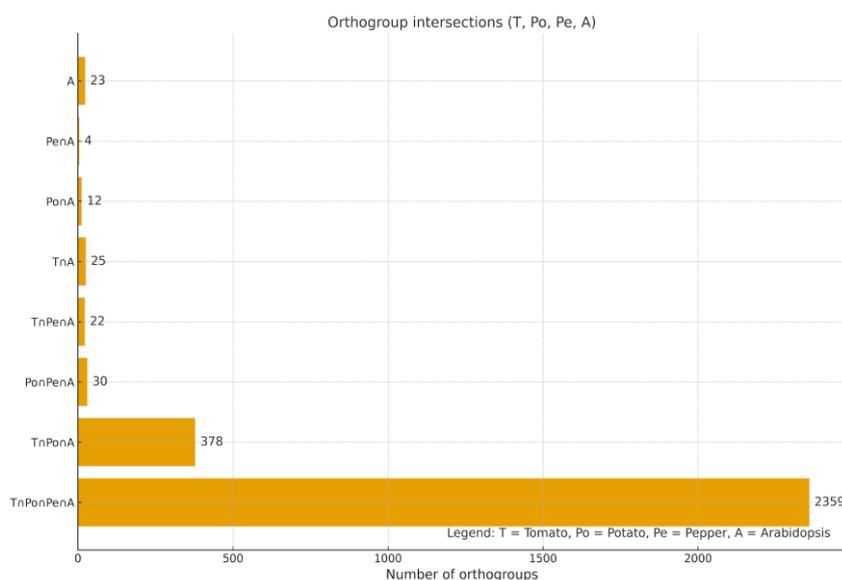

Figure 8. Gene Copy Number Distribution across Orthogroups in *C. annuum*.

Figure 9. Cumulative Distribution of Gene Copy Numbers in Pepper Orthogroups.

Figure 10. Shared and Species-Specific Orthogroups in Tomato, Potato, Pepper, and Arabidopsis *TnPonPenA* - Tomato, Potato, Pepper, Arabidopsis *TnPonA* = Tomato, Potato, Arabidopsis *PonPenA* = Potato, Pepper, Arabidopsis *TnPenA* = Tomato, Pepper, Arabidopsis *TnA* = Tomato, Arabidopsis *PonA* = Potato, Arabidopsis *PenA* = Pepper, Arabidopsis *A* = Arabidopsis.

Histogram of Gene Copy Number per Orthogroup

The majority of pepper orthogroups contained either a single gene copy or none at all (***Figure 8***). Within the 2,359 conserved COSII orthogroups, 1,839 were represented by a single-copy gene, constituting the dominant category. Additionally, 438 orthogroups lacked any corresponding gene in pepper, likely reflecting annotation gaps or lineage-specific gene loss, while 407 orthogroups contained two copies. Higher copy numbers were progressively rarer: 75 orthogroups contained three copies, 23 contained four, and only a few exhibited more than five paralogues. A small number of extreme cases were observed, including orthogroups with 14, 20, and up to 49 paralogues, corresponding to substantially expanded gene families. This right-skewed distribution, characterized by a dominant single-copy peak and a long tail of low-frequency multi-copy orthogroups, is typical of plant genomes. It reflects the combined effects of purifying selection on conserved, low-copy genes and targeted duplications of specific gene families potentially linked to adaptive functions in

C.annuum. Chromosomal mapping confirmed that these multi-copy loci were primarily clustered on Chr03, Chr02, and Chr01, with additional representation on CA00**,** consistent with the localization of segmental duplications identified in previous genomic studies [41].

Cumulative Distribution of Gene Copy Numbers

The cumulative distribution curve of gene copy numbers in pepper (***Figure 9***) reveals a pronounced skew toward low copy numbers. More than 95% of orthogroups contain three or fewer gene copies**,** and the cumulative curve approaches saturation rapidly, indicating that high copy orthogroups are rare. The small subset of orthogroups with ≥10 copies represent exceptional, lineage-specific expansions within *Capsicum*. This cumulative pattern confirms that the *C. annuum* genome is dominated by single-copy and low-copy orthogroups**,** reflecting the overall genomic stability of the COSII framework. The limited but distinct high-copy fraction likely corresponds to lineage-specific duplications associated with defense mechanisms, stress responses, or specialised metabolic pathways a trend previously noted in Solanaceae [32,42]. The steep initial slope and rapid plateau of the cumulative distribution emphasise that low copy numbers overwhelmingly dominate within the COSII subset, reinforcing its value as a low-duplication marker system. This cumulative distribution underscores the highly constrained nature of the COSII genomic subset. To place these patterns in a broader comparative framework, we next assessed the coverage and conservation of COSII orthogroups across the three Solanaceae genomes.

Comparative Coverage and Conservation

Species-level coverage analyses using independent tomato and potato orthogroup lists demonstrated complete concordance with both the COSII universe (U = 2,853) and the Solanaceae shared subset (S = 2,359). Both tomato and potato recovered the entire conserved set (Tomato ∩ S = 2,359; Potato ∩ S = 2,359), with no species-specific or missing orthogroups within the COSII panel. All shared orthogroups (S) were also present in *A.thaliana* (S ⊂ A; S \ A = 0), confirming that COSII loci represent ancient, deeply conserved euasterid genes (***Figure 10***). The complete recovery of these loci by *Solanum* species, coupled with their consistent single-copy status, underscores the stability and evolutionary robustness of COSII loci as anchoring elements for comparative genomics and phylogenomics**.** In contrast, copy number variation observed in pepper primarily reflects lineage-specific duplication events rather than broad-scale genome restructuring across Solanaceae. The results obtained are shown in Figures 11. This structural flexibility within *C. annuum* likely underpins its unique phenotypic diversification, while maintaining the conserved COSII genomic backbone shared across Solanaceae species. Overall, these observations reinforce the dual evolutionary pattern of the Solanaceae genome — structural dynamism within *C. annuum* juxtaposed with a deeply conserved COSII core shared across the Solanaceae–Arabidopsis lineage. Several of the pepper multi-copy COSII orthogroups correspond to gene families known to participate in abiotic stress responses, including kinase-mediated signalling, ROS detoxification, and osmoprotective pathways. Although COSII genes are overall conserved and low-copy, these localized expansions may reflect lineage-specific adaptations to drought and heat stress in Capsicum.

Summary and Evolutionary Implications

This comprehensive in silico analysis of COSII orthologues across *S. lycopersicum*, *S. tuberosum*, and *C. annuum* provides new insights into genome conservation, gene family dynamics, and lineage-specific evolutionary trajectories within Solanaceae. We identified a total of 2,852 COSII orthogroups, of which 2,358 (82.7%) were shared among all the three species, reflecting a highly conserved genomic backbone. The majority of these loci (1,839; 64.5%**)** were present as single-copy orthologues, confirming the stability and suitability of COSII genes for comparative genomics and phylogenetic reconstruction. Tomato and potato exhibited nearly identical orthogroup coverage and uniformly single-copy status, whereas pepper displayed moderate copy number variation, averaging 1,351

genes per orthogroup and comprising 520 multi-copy orthogroups (22%). These duplications were clustered primarily on chromosomes 1–3 and CA00, consistent with localised segmental rearrangements rather than large-scale gene family amplifications. The HOG analyses revealed major gene family expansions at the ancestral Solanaceae node (N1), followed by progressive diversification within *Solanum* (N2–N3). Pepper-specific expansions were concentrated at N1, while tomato exhibited markedly higher numbers of lineage-specific orthogroups at Solanum specific nodes, suggesting distinct evolutionary trajectories within the family. The resulting two-phase model comprising an early burst of ancestral gene expansion followed by lineage specific functional specialisation captures the evolutionary dynamics that have shaped the modern Solanaceae genomes. Presence/absence heatmaps and synteny analyses further illustrated the dual genomic signature of deep conservation and structural innovation. Tomato and potato shared extensive collinear blocks, whereas pepper displayed localised rearrangements and uneven chromosomal clustering of COSII loci. Copy number variation analyses confirmed that the pepper genome is dominated by single-copy and low-copy orthogroups, with a right-skewed distribution reflecting a small number of expanded, adaptively relevant gene families. Collectively, these findings highlight a dual evolutionary signature within Solanaceae: a stable, deeply conserved COSII core shared across the euasterid lineage, coupled with lineage-specific structural and functional innovations driving diversification. COSII markers therefore provide a robust, evolutionarily stable scaffold for comparative genomics, phylogenetic inference, and molecular breeding in major Solanaceae crops. A total of 2,852 orthogroups were identified across *S. lycopersicum*, *S. tuberosum*, and *C. annuum*. Among these, 2,358 orthogroups (82.7%) were shared among all three species, forming a deeply conserved genomic backbone within Solanaceae. Within this set, 1,839 orthogroups (64.5%) represent strict single-copy orthologues, which are especially valuable for comparative genomics and phylogenetic reconstruction as they minimise paralogy-related ambiguity. Orthology inference across Solanaceae and *A. thaliana* identified 2,359 shared orthogroups (S∩A), confirming that nearly all Solanaceae COSII loci correspond to euasterid-wide single-copy genes. Tomato and potato were strictly single copy across this core set, whereas pepper exhibited moderate copy number variation, averaging 1,351 genes per orthogroup. Specifically, *C. annuum* contained 520 multi-copy orthogroups (22%), consistent with local gene family duplications rather than whole-genome amplification. Chromosomal mapping revealed that pepper COSII loci (*Figure 7*) were enriched on chromosomes 3, 2, and 1, with additional accumulation on CA00 unplaced scaffolds, in line with previously described structural rearrangements and assembly gaps in the Capsicum genome [7,41]. The absence of Solanaceae-shared COSII orthogroups missing from Arabidopsis ($|S \setminus A| = 0$) underscores the deep evolutionary conservation of the COSII marker set. Notably, the top Solanaceae-expanded and contracted gene families (n = 20 and 21 orthogroups, respectively) showed no overlap with the COSII universe, reinforcing that COSII loci form a stable, low-copy genomic framework largely independent of lineage-specific gene family dynamics. Species-wise coverage analyses further confirmed full concordance between the COSII universe (U = 2,853) and the Solanaceae-shared subset (S = 2,359): both tomato and potato recovered the complete S set (Tomato ∩ S = 2,359; Potato ∩ S = 2,359), with no species-specific or missing orthogroups. Collectively, these results demonstrate that COSII orthogroups define a highly conserved, low-copy genomic scaffold across Solanaceae, while the limited copy number variation observed in *C. annuum* reflects lineage-specific duplications localised to a few chromosomal regions.
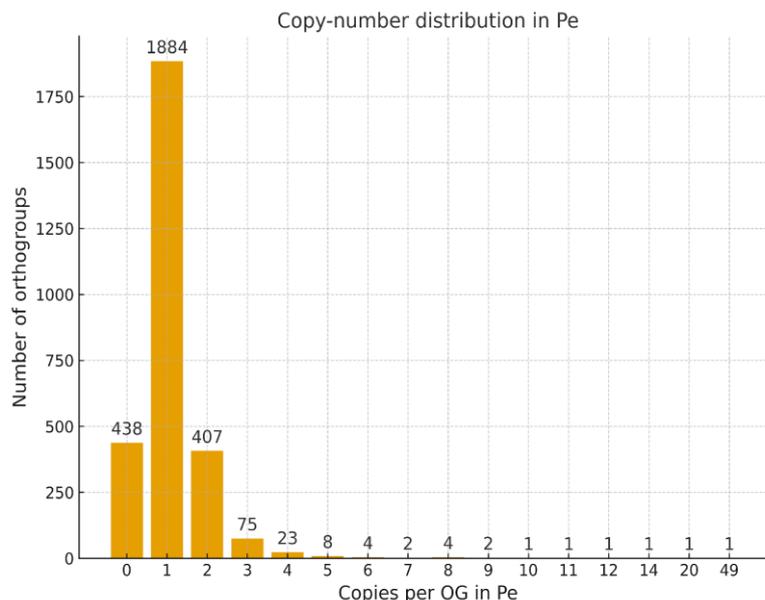
Figure 11. Copy-Number Distribution of COSII Orthogroups in *Capsicum annuum* (Pe). Histogram showing the frequency of orthogroups according to the number of gene copies identified in the *C. annuum* genome. The majority of orthogroups are represented by a single gene copy (n = 1,884), while smaller fractions contain zero (n = 438) or two copies (n = 407). Only a few orthogroups exhibit three or more copies, indicating that most COSII loci remain single- or low-copy in the pepper genome.

4. Discussion

Our comparative analysis of COSII markers across *S. lycopersicum*, *S. tuberosum*, and *C. annuum* provides a genome-wide view of orthology, gene family evolution, and structural divergence within Solanaceae. By integrating reciprocal BLAST searches, OrthoFinder-based orthology inference, and synteny analysis, we resolved both deeply conserved loci and lineage-specific variations that have shaped Solanaceae genome architecture.

Conservation and Divergence of COSII Orthologs

OrthoFinder analysis significantly improved ortholog detection compared with reciprocal BLAST alone, particularly for historically duplicated families. Ambiguous reciprocal hits were confidently assigned to single copy orthogroups (1,839 total), highlighting the reliability of phylogeny aware orthology inference [19]. These single copy COSII orthogroups form a robust framework for comparative mapping, phylogenomics, and evolutionary rate estimation within Solanaceae.

Chromosomal Organization and Synteny

Mapping of COSII loci revealed a broad yet structured chromosomal distribution. In both tomato and potato, loci were evenly distributed with moderate clustering on chromosomes 1 and 9, while in pepper, COSII loci were enriched on chromosomes 3, 2, and 1, and on unplaced scaffolds (CA00), consistent with segmental duplications and incomplete assembly. Synteny analyses revealed extensive collinearity between tomato and potato, while Capsicum exhibited localised rearrangements, particularly on chromosomes 3, 5, and 11 [44]. One notable example involves a COSII locus adjacent to the *MYB12* regulatory gene on tomato chromosome 1, which remains conserved in potato and pepper but exhibits structural rearrangements in *Capsicum*. Such loci coincide with evolutionary hotspots for fruit pigmentation and secondary metabolism [7,18]. Many

expanded COSII loci fall within genomic regions known to harbor stress-response and signaling gene families, suggesting targeted lineage-specific gene retention in pepper.

Evolutionary Dynamics of Gene Families

The distribution of HOGs across phylogenetic nodes clarifies the tempo and mode of gene family diversification. The largest expansion (40,499 HOGs) occurred at the ancestral Solanaceae node (N1), preceding the divergence of *Capsicum* and *Solanum*. This reflects a major wave of gene family expansion—particularly receptor-like kinases (RLKs), receptor-like proteins (RLPs), and NBS-LRR genes [31,32,34] that established the functional diversity seen in extant Solanaceae. Later nodes **(N2–N3)** represent Solanum-specific diversification, including tuber formation, fruit pigmentation, and developmental regulation [2,36]. The conserved genomic backbone at N0 (32,104 HOGs) corresponds to essential euasterid genes shared with *A. thaliana*. Together, these results support a two-phase evolutionary model**:** (1) early Solanaceae-wide gene family expansion (N0–N1) forming a versatile genomic scaffold, followed by (2) lineage-specific specialization (N2–N3) shaping modern phenotypes and adaptive capacities. Notably, several of the expanded HOGs identified in *C.annuum* contain domains and functional annotations associated with abiotic stress responses, including drought tolerance, osmotic adjustment, and heat stress signaling. This observation aligns with the moderate but biologically meaningful increase in gene copy number detected in pepper relative to the strictly single-copy Solanum genomes. Such lineage-specific expansions may have contributed to the adaptive plasticity of Capsicum, supporting more efficient responses to water deficit and fluctuating environmental conditions.

Copy Number Variation and Structural Patterns

Tomato and potato retained almost complete one-to-one orthology for these loci, consistent with previous reports of their conserved genome structure. In contrast, *C. annuum* displayed moderate but notable copy-number variation [7,8,41,45]. Across 2,359 shared orthogroups (S∩A), tomato and potato were strictly single copy, whereas pepper exhibited an average of 1,351 genes per orthogroup, corresponding to 520 multi-copy orthogroups (22%)**.** These duplications were primarily confined to a few chromosomal regions and did not overlap with the 20 most expanded Solanaceae gene families, confirming that COSII markers represent a low-copy, evolutionarily stable subset of the genome. Histograms and cumulative distributions further demonstrated that >95% of pepper orthogroups contain ≤3 gene copies, while high-copy expansions (≥10 copies) are extremely rare. This right-skewed distribution indicates strong purifying selection on single-copy genes and targeted duplication of adaptive gene families, particularly those linked to stress tolerance and defence [32,42]. Together, these structural and copy-number patterns highlight the interplay between deep genomics conservation and lineage-specific innovation within *C. annuum*.Importantly, chromosomes 1–3 of Capsicum where COSII-associated duplications are most strongly enriched partially overlap with published QTL regions for drought tolerance, osmotic stress regulation, and heat stress response. This suggests that the expanded multicopy orthogroups within these regions may represent adaptive modules that evolved in Capsicum under abiotic stress conditions, without compromising the stability of the low-copy COSII genomic core. The expansion of RLK, HSP, LEA, and NBS-LRR–related orthogroups at N1 and in pepper CNV subsets suggests that abiotic stress responses may have been a key driver of early Solanaceae diversification. Capsicum, in particular, shows lineage-specific duplication in gene families associated with drought tolerance, cuticle biosynthesis, and osmotic stress signaling, consistent with its ecological adaptation to hotter and drier environments.

Limitations and Future Perspectives

While our in silico approach provides a systematic framework for COSII orthologues identification, several factors may influence the results. Differences in genome assembly quality, annotation completeness, and gene model predictions can affect orthology inference. Additionally,

COSII markers were originally derived from expressed sequence tags (ESTs), which may not fully capture recent gene family expansions or lineage-specific duplications [13]. Integrating transcriptomic data and applying synteny-aware gene model curation will help refine orthologue assignments, particularly in pepper, where assembly fragmentation remains a challenge. The near-complete embedding of Solanaceae-shared COSII orthogroups within *A. thaliana* (S ⊂ A; S \ A = 0) underscores their ancient and deeply conserved nature. The strict single-copy status observed in *Solanum* and the localized multi-copy expansions in *Capsicum* (~22%) indicate lineage-specific duplication events rather than genome-wide copy number shifts [7,43]. These features make COSII markers exceptionally stable genomic anchors for comparative genomics, QTL mapping, and phylogenetic reconstruction. Moreover, the pepper specific multi-copy orthogroups identified here offer complementary opportunities to investigate recent genomic innovations and structural variation within Solanaceae. Future studies integrating COSII-based comparative analyses with transcriptomic datasets under abiotic stress (e.g., drought, salinity, and heat stress) would further clarify whether the pepper-expanded orthogroups participate in stress-responsive pathways. The evolutionary stability and predominantly single-copy nature of COSII loci make them ideal reference anchors for such expression-based investigations.

Evolutionary Implications and Marker Utility

The observed patterns of nucleotide diversity among COSII loci closely reflect the phylogenetic relationships of the analysed species. Average sequence identity between tomato and potato orthologues is high (~95%), whereas tomato–pepper orthologues exhibit lower identity (~87%), consistent with divergence time estimates of 19–23 Mya for the tomato–pepper split [15,28]. A subset of COSII loci showed elevated divergence in pepper, potentially reflecting lineage-specific adaptive processes or the effects of domestication bottlenecks [46]. Due to their conserved, predominantly single-copy nature, COSII markers remain highly effective tools for comparative genomics, map-based cloning, and phylogenetic inference [13,47]. The conserved orthologous sets, identified in this study, provide stable anchor points for aligning genetic and physical maps, thereby facilitating the efficient transfer of genomic information across species, a key objective for Solanaceae crop improvement. Because of their high evolutionary conservation and low duplication rate, COSII loci provide an ideal genomic framework for mapping lineage-specific expansions related to abiotic stress responses, enabling integrated analyses of functional adaptation and genome evolution in Solanaceae.

Gene Family Evolution in Solanaceae

The distribution of HOGs across phylogenetic nodes provides valuable insights into the timing and mechanisms of gene family diversification within Solanaceae. The largest expansion occurred at the early ancestral node N1, preceding the divergence of *Capsicum* and *Solanum* (***Figure 1***). This pattern is consistent with previous reports of large-scale expansions in receptor-like kinase (RLK) and receptor-like protein (RLP) gene families during early Solanaceae diversification [31,32,34]. Such expansions likely provided the genetic substrate for the evolution of novel signalling and defence mechanisms. The high number of orthogroups detected at N1 (40,499 HOGs) reflects both the retention of ancient genes and extensive gene duplication associated with adaptive radiation. This finding supports the hypothesis that Solanaceae experienced a period of rapid genomic innovation linked to ecological diversification, host pathogen interactions, and pollinator specialisation [33,34]. In contrast, the reduced number of orthogroups at nodes N2 and N3 corresponds to more recent lineage-specific innovations that arose after the Capsicum–Solanum split, including traits such as tuber formation and specialised secondary metabolism [2,36]. The conserved genomic core represented at N0 (32,104 HOGs) highlights the deep evolutionary backbone shared with *A. thaliana*, comprising essential housekeeping genes and fundamental metabolic functions [35,39]. In this study the *A.thaliana* genome was used due to the fact that it has been very thoroughly studied. Together, these patterns support a two-phase model of Solanaceae genome evolution: (1) early expansions at

N0–N1 established a versatile genomic framework, and (2) subsequent diversification at N2–N3 refined this repertoire, underpinning the phenotypic and ecological diversity observed in modern Solanaceae crops. This two-phase framework, characterised by ancestral expansion followed by lineage-specific refinement, encapsulates the evolutionary forces that shaped the genomic diversity of modern Solanaceae crops [47–50].

COSII orthologues as a stable genomic framework for investigating abiotic stress responses

Although COSII loci are not specifically enriched for abiotic stress-related gene families, their deeply conserved single-copy nature makes them an exceptionally stable genomic framework for comparative studies of stress adaptation, including drought tolerance. Many stress-responsive pathways, such as ROS detoxification, ABA signaling, osmoprotection, and membrane stability, rely on housekeeping metabolic and regulatory components that show strong evolutionary conservation across angiosperms. Because COSII orthologues represent this highly conserved functional layer, they provide reliable anchor points for integrating transcriptomic and functional genomic datasets generated under abiotic stress conditions. Furthermore, lineage-specific copy number variation detected in *C. annuum*, although limited in scope, may intersect with gene families involved in stress perception and signalling (e.g., receptor-like kinases, hormone-related regulators). Such localized expansions could reflect adaptive processes linked to the ecological specialization of pepper in warmer and more drought-prone environments. The stable COSII core combined with lineage-specific variability in non COSII gene families offers a useful framework for distinguishing ancestral metabolic constraints from recently evolved stress-response innovations. In future studies, the COSII-based orthogroup architecture defined here can serve as a genomic scaffold for mapping transcriptional responses to drought and other abiotic stresses across Solanaceae species.

5. Conclusions

In this study, we provide the most comprehensive in silico reassessment of COSII loci across tomato, potato, and pepper using modern high-quality genome assemblies and graph-based orthology frameworks. Our analyses show that most COSII genes form a deeply conserved single-copy core across Solanaceae, confirming their longstanding value as stable markers for phylogeny, comparative genomics, and genetic mapping. At the same time, we identify clear lineage-specific innovations, particularly in *C. annuum*, where moderate but non-random copy-number expansion is associated with genomic rearrangements and regions enriched for stress-related gene families. By integrating orthogroup inference, hierarchical orthogroup reconstruction, and chromosome-scale synteny analysis, this work bridges classical COSII marker studies with contemporary genome-scale resources. The results refine our understanding of COSII evolutionary dynamics and highlight the dual genomic architecture of Solanaceae: a stable, conserved ortholog backbone complemented by localized structural divergence in specific lineages. Overall, this updated COSII framework enhances the resolution of comparative analyses and provides a robust foundation for evolutionary, phylogenomic, and breeding applications in Solanaceae. The conserved COSII single-copy set and the identified lineage-specific expansions offer valuable molecular resources for future studies focused on genome evolution, trait diversification, and adaptive responses in these economically important crops.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

HOG	Hierarchical orthogroup
CNV	Copy Number Variation

References

1. Knapp, S. Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. *Taxon* 2002, 51: 45–61. doi: 10.2307/1554965. https://pubmed.ncbi.nlm.nih.gov/12324525/

2. Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature 2012,* 485: 635–641. doi: 10.1038/nature11119. https://pubmed.ncbi.nlm.nih.gov/22660326/

3. Zhou, L.; Feng, T., Xu, S., Gao, F., Lam, T.T., Wang, Q., Wu, T., Huang, H., Zhan, L., Li, L. ggmsa: A visual exploration tool for multiple sequence alignment and associated data. *Briefings in Bioinformatics* 2022, 23, 4: bbac222. https://doi.org/10.1093/bib/bbac222

4. Rivera-Silva, R.; Chávez Montes, R.A., Jaimes-Miranda, F. Gene ontology functional annotation datasets for the ITAG3.2 and ITAG4.0 tomato (*Solanum lycopersicum*) genome annotations. *Data Brief.* 2024, 6;54:110401. doi: 10.1016/j.dib.2024.110401. PMID: 38646191; PMCID: PMC11033075.

5. Hardigan, M.A.; Laimbeer, F.P.E., Newton, L., Crisovan, E., Hamilton, J.P., Vaillancourt, B., Wiegert-Rininger, K., Wood, J.C., Douches, D.S., Farré, E.M., Veilleux, R.E., Buell, C.R. Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication. *Proc. Natl. Acad. Sci. USA*, 2017, *114*, E9999–E10008. doi: 10.1073/pnas.1714380114. https://pubmed.ncbi.nlm.nih.gov/29087343/

6. Pham, G.M.; Hamilton, J.P., Wood, J.C., Burke, J.T., Zhao, H., Vaillancourt, B., Ou, S., Jiang, J., Buell, C.R. Construction of a chromosome-scale long-read reference genome assembly for potato. *GigaScience* 2020, 9: giaa100. doi: 10.1093/gigascience/giaa100. https://pubmed.ncbi.nlm.nih.gov/32964225/

7. Kim, S.; Park, M., Yeom, S.I., Kim, Y.M., Lee, J.M., Lee, H.A., Seo, E., Choi, J., Cheong, K., Kim, K.T., Jung, K., Lee, G.W., Oh, S.K., Bae, C., Kim, S.B., Lee, H.Y., Kim, S.Y., Kim, M.S., Kang, BC, Jo, Y.D., Yang, H.B., Jeong, H.J., Kang, W.H., Kwon, J.K., Shin, C., Lim, J.Y., Park, J.H., Huh, J.H., Kim, J.S., Kim, B.D., Cohen, O., Paran, I., Suh, M.C., Lee, S.B., Kim, Y.K., Shin, Y., Noh, S.J., Park, J., Seo, Y.S., Kwon, S.Y., Kim, H.A., Park, J.M., Kim, H.J., Choi, S.B., Bosland, P.W., Reeves, G., Jo, S.H., Lee, B.W., Cho, H.T., Choi, H.S., Lee, M.S., Yu, Y., Do Choi, Y., Park, B.S., van Deynze, A., Ashrafi, H., Hill, T., Kim, W.T., Pai, H.S., Ahn, H.K., Yeam, I., Giovannoni, J.J., Rose, J.K., Sørensen, I., Lee, S.J., Kim, R.W., Choi, I.Y., Choi, B.S., Lim, J.S., Lee, Y.H., Choi, D. Genome sequence of *Capsicum annuum* reveals the dynamic genome evolution of pepper. *Nat. Genet.* 2014, *46*, 270–278. https://pubmed.ncbi.nlm.nih.gov/24441736/

8. Hulse-Kemp, A.M.; Maheshwari, S., Stoffel, K., Hill, T.A., Jaffe, D., Williams, S.R., Weisenfeld, N., Ramakrishnan, S., Kumar, V., Shah, P., Schatz, MC, Church DM, Van Deynze A (2018). Reference-quality

assemblies of pepper genomes reveal structural variation and diversification of disease-resistance genes**.** *Genome Biol*, *19*, 224. doi: 10.1038/s41438-017-0011-0. https://pubmed.ncbi.nlm.nih.gov/29423234/

9. Ou, L.; Li, D., Lv, J., Chen, W., Zhang, Z., Li, X., Yang, B., Zhou, S., Yang, S., Li, W., Gao, H., Zeng, Q., Yu, H., Ouyang, B., Li, F., Liu, F., Zheng, J., Liu, Y., Wang, J., Wang, B., Dai, X., Ma, Y., Zou, X. Pan-genome of cultivated pepper (Capsicum) and its use in gene presence-absence variation analyses. *New Phytology* 2018 220 (2):360-363. doi: 10.1111/nph.15413. Epub 2018 Aug 21. PMID: 30129229.

10. Peters, S.A.; Bargsten, J.W., Szinay, D., van de Belt, J., Visser, R.G., Bai, Y., de Jong H. Structural homology in the Solanaceae: analysis of genomic regions in support of synteny studies in tomato, potato and pepper. *Plant Journal* 2012, 71(4):602-14. doi: 10.1111/j.1365-313X.2012.05012.x. Epub 2012 Jun 12. PMID: 22463056.

11. Choe, J., Kim, J.E., Lee, B.W., Lee, J.H., Nam, M., Park, Y.I., Jo, S.H. A comparative synteny analysis tool for target-gene SNP marker discovery: connecting genomics data to breeding in Solanaceae. *Database (Oxford)*.2018:bay047. doi: 10.1093/database/bay047. PMID: 29873704; PMCID: PMC6007222.

12. Wei, K.; Stam, R., Tellier, A., Silva-Arias, G.A. **C**opy number variations shape genomic structural diversity underpinning ecological adaptation in the wild tomato *Solanum chilense.* bioRxiv 2023. preprint, **doi:** https://doi.org/10.1101/2023.07.21.549819

13. Wu, F.; Mueller, L.A., Crouzillat, D., Pétiard, V., Tanksley, S.D. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. *Plant Physiology 2006,* 141: 1186–1200. doi: 10.1104/pp.106.082644. https://pmc.ncbi.nlm.nih.gov/articles/PMC1667096/

14. Rodríguez, G.R.; Moyseenko, J.B., Robbins, M.D., Morejón, N.H., Francis, D.M, Oost, K., van der Knaap, E. Tomato Analyzer 3.0: New tools for digital phenotyping**.** *Plant Physiology* 2009, *150*, 842–853. doi: 10.3791/1856. https://pubmed.ncbi.nlm.nih.gov/20234339/

15. Wu, F.; Tanksley, S.D. Chromosomal evolution in the plant family Solanaceae. *BMC Genomics* 2010 11: 182. doi: 10.1186/1471-2164-11-182. https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-11-182 https://pubmed.ncbi.nlm.nih.gov/20236516/

16. Wu, F.; Eannetta, N.T., Xu, Y., Durrett, R., Mazourek, M., Jahn, M.M., Tanksley, S.D. A COSII genetic map of the pepper genome provides a detailed picture of synteny with tomato and new insights into recent chromosome evolution in the genus *Capsicum. Theoretical and Applied Genetics* 2009, 118: 1279–1293. doi: 10.1007/s00122-009-0980-y. https://pubmed.ncbi.nlm.nih.gov/19229514/

17. Lindqvist-Kreuze, H.; Gastelo, M., Perez, W., Forbes, G.A., de Koeyer, D., Bonierbale, M. Phenotypic stability and genome-wide association study of late blight resistance in potato genotypes adapted to the tropical highlands. *Phytopathology* 2014, 104(6):624-33. doi: 10.1094/PHYTO-10-13-0270-R. PMID: 24423400.

18. Lefebvre-Pautigny, F.; Wu, F., Philippot, M., Rigoreau, M., Priyono, Zouine M, et al. High resolution synteny maps allowing direct comparisons between the coffee and tomato genomes. *Tree Genetics & Genomes*, 2010 6: 565–577. doi: 10.1007/s11295-010-0272-3. https://link.springer.com/article/10.1007/s11295-010-0272-3

19. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics**.** *Genome Biol*, 2019, *20*, 238. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1832-y

20. Hosmani, P.S.; Gonzalez, M.F., van de Geest, H., Maumus, F., Bakker, L.V., Schijlen, E., Haarst, J., Cordewener, J., Sanchez-Perez, G., Peters, S., FeI, Zh., Giovannoni, J.J., Mueller, L.A., Saha, S. An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. *bioRxiv* 2019, **doi:** https://doi.org/10.1101/767764. https://www.biorxiv.org/content/10.1101/767764v1.

21. Mi, H.; Poudel, S., Muruganujan, A., Casagrande, J.T., Thomas, P.D. PANTHER version 17: Expanded protein families and improved gene-tree inference**.** *Nucleic Acids Res., 2016, 51*, D468–D476. doi: 10.1093/nar/gkv1194. https://pubmed.ncbi.nlm.nih.gov/26578592/

22. Camacho, C.; Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L. BLAST+: architecture and applications. *BMC Bioinformatics* 2009, 10: 421. doi: 10.1186/1471-2105-10-421. https://pubmed.ncbi.nlm.nih.gov/20003500/

23. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 2013, 30: 772–780. doi: 10.1093/molbev/mst010. https://pubmed.ncbi.nlm.nih.gov/23329690/

24. Capella-Gutiérrez, S.; Silla-Martínez, J.M., Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics 2009,* 25: 1972–1973. doi: 10.1093/bioinformatics/btp348. https://pubmed.ncbi.nlm.nih.gov/19505945/

25. Minh, B.Q.; Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D, von Haeseler, A., Lanfear, R. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 2020, 37: 1530–1534. doi: 10.1093/molbev/msaa015. https://pubmed.ncbi.nlm.nih.gov/32011700/

26. Kalyaanamoorthy, S.; Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 2017, 14: 587–589. doi: 10.1038/nmeth.4285. https://pubmed.ncbi.nlm.nih.gov/28481363/

27. Hoang, D.T.; Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S. UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 2018, 35: 518–522. doi: 10.1093/molbev/msx281. https://pubmed.ncbi.nlm.nih.gov/29077904/

28. Sato, S.; Tabata, S., Hirakawa, H., Asamizu, E., Shirasawa, K., Isobe, S., et al. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 2012 485: 635–641. doi: 10.1038/nature11119. https://pmc.ncbi.nlm.nih.gov/articles/PMC3378239/

29. Krzywinski, M.; Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A. Circos: an information aesthetic for comparative genomics. *Genome Research* 2009, 19: 1639–1645. doi:10.1101/gr.092759.109. https://pubmed.ncbi.nlm.nih.gov/19541911/

30. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v6: recent updates. *Nucleic Acids Research* 2024, 52: 78–82. doi: 10.1093/nar/gkae268. https://pubmed.ncbi.nlm.nih.gov/38613393/

31. Shiu, S.H.; Karlowski, W.M., Pan, R., Tzeng, Y.H., Mayer, K.F., Li, W.H. Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. *Plant Cell* 2004, 16: 1220–1234. doi: 10.1105/tpc.020834. https://pubmed.ncbi.nlm.nih.gov/15105442/

32. Andolfo, G.; Sanseverino, W., Rombauts, S., Van de Peer, Y., Bradeen, J.M., Carputo, D., Frusciante, L., Ercolano, M.R. Overview of tomato (*Solanum lycopersicum*) candidate pathogen recognition genes reveals important Solanum R locus dynamics. *New Phytol.* 2013, 197(1):223-237. doi: 10.1111/j.1469-8137.2012.04380.x, https://pubmed.ncbi.nlm.nih.gov/23163550/

33. Andolfo, G.; Ferriello, F., Tardella, L., Ferrarini, A., Sigillo, L., Frusciante, L., Ercolano, M.R. Evolutionary dynamics and functional specialization of plant NLR gene families. *Plant Biotechnology Journal* 2014. 12: 1–12. doi: 10.1111/pbi.12165. https://pubmed.ncbi.nlm.nih.gov/23163550/

34. Kang, W.H.; Yeom, S.I. Genome-wide analysis of disease resistance genes in pepper (*Capsicum annuum*). *Frontiers in Plant Science* 2018, 9: 1012. doi: 10.3389/fpls.2018.01012

35. Bombarely, A.; Rosli, H.G., Vrebalov, J., Moffett, P., Mueller, L.A., Martin, G.B. A draft genome sequence of Nicotiana benthamiana to enhance molecular plant-microbe biology research. Mol Plant Microbe Interact. 2012 Dec; 25(12):1523-30. doi: 10.1094/MPMI-06-12-0148-TA. PMID: 22876960.

36. Fernandez-Pozo, N.; Menda, N., Edwards, J.D., Saha, S., Tecle, I.Y., Strickler, S.R., Bombarely, A., Fisher-York, T., Pujar, A., Foerster, H., Yan, A., Mueller, L.A. The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. *Nucleic Acids Research 2012,* 40: D1036–D1044. doi: 10.1093/nar/gkr1195. https://pubmed.ncbi.nlm.nih.gov/25428362/

37. Potato Genome Sequencing Consortium.; Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., Zhang, G., Yang, S., Li, R., Wang, J., Orjeda, G., Guzman, F., Torres, M., Lozano, R., Ponce, O., Martinez, D., De la Cruz, G., Chakrabarti, S.K., Patil, V.U., Skryabin, K.G., Kuznetsov, B.B., Ravin, N.V., Kolganova, T.V., Beletsky, A.V., Mardanov, A.V., Di Genova, A., Bolser, D.M., Martin, D.M., Li, G., Yang, Y., Kuang, H., Hu, Q., Xiong, X., Bishop, G.J., Sagredo, B., Mejía, N., Zagorski, W., Gromadka, R., Gawor, J., Szczesny, P., Huang, S., Zhang, Z., Liang, C., He, J., Li, Y., He, Y., Xu, J., Zhang, Y., Xie, B., Du, Y., Qu, D., Bonierbale, M., Ghislain, M., Herrera Mdel, R., Giuliano, G., Pietrella, M., Perrotta, G., Facella, P., O'Brien, K., Feingold, S.E., Barreiro, L.E., Massa, G.A., Diambra, L., Whitty, B.R., Vaillancourt, B., Lin, H., Massa, A.N., Geoffroy,

M., Lundback, S., DellaPenna, D., Buell, C.R., Sharma, S.K., Marshall, D.F.; Waugh, R., Bryan, G.J., Destefanis, M., Nagy, I., Milbourne, D., Thomson, S.J., Fiers, M., Jacobs, J.M., Nielsen, K.L., Sønderkær, M., Iovene, M., Torres, G.A., Jiang, J., Veilleux, R.E., Bachem, C.W., de Boer, J., Borm, T., Kloosterman, B., van Eck, H., Datema, E., Hekkert, Bt., Goverse, A., van Ham, R.C., Visser, R.G. Genome sequence and analysis of the tuber crop potato. *Nature* 2011, 475: 189–195. doi: 10.1038/nature10158. https://pubmed.ncbi.nlm.nih.gov/21743474/

38. Särkinen, T.; Bohs, L., Olmstead, R.G., Knapp, S.A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evolutionary Biology*, 2013, 13: 214. doi: 10.1186/1471-2148-13-214. https://pubmed.ncbi.nlm.nih.gov/24283922/

39. Aflitos, S.; Schijlen, E., de Jong, H., de Ridder, D., Smit, S., Finkers, R., Wang, J., Zhang, G., Li, N., Mao, L., Bakker, F., Dirks, R., Breit, T., Gravendeel, B., Huits, H., Struss, D., Swanson-Wagner, R., van Leeuwen, H., van Ham, R.C., Fito, L., Guignier, L., Sevilla, M., Ellul, P., Ganko, E., Kapur, A., Reclus, E., de Geus, B., van de Geest, H., Lintel, T., Hekkert, B., van Haarst, J., Smits, L., Koops, A., Sanchez-Perez, G., van Heusden, A.W., Visser, R., Quan, Z., Min, J., Liao, L., Wang, X., Wang, G., Yue, Z., Yang, X., Xu, N., Schranz, E., Smets, E., Vos, R., Rauwerda, J., Ursem, R., Schuit, C., Kerns, M., van den Berg, J., Vriezen, W., Janssen, A., Datema, E., Jahrman, T., Moquet, F., Bonnet, J., Peters, S. Exploring genetic variation in the tomato (*Solanum* section, *Lycopersicon*) clade by whole-genome sequencing. *Plant Journal*, 2014, 80: 136–148. doi: 10.1111/tpj.12616. https://pubmed.ncbi.nlm.nih.gov/25039268/

40. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000, 408: 796–815. doi: 10.1038/35048692. https://pubmed.ncbi.nlm.nih.gov/11130711/

41. Soltis, D.E.; Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., Depamphilis, C.W., Wall, P.K, Soltis, P.S. Polyploidy and angiosperm diversification. *American Journal of Botany* 2009, 96: 336–348. doi: 10.3732/ajb.0800079. https://pubmed.ncbi.nlm.nih.gov/21628192/

42. Qin, C.; Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., Cheng, J., Zhao, S., Xu, M., Luo, Y., Yang, Y., Wu, Z., Mao, L., Wu, H., Ling-Hu, C., Zhou, H., Lin, H., González-Morales, S., Trejo-Saavedra, D.L., Tian, H., Tang, X., Zhao, M., Huang, Z., Zhou, A., Yao, X., Cui, J., Li, W., Chen, Z., Feng, Y., Niu, Y., Bi, S., Yang, X., Li, W., Cai, H., Luo, X., Montes-Hernández, S., Leyva-González, M.A., Xiong, Z., He, X., Bai, L., Tan, S., Tang, X., Liu, D., Liu, J., Zhang, S., Chen, M., Zhang, L., Zhang, L., Zhang, Y., Liao, W., Zhang, Y., Wang, M., Lv, X., Wen, B., Liu, H., Luan, H., Zhang, Y., Yang, S., Wang, X., Xu, J., Li, X., Li, S., Wang, J., Palloix, A., Bosland, P.W., Li, Y., Krogh, A., Rivera-Bustamante, R.F., Herrera-Estrella, L., Yin, Y., Yu, J., Hu, K., Zhang, Z. Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proceedings of the National Academy of Sciences USA* 2014, 111: 5135–5140. doi: 10.1073/pnas.1400975111. https://pubmed.ncbi.nlm.nih.gov/24591624/

43. Fischer, I.; Diévart, A., Droc, G., Dufayard, J.F., Chantret, N. Evolutionary dynamics of RLK/Pelle gene family in land plants. *Genome Biology and Evolution* 2016, 13: evab058. doi: 10.1093/gbe/evab058. https://agritrop.cirad.fr/579776/1/579776.pdf

44. Livingstone, K.D.; Lackney, V.K., Blauth, J.R., van Wijk, R., Jahn, M.K. Genome mapping in *Capsicum* and the evolution of genome structure in the Solanaceae. *Genetics* 1999, 152: 1183 1202. doi: 10.1093/genetics/152.3.1183. https://pmc.ncbi.nlm.nih.gov/articles/PMC1460652/

45. He, S.; Weng, D., Zhang, Y., Kong, Q., Wang, K., Jing, N., Li, F., Ge, Y., Xiong, H., Wu, L., Xie, D.Y., Feng, S., Yu, X., Wang, X., Shu, S., Mei, Z. A telomere-to-telomere reference genome for pepper provides insight into genome expansion and evolution. *Nat. Plants*, 2023, *9*, 1152–1165. https://pubmed.ncbi.nlm.nih.gov/37927407/

46. Gebhardt C.; The historical role of species from the Solanaceae plant family in genetic research. Theor Appl Genet. 2016 Dec;129(12):2281-2294. doi: 10.1007/s00122-016-2804-1. Epub 2016 Oct 15. PMID: 27744490; PMCID: PMC5121179.

47. Simko, I.; Jia, M., Venkatesh, J., Kang, B.C., Weng,Y., Barcaccia, G., Lanteri, S., Bhattara, G., Foolad, M.R. Genomics and marker-assisted improvement of vegetable crops. *Critical Reviews in Plant Sciences* 2021, 40:4, 303-365, DOI: 10.1080/07352689.2021.1941605.

48.   Wenke, T.; Seibt, K.M., Döbel, T., Muders, K., Schmidt, T. Inter-SINE Amplified Polymorphism (ISAP) for rapid and robust plant genotyping. Methods Mol Biol. 2015; 1245:183-92. doi: 10.1007/978-1-4939-1966-6_14. PMID: 25373758.

49.   Seibt, K.M.; Wenke, T., Wollrab, C., Junghans, H., Muders, K., Dehmer, K.J., Diekmann, K., Schmidt, T. Development and application of SINE-based markers for genotyping of potato varieties. Theor Appl Genet. 2012 Jun;125(1):185-96. doi: 10.1007/s00122-012-1825-7. Epub 2012 Feb 28. PMID: 22371142.

50.   Seibt, K.M., Wenke, T., Muders, K., Truberg, B., Schmidt, T. Short interspersed nuclear elements (SINEs) are abundant in Solanaceae and have a family-specific impact on gene structure and genome organization. Plant J. 2016 May;86(3):268-85. doi: 10.1111/tpj.13170. PMID: 26996788.

51.   Denna, R.; Barboza, G.E., Bohs, L., Dodsworth, S., Gagnon, E., Giacomin, L.L., Knapp, S., Orejuela, A., Poczai, P., Smith, S.D., Olmstead, R.G. A new phylogeny and phylogenetic classification for Solanaceae. bioRxiv preprint 2025, https://doi.org/10.1101/2025.07.10.663745. https://www.biorxiv.org/content/10.1101/2025.07.10.663745v1