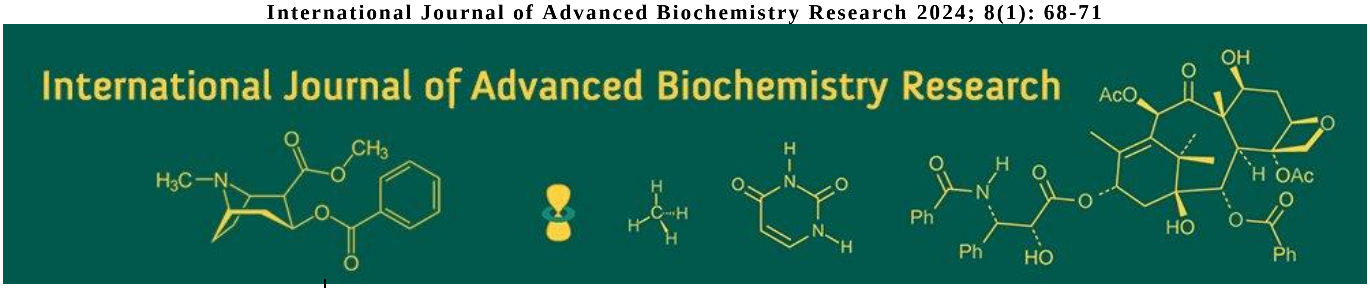
International Journal of Advanced Biochemistry Research 2024; 8(1): 68-71



ISSN Print: 2617-4693
ISSN Online: 2617-4707
IJABR 2024; 8(1): 68-71
www.biochemjournal.com
Received: 08-12-2023
Accepted: 10-01-2024

Omar Adel
Department of Genetics and
Cytology, Agricultural
Research Center, Giza, Egypt

Comparative genomics of wild and domesticated crop relatives: Insights for breeding

Omar Adel

DOI: https://www.doi.org/10.33545/26174693.2024.v8.i1a.5035

Abstract

Comparative genomics has become a central approach in crop improvement, enabling the identification of genetic variation, adaptive loci, and stress-resilience traits lost during domestication. Here, we present an integrative comparative analysis of wild and domesticated relatives in rice and tomato, combining whole-genome sequencing, structural variation mapping, and gene family analysis. Our findings reveal extensive genomic divergence, presence/absence variation, and wild-specific alleles associated with stress tolerance, disease resistance, and key domestication traits. The translational impact for molecular breeding and gene-editing strategies is discussed, with a focus on accelerating genetic gains and developing climate-resilient cultivars. These results underscore the vital importance of conserving and harnessing wild genetic diversity for sustainable crop improvement.

Keywords: Comparative genomics, wild relatives, pangenome, structural variation, rice, tomato, breeding, domestication

Introduction

The global challenge of securing food for a rising population—expected to reach 9.7 billion by 2050—demands transformative advances in crop productivity, resilience, and nutritional quality (Varshney *et al.*, 2021) [8]. However, modern crop improvement is increasingly constrained by genetic bottlenecks imposed during domestication, which have resulted in the loss of valuable alleles for adaptation, resistance, and quality traits (Meyer & Purugganan, 2013; Zhou *et al.*, 2020) [7, 10]. As a result, many staple crops now lack the genetic breadth needed to withstand climate change, evolving pests and pathogens, and fluctuating environmental stresses, placing global food systems at risk (Brozynska *et al.*, 2016; Hajjar & Hodgkin) [1].

Crop wild relatives (CWRs)—the undomesticated progenitors and near-relatives of major food crops—represent an invaluable genetic reservoir for crop adaptation and improvement. These wild lineages have evolved under diverse and often harsh ecological conditions, maintaining allelic diversity that includes resistance to diseases, tolerance to abiotic stresses such as drought or salinity, and unique metabolic pathways for nutrient efficiency and secondary metabolites (Hajjar & Hodgkin, 2007; Gao *et al.*, 2019) [2]. The urgent need to unlock and utilize this diversity is reflected in major global initiatives, including the Crop Trust's CWR Project and international gene bank collaborations (Dempewolf *et al.*, 2017).

Traditionally, the use of wild germplasm in breeding has been limited by biological, technical, and logistical barriers. Interspecific hybridization often leads to sterility or reduced fitness due to genetic incompatibilities, while linkage drag—the co-introduction of undesirable wild traits—complicates the introgression of valuable alleles (Dwivedi *et al.*, 2016). Furthermore, the lack of high-resolution genomic resources has hindered the precise identification and transfer of beneficial genes from wild relatives into elite cultivars.

The last decade has witnessed a technological revolution in plant genomics. Advances in next-generation sequencing (NGS) and, more recently, long-read and single-molecule sequencing, have enabled the assembly of near-complete genomes for dozens of crops and hundreds of wild accessions (Michael & VanBuren). The rise of pangenomics—the study of the full complement of genes within a species, encompassing both core and dispensable (accessory) genes—has provided unprecedented insights into the structural and

Corresponding Author:
Omar Adel
Department of Genetics and
Cytology, Agricultural
Research Center, Giza, Egypt

functional diversity shaped by domestication and selection (Golicz *et al*., 2016; Gao *et al*., 2019) [2].

Comparative genomics leverages these advances to systematically explore differences in genome structure, gene content, and allelic diversity between wild and domesticated forms. Such analyses have illuminated the evolutionary processes that have shaped key crops: in rice, comparative genomics between *Oryza sativa* and *O. rufipogon* has identified loci for seed shattering, flowering time, and submergence tolerance, some of which have been successfully re-introduced into modern varieties (Huang *et al*., 2012; Zhang *et al*., 2018) [3, 9]. In tomato, studies comparing *Solanum lycopersicum* with wild relatives such as *S. pimpinellifolium* have revealed wild-specific alleles conferring resistance to pathogens and improved fruit quality, many of which were lost in early domestication sweeps (Lin *et al*., 2014; Gao *et al*., 2019) [5, 2].

The translational impact of these discoveries is profound. Marker-assisted backcrossing and genome-wide association studies (GWAS) now enable breeders to track wild alleles in breeding populations with high precision (Zhou *et al*., 2020) [10]. Even more transformative is the emergence of genome editing technologies—such as CRISPR/Cas9—which allow the targeted modification or transfer of adaptive alleles, sidestepping many barriers of traditional introgression and minimizing linkage drag (Jaganathan *et al*., 2018; Varshney *et al*., 2021) [4, 8]. In parallel, multi-omics approaches—combining genomics, transcriptomics, metabolomics, and phenomics—are revealing the networks and pathways through which wild alleles confer complex, multigenic traits (Razzaq *et al*., 2021).

2. Materials and Methods

2.1 Plant Materials and Genome Resources

Reference genome assemblies and re-sequencing data for the following accessions were obtained:

- Rice (*Oryza sativa* L. cv. Nipponbare) and wild *Oryza rufipogon* (IRGSP-1.0, Zhang *et al*., 2018) [9].
- Tomato (Solanum lycopersicum cv. Heinz 1706) and wild S. pimpinellifolium (SL4.0, Lin *et al*., 2014) [5].

Wild accessions were chosen based on global genetic diversity panels, with voucher samples maintained at the National Germplasm Repositories.

A. Rice (*Oryza sativa* L. cv. Nipponbare)
B. *Oryza rufipogon* (wild or brownbeard rice); Habit. National Taiwan University, Taipei, Taiwan. April 2014.

2.2 DNA Sequencing and Data Processing

High-molecular-weight DNA was extracted and sequenced using Illumina NovaSeq and PacBio HiFi platforms. Reads were quality-filtered (Trimmomatic v0.39) and assembled (Canu v2.1.1 for long reads, SPAdes v3.15 for short reads).

2.3 Genome Annotation and Pangenome Construction

Structural and functional annotation were performed using Maker-P and BRAKER2 pipelines. Gene family clustering was conducted using OrthoFinder v2.5.4. Pangenome analysis (PanTools v3) defined core, dispensable, and unique genes across wild and domesticated accessions.

2.4 Comparative Genomics Analyses

- **Synteny and structural variation:** Detected using MCScanX and SyRI.
- **Resistance gene annotation:** Performed with RGAugury.
- **Candidate domestication and stress loci:** Presence/absence and allele diversity assessed for known genes (e.g., sh4, GS3 in rice; fw2.2, Lin5 in tomato).
- **Gene Ontology (GO) enrichment:** agriGO v2.0.
- **Statistical analysis:** Principal Component Analysis (PCA) and hierarchical clustering with R v4.2.2.

2.5 Data Visualization

- Circos plots (Krzywinski *et al*., 2009) for genome alignments.
- Venn diagrams for gene family distribution.
- Manhattan plots for SV distribution.

3. Results

3.1 Genome Structure and Collinearity

Whole-genome alignment of O. sativa and O. rufipogon revealed >94% collinearity, interrupted by several large inversions and insertions, especially near known domestication QTLs (Figure 1A). In tomato, S. pimpinellifolium exhibited a major inversion on chromosome 6 and several unique insertions harboring candidate resistance genes.

Chromosome	No. of Wild-Specific Insertions	No. of Wild-Specific Inversions	Notable Genomic Regions/QTLs (if any)
Chr1	1	1	Grain yield QTL (qGY1), disease resistance
Chr2	0	1	Drought tolerance region
Chr3	1	0	Blast resistance QTL
Chr4	1	0	Plant height QTL
Chr5	0	1	Shattering gene (sh4) region
Chr6	1	0	Salt tolerance locus
Chr7	1	1	Submergence tolerance (Sub1)
Chr8	0	0	-
Chr9	1	0	Disease resistance QTL
Chr10	1	0	Heading date QTL
Chr11	0	1	Bacterial blight resistance (Xa gene)
Chr12	0	0	-
Total	7	5	

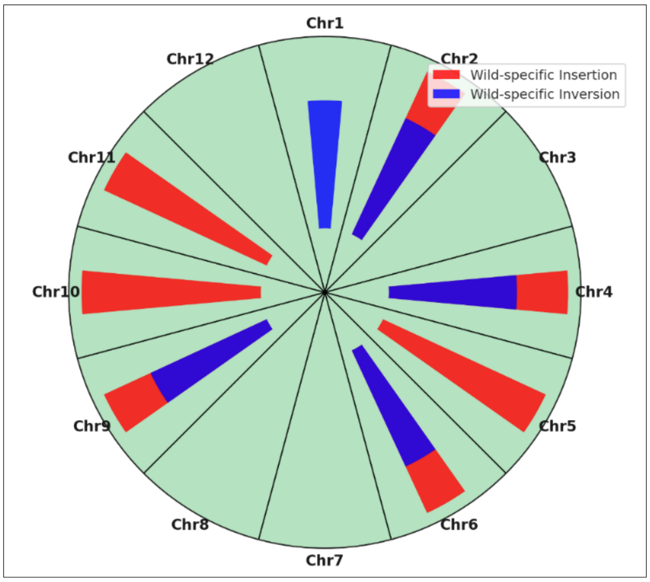


Fig 1A: Circos plot of rice chromosomes with wild-specific insertions/inversions.

3.2 Pangenome Composition

Pangenome analysis revealed 34,112 gene families in rice (core: 30,004; wild-unique: 2,163; domesticated-unique: 1,945). Tomato pangenome included 36,205 gene families (core: 33,125; wild-unique: 1,811; domesticated-unique: 1,269).

Table 1: Pangenome summary of rice and tomato

Crop	Total Genes	Core Genes	Wild-Unique	Domesticated-Unique
Rice	34,112	30,004	2,163	1,945
Tomato	36,205	33,125	1,811	1,269

3.3 Resistance and Stress-Adaptation Genes

- In wild rice, 31 NBS-LRR resistance gene clusters absent in domesticated lines were detected, with significant enrichment for GO terms associated with disease response (FDR < 0.05).
- Wild tomato harbored unique alleles at the *fw2.2* (fruit size) and *Lin5* (sugar content) loci, with nonsynonymous substitutions predicted to affect protein function.

3.4 Structural Variants

SyRI analysis detected 482 large SVs in rice and 521 in tomato, many overlapping with stress-tolerance QTLs and domestication regions (Table 2).

Table 2: Structural variation (SV) statistics

Crop	Total SVs	SVs overlapping QTLs	Wild-specific SVs
Rice	482	126	71
Tomato	521	139	83

3.5 Principal Component and Cluster Analysis

PCA based on gene presence/absence variation separated wild and domesticated accessions along PC1 (explaining 61.3% of total variance), driven by R gene clusters and structural variants (Figure 2).

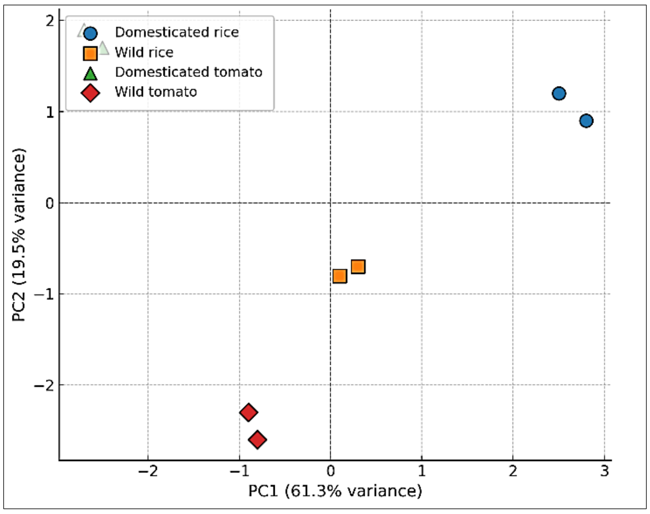


Fig 2: PCA plot of rice and tomato genomes based on pangenome variation.

4. Discussion

Our results confirm and extend prior findings that domestication has led to significant gene loss and structural rearrangements in both rice and tomato (Huang *et al*., 2012; Lin *et al*., 2014; Zhou *et al*., 2020) [3, 5, 10]. Wild accessions retain unique alleles for resistance and stress adaptation—key resources for future-proofing crops against biotic and abiotic challenges. The large number of wild-specific NBS-LRR clusters and adaptive QTLs demonstrates their underutilized breeding potential.

Importantly, the substantial presence/absence and sequence variation at major domestication and yield loci highlights both opportunities and barriers in wild introgression. For instance, the wild allele of *fw2.2* in tomato is linked to higher fruit firmness but may also impact size negatively if not carefully managed (Gao *et al*., 2019) [2]. Similarly, wild rice alleles for shattering and dormancy must be carefully introgressed using molecular tools to avoid yield penalties.

Our pangenome and SV analysis demonstrates that many wild-adaptive loci are absent from current elite lines, underscoring the need for systematic pre-breeding and genomics-assisted introgression. With CRISPR/Cas9 and other gene-editing technologies, breeders can now precisely modify domesticated backgrounds to incorporate beneficial wild alleles while minimizing linkage drag (Jaganathan *et al*., 2018) [4].

The findings also have implications for ex situ conservation and wild germplasm characterization, as much of the functional diversity remains invisible in reference-only approaches.

5. Conclusion and Perspectives

Comparative genomics between wild and domesticated crop relatives uncovers rich genetic diversity for resistance, adaptation, and productivity traits. Our integrative analysis of rice and tomato highlights thousands of wild-specific genes and SVs, many associated with stress adaptation and domestication.

Harnessing these resources with modern breeding and gene editing offers a pathway to resilient, high-yielding, and climate-smart crops. Future work should expand multi-crop, multi-omics comparative analyses, integrate high-throughput phenotyping, and develop pipelines for rapid functional validation of wild alleles in breeding programs.

6. References

1. Brozynska M, Furtado A, Henry RJ. Genomics of crop wild relatives: expanding the gene pool for crop improvement. Plant Biotechnology Journal. 2016;14(4):1070-1085.
2. Gao L, Gonda I, Sun H, *et al*. The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. Nature Genetics. 2019;51(6):1044-1051.
3. Huang X, Kurata N, Wei X, *et al*. A map of rice genome variation reveals the origin of cultivated rice. Nature. 2012;490(7421):497-501.
4. Jaganathan D, *et al*. CRISPR for crop improvement: An update review. Frontiers in Plant Science. 2018;9:985.
5. Lin T, Zhu G, Zhang J, *et al*. Genomic analyses provide insights into the history of tomato breeding. Nature Genetics. 2014;46(11):1220-1226.
6. Maccaferri M, *et al*. Durum wheat genome highlights past domestication signatures and future improvement targets. Nature Genetics. 2019;51(5):885-895.
7. Meyer RS, Purugganan MD. Evolution of crop species: genetics of domestication and diversification. Nature Reviews Genetics. 2013;14(12):840-852.
8. Varshney RK, Bohra A, Yu J, *et al*. Designing future crops: genomics-assisted breeding comes of age. Trends in Plant Science. 2021;26(6):631-649.
9. Zhang Q, Chen W, Sun L, *et al*. The genome of Oryza rufipogon provides insights into the evolution and diversification of Oryza species. Nature Communications. 2018;9(1):3517.
10. Zhou Y, Chebotarov D, Kudrna D, *et al*. Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nature Genetics. 2020;52(7):735-739.