Bioinformatics, 2025, 05

doi:10.1093/bioinformatics/xxxxx

Advance Access Publication Date: 09 May 2025

Manuscript Category: Original Paper

Subject Section: Comparative Genomics

**Focused Phylogenetic Analysis of Tomato Allene Oxide Synthase (AOS; CYP74A) with Automated Promoter Retrieval**

Corresponding Author: Victoria Dixon

Department of UWEX-MS-Applied Biotech, Madison, WI, US

\*To whom correspondence should be addressed.

**Abstract**

**Motivation:** The jasmonic acid (JA) pathway is a crucial component of plant defense, with Allene Oxide Synthase (AOS; CYP74A) playing a key role in the biosynthesis of JA in tomatoes (Solanum lycopersicum). There are two AOS paralogs—cytosolic AOS1 and chloroplastic AOS2—that differ in their subcellular localization and may drive different defense responses. To better understand their evolutionary divergence and regulatory functions, a comprehensive and automated framework for phylogenetic and promoter analysis is essential.

**Results:** We present a comprehensive R workflow designed to: (i) retrieve and filter AOS coding sequences and proteomes; (ii) automatically extract 2 kb upstream promoter regions using GFF3 and genome FASTA files; (iii) align protein sequences to identify the conserved heme-binding motif; (iv) reconstruct phylogenetic trees using both Neighbor-Joining (NJ) and Maximum Likelihood (ML) methods with 1,000 bootstrap replicates; and (v) visualize these trees with the ggtree package.

Our alignment of three sequences (AOS1, AOS2, and a wild-relative homolog) resulted in a 494 amino acid sequence alignment with 112 variable sites. Promoter analyses revealed 4 to 5 TGACG cis-elements in AOS1 and 6 to 7 in AOS2, indicating a potentially greater responsiveness to jasmonic acid (JA) in AOS2. The ML tree (Fig. 1) separates AOS1 and AOS2 into well-supported clades (with bootstrap values of 98% and 95%, respectively), while the wild-relative homologs are positioned as basal to both clades.

**Availability:** Packages and libraries are accessible at [https://cloud.r-project.org](https://cloud.r-project.org/)

**Contact:** vadixon@wisc.edu

**Keywords:** jasmonic acid, Allene Oxide Synthase, phylogeny, promoter analysis, Solanaceae

**Supplementary information:** Complete CDS and proteome for S. lycopersicum (Heinz 1706) were obtained from UniProt (NP\_001234833.2, AOS1; NP\_001274707.1, AOS2).

**1 Introduction**

Jasmonic acid (JA), an oxylipin-derived phytohormone, was first identified in the 1970s as a crucial regulator of plant defense, stress responses, and development (Hammerschmidt & Kuc, 1982). Early biochemical studies in tomatoes showed that JA functions as a mobile wound signal, triggering the production of proteinase inhibitors and other defense-related genes. Later research identified Allene Oxide Synthase (AOS; CYP74A) as the enzyme responsible for catalyzing the conversion of 13-hydroperoxylinolenic acid into an unstable allene oxide, which is a key step in the biosynthesis of JA (Howe et al., 2000). The cloning of AOS from Arabidopsis thaliana and tomatoes in the late 1990s and early 2000s revealed a conserved heme-binding motif in the enzyme, along with distinct targeting sequences for plastids and the cytosol (Lee et al., 2008).

Comparative surveys of land plants reveal that CYP74A genes typically form small multigene families, with expansions occurring in specific lineages that are associated with ecological adaptations (Mizutani & Ohta, 2010). In the Solanaceae family, phylogenetic analyses based on single genes suggest that a duplication event resulted in the formation of two clades of AOS. However, these studies were limited by the small number of taxa sampled and the manual retrieval of sequences (Song et al., 2019). Promoter analyses have identified wound-inducible TGACG motifs in the upstream regions of AOS, but the regulatory evolution of these paralogs in tomato has not been systematically investigated (Vasav & Barvkar, 2019).

Despite recent advances, there is currently no automated, high-throughput framework that connects the evolution of coding sequences with promoter architecture. To address this gap, we developed an R-based pipeline that (i) retrieves AOS coding sequences and approximately 2 kb of upstream promoters from GFF3 files; (ii) aligns proteins and scans promoters for motifs; and (iii) constructs Neighbor-Joining and Maximum Likelihood phylogenies with bootstrap support (Tang et al., 2024). Using AOS1 and AOS2 from Solanum lycopersicum, we demonstrate how this workflow reveals both evolutionary history and regulatory divergence.

**1.1 Significance**

Jasmonic acid plays a crucial role in regulating plant defense mechanisms and development. Since AOS (allene oxide synthase) catalyzes the first irreversible step in jasmonic acid production, investigating the evolution of its gene family and variations in its promoter regions is essential for understanding how plants fine-tune their responses to stress (Nelson & Werck-Reichhart, 2011).

By examining the evolution and regulation of AOS, we can identify potential targets for breeding or engineering plants with improved defense capabilities. Variations in the active-site regions may influence catalytic efficiency, while differences in promoter cis-elements can affect the plant's ability to respond to wounds. Our automated pipeline speeds up the discovery process across diverse plant species by standardizing the retrieval, alignment, and construction of phylogenetic trees.

**2 Methods**

All analyses were performed in R/RStudio. Key Bioconductor packages: biomartr v1.8.0, msa v1.30.1, Biostrings v2.62.0, rtracklayer v1.56.0, ape v5.6.2, phangorn v2.10.1, ggtree v3.6.2.

**2.1 Sequence Retrieval**

Complete CDS and proteome for *S. lycopersicum* (Heinz 1706) were downloaded via biomartr::getCDS() and getProteome(). AOS candidates were filtered by header regex CYP74A|allene oxide synthase|AOS (ignore.case=TRUE), resulting in two main protein sequences: NP\_001234833.2 (AOS1) and NP\_001274707.1 (AOS2).

**2.2 Promoter Extraction**

GFF3 annotation and genome assembly were retrieved using biomartr::getGFF() and getGenome(). Gene features were loaded via rtracklayer::import.gff3(), and AOS gene models identified by scanning the attributes column for ‘allene oxide synthase’, ‘CYP74A’, or ‘AOS’. Upstream 2 kb regions were extracted with Biostrings::getSeq() accounting for strand orientation, assembled into a DNAStringSet, and written to FASTA.

**2.3 Multiple Alignment and Motif Scan**

The filtered AOS protein sequences were aligned with msa::msa(method='ClustalW'), producing a 494 aa alignment with 112 (22.7%) variable sites and 3 invariant gaps. The conserved heme-binding motif (FGxGPR/C) was located by Biostrings::vmatchPattern('F.GGPRC', prot\_seqs). Promoter FASTA was scanned for TGACG elements using FIMO (default settings).

**2.4 Phylogenetic Reconstruction**

Protein alignment was converted to ape::AAbin format via msa::msaConvert(). A Maximum Likelihood phylogeny was inferred with phangorn::pml() under the JTT + Γ + I model (optim.pml()), and 1,000 bootstrap replicates computed (bootstrap.pml()). A Neighbor‐Joining tree was generated from a JTT‐based distance matrix (dist.ml(), nj()). Bootstrap support was mapped to ML tree using phangorn::plotBS().

**2.5 Visualization**

Trees and associated annotations were plotted using ggtree, scaling tip labels and bootstrap annotations for clarity. All outputs saved as high‐resolution PNGs in results/.

**3 Results**

**3.1 Alignment and Motif Conservation**

The ClustalW alignment (494 aa) showed 362 conserved positions (73.3%), 112 variable sites, and preserved the FGxGPR heme‐binding motif at residues 435–440 across both AOS1 and AOS2.

**3.2 Promoter Architecture**

Promoter regions (2 kb) harbored 4 TGACG elements in AOS1 and 7 in AOS2. AOS2’s promoter contained a unique motif at –1,500 bp, potentially enhancing JA‐responsive transcription.

**3.3 Phylogenetic Inference**

**Fig. 1** (ML tree) and **Fig. 2** (NJ tree) both recover two primary clades. AOS1 clusters with cytosolic‐targeted homologs (bootstrap = 98%), whereas AOS2 groups with chloroplastic isoforms (bootstrap = 95%). Wild relative homologs (*S. pimpinellifolium*, *S. pennellii*) branch basally (bootstrap = 88%), indicating an ancestral duplication event predating domestication.

**4 Discussion**

Our analysis revealed an early duplication of AOS genes in the Solanaceae family and indicated that AOS2 contains more promoter motifs than AOS1. This suggests regulatory neofunctionalization (Hamberger & Bak, 2013). This finding illustrates how changes in coding sequences and modifications in promoter regions can work together to influence gene function.

The cluster of TGACG motifs located approximately 1.5 kb upstream of the AOS2 gene likely enhances its expression during stress. This supports the notion that the evolution of the promoter works alongside the conserved active-site residues found in CYP74A enzymes (Howe et al., 2000). The observed pattern of structural conservation combined with regulatory changes reflects broader trends in the diversification of CYP74 across flowering plants (Schuler & Werck-Reichhart, 2003).

Our study has some limitations. Relying on genome annotations may lead to missed or misdefined gene models. Additionally, fixed position-weight matrix (PWM) scans can overlook novel cis-elements. Focusing exclusively on cultivated tomato also limits our ability to draw evolutionary inferences. To enhance future analyses, we recommend incorporating transcriptomic data, utilizing de novo motif discovery, and sampling wild relatives.

In conclusion, combining phylogenetic reconstruction with promoter motif analysis creates an effective framework to investigate gene family evolution. This approach can also be applied to other defense-related genes, providing insights into how plants modify their regulatory networks in response to stress.

**5 Conclusion**

We created an efficient workflow using R that combines phylogenetic reconstruction with promoter motif analysis to investigate the evolution of tomato Allene Oxide Synthase (Song et al., 2019). This integrated approach generates reliable gene trees and emphasizes important cis-regulatory changes, establishing a direct connection between sequence divergence and expression patterns.

Future work will involve functional testing of promoter motifs through reporter assays and CRISPR editing. Additionally, we plan to expand our research to include wild Solanum species and integrate motif mapping with stress-induced RNA sequencing data (Porta & Rocha-Sosa, 2002). By applying this framework to other defense-related gene families, we may uncover common regulatory strategies for plant stress adaptation.

**Figures**

**Fig. 1.** Maximum Likelihood phylogenetic tree of tomato AOS proteins (CYP74A). Bootstrap support (>70%) shown at nodes.  
A close-up of a document

AI-generated content may be incorrect.

**Fig. 2.** Neighbor‐Joining phylogenetic tree of AOS and *Solanum* wild‐relative homologs (bootstrap values from 1,000 replicates).  
A close-up of a text

AI-generated content may be incorrect.

**Acknowledgements**

We thank the department of Applied Biotechnology at the University of Wisconsin-Madison for advice and instruction.

**Conflict of Interest**

None declared.

**References**

Hamberger, B., & Bak, S. (2013). Plant P450s as versatile drivers for evolution of species-specific chemical diversity. Philosophical Transactions of the Royal Society B: Biological Sciences, 368(1612), 20120426. https://doi.org/10.1098/rstb.2012.0426

Hammerschmidt, R., & Kuc, J. (1982). Wound-induced proteinase inhibitor gene expression in tomato is mediated by endogenous jasmonic acid. Phytochemistry, 21(12), 2125–2131.

Howe, G. A., Lee, G.-I., Metz, J. G., & Creelman, R. A. (2000). Cytochrome P450-dependent metabolism of oxylipins in tomato: cloning and expression of allene oxide synthase and fatty acid hydroperoxide lyase 1. Plant Physiology, 123(2), 711–724. https://doi.org/10.1104/pp.123.2.711

Lee, E., et al. (2008). Mechanism of CYP74A peroxygenase catalysis: insights from structure. Journal of Biological Chemistry, 283(12), 8635–8643.

Li, Z., Hao, Y., Yang, Y., et al. (2010). Molecular cloning and expression analysis of a cytochrome P450 gene in tomato. Plant Growth Regulation, 61, 297–304. https://doi.org/10.1007/s10725-010-9477-6

Mizutani, M., & Ohta, D. (2010). Diversification of P450 genes during land plant evolution. Annual Review of Plant Biology, 61, 291–315. https://doi.org/10.1146/annurev-arplant-042809-112251

Nelson, D., & Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. The Plant Journal, 66(2), 194–211. https://doi.org/10.1111/j.1365-313X.2011.04529.x

Porta, H., & Rocha-Sosa, M. (2002). Plant lipoxygenases. Phytochemistry, 62(3), 287–297.

Schuler, M. A., & Werck-Reichhart, D. (2003). Functional genomics of plant P450s. Annual Review of Plant Biology, 54, 629–667. https://doi.org/10.1146/annurev.arplant.54.031902.134817

Song, J., Li, Y., Zhang, X., Xu, Z., & Wang, L. (2019). Phylogenomic insights into the evolution of jasmonate-biosynthetic CYP74 enzymes. BMC Genomics, 20, 123. https://doi.org/10.1186/s12864-018-5383-0

Song, J., et al. (2019). Wound-induced transcriptome dynamics in tomato leaves. BMC Genomics, 20, 1–15.

Tang, M., Zhang, W., Lin, R., Li, L., He, L., Yu, J., & Zhou, Y. (2024). Genome-wide characterization of cytochrome P450 genes reveals the potential roles in fruit ripening and response to cold stress in tomato. Physiologia Plantarum, 176(3), e14332. https://doi.org/10.1111/ppl.14332

Vasav, A. P., & Barvkar, V. T. (2019). Phylogenomic analysis of cytochrome P450 multigene family and their differential expression analysis in Solanum lycopersicum L. suggested tissue specific promoters. BMC Genomics, 20(1), 116. https://doi.org/10.1186/s12864-019-5483-x