Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Plant Biology

How to utilize comparative transcriptomics to dissect morphological diversity in plants

Siyu Li[1], Hokuto Nakayama[1,2] and Neelima R. Sinha[1]

Abstract

Comparative transcriptomics has emerged as a powerful approach that allows us to unravel the genetic basis of organ morphogenesis and its diversification processes during evolution. However, the application of comparative transcriptomics in studying plant morphological diversity addresses challenges such as identifying homologous gene pairs, selecting appropriate developmental stages for comparison, and extracting biologically meaningful networks. Methods such as phylostratigraphy, clustering, and gene co-expression networks are explored to identify functionally equivalent genes, align developmental stages, and uncover gene regulatory relationships. In the current review, we highlight the importance of these approaches in overcoming the complexity of plant genomes, the impact of heterochrony on stage alignment, and the integration of gene networks with additional data for a comprehensive understanding of morphological evolution.

Addresses

[1] Department of Plant Biology, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

[2] Graduate School of Science, Department of Biological Sciences, The University of Tokyo, Science Build. #2, 7-3-1 Hongo Bunkyo-ku Tokyo, 113-0033, Japan

Corresponding author: Sinha, Neelima R. (nrsinha@ucdavis.edu)

Current Opinion in Plant Biology 2023, **76:**102474

Edited by **Zachary L Nimchuk** and **Ikram Blilou**

This review comes from a themed issue on **Growth and development 2023**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.pbi.2023.102474

1369-5266/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords

Comparative transcriptomics, Evo-devo, Homology, Network analysis.

BOX.

Homologous genes: Genes that are similar in sequence and structure because they share a common ancestor.

Orthologous genes: Homologous genes found in different species that evolved from a single gene in the last common ancestor of those species.

Paralogous genes: Homologous genes found in the same species that have evolved through gene duplication events.

Developmental toolkit genes A set of genes that are shared among different organisms and play critical roles in developmental processes. Such as *Hox* genes in animals and *WOX* genes in plants.

Gene co-expression networks: Networks of genes that are co-expressed across different samples or conditions, providing a more holistic view of gene expression patterns without directionality.

Gene regulatory networks: Networks of genes that are co-expressed and linked by regulatory interactions, providing a more detailed understanding of the mechanisms underlying gene expression with directionality.

Phylostratigraphy: A method for dating the origin of genes based on the age of the earliest known homolog in other species.

Homologous organs: anatomical structures in different species that have evolved from a common ancestral structure but may or may not serve different functions.

Hub gene: A gene that is highly connected to other genes in a network, and is thought to be important for the regulation of the network as a whole.

Introduction

A major challenge in biology is to understand how organs are formed in ontogeny and how they diversified during evolution [1]. In plants, the molecular mechanisms of organ formation are well-investigated in several model species [2]. Based on this knowledge, researchers have tried to identify conserved modules during many developmental processes and examine how these conserved modules may have generated morphological diversity over time. Evolutionary developmental biology (Evo-devo) has tackled this open question [3,4] by comparing molecular developmental mechanisms among model species and non-model species [5,6] or distantly related species [7,8].

Comparative transcriptomics allows us to capturing changes in gene expression along a developmental trajectory and it has contributed to our understanding of the molecular and genetic mechanisms underlying diversification processes. Large-scale comparisons of gene profiles among phyla have been made in metazoans [9,10]. Transcriptomics data from roundworms (*Caenorhabditis elegans*), fruit flies (*Drosophila melanogaster*), and humans (*Homo sapiens*) reveal a common phylotypic stage during embryogenesis where all conserved modules show the lowest expression divergence [9,10]. A comparative analysis of this type has the potential to explain how homologous organs are formed and diversified at the molecular level , unraveling conserved gene expression patterns and identifying key factors driving organ development. However, in plants, such research has not yet flourished, and many comparable studies have been conducted only within the same genus or the same family [11–14]. Unlike metazoans, plant genomes are often more complex, having been subjected to polyploidy and whole genome duplication events [15]. This has made identification of homologous and functionally equivalent genes a significant issue. Also, many plant morphological traits have arisen through multiple events of convergent evolution [16,17], making it challenging to select the "right" organ, tissue, and stage to compare across distantly related species.

In this review, we discuss three issues 1) identification of orthologous 2) stage alignment during development, and 3) extraction of biologically meaningful networks when comparing transcriptomes across species by providing examples and possible alternatives.

What to compare? Ortholog conjecture and functional homology

Comparative analysis of gene expression across species begins by identifying orthologous (Figure 1). Orthologous are assumed to retain their original function more often than non-orthologous. This is known as the ortholog conjecture [18]. Correct identification of orthologous is essential for comparing functionally equivalent across different species, particularly in plants with limited genome annotation and functional studies. Identifying orthologs allows annotation transfer from a well-studied model species such as Arabidopsis (*Arabidopsis thaliana*) to uncover the putative function of previously uncharacterized genes in non-model species, leading to a more comprehensive understanding of gene function in plants [18]. The degree of relatedness between species impacts the sequence similarity and functional conservation of genes, necessitating different methods for orthologous gene identification to account for these variations. For closely related species, methods such as BLAST [19], sequence-based clustering [20], synteny [21,22], and integrative approaches [23] can be used to identify orthologous. These straightforward methods are based on sequence similarity and genome structure, offering advantages in term of computational efficiency and ease of implementation. However, for more distantly related species, phylogenetic-tree-based methods such as OrthoFinder [24] can be more effective in identifying orthologous. Nevertheless, it's important to note that OrthoFinder can be computationally intensive, may require parameter tuning, and may not distinguish between different types of duplicated genes effectively, which can necessitate additional manual curation.

While the orthology approach is a powerful tool for comparative transcriptomics, there are also limitations to its use. Firstly, it is often challenging to identify gene orthologs due to computational limitations and genome complexity [25]. Many developmental toolkit genes involved in morphological changes, such as *CUC* genes, belong to large transcription factor families. Their family members vary in the number of paralogs, sequences and function even between relatively closely related species [26]. Secondly, the orthology conjecture is not always applicable. Orthologous are not always functionally equivalent, particularly in cases of gene duplication followed by sub-functionalization or neo-functionalization (for instance, α and (α_1 and α_2) in Figure 1A) [27]. A functional prediction analysis of two pairs of mammalian and fungal species suggested that in some cases paralogs might lead to better functional prediction than orthologs alone [28]. The researchers utilized gene family trees to infer duplication events and specify different homologous relationships: orthologs, in-paralogs and out-paralogs (Figure 1). Notably, they observed that functional similarity based on gene ontology annotations had the best performance when all homologs were used. The removal of orthologs or paralogs, especially the removal of in-paralogs resulted in a significant decrease in prediction performance across all ontologies Therefore, they concluded that maximizing the amount of data used for function prediction, regardless of whether it comes from orthologs or paralogs, is crucial for achieving higher prediction

Figure 1

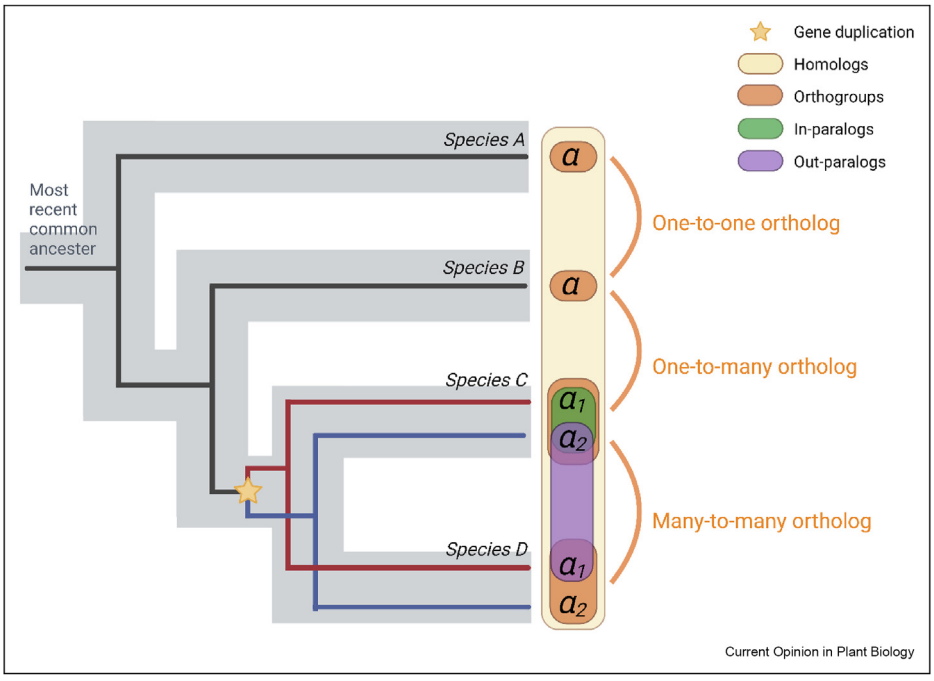


Illustration of phylogenetic relationships and gene duplications through hypothetical gene trees. The gray tree represents the phylogenetic relationships among four hypothetical taxa. The embedded trees illustrate the evolution of a hypothetical protein α, highlighting the concepts of homologs, orthologs, in-paralogs, and out-paralogs. The yellow star indicates a gene duplication event. Yellow boxes represent homologous originating from a common ancestor. Orange boxes represent orthogroups resulting from speciation events. The orange lines on the right denotes different types of orthologous relationships. Green boxes represent in-paralogouss arising from duplication events within the same species. Purple boxess represent out-paralogous resulting from duplication events in different lineages.

accuracy. Although no similar studies have been conducted in plants yet, the abundance of whole genome duplications and local duplications in plants suggests that this may also be true in plants.

To address the limitations of orthology-based approaches in comparative transcriptomics, two commonly employed alternatives exist. The first alternative is to use orthogroups, which refer to groups of genes that include both orthologous and paralogous derived from a single ancestral gene, and compare them as a whole. Many algorithms such as OrthoFinder generate orthogroups [24], although often only one set of orthologs is selected for further comparison to decrease the computational burden. This approach can be implemented using clustering methods and network construction methods [20], which allow for the inclusion of both orthologs and paralogs at the outset. By considering orthogroups as groups of genes sharing common functions at the beginning step, we can compare their expression patterns across species, even in the absence of direct orthologous pairs [24,29,30]. After cross-species comparison, we can examine transcripts within the orthogroups of interest in a particular species, considering their sequence conservation, gene expression patterns, and functional annotations to determine their detailed orthologous relationship. This allows us to identify potential sub- and neo-functionalization events within the orthogroups and gain insights into the evolutionary dynamics of gene function. This approach is particularly valuable for studying evolutionary phenomena like sub-functionalization and neo-functionalization. Expression patterns often differ between paralogs that have undergone these processes, allowing us to discern their functional distinctions. Thus, this approach not only addresses the challenge of functional divergence but also provides insights into functional innovation within plant gene families. Another approach is to incorporate other types of information, such as phylostratigraphy and single-cell RNA sequencing data to identify functionally equivalent genes that may not be orthologous but have similar functions during specific developmental processes [31]. A comparative study of the transcriptomes of ten phylogenetically representative land plant species showed that most organ transcriptomes are conserved across land plants and reported the identity of hundreds of organ-specific orthogroups [32]. Such information can be used to identify homologous that are functionally equivalent in a particular organ or developmental process. Functionally equivalent homologs

are also likely to have similar expression patterns or co-expression relationships [33,34]. However, expression levels measured in one organ or tissue are an average of expression levels of the cell types constituting this tissue. Recent reports from single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics can provide even more detailed insights into the molecular programs of individual cells and tissues and can be helpful to identify genes with similar expression patterns in specific cell types [35]. Indeed, in a comparative analysis of three grass species — *Zea mays*, *Sorghum bicolor* and *Setaria virdis*, the researchers successfully identified orthologs that can serve as cell-type specific markers with scRNA-seq and single-nucleus RNA-seq data, providing valuable insights into the conservation and evolutionary relationships of genes across these species [36].

When to compare? Heterochrony and stage alignment

Unlike animals, where organogenesis is mostly completed during embryogenesis, plants often undergo post-embryonic and iterative development, allowing for a prolonged window to study organ formation and the generation of morphological diversity [37]. This extended developmental process provides a valuable opportunity to focus on the morphogenesis of individual organs and understand how morphological diversity emerges over time. By examining the dynamic changes during post-embryonic development, we can gain insights into the regulatory mechanisms underlying organ formation and the evolutionary processes that shape plant morphology (Figure 2A). However, studying plant development presents a challenge due to heterochrony - the variability in the timing and duration of developmental processes among different species [3]. Patterns of heterochrony can be discerned within transcriptomic data. Expression levels of certain genes obtained by sampling across time points can provide insights into the temporal state of organs [38,39]. To minimize the effects of heterochrony and increase comparability, one can utilize the expression levels of a set of "feature genes" to divide a developmental trajectory into distinct developmental periods for comparison [36,40,41] or identify stages solely based on their stage-specific expression patterns [42] (Figure 2B). Feature genes are often the most variable orthogroups during a specific morphogenesis process. The selection of feature genes depends on the nature of datasets, the evolutionary relationships between the species under study, and the existing knowledge of the relevant developmental processes [3,38,41]. Theoretically, when comparing more distantly related species, the selected feature genes may exhibit a higher degree of conservation. Aligning stages with feature genes enables a focused investigation of specific developmental genes instead of analyzing all differentially expressed genes.

In a study comparing inflorescence development in *Z. mays* and *Sorghum bicolor*, two closely-related species with distinct morphology, the authors applied the random forest method to analyze the 5,000 most variable transcripts for each species [38]. They selected the top 3,000 informative, stage-specific genes as feature genes for sample alignment. By categorizing 40 maize tassels and 47 sorghum panicles into their respective developmental stages, the researchers identified expression shifts in key regulators that contributed to the morphological differences between the species. Recent research in phylostratigraphy and comparative functional genomics has revealed tissue-specific and lineage-specific genes [32,43]. These genes could serve as suitable feature gene candidates when comparing distantly-related species.

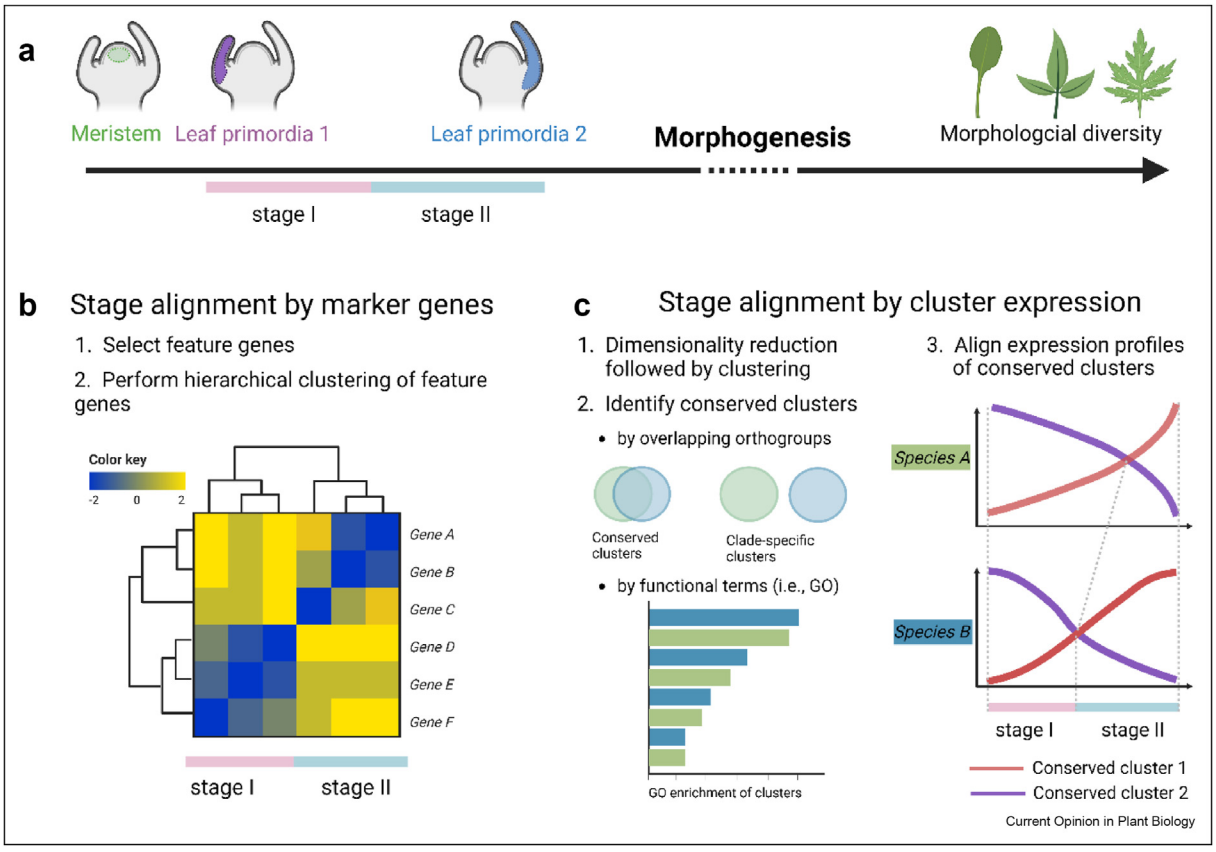
To achieve accurate stage alignment in comparative transcriptomics, it is crucial to prioritize sufficient sampling that effectively captures the complex morphogenesis patterns specific to each species. This necessitates the collection of a comprehensive set of samples, encompassing diverse developmental time points and tissue types. Such an approach ensures a thorough representation of the dynamic morphological changes occurring throughout development, enabling meaningful statistical comparisons and enhancing the reliability of transcriptomic analyses in a comparative context.

How to compare? Genes and connection

The analysis of differential gene expression (DGE) patterns is routinely the first step in comparative transcriptomics examining the molecular basis of morphological diversity [44]. However, DGE analysis alone does not provide a comprehensive view of how genes and gene networks contribute to morphological evolution [44]. Given the complexity of morphogenesis and the involvement of numerous genes, it is crucial to employ suitable algorithms for visualizing gene profiles and integrate additional data (like organ-specific gene sets and phylostratigraphic analysis) to identify key genes and facilitate meaningful comparisons. The examples provided below may be limited in their direct relevance to the topic of plant morphological diversity. However, they serve as valuable references from which we can draw inspiration and learn methodological approaches that can be adapted and applied to study plant morphological diversity.

Gene co-expression networks are networks composed of nodes (genes) and edges (connections) that indicate the degree of co-expression between genes [45,46], and the choice of input data and biological context is essential for generating biologically meaningful networks [44]. Different GCNs can be created by using different data matrices and selecting various stages, which can offer insights into putative gene regulation and functional

Figure 2

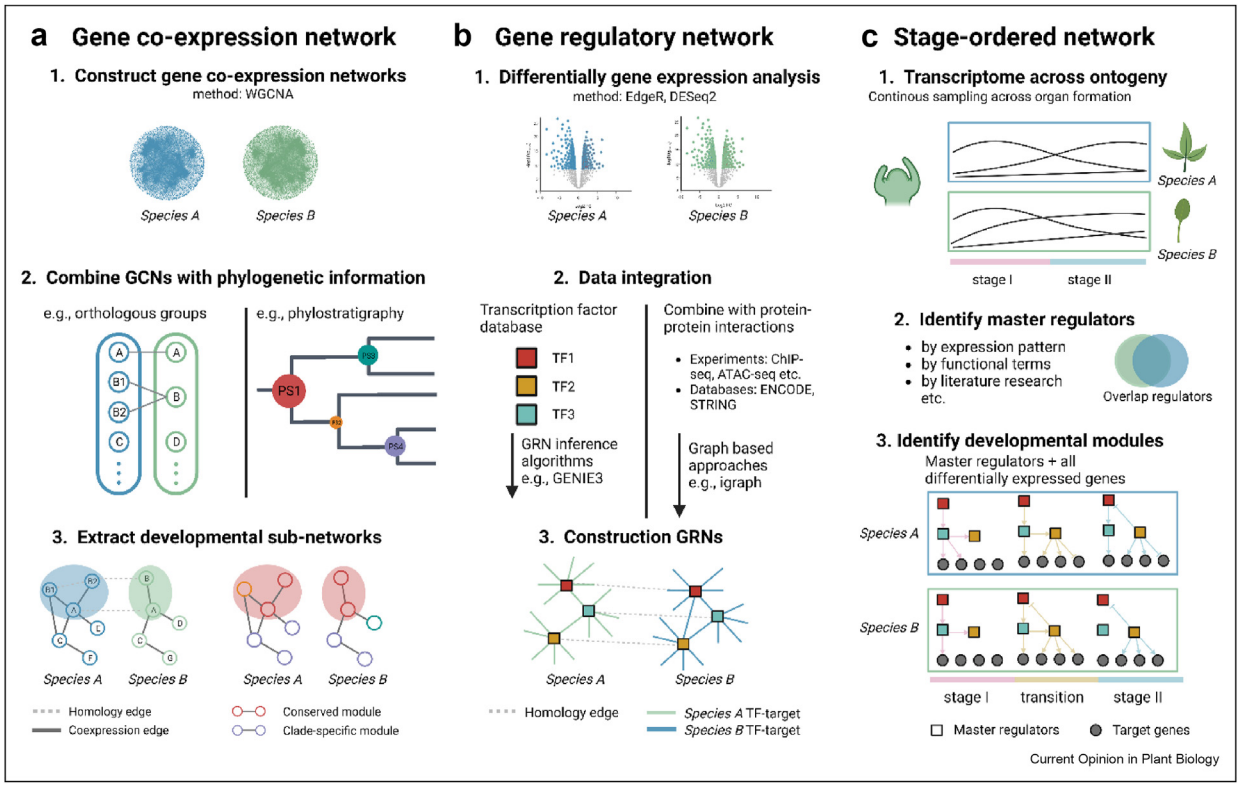


Two stage-alignment approaches of cross-species comparative transcriptomic analysis. (a) Developmental process illustrates the progression from the meristem to mature organ stages. **(b)** Stage alignment using hierarchical clustering of specific feature genes. **(c)** Stage alignment based on expression patterns of conserved genes. Pink and blue bars indicate two distinct developmental stages identified by transcriptomics during morphogenesis. The figure was created using BioRender.

connections in different biological contexts. Integrating GCN analysis with other types of data can facilitate understanding of the genetic basis of morphological diversity, while dissection and stage-by-stage analysis of GCNs can help understand how differences in transcriptional regulation affect morphological traits over time [47]. One common approach is to combine GCNs with phylogenetic data (i.e., phylostraitigraphic and orthogroups information) to reveal conserved modules and clade-specific interactions [48–51]. This approach can identify genes and pathways that have been conserved throughout evolution and can provide insights into the evolutionary relationships between different species (Figure 3A). Another approach is to integrate GCNs with additional experimental data such as chromatin accessibility [52] or use GRN inference algorithms such as GRNBoost2 [53] and GENIE3 [54] to extract gene regulatory networks (GRNs) to reduce complexity. This approach can provide a more detailed understanding of the mechanisms underlying morphological diversity by identifying the regulatory relationships between different genes and proteins. The identification of hub genes that are highly connected within the network can also provide insights into key regulators of morphological development [55] (Figure 3B). Transcriptome profiling of ontogeny or time-course data across two species can also be used to investigate how core network components change over space and time [38,56]. This approach can identify genes that are conserved between species and those that have diverged, providing insights into the genetic basis of morphological differences between species (Figure 3C).

The choice of approach will depend on the genetic data available for the species, the sampling method, and the quality and depth of the transcriptomics data. Ultimately, the choice of approach will depend on the specific research question and the available genetic data for each species. For example, if the goal is to understand the evolutionary relationships between different species, a phylogenetically-informed approach may be most appropriate. Ruprecht et al., presented a method that

Figure 3



Overview of pairwise comparison of cross-species comparative transcriptomic analysis. The three approaches aimed to construct and compare **(a)** gene co-expression networks, **(b)** gene regulatory networks and **(c)** stage-ordered networks to explain morphological differences during organ development between *Species A* and *Species B*. **(a)** Combines phylogenetic information, such as orthologous groups or phylostratigraphic data, with gene co-expression networks to extract developmental sub-networks based on shared patterns of gene expression among orthologs. This approach highlights the significance of evolutionary conservation in developmental processes. **(b)** Integrates transcriptomics data with inferred transcriptional regulation relationships (e.g., promoter binding sites) or experimentally validated protein–protein interactions to construct species-specific gene regulatory networks, providing insights into the molecular mechanisms underlying organ development. **(c)** Focuses on a set of master regulators, key genes controlling developmental processes, and uses ontogeny transcriptomes as datasets to capture GRN changes over a developmental phase. Stage-ordered networks are constructed by aligning the expression patterns of these master regulators, facilitating the identification of conserved and divergent regulatory mechanisms. The figure was created using BioRender.

uses genomic and phylogenetic data with gene co-expression networks to study the evolutionary make-up of modules in moss and two angiosperms [48]. The analysis revealed modules that emerged at a specific time in plant evolution, and added phylogenetic information that revealed duplication and speciation events on the module level to uncover the evolutionary relationships of the conserved modules across the plant kingdom. On the other hand, if the goal is to identify specific genes or gene networks that are involved in shaping morphological differences between species, a more targeted approach that integrates protein–protein interaction data or focuses on core modules may be more appropriate. In a comparative network-based analysis of six angiosperm species, the authors integrated transcriptomic datasets with proteomic datasets to identify tissue-specific conserved modules and clade-specific gene sets to explain different species-specific phenotypic traits [43].

Conclusion

Comparative transcriptomics is a valuable tool for understanding the molecular mechanisms of organ formation and diversification during evolution. However, when applying this approach to plant transcriptome data, there are challenges in identifying homologous gene pairs and functional orthologs, selecting the appropriate developmental stage to compare, and extracting biological networks. To overcome these challenges, researchers can utilize methods such as orthogroups and phylostratigraphy to identify functionally equivalent genes that may not be orthologous but have similar functions. Additionally, network construction methods and combining other layered data can be employed to identify gene regulatory networks and infer gene function.

Despite the challenges, comparative transcriptomics has the huge potential to provide new insights into the

evolution of plant morphology and the genetic basis of plant development. Further research in this area including developing new algorithms and integration of AI technologies can help to elucidate the conserved/species-specific modules and key regulatory genes that underlie the formation of homologous organs across plant species and shed light on the mechanisms of morphological diversification.

Funding

This work is funded by NSF IOS-211980 grant. S.L. was supported by Plant Biology graduate program fellowship, University of California, Davis. This work was supported by JSPS KAKENHI (JP19K23742, JP20K06682, and JP20KK0340 to H.N.).

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT by OpenAI to assist in writing the content. The purpose of using this tool was to enhance the clarity, coherence, and readability of the manuscript. After utilizing the AI language model, the authors thoroughly reviewed and edited the generated content as necessary, ensuring its accuracy and appropriateness for publication. The authors take full responsibility for the final content of the publication.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Neelima R Sinha reports financial support was provided by US National Science Foundation.

Data availability

No data was used for the research described in the article.

Acknowledgment

We thank Dr. Aaron R. Leichty and two anonymous reviewers for the critical reading of this manuscript and for valuable feedback.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Carroll SB: **Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution**. *Cell* 2008, **134**:25–36.

2. Nardmann J, Werr W: **The evolution of plant regulatory networks: what Arabidopsis cannot say for itself**. *Curr Opin Plant Biol* 2007, **10**:653–659.

3. Roux J, Rosikiewicz M, Robinson-Rechavi M: **What to compare and how: comparative transcriptomics for Evo-Devo**. *J Exp Zool B Mol Dev Evol* 2015, **324**:372–382.

4. Delaux P-M, Hetherington AJ, Coudert Y, Delwiche C, Dunand C, Gould S, Kenrick P, Li F-W, Philippe H, Rensing SA, *et al.*: **Reconstructing trait evolution in plant evo–devo studies**. *Curr Biol* 2019, **29**:R1110–R1118.

5. Gan X, Hay A, Kwantes M, Haberer G, Hallab A, Ioio R Dello, Hofhuis H, Pieper B, Cartolano M, Neumann U, *et al.*: **The Cardamine hirsuta genome offers insight into the evolution of morphological diversity**. *Nat Plants* 2016, **2**.

6. Jhanwar S, Malkmus J, Stolte J, Romashkina O, Zuniga A, Zeller R: **Conserved and species-specific chromatin remodeling and regulatory dynamics during mouse and chicken limb bud development**. *Nat Commun* 2021:12.

7. Fukushima K, Pollock DD: **Amalgamated cross-species transcriptomes reveal organ-specific propensity in gene expression evolution**. *Nat Commun* 2020:11.

8. Cardoso-Moreira M, Halbert J, Valloton D, Velten B, Chen C, Shao Y, Liechti A, Ascenção K, Rummel C, Ovchinnikova S, *et al.*: **Gene expression across mammalian organ development**. *Nature* 2019, **571**:505–509.

9. Gerstein MB, Rozowsky J, Yan KK, Wang D, Cheng C, Brown JB, Davis CA, Hillier L, Sisu C, Li JJ, *et al.*: **Comparative analysis of the transcriptome across distant species**. *Nature* 2014, **512**:445–448.

10. Cardoso-Moreira M, Halbert J, Valloton D, Velten B, Chen C, Shao Y, Liechti A, Ascenção K, Rummel C, Ovchinnikova S, *et al.*: **Gene expression across mammalian organ development**. *Nature* 2019, **571**:505–509.

11. Jin G, Ma PF, Wu X, Gu L, Long M, Zhang C, Li DZ: **New genes interacted with recent whole-genome duplicates in the fast stem growth of bamboos**. *Mol Biol Evol* 2021, **38**:5752–5768.

12. Yu Y, Hu H, Doust AN, Kellogg EA: **Divergent gene expression networks underlie morphological diversity of abscission zones in grasses**. *New Phytol* 2020, **225**:1799–1815.

13. McCarthy EW, Landis JB, Kurti A, Lawhorn AJ, Litt A: **The genetic basis of flower color differences in nicotiana tabacum**. In *The tobacco plant genome*. Edited by Ivanov NV, Sierro N, Peitsch MC: Springer International Publishing; 2020:175–193.

14. Ichihashi Y, Aguilar-Martínez JA, Farhi M, Chitwood DH, Kumar R, Millon LV, Peng J, Maloof JN, Sinha NR: **Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape**. *Proc Natl Acad Sci U S A* 2014:111.

15. Clark JW, Donoghue PCJ: **Whole-genome duplication and plant macroevolution**. *Trends Plant Sci* 2018, **23**:933–945.

16. Hetherington AJ, Berry CM, Dolan L: **Multiple origins of dichotomous and lateral branching during root evolution**. *Nat Plants* 2020, **6**:454–459.

17. Bharathan G, Goliber TE, Moore C, Kessler S, Pham T, Sinha NR: **Homologies in leaf form inferred from KNOXI gene expression during development**. *Science (1979)* 2002, **296**:1858–1860.

18. Gabaldón T, Koonin EV: **Functional and evolutionary implications of gene orthology**. *Nat Rev Genet* 2013, **14**:360–366.

19. Altschup SF, Gish W, Miller W, Myers EW, Lipman DJ: *Basic local alignment search tool*. 1990.

20. Li L, Stoeckert CJ, Roos DS: **OrthoMCL: identification of ortholog groups for eukaryotic genomes**. *Genome Res* 2003, **13**:2178–2189.

21. Haug-Baltzell A, Stephens SA, Davey S, Scheidegger CE, Lyons E: **SynMap2 and SynMap3D: web-based whole-genome synteny browsers**. In *Bioinformatics*. Oxford University Press; 2017:2197–2198.

22. Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D, Goodstein DM, Schmutz J: **GENESPACE tracks regions of interest and gene copy number variation across multiple genomes**. *Elife* 2022, **11**, e78526.

23. Van Bel M, Silvestri F, Weitz EM, Kreft L, Botzki A, Coppens F, Vandepoele K: **Plaza 5.0: extending the scope and power of**

comparative and functional genomics in plants. *Nucleic Acids Res* 2022, **50**:D1468–D1474.

24. Emms DM, Kelly S: **OrthoFinder: phylogenetic orthology inference for comparative genomics**. *Genome Biol* 2019, **20**.

25. Sun Y, Shang L, Zhu QH, Fan L, Guo L: **Twenty years of plant genome sequencing: achievements and challenges**. *Trends Plant Sci* 2022, **27**:391–401.

26. Floyd SK, Bowman JL: **The ancestral developmental tool kit of land plants**. *Int J Plant Sci* 2007, **168**:1–35.

27. Sémon M, Wolfe KH: **Consequences of genome duplication**. *Curr Opin Genet Dev* 2007, **17**:505–512.

28. Stamboulian M, Guerrero RF, Hahn MW, Radivojac P: **The ortholog conjecture revisited: the value of orthologs and paralogs in function prediction**. *Bioinformatics* 2020, **36**: I219–I226.

29. Alejo-Jacuinde G, González-Morales SI, Oropeza-Aburto A, Simpson J, Herrera-Estrella L: **Comparative transcriptome analysis suggests convergent evolution of desiccation tolerance in Selaginella species**. *BMC Plant Biol* 2020:20.

30. Yan KK, Wang D, Rozowsky J, Zheng H, Cheng C, Gerstein M: **OrthoClust: an orthology-based network framework for clustering data across multiple species**. *Genome Biol* 2014, **15**:R100.

31. Domazet-Lošo T, Brajković J, Tautz D: **A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages**. *Trends Genet* 2007, **23**:533–539.

32. ** Julca I, Ferrari C, Flores-Tornero M, Proost S, Lindner AC, Hackenberg D, Steinbachová L, Michaelidis C, Gomes Pereira S, Misra CS, *et al.*: **Comparative transcriptomic analysis reveals conserved programmes underpinning organogenesis and reproduction in land plants**. *Nat Plants* 2021, **7**:1143–1159.

This comparative study utilized transcriptomic data and phylostratigraphic analysis to examine 9 organs across 23 land plant species. This study provide a list of organ-specific orthogroups could be used as feature genes for stage alignment.

33. van Noort V, Snel B, Huynen MA: **Predicting gene function by conserved co-expression**. *Trends Genet* 2003, **19**:238–242.

34. Serin EAR, Nijveen H, Hilhorst HWM, Ligterink W: **Learning from co-expression networks: possibilities and challenges**. *Front Plant Sci* 2016, **7**.

35. Tarashansky AJ, Musser JM, Khariton M, Li P, Arendt D, Quake SR, Wang B: **Mapping single-cell atlases throughout metazoa unravels cell type evolution**. *Elife* 2021:10.

36. Guillotin B, Rahni R, Passalacqua M, Mohammed MA, Xu X, Raju SK, Ramírez CO, Jackson D, Groen SC, Gillis J, *et al.*: **A pan-grass transcriptome reveals patterns of cellular divergence in crops**. *Nature* 2023, https://doi.org/10.1038/s41586-023-06053-0.

37. Drost HG, Janitza P, Grosse I, Quint M: **Cross-kingdom comparison of the developmental hourglass**. *Curr Opin Genet Dev* 2017, **45**:69–75.

38. ** Leiboff S, Hake S: **Reconstructing the transcriptional ontogeny of maize and sorghum supports an inverse hourglass model of inflorescence development**. *Curr Biol* 2019, **29**: 3410–3419.e3.

This paper describes how comparative transcriptomics can be used to compare inflorescence development across species. The authors utilized a pseudotime metric to align developmental stages and reconstruct the ontogeny of Maize and Sorghum inflorescences.

39. Calderwood A, Hepworth J, Woodhouse S, Bilham L, Jones DM, Tudor E, Ali M, Dean C, Wells R, Irwin JA, *et al.*: **Comparative transcriptomics reveals desynchronisation of gene expression during the floral transition between Arabidopsis and Brassica rapa cultivars**. *Quant Plant Biol* 2021:2.

40. Lemmon ZH, Park SJ, Jiang K, Van Eck J, Schatz MC, Lippman ZB: **The evolution of inflorescence diversity in the nightshades and heterochrony during meristem maturation**. *Genome Res* 2016, **26**:1676–1686.

41. * Meir Z, Aviezer I, Chongloi GL, Ben-Kiki O, Bronstein R, Mukamel Z, Keren-Shaul H, Jaitin D, Tal L, Shalev-Schlosser G, *et al.*: **Dissection of floral transition by single-meristem transcriptomes at high temporal resolution**. *Nat Plants* 2021, **7**: 800–813.

This paper presented a statistical method that enables the alignment and visualization of developmental stages using feature genes.

42. Vercruysse J, Van Bel M, Osuna-Cruz CM, Kulkarni SR, Storme V, Nelissen H, Gonzalez N, Inzé D, Vandepoele K: **Comparative transcriptomics enables the identification of functional orthologous genes involved in early leaf growth**. *Plant Biotechnol J* 2020, **18**:553–567.

43. * Shin J, Marx H, Richards A, Vaneechoutte D, Jayaraman D, Maeda J, Chakraborty S, Sussman M, Vandepoele K, Ané JM, *et al.*: **A network-based comparative framework to study conservation and divergence of proteomes in plant phylogenies**. *Nucleic Acids Res* 2021:49.

The paper presented a comparative analysis of transcriptome and proteome data from six plant species, employing clustering and correlation methods.

44. * Julca I, Tan QW, Mutwil M: **Toward kingdom-wide analyses of gene expression**. *Trends Plant Sci* 2023, **28**:235–249.

This review comprehensively listed major resources and outlined general pipelines for comparative transcriptomics.

45. Koutrouli M, Karatzas E, Paez-Espino D, Pavlopoulos GA: **A guide to conquer the biological network era using graph theory**. *Front Bioeng Biotechnol* 2020, **8**.

46. Nakayama H, Rowland SD, Cheng Z, Zumstein K, Kang J, Kondo Y, Sinha NR: **Leaf form diversification in an ornamental heirloom tomato results from alterations in two different HOMEOBOX genes**. *Curr Biol* 2021, **31**:4788–4799.e5.

47. Ovens K, Eames BF, McQuillan I: **Comparative analyses of gene Co-expression networks: implementations and applications in the study of evolution**. *Front Genet* 2021:12.

48. Ruprecht C, Proost S, Hernandez-Coronado M, Ortiz-Ramirez C, Lang D, Rensing SA, Becker JD, Vandepoele K, Mutwil M: **Phylogenomic analysis of gene co-expression networks reveals the evolution of functional modules**. *Plant J* 2017, **90**: 447–465.

49. Jin G, Ma P-F, Wu X, Gu L, Long M, Zhang C, Li D-Z: **New genes interacted with recent whole-genome duplicates in the fast stem growth of bamboos**. *Mol Biol Evol* 2021, **38**:5752–5768.

50. Yu Y, Hu H, Doust AN, Kellogg EA: **Divergent gene expression networks underlie morphological diversity of abscission zones in grasses**. *New Phytol* 2020, **225**:1799–1815.

51. * Curci PL, Zhang J, Mähler N, Seyfferth C, Mannapperuma C, Diels T, Van Hautegem T, Jonsen D, Street N, Hvidsten TR, *et al.*: **Identification of growth regulators using cross-species network analysis in plants**. *Plant Physiol* 2022, **190**:2350–2365.

The authors performed cross-species network comparison by extracting leaf developmental gene co-expression networks from three angiosperm species, integrating them with orthologous information to predict regulator function.

52. Reynoso MA, Kajala K, Bajic M, West DA, Pauluzzi G, Yao AI, Hatch K, Zumstein K, Woodhouse M, Rodriguez-Medina J, *et al.*: **Evolutionary flexibility in flooding response circuitry in angiosperms**. *Science (1979)* 2019, **365**:1291–1295.

53. Moerman T, Aibar Santos S, Bravo González-Blas C, Simm J, Moreau Y, Aerts J, Aerts S: **GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks**. *Bioinformatics* 2019, **35**:2159–2161.

54. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P: **Inferring regulatory networks from expression data using tree-based methods**. *PLoS One* 2010, **5**.

55. Das Gupta M, Tsiantis M: **Gene networks and the evolution of plant morphology**. *Curr Opin Plant Biol* 2018, **45**:82–87.

56. Chang YM, Lin HH, Liu WY, Yu CP, Chen HJ, Wartini PP, Kao YY, Wu YH, Lin JJ, Lu MYJ, *et al.*: **Comparative transcriptomics method to infer gene coexpression networks and its applications to maize and rice leaf transcriptomes**. *Proc Natl Acad Sci U S A* 2019, **116**:3091–3099.