American Journal of Botany
BOTANICAL SOCIETY OF AMERICA
leading scientists and educators
since 1893

1, 2, 3, GO! Venture beyond gene ontologies in plant evolutionary research

Cody Coyotee Howard[1,2,7,*], Carrie M. Tribble[3,*], Jesús Martínez-Gómez[4,*], Emily B. Sessa[2], Chelsea D. Specht[4], and Nico Cellinese[1,5,6]

Manuscript received 21 July 2020; revision accepted 28 September 2020.

[1] Florida Museum of Natural History, University of Florida, Gainesville, FL, USA

[2] Department of Biology, University of Florida, Gainesville, FL, USA

[3] University Herbarium and Department of Integrative Biology, University of California, Berkeley, CA, USA

[4] School of Integrative Plant Science, Section of Plant Biology and the L. H. Bailey Hortorium, Cornell University, Ithaca, NY, USA

[5] Biodiversity Institute, University of Florida, Gainesville, FL, USA

[6] Genetics Institute, University of Florida, Gainesville, FL, USA

[7] Author for correspondence (e-mail: cchoward@ufl.edu)

*These authors contributed equally to the writing of this manuscript.

Citation: Howard, C. C., C. M. Tribble, J. Martínez-Gómez, E. B. Sessa, C. D. Specht, and N. Cellinese. 2021. 1, 2, 3, GO! Venture beyond gene ontologies in plant evolutionary research. *American Journal of Botany* 108(3): 361–365.

doi:10.1002/ajb2.1622

KEY WORDS evolution; morphology; ontology; phylogeny; reproducibility.

Over the past few decades evolutionary biologists have vastly improved computational methods for streamlining the collection and analysis of genetic data from nonmodel systems. However, tools for phenotypic data are lagging despite the critical need for such data to understand how traits have evolved at micro- and macroevolutionary scales. Compared to data in standardized genetic databases (NCBI, etc.), phenotypic data are often scattered across non-standard repositories (e.g., supplementary material), are not publicly available, or are not comparable across studies due to subjectivity of the terms used, inability to test homology given missing data, and variability of approaches to coding characters and character states. Subjective terminology for characterizing phenotypes hampers our ability to leverage existing data for large-scale meta-analyses in evolutionary or morphological research or determine whether results across studies are comparable. These problems are magnified in studies that include complex or poorly characterized morphologies where terminology is often imprecise and conflicting, or specific to the group of study. To increase data clarity and scientific reproducibility, we advocate for the use of existing ontologies and continued improvement of plant ontologies.

ONTOLOGIES: FROM CHAOS TO ORDER

Ontologies are used in many fields of comparative biology to organize terminology (see Ontobee [www.ontobee.org] for a listing of available ontologies). An ontology acts as a roadmap, where roads (i.e., edges) connect cities (i.e., terms), which allows a traveler (i.e., a computer) to understand how places are connected and/or nested. This hierarchical structure provides a way to consistently describe the relationship(s) between terms for both humans and computers. Perhaps the best known ontology is the Gene Ontology (GO) (Ashburner et al., 2000), a database connecting organisms, their parts, developmental processes, and involved genes. For example, GO specifies that the gene *starch synthase 3* is involved in starch synthase activity (molecular function) in the chloroplast (cellular component). Users can also query GO for all genes involved in starch synthase activity.

Although subject-specific ontologies address questions at different scales or contexts (e.g., genes, traits, ecological interactions), many ontologies can be linked through shared terminology, allowing for data integration and comparisons across databases and/or biological levels (e.g., cellular, phenotypic, ecological) (Fig. 1). As more data types are obtained, the integration of diverse data opens the potential to address complex biological questions (such as Fig. 1) (Parr et al., 2012). However, integrating such complex data can only be achieved with targeted efforts to develop new ontologies and improve existing ones.

ONTOLOGIES AND PHYLOGENIES

The incompleteness of large data sets stymies comparative research. However, relationships formalized through ontologies can be used to populate sparse character matrices (Dececchi et al., 2015). Jackson et al. (2018) investigated fin evolution across 12,500+ teleost fishes, but many taxa were not coded for some traits. For taxa with missing data, ontologies were used to infer the presence or absence of paired fins if a taxon was known to have a dependent feature, such as a girdle (Jackson et al., 2018), thereby permitting the testing of broad evolutionary patterns across diverse taxa by leveraging our knowledge of the nestedness of morphological terms within an ontological framework.

Recently, Tarasov et al. (2019) developed a pipeline (PARAMO) for leveraging anatomy-based ontologies to investigate morphological evolution across lineages within a phylogenetic context. PARAMO uses comparative methods to investigate the evolutionary rates and tempos for developmentally connected traits as units by arranging individually scored characters (e.g., tuberous root shape and color, leaf phyllotaxy, leaf shape) into higher-level organizations (e.g., root, leaf) for analysis (Fig. 2). Ontology-informed phylogenetic methods not only allow us to parse the evolution of different systems (for example, is the root system evolving at a different rate than the shoot system across clades?), but also to integrate these interconnected systems into one cohesive evolutionary framework (Fig. 2), which opens the door to tackling organismal and inter-organismal questions addressing adaptation and diversification in comparative biology.

THE STATE OF PLANT TRAIT ONTOLOGIES

The foundation of any ontology is the breadth and depth of terms that can be used to describe content, detail properties, and demonstrate relationships. A critical first step before building an ontology is establishing a controlled vocabulary agreed upon by subject experts. This vocabulary can then be used as the basis to construct the hierarchical relationship between terms that ontologies require. However, specifications for many plant trait ontologies are often limited in taxonomic scope or details such that organismal complexity is not fully considered (e.g., the situation with geophytes, Howard et al., 2020; see Tribble et al., 2021, in this issue). For example, the current definition of bulb in the Plant Ontology (PO) places it under shoot system and lacks descriptors for diverse morphologies and processes associated with bulbs, including foliage leaf bases, swollen leaf bases and/or leaf scales, and leaf thickening. Variations in these traits underlie differences between key morphological features with evolutionary and ecological importance, e.g., imbricate vs. tunicate bulbs (see Tribble et al., 2021, in this issue), and demonstrate the need to further develop the Plant Ontology. In addition to capturing morphological diversity of particular taxa, ontologies can focus on specific traits such as wood anatomy (Lens et al., 2012). For example, a tracheary element is a plant cell and part of the xylem, which can be further designated as primary or secondary xylem. Heartwood, growth rings, and reaction wood are all standardized and computable elements of secondary xylem, forming a wood ontology. Despite broad applicability, innovative applications of ontologies in the plant sciences (e.g., large-scale integrations across data sets) are hampered by a lack of comprehensive ontological data sets that span taxa and tissue types (but see Walls et al., 2019).

“ONTOLOGIZING” THE BOTANICAL COMMUNITY

Given the important insights into plant evolution that ontologies can offer, it is concerning that ontologies are not more common in standard comparative research workflows. Ontologies have been used to hypothesize that the angiosperm inner integument (seed coat) is homologous with the gymnosperm integument (Stevenson and Zumajo-Cardona, 2018) or to explore gene expression patterns across developmental stages (Walls et al., 2019), the results of which can be tested with additional comparative analyses. To encourage the adoption of ontologies by botanists, we provide suggestions for engaging colleagues and standardizing practices.

Working collaboratively to improve plant ontologies

Ontologies are built and improved through collaborations between subject-specific experts (e.g., ecologists, developmental biologists, morphologists, and taxonomists) and ontology curators, many of whom are eager to improve ontologies by actively seeking input from experts (Stucky et al., 2019; Walls et al., 2019). The lack of detailed ontologies for geophytic traits, for example, comes from having few geophytic model organisms to pull terms from, a lack of agreement of how terms are defined by individuals and in the literature, and/or a lack of community buy-in for the use of ontologies in organismal studies. Substantial effort is required, beginning with a detailed knowledge and vocabulary of plant morphology and diversity to fully capture geophyte terminology and to place these terms and the diversity they represent into an ontological framework. Fortunately, an excellent resource (Planteome; www.planteome.org) provides a tool to reevaluate and generate ontologies for defined morphological traits. For other disciplines, we encourage readers to visit Ontobee (www.ontobee.org) and find an ontology that fits with their organismal, evolutionary, or ecological interests. Researchers today can contribute by including unique identifiers associated with terms (Fig. 1) used in their publication, thereby reducing ambiguity and improving discoverability and reproducibility of their results. If a term is not available or is insufficiently defined, we encourage actively working with ontology curators (e.g., Request Terms on Plant Ontology) to refine and add to an ontology as part of the publication process. Workshops on subject-specific ontologies (e.g., Wood Anatomy Ontology meeting) may be an effective way to leverage collective

15372197, 2021, 3, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1622 by Cornell University Library, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

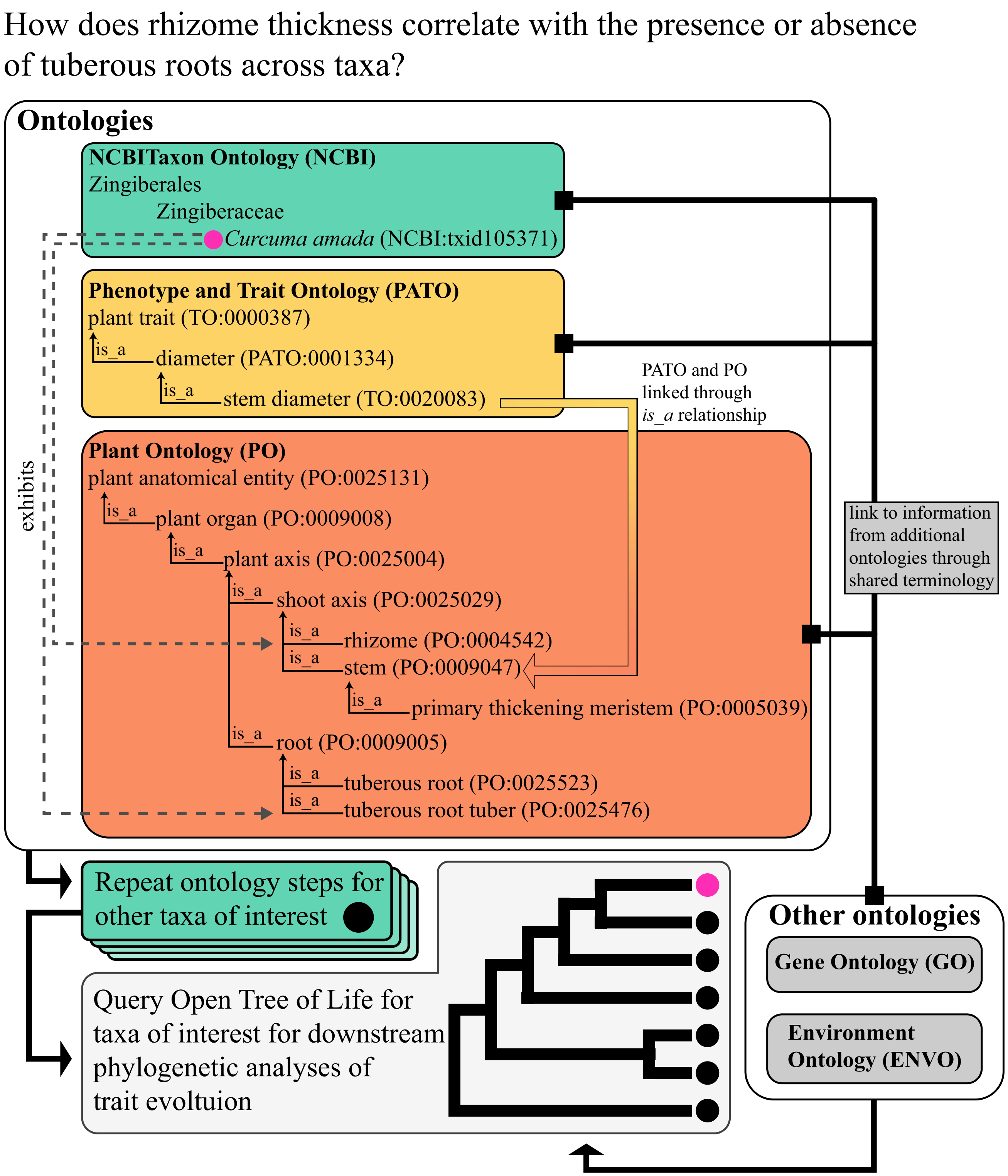

FIGURE 1. Ontologies allow for data integration across disciplines, which drives innovative research on a wide breadth of biological questions. Using currently available ontologies to associate data with various taxa, one could ask, "How does rhizome thickness correlate with the presence or absence of tuberous roots across taxa?" The NCBITaxon ontology (NCBI, www.ncbi.nlm.nih.gov/taxonomy) can be used to associate data with *Curcuma amada* (Zingiberaceae), which has a rhizome and tuberous roots. Stem diameter of the rhizome can be captured through the Phenotype and Trait Ontology (PATO), which can be linked to the Plant Ontology (PO) by the term stem. These steps can then be repeated for other taxa of interest. Additional information such as habitat types or gene expression patterns may be linked through shared terminology in other ontologies, including the Gene Ontology (GO) or the Environment Ontology (ENVO). The analysis of all of this information would provide a holistic view of rhizome and tuberous root evolution.

15372197, 2021, 3, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1622 by Cornell University Library, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

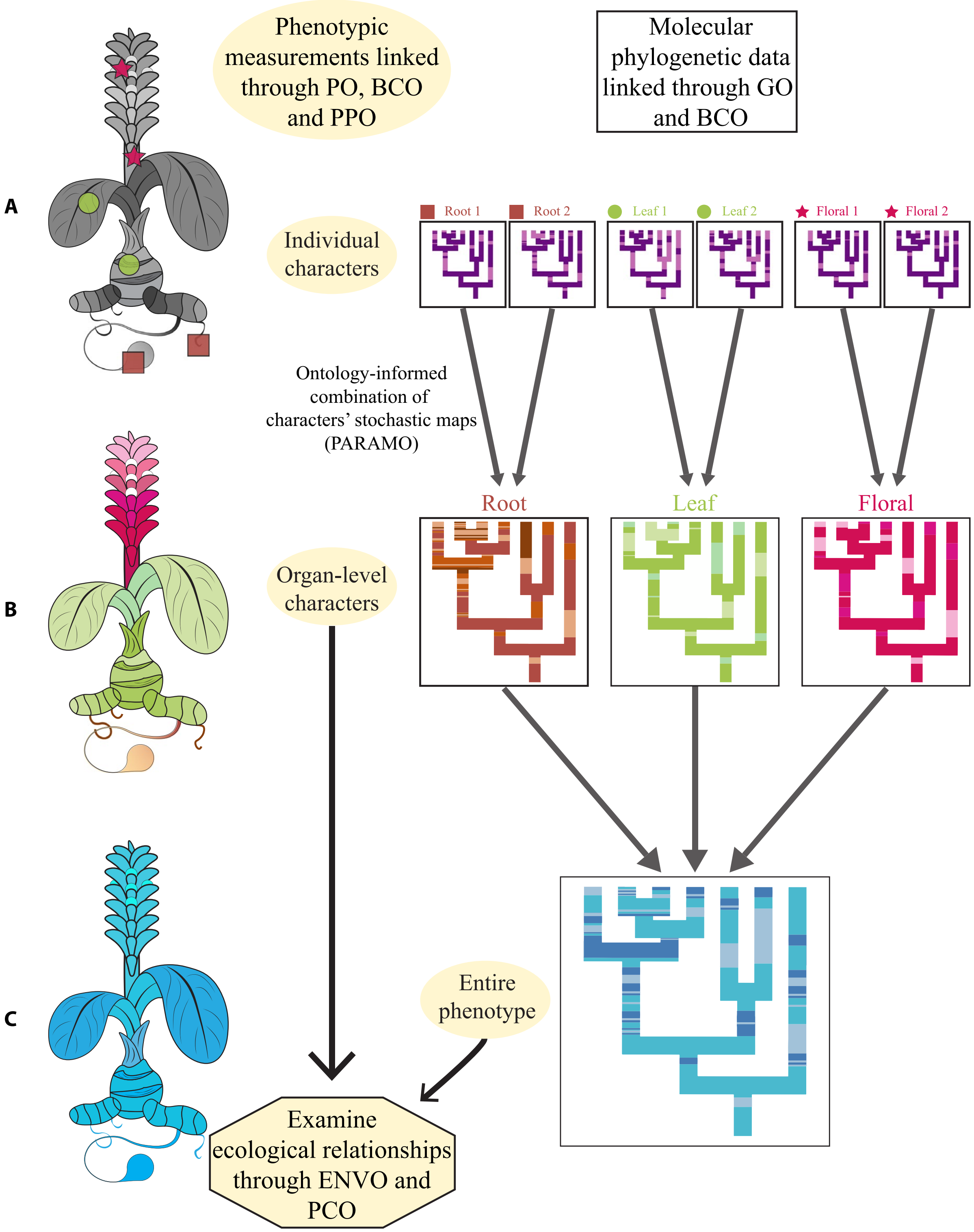

FIGURE 2. (A) Individual characters, here illustrated on *Curcuma* (Zingiberaceae), can be coded and stored using standardized terminology found in ontology databases, such as the Plant Ontology (PO). Using other ontologies, data such as the genes used to build the phylogeny (GO; Gene Ontology), collection techniques (BCO; Biological Collections Ontology), herbarium specimen and phenology information (PPO; Plant Phenology Ontology) can also be recorded. (B) Using ontology-based amalgamation of characters (PARAMO), the individual characters can then be combined into their respective higher level organizations, such as root, leaf and floral characters (organ-level characters) and mapped onto a phylogeny to visualize character changes along the branches. (C) The evolutionary history of the higher levels can then be examined from the viewpoint of the entire plant phenotype. Additional factors, such as ecological interactions could be examined at each level of organization by integrating data with the Environment Ontology (ENVO) or the Population and Community Ontology (PCO). Figure adapted from Tarasov et al. (2019).

15372197, 2021, 3, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1622 by Cornell University Library, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

expertise and bring together subject experts and ontology developers. These subject- or clade-specific ontologies may provide a better basis for term integration with more expansive ontologies by allowing complex traits to be more accurately described and ontologized.

Create user-friendly ontology interfaces

Many scientists working with ontologies are computational biologists or bioinformaticians adept at querying and retrieving data. For many organismal biologists, ontologies might seem like a time-consuming approach to describe trait data. Therefore, we advocate for the development of user-friendly ontology interfaces with options to explore, retrieve, and integrate data across taxonomic scales, similar to Phenoscape (Mabee et al., 2018). These interfaces should be developed through collaborations between web developers, ontology curators, and botanical experts. User-friendly interfaces for phylogeny reconstruction are readily available (e.g., OneTwoTree [Drori et al., 2018]; www.onetwotree.tau.ac.il) and lower the barriers associated with phylogenetic reconstruction. Corresponding interfaces for ontologies, however, do not exist in the plant sciences. Platforms with step-by-step guides for different subjects would likely increase their use. Additionally, ontologies are hierarchical, like a phylogeny. Interactively zooming in and clicking on terms of interest, similarly to the OneZoom Tree of Life Explorer (www.onezoom.org), could facilitate an understanding of how ontologies and traits are organized. Moving forward, user-friendly platforms are critical, especially as plant and related ontologies become more elaborate due to our continually improved understanding of plant morphological development, form, and function.

CONCLUSIONS

Building upon available ontologies to accurately convey plant form and function will improve our ability to make inferences about complex evolutionary processes. But we are hampered by limited awareness of ontologies and by the lack of user-friendly tools for inputting, manipulating, and managing such data. Given the need for broad syntheses of complex biological questions by means of integrating heterogeneous data sets, creating or refining ontological terms should become an integral component of our scientific workflows. Large-scale analyses act as hypothesis generators; the generalizability of the resulting patterns can then be tested at different scales. In a similar vein, large-scale ontologies provide opportunities to generate hypotheses about plant form and function that span the plant tree of life as well as ease incorporating additional information at smaller scales, such as gene expression or environmental interactions. Ontologies are an exciting step forward for enhancing our understanding of the intricacies of plant evolution and improving scientific reproducibility.

ACKNOWLEDGMENTS

We thank Pamela Diggle and two anonymous reviewers for their constructive feedback on earlier versions of this manuscript.

AUTHOR CONTRIBUTIONS

C.C.H., C.M.T., and J.G.-M. initiated the ideas laid out in the manuscript; all authors improved the concepts introduced and helped with writing the manuscript.

LITERATURE CITED

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, et al. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics* 25: 25–29.

Dececchi, T. A., J. P. Balhoff, H. Lapp, and P. M. Mabee. 2015. Toward synthesizing our knowledge of morphology: using ontologies and machine reasoning to extract presence/absence evolutionary phenotypes across studies. *Systematic Biology* 64: 936–952.

Drori, M., A. Rice, M. Einhorn, O. Chay, L. Glick, and I. Mayrose. 2018. OneTwoTree: an online tool for phylogeny reconstruction. *Molecular Ecology Resources* 18: 1492–1499.

Howard, C. C., J. B. Landis, J. M. Beaulieu, and N. Cellinese. 2020. Geophytism in monocots leads to higher rates of diversification. *New Phytologist* 225: 1023–1032.

Jackson, L. M., P. C. Fernando, J. S. Hanscom, J. P. Balhoff, and P. M. Mabee. 2018. Automated integration of trees and traits: a case study using paired fin loss across teleost fishes. *Systematic Biology* 67: 559–575.

Lens, F., L. Cooper, M. A. Gandolfo, A. Groover, P. Jaiswai, B. Lachenbruch, R. Spicer, et al. 2012. An extension of the Plant Ontology project supporting wood anatomy and development research. *IAWA Journal* 33: 113–117.

Mabee, P. M., W. M. Dahdul, J. P. Balhoff, H. Lapp, P. Manda, J. Uyeda, T. Vision, and M. Westerfield. 2018. Phenoscape: semantic analysis of organismal traits and genes yields insights in evolutionary biology. *PeerJ Preprints* 6: e26988v1.

Parr, C. S., R. Guralnick, N. Cellinese, and R. D. M. Page. 2012. Evolutionary informatics: unifying knowledge about the diversity of life. *Trends in Ecology & Evolution* 27: 94–103.

Stevenson, D., and C. Zumajo-Cardona. 2018. From Plant Ontology to Gene Ontology and back. *Current Plant Biology* 14: 66–69.

Stucky, B. J., J. P. Balhoff, N. Barve, V. Barve, L. Brenskelle, M. H. Brush, G. A. Dahlem, et al. 2019. Developing a vocabulary and ontology for modeling insect natural history data: example data, use cases, and competency questions. *Biodiversity Data Journal* 7: e33303.

Tarasov, S., I. Mikó, M. J. Yoder, and J. C. Uyeda. 2019. PARAMO: A pipeline for reconstructing ancestral anatomies using ontologies and stochastic mapping. *Insect Systematics and Diversity* 3: 1–7.

Tribble, C. M., J. Martínez-Gómez, C. C. Howard, J. Males, V. Sosa, E. B. Sessa, N. Cellinese, and C. D. Specht. 2021. Get the shovel: morphological and evolutionary complexities of belowground organs in geophytes. *American Journal of Botany* 108: in press.

Walls, R. L., L. Cooper, J. Elser, M. A. Gandolfo, C. J. Mungall, B. Smith, D. W. Stevenson, and P. Jaiswal. 2019. The plant ontology facilitates comparisons of plant development stages across species. *Frontiers in Plant Science* 10: 631.

15372197, 2021, 3, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1622 by Cornell University Library, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License