Multi-scale Digital Twins for Plants: Integrating Phenotypic and Molecular Responses Across Time and Space

Authors: Kriti Chopra, Vanessa Lopez-Marrero, Xin Dai and David Park (kchopra,vlopezmar,xdai,dpark1@bnl.gov); Brookhaven National Laboratory.

Topic: Digital Twins

Challenges: The rapid evolution of plant biology research demands novel approaches to modeling and prediction. Current models struggle to integrate multiple layers of biological organization, from molecular mechanisms to visible phenotypes, hindering our ability to predict whole-plant responses. Complex, non-linear relationships between molecular responses, such as metabolic flux reorganization without corresponding transcriptional changes, challenge our ability to build accurate predictive models [1]. The inherent uncertainty in biological measurements and complex feedback loops make establishing reliable cause-effect relationships between molecular perturbations and phenotypic responses difficult. Most current modeling approaches provide steady-state solutions, failing to capture the dynamic nature of plant responses to rapidly changing environmental conditions. While high-throughput phenotyping and multi-omics technologies provide unprecedented data, integrating diverse data types across different experiments, resolutions, and time scales remains a significant hurdle.

Opportunity: The rapid advancement of AI and machine learning, coupled with the increasing complexity of plant biology research, necessitates a paradigm shift in how we model and understand plant systems. Digital twin technology [2,3], if applied to plant systems, offers a promising solution to the above stated challenges. By creating comprehensive virtual representations of plants, we can simulate and predict responses to various stimuli across multiple biological scales. Such plant digital twins offer numerous advantages. For instance, they can enable researchers to: a) simulate complex plant behaviors *in silico*, reducing the need for time-consuming and resource-intensive experiments, b) allow for the integration of diverse data types and the prediction of plant responses to environmental stresses and potential genetic modifications, and c) realize optimal goals, such as crop breeding strategies and agricultural practices.

To enable this potential for plant digital twins, we must develop new computational frameworks that can handle the scale and complexity of plant systems. This will require advances in computational methods that allow for the judicious integration/coupling of a range of models, encompassing various degrees of fidelities and spatio-temporal scales, as well as the incorporation of uncertainty quantification, validation and verification, and data assimilation techniques [2,3], and necessitates new research on:

a. Multi-scale modeling techniques that seamlessly transition between molecular, cellular, and whole-plant levels. This includes new methods for coupling advanced Flux Balance Analysis (FBA) techniques [4,5] for metabolic modeling with mechanistic models of ordinary/partial differential equations for capturing temporal/spatio-temporal dynamics of signaling molecules, nutrients, growth patterns, and other biological processes [6–9].
b. AI algorithms and architectures that can learn from sparse, heterogeneous biological data, capture long-range dependencies in genomic data, and allow the optimization of intervention strategies. This can entail new research on graph neural networks, implicit neural representations, transformers, and reinforcement learning [10], among others.
c. Uncertainty quantification methods, such as Bayesian approaches and ensemble modeling [11], to assess model predictions and guide experimental design.
d. High-performance computing solutions, including distributed computing and GPU acceleration, to manage the computational demands of these complex models.

This integrative approach has the potential to uncover novel biological insights and generate new hypotheses. By simulating complex interactions across multiple scales, digital twins may reveal

emergent properties and non-intuitive relationships that are not apparent when studying individual components in isolation. For instance, this framework could illuminate how seemingly unrelated genetic modifications might synergistically enhance drought resistance, or how subtle changes in metabolic fluxes could cascade into significant alterations in plant architecture. Such discoveries could guide targeted experiments, accelerate the pace of scientific discovery, and potentially lead to breakthroughs in our fundamental understanding of plant biology.

Timeliness or maturity: The convergence of high-throughput phenotyping technologies, multi-omics data generation, and advancements in computational modeling and AI creates a unique opportunity to develop integrated plant digital twins. Recent breakthroughs in metabolic modeling, such as genome-scale models for crop species, and AI, particularly in generative foundation models pre-trained on large biological datasets, provide a solid foundation for capturing complex plant processes across different scales and modalities of data. The increasing urgency of global challenges further underscores the timeliness of this approach. Climate change adaptation, food security, sustainable agriculture, and rapid response to emerging threats all demand innovative solutions that plant digital twins can provide. The maturity of component technologies, combined with the critical need for these solutions, positions this framework at the cusp of transformative impact. By integrating mechanistic understanding with data-driven insights and rigorous uncertainty quantification across disciplines, we can create a powerful tool for advancing plant biology and developing more resilient and sustainable agricultural systems. This approach aligns with broader trends in scientific research, emphasizing interdisciplinary collaboration and systems-level understanding in biology.

References

[1] H. Shi *et al*.. Mechanisms of metabolic adaptation in the duckweed Lemma gibba: an integrated metabolic, transcriptomic and flux analysis. *BMC Plant Biology*, 2023.

[2] National Academy of Engineering and National Academies of Sciences, Engineering, and Medicine. *Foundational Research Gaps and Future Directions for Digital Twins*. The National Academies Press, Washington, DC, 2024.

[3] K. Willcox and B. Segundo. The role of computational science in digital twins. *Nat Comput Sci*, 4, 2024.

[4] H. Yuan *et al.* The effect of biomass composition and model structure on model predictions. *Frontiers in Plant Science*, 7, 2016.

[5] A. Sahu *et al.* Advances in flux balance analysis by¨ integrating machine learning and mechanism-based models. Comput. Struct. Biotechnol. J. , 19:4129–4138, 2021.

[6] L. Edelstein-Keshet. *Mathematical Models in Biology*. Classics in Applied Mathematics. SIAM, Philadelphia, PA, 2005.

[7] J. Elias and J. Clairambault. Reaction–diffusion systems for spatio-temporal intracellularˇ protein networks: A beginner's guide with two examples. Comput. Struct. Biotechnol. J. , 10(16):12–22, 2014.

[8] W. L. Schroeder and R. Saha. Introducing an optimization- and explicit Runge-Kutta- based approach to perform dynamic flux balance analysis. *Scientific Reports*, 10(1), 2020.

[9] M. Leclerc *et al.* Imaging with spatio-temporal modelling to characterize the dynamics of plant-pathogen lesions. *PLOS Computational Biology*, 19(11).

[10] J. L. Ballard *et al.* Deep learning-based approaches for multi-omics data integration and analysis. *BioData Mining*, 17(1):1–22, 2024.

[11] R. C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. Series in Computational Science & Engineering. SIAM, Philadelphia, PA, 2013.