

Perspective

PLANeT: Understanding and leveraging the genome of land plants for a sustainable future

Li Wang,¹ Mohamed Amar,² Carlos I. Arbizu,³ John L. Bowman,⁴ Charles H. Cannon,⁵ David J. Cantrill,⁶ Xiaofeng Cao,⁷ Shilin Chen,⁸ Xiaoya Chen,⁹ Xuemei Chen,¹⁰ Kang Chong,¹¹ Xingwang Deng,¹⁰ Weinan E,¹⁰ Hongwei Guo,¹² Chuan He,¹³ Peter Hollingsworth,¹⁴ Qing Hu,¹² Hongzhi Kong,¹¹ Ilia J. Leitch,¹⁵ Dezhu Li,¹⁶ Jiayang Li,⁷ Hongtao Liu,¹⁷ William Lucas,¹⁸ Eshchar Mizrahi,¹⁹ Mijoro Rakotoarinivo,²⁰ Hai Ren,²¹ Loren Rieseberg,²² Jue Ruan,¹ Rosa Scherson,²³

(Author list continued on next page)

¹Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China

²Center for Converging Sciences and Emerging Technology (CoSET), Benha National University, Al Obour, Egypt

³Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas (UNTRM), Chachapoyas, Peru

⁴Monash University, Melbourne, VIC, Australia

⁵Singapore Botanic Gardens, Singapore, Singapore

⁶Royal Botanic Gardens Victoria, Melbourne, VIC, Australia

⁷Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

⁸Chengdu University of Traditional Chinese Medicine, Chengdu, China

⁹CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai, China

¹⁰Peking University, Beijing, China

¹¹Institute of Botany, Chinese Academy of Sciences, Beijing, China

¹²New Cornerstone Science Laboratory, Shenzhen Key Laboratory of Plant Genetic Engineering and Molecular Design, Institute of Plant and Food Science, Department of Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen, China

¹³University of Chicago, Chicago, IL, USA

¹⁴Royal Botanic Garden Edinburgh, Edinburgh, UK

¹⁵Royal Botanic Gardens, Kew, Richmond, UK

¹⁶Shandong Agriculture University, Tai'an, China

¹⁷Shenzhen University, Shenzhen, China

¹⁸University of California, Davis, Davis, CA, USA

¹⁹University of Pretoria, Pretoria, South Africa

²⁰Université d'Antananarivo, Antananarivo, Madagascar

²¹South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China

²²University of British Columbia, Vancouver, BC, Canada

²³Universidad de Chile, Santiago, Chile

²⁴Xishuangbanna Botanical Garden, Chinese Academy of Sciences, Mengla, China

(Affiliations continued on next page)

SUMMARY

Land plants underpin civilization and planetary health, yet their genomic diversity remains largely uncharted. Current resources are unstandardized and scarce, lacking reference genomes for 95% of genera, 70% of families, and 51% of orders, impeding evolutionary and functional insight. We thus propose the PLANeT initiative, an international effort to generate high-quality, standardized genomes across the plant tree of life. Integrating artificial intelligence (AI) with genomics, we will decode conserved principles to advance fundamental plant biology, biodiversity conservation, crop improvement, and natural product discovery. Engaging around 100 labs to train 1,000 scientists, we will tackle pivotal questions for a sustainable future.

INTRODUCTION

Land plants (embryophytes), with ~450,000 species, form the foundation of terrestrial life and human civilization.^{1,2} They display a staggering breadth of morphological, karyotypic, ecological, and physiological diversity: genome sizes range from under 0.1 to over 160 Gb; life histories can be dominated

by haploid gametophyte or diploid sporophyte phases; and complex organs such as roots, leaves, flowers, and fruits were gained during their evolution but can be lost or become highly modified.^{3,4} Their abundant phenotypic diversity enables land plants to succeed in an enormous variety of habitats, including marine, arctic, alpine, and desert ecosystems,⁵ but many species are progressively threatened by global climate change



Harald Schneider,²⁴ Douglas E. Soltis,²⁵ Hang Sun,²⁶ Marcio C. Silva-Filho,²⁷ Stephen Smith,²⁸ Alex Twyford,²⁹ Björn Usadel,^{30,31} Yves Van de Peer,^{19,32,33,34} Cássio Van den Berg,³⁵ Rajeev K. Varshney,³⁶ Lyderson Facio Viccini,³⁷ Qingfeng Wang,³⁸ Xiao-Quan Wang,¹¹ Yadong Wang,³⁹ Jonathan Wendel,⁴⁰ Rod Wing,⁴¹ Gane Ka-Shu Wong,⁴² Yaowu Xing,²⁴ Xun Xu,⁴³ Nieng Yan,⁴⁴ Huanming Yang,^{43,45} Shouzhou Zhang,⁴⁶ Yuxian Zhu,⁴⁷ Pamela S. Soltis,^{25,*} Detlef Weigel,^{48,49,*} and Sanwen Huang^{1,*}

²⁵University of Florida, Gainesville, FL, USA

²⁶Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China

²⁷Universidade de São Paulo, Piracicaba, Brazil

²⁸University of Michigan, Ann Arbor, MI, USA

²⁹University of Edinburgh, Edinburgh, UK

³⁰Institute for Bio- and Geosciences (IBG-4: Bioinformatics), CEPLAS, Forschungszentrum Jülich, Jülich, Germany

³¹Institute for Biological Data Science, Heinrich Heine University, Düsseldorf, Germany

³²Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium

³³VIB Center for Plant Systems Biology, VIB, 9052 Ghent, Belgium

³⁴Nanjing Agricultural University, Nanjing, China

³⁵Universidade Estadual de Feira de Santana, Feira de Santana, Brazil

³⁶WA State Biotechnology Centre, Centre for Crop and Food Innovation, Murdoch University, Murdoch, Perth, WA, Australia

³⁷Universidade Federal de Juiz de Fora, Juiz de Fora, Brazil

³⁸Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, China

³⁹Harbin Institute of Technology, Harbin, China

⁴⁰Iowa State University, Ames, IA, USA

⁴¹King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

⁴²University of Alberta, Edmonton, AB, Canada

⁴³BGI-Shenzhen, Shenzhen, China

⁴⁴Institute of Bio-Architecture and Bio-Interactions, Shenzhen Medical Academy of Research and Translation, Shenzhen, China

⁴⁵James D. Watson Institute of Genome Sciences, Hangzhou, China

⁴⁶Fairy Lake Botanical Garden, Chinese Academy of Sciences, Shenzhen, China

⁴⁷Wuhan University, Wuhan, China

⁴⁸Max Planck Institute for Biology Tübingen, Tübingen, Germany

⁴⁹Institute for Bioinformatics and Medical Informatics (IBMI), University of Tübingen, Tübingen, Germany

*Correspondence: psoltis@flmnh.ufl.edu (P.S.S.), detlef.weigel@tuebingen.mpg.de (D.W.), huangsanwen@caas.cn (S.H.)

<https://doi.org/10.1016/j.cell.2026.01.026>

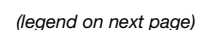
and habitat deterioration,^{6–8} posing a risk to human health and well-being. At the same time, much of land plant diversity remains uncharacterized.

A reference genome provides essential genetic information for a species—it not only captures its shared history and its unique traits but is also the platform for understanding intraspecific variation. Following the release of the first plant genome assembly in 2000 by the Arabidopsis Genome Initiative,⁹ genomic biology is helping us to understand, utilize, and preserve plant diversity.¹⁰ Nevertheless, the quality of genome assemblies and annotations, which is partially a consequence of the complex structure and history of plant genomes, is very uneven,¹¹ hampering progress in both comparative and functional genomics. Furthermore, although expansive “omics” datasets have established foundational knowledge islands, the sampled species are a poor reflection of overall embryophyte diversity, such that, as of 2024, 95% of genera, 70% of families, and 51% of orders do not contain species with a reference genome.^{12–14} Although the angiosperms, or flowering plants, are best studied, reference genomes are lacking for more than half of all angiosperm families and almost one out of five orders, some of which represent basal lineages (Figure 1). This is limiting the full resolution of their evolutionary history and thereby our understanding of the principal drivers of diversification at multiple phylogenetic levels, especially for plants from the Southern Hemisphere, where genomic diversity is largely untapped.¹⁵ The gaps described above make it difficult to transfer knowledge gained from model and

crop species to wild species, which are crucial sources of new medicines and crops, key players in ecosystems, and essential contributors to the planet’s and our own history and vice versa.

A primary challenge in making use of genomic information is to understand how genomes are organized and how genome organization relates to dynamic gene regulatory networks. Current efforts, which rely heavily on comparative genomics across limited model organisms, have struggled to capture the full spectrum of functionally relevant sequences in a systematic way. Emerging artificial intelligence (AI) algorithms have begun to offer powerful solutions, enabling, for example, the high-throughput exploration of regulatory diversity^{18,19} and the prediction of function from DNA sequences.²⁰ The power of AI models scales with the quality, diversity, and volume of training data,²¹ as shown by recent studies.^{22,23} Notably, Pavlopoulos and colleagues²² uncovered billions of novel protein sequences from metagenome data that bear no similarity to any known proteins. We therefore urgently need datasets encompassing the entire breadth of plant diversity, as genomes from a limited number of species are unlikely to provide sufficient sequence diversity for robust parameter learning.²⁴

To address the above-mentioned challenges and to ensure that humanity continues to live on a planet with not only a wide range of diverse plants but also genetically diverse plant populations, we propose an international collaborative effort, PLANeT, with the goal of characterizing the full spectrum of land plant genomic diversity and revealing the drivers that determine how



this diversity is generated, maintained, or lost. The proposed sampling prioritizes orders, families, and genera that lack a reference genome, aiming to resolve phylogenetic uncertainties and establish a robust framework for investigating trait evolution. Our vision is to work collaboratively to compile a digital library of the collective biological information of many thousands of land plant reference genomes,^{1,2} representing approximately 470 million years of evolution.²⁵ Guided by open science principles, this initiative will democratize genome accessibility across the diversity of plants and leverage AI to decipher conserved genomic principles underlying the regulation of genomic activity—including regulatory grammar, genomic organization, and molecular interactions—using scalable, diverse, and standardized datasets. Reference genomes also serve as a platform for the discovery of intraspecific diversity, which, in turn, can be a resource for enriching the genetic basis of cultivated crops, aiding policymaking on biodiversity preservation, and elucidating biosynthetic pathways of bioactive compounds (Box 1).

Based on extensive community discussion, this perspective lays out priorities, opportunities, and pressing questions across both the fundamental and applied disciplines of plant biology. An integral part of the project will be the establishment of bioinformatic platforms utilizing machine learning and other AI applications (Box 1). We also propose a list of questions that can be explored using the plethora of genome sequences we will generate (Note S1); discuss PLANeT's architecture and implementations (Note S2); and propose standard operating procedures for sample collection, data generation, and genome assembly (Note S3). These combined efforts will have far-reaching impacts on global food security, helping to deliver a sustainable future, improve human health, and train and empower a new generation of plant scientists.

FUNDAMENTAL PLANT BIOLOGY

Scientists have long been fascinated by the incredible diversity and adaptation of plants through deep time, a pursuit dating back to Darwin's theory of natural selection. Currently, the overarching challenge is to tease out the complex evolutionary paths behind key trait innovations and to reveal the mechanisms that drove the explosive diversification of land plants. A central puzzle is reconstructing the land plant "tree of life" with genomic data. In angiosperms, for example, a consensus recognizes the ANA grade (Amborellales, Nymphaeales, and Austrobaileyales) and five main lineages: eudicots, monocots, magnoliids, Chloranthales, and Ceratophyllales (Figure 1).^{16,37} However, the phylogenetic relationships between the five lineages remain controversial. This incongruence is linked to Darwin's "abominable mystery"—the rapid radiation of angiosperms. Both incomplete lineage sorting (ILS) and ancient hybridizations may have contributed to this phylogenetic uncertainty.^{38,39} With newly *de novo*-assembled genomes filling phylogenetic gaps,

along with published genomes from representative orders, families, and genera, a time-calibrated phylogenetic tree of land plants can be constructed.

This tree will serve as the basis for large-scale comparative genomics, revealing at scale the balance between purifying selection, which maintains the function of universally important DNA sequences (constraint), and positive selection, which supports evolutionary innovation. By providing chromosome-level genomes for over 90% of land plant families, we seek to unravel the evolutionary constraints and genomic innovations associated with each significant node in the land plant phylogenetic tree. This is akin to connecting disparate islands of knowledge (from accumulated omics data for a limited number of crops and model species) with "stepping-stones and milestones," revealing the entire "ocean" of land plant evolutionary journeys (Figure 2).

The greatly improved phylogenetic tree of land plants will also enable insights into the genetic basis of major evolutionary transitions, such as the origins of seeds and flowers, as well as trait variation. Genome size varies more than 2,600-fold, from 61 Mb/1C (*Genlisea tuberosa*) to 160,000 Mb/1C (*Tmesipteris oblanco-lata*).³ Genome size determines the time to complete the cell cycle due to its impact on the duration of mitosis and meiosis and DNA synthesis (S-phase), as well as the minimum nuclear, and hence cell, size via the biophysical constraints imposed by fitting DNA into the nucleus. In turn, these nuclear and cellular constraints influence phenotypes, physiological and metabolic functions, and ultimately species distributions and extinction risk.^{4,40–46} What are the forces driving the evolution of very small or very large genomes? How do polyploidization and rediploidization, genome downsizing, repetitive DNA amplification, and DNA removal contribute to genome size evolution? Given the "large genome constraint hypothesis,"⁴⁷ do adaptive potential and extinction risk differ between species with large and small genomes?

Another important set of questions concerns heterotrophic plants that exploit fungi (mycoheterotrophy), other plants (parasitism), or animals (carnivorism) to obtain minerals and carbohydrates. Each type of heterotrophy is found in at least six orders, indicating convergent evolution (Figure 3), but the underlying genetic basis is unknown. Extensive genomic coverage of heterotrophic plants will facilitate the elucidation of evolutionary parallelism in both physiology and underlying molecular networks, which will enrich our understanding of ecological interactions between angiosperms and other multicellular organisms.

BIODIVERSITY CONSERVATION

The genomic resources made available through PLANeT will not only facilitate research on plant and trait evolution but also assist in biodiversity conservation. A major obstacle in conservation biology is the limited assessment of conservation status and extinction risk for different plants.⁴⁸ This lacuna can lead to

Figure 1. Phylogeny of 64 angiosperm orders

Bold indicates orders without public genomes as of December 31, 2024. The pie charts on the outside indicate the percentages of families with at least one reference genome in the corresponding orders. The numbered pictures denote species selected for genome sequencing and assembly in PLANeT to represent orders without reported genomes. The tree topology, based on chloroplast sequences, is modified from Angiosperm Phylogeny Group IV classification¹⁶ and Li et al.¹⁷

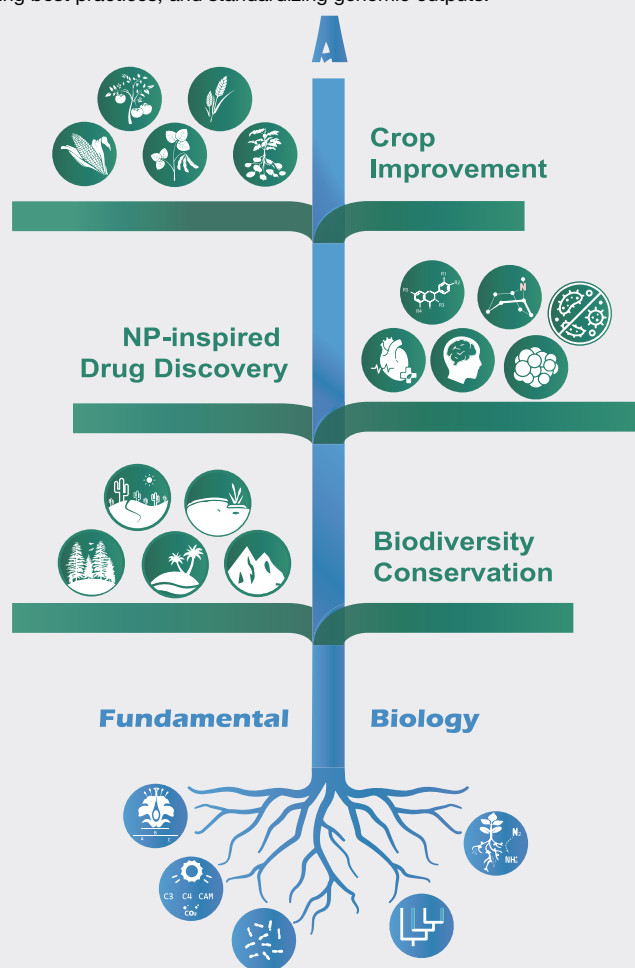
Box 1. Vignettes of PLANeT

COMPLEMENTATION TO OTHER LARGE-SCALE GENOME INITIATIVES

PLANeT builds on the success of other large-scale initiatives, which have already demonstrated the power of genomic data across phylogenetic scales to address long-standing biological questions.^{26–31} Plants are included in several other, broader sequencing efforts, such as the Earth BioGenome (EBP),¹⁰ the Darwin Tree of Life (DTOL),³² and the African Biogenome Projects.^{33–35} The latter two have, however, a regional focus and do not explicitly prioritize the targeting of phylogenetic gaps among their samples, whereas EBP is a federated effort that is not centrally supported by any funding agency in its entirety, which has limited its progress, particularly for plants. The community lacks a large-scale genome project that encompasses a substantial fraction of land plant lineages, aiming for taxonomic and geographic balance across samples. With its phylogenetically balanced sampling of all land plant lineages, PLANeT complements existing projects (Table S1). The PLANeT consortium includes colleagues from the previously launched 10,000 Plant Genomes Project,³⁶ the EBP, and DTOL to achieve shared goals and realize a shared vision of genomics for the benefit of humanity.

MISSION

Our mission is to decode conserved principles governing land plants using evolutionary and functional genomics, with profound implications for securing global food supplies and enabling a sustainable future. Our vision is to establish a genomic infrastructure that will empower research in fundamental plant biology, support biodiversity conservation, boost breeding efforts for crops, and fuel NP-inspired drug discovery, as well as enable young scientists with creative thinking and bold actions. PLANeT will serve as a central hub for ongoing projects and researchers worldwide, minimizing redundancy, disseminating best practices, and standardizing genomic outputs.



(Continued on next page)

Box 1. Continued

MILESTONES AND TIMELINES

We aim for chromosome-level reference genomes. Within 3 years, genomes will be available for >95% of flowering plant families and >95% of non-flowering plant orders. Within ca. 12 years, genomes will be available for >90% of genera of flowering plants and >90% of non-flowering plant families, and genomes will have been explored to answer fundamental questions. Within a decade, PLANeT will involve at least 100 labs, engaged in the training and/or mentoring of at least 1,000 plant scientists, to collectively generate answers to hundreds of relevant questions. These milestones and goals will be achieved by 2036, memorializing the 200-year anniversary of the successful ending of the HMS Beagle voyage.

INVITATION TO JOIN

We extend a wholehearted invitation to institutions and researchers passionate about land plant genomes—our common heritage—or enthusiastic about answering fascinating biological questions with the envisioned dataset. Please join us in making the benefits of PLANeT everlasting and beyond our collective imagination!

critical discrepancies, where some species designated as “endangered” show no signs of genetic decline, while numerous species that have not yet been assessed for their extinction risk are experiencing a loss of genetic diversity.⁴⁹ Genomic data can address this gap by providing efficient estimates of genetic diversity and historical population size, which are crucial indicators of a species’ adaptive potential and extinction risk. Reference genomes will serve as fundamental resources for biodiversity assessments, robust conservation decisions, and the development of appropriate restoration plans.^{50,51} For instance, by comparing the genomes of severely damaged and fully healthy beech trees in close proximity, loci for drought resistance have been identified, and this information has been used to select trees better adapted to drought.⁵² A collection of reference genomes of species that are extinct or have become recently extinct, and for which well-preserved specimens are available, will serve as a digital Noah’s ark for the planet that we share with all other living creatures.

CROP IMPROVEMENT

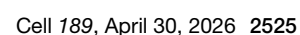
A serious challenge in crop improvement is the constrained genetic diversity accessible within domesticated populations. Many modern crop varieties suffer from domestication bottlenecks, reduced genetic diversity, and an unintended increase in genetic load.^{53–58} Although wild plants may help restore genetic diversity that was lost during crop domestication,⁵⁹ they are profoundly underrepresented among species with high-quality genome assemblies. A better understanding of the breadth of genomic diversity will enable us to harness key phenotypic traits from wild plants, particularly those relating to environmental adaptation, for crop improvement. Awareness of how different wild plants have convergently adapted to similar environmental and climatic conditions will help us identify genes and alleles that can be ported to crops.⁶⁰ Therefore, expanding our knowledge of wild plant genomes is not merely an academic pursuit but a critical step for accelerating breeding and securing a resilient food supply in the face of climate change and rapid changes in the dietary preferences of humans.

Another difficulty is knowing which genomic variants influence crop traits. Although genome-wide association studies (GWASs) have linked genetic variants to specific traits in

many species, these methods often struggle to identify the true causal variants amid hundreds of candidates.⁶¹ An evolutionary genomic perspective can help address this challenge via identifying genomic variants that have mutated or been retained at rates significantly different from those expected under neutral evolution. By pinpointing nucleotide positions that have undergone minimal changes across many millions of years of evolution, we can compile a catalog of constrained loci among distantly related species with diverse evolutionary histories (Figure 4).^{62,63} Mutations occurring at such loci in crops will often be deleterious and represent targets for enhancing fitness-related traits such as yield. The rich genomic data for popular crops, combined with a genomic perspective of distantly related taxa, function like a “microscope” (with fine-grained functional resolution) fortified with a “telescope” (offering deep evolutionary context) (Figure 4). This approach aids in identifying valuable functional variation, based on evolutionary constraints and footprints of adaptation identified from pangenomes of crops and crop wild relatives (CWRs), allowing for hypothesis testing in key crops (Figure 4). However, numerous questions remain unanswered, and these explicitly constitute the objectives of the initiative (Note S1). For instance, how can data across different phylogenomic scales, from reference genomes of distantly related plant species to population genomic data of crops and CWRs, be integrated?

NATURAL-PRODUCT-INSPIRED DRUG DISCOVERY

Beyond visible adaptive traits, wild plants are a source of myriad natural products (NPs), which are typified by massive chemical scaffold diversity and complexity, often encompassing uncommon functional groups. Plant NPs have played key roles in the discovery of important drugs,⁶⁴ such as the antimalarial medicine artemisinin from sweet wormwood (*Artemisia annua*)⁶⁵ and the first-line chemotherapeutic paclitaxel from the bark of *Taxus* trees.⁶⁶ NPs usually exhibit spatial conformations and 3D “bulky” structures,⁶⁷ often making them difficult to synthesize chemically. Other major differences compared with synthetic drugs include NPs’ higher molecular weight (>500 Da), higher hydrophilicity, and greater molecular rigidity, all of which can be advantageous for designing oral drugs that are not easily inactivated by protein-protein interactions.⁶⁷



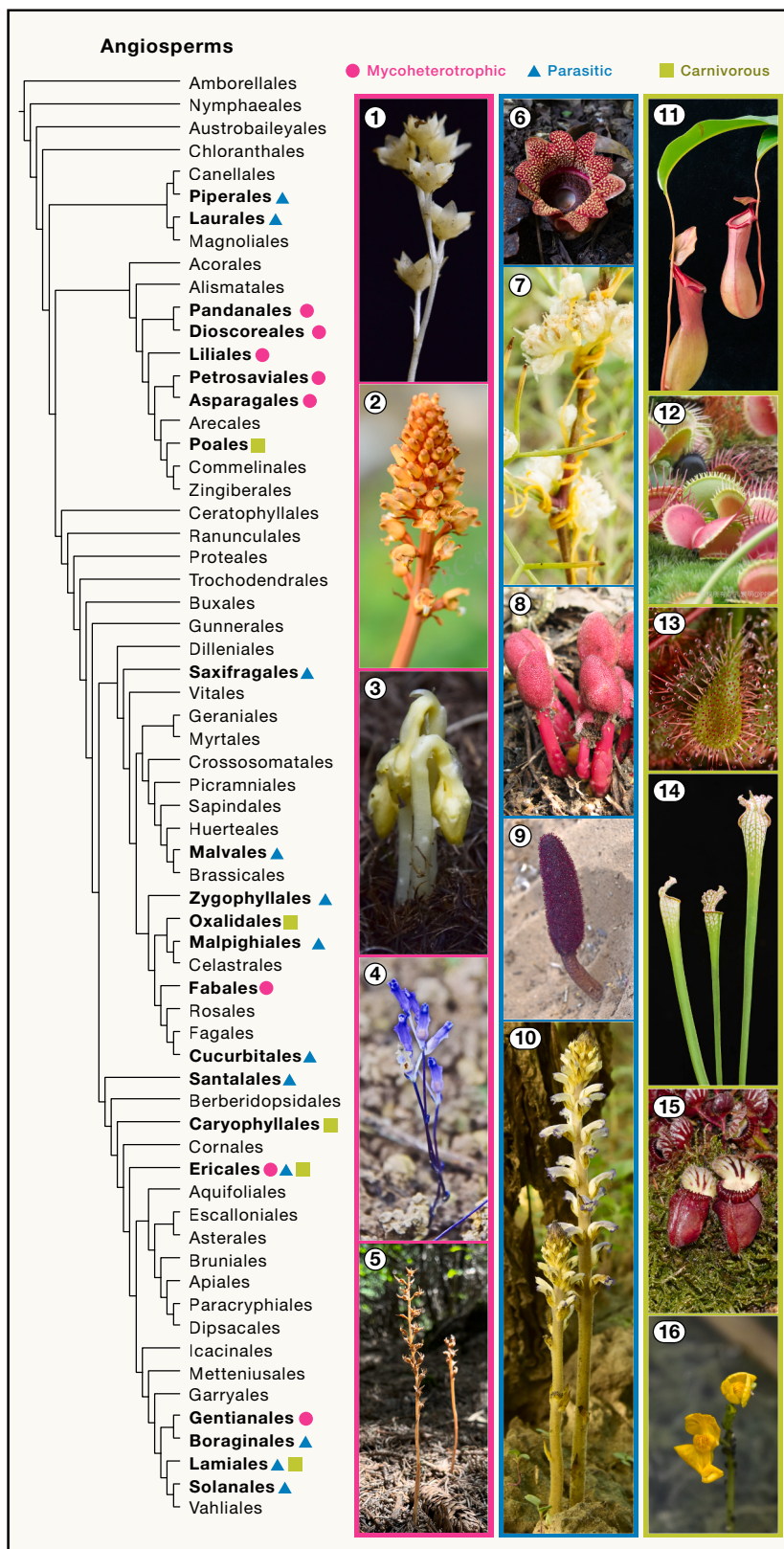


Figure 3. Distribution of heterotrophic plants across the angiosperm phylogeny

The orders in bold contain heterotrophic plants, with the types of heterotrophies indicated by the colored symbols. The numbered pictures show *Petrosavia sinii* (1), *Gastrodia elata* (2), *Hypopitys monotropa* (3), *Burmannia itoana* (4), *Neottia acuminata* (5), *Sapria himalayana* (6), *Cuscuta chinensis* (7), *Balanophora fungosa* (8), *Cynomorium songaricum* (9), *Orobancha sordida* (10), *Nepenthes mirabilis* (11), *Dionaea muscipula* (12), *Drosera dielsiana* (13), *Sarracenia leucophylla* (14), *Cephalotus follicularis* (15), and *Utricularia aurea* (16). The tree topology is adopted from Li et al.¹⁷

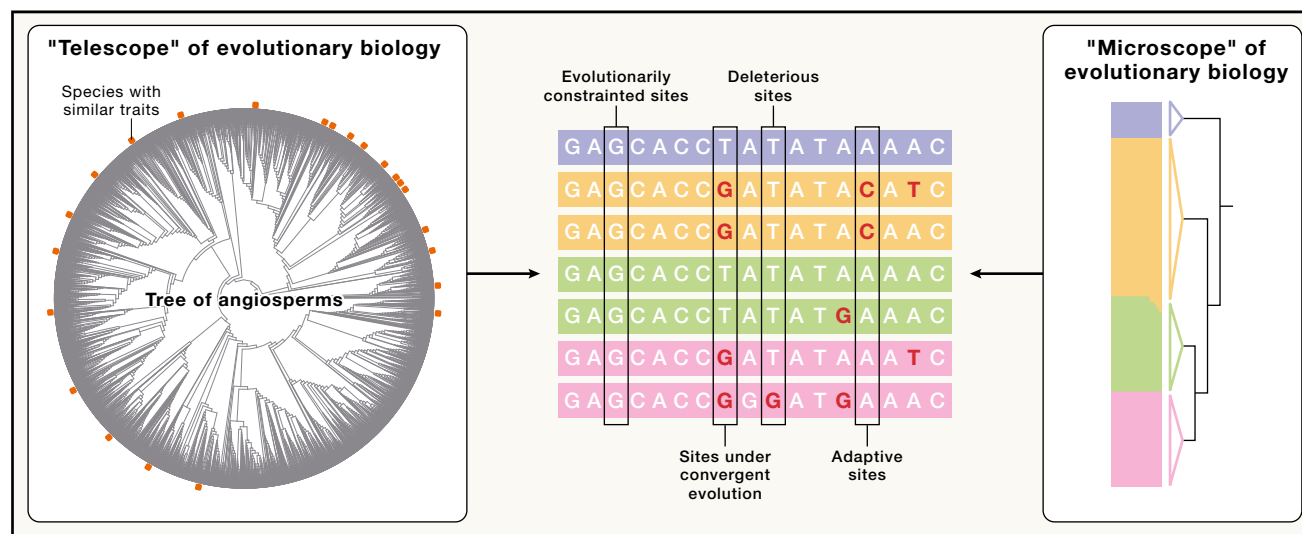


Figure 4. A diagram exemplifying how phylogenomics (the “telescope” of evolutionary biology) and pangenomes (the “microscope” of evolutionary biology) can be combined to identify constrained, adaptive, and deleterious sites in the genomes as potential targets for molecular breeding

cells instead of pure chemical synthesis. Complete reference genomes will also support the large-scale discovery of biosynthetic gene clusters (BGCs)⁶⁹ across the entire diversity of land plants.⁷⁰ Investigating genes and genomic variation underpinning metabolic traits will provide insights into the evolutionary history of chemodiversity in a wide range of taxa,⁷¹ which should inspire drug discovery programs based on the modification of bioactive compounds, ultimately empowering new drug design and metabolic engineering strategies.⁷²

AI FOR GENOMICS

The development and application of AI tools are woven into the fabric of the four areas discussed above. AI approaches hold significant potential for revealing the causes of important biological patterns, and several foundation models have demonstrated potential to decipher the complex “language” in biology, both for RNA⁷³ and DNA.^{74–76} The highly diverse collection of plant genomes generated in PLANET will serve as the basis for a foundation model that allows fine-tuning and/or zero-shot inference, flanked by task-specific models at multiple levels, including structural annotation of noncoding sequences, transcription units, and coding sequences, along with effective prediction for both coding and noncoding variants (level 1); functional annotation of DNA-RNA/protein interactions, expression patterns, and subcellular colocalization (level 2); detection of purifying and diversifying selection, neutral divergence, and introgression (level 3); and discovery of functional gene modules, circuits for key traits including NP biosynthesis, and species to be prioritized in biodiversity conservation (level 4) (Figure 5).

The feasibility of leveraging deep learning is nicely demonstrated with Helixer,⁷⁷ a fast and user-friendly gene structural annotation tool, and with a recently developed gene expression prediction model.⁷⁸ The prediction of the effects of variants using

machine learning has a long history,⁷⁹ although there are still limits to these approaches.⁸⁰ The repertoire of small proteins, also known as microproteins, is likely to be much larger than appreciated,⁸¹ and deep-learning methods are excellent at discovering these from sequence data.⁸² Machine-learning methods have also been very successful in identifying new antimicrobial peptides (AMPs) in the global microbiome,^{83,84} and such approaches can even be extended to ancient microbiomes and proteomes.^{85,86} Plants do not have specialized immune cells as animals do, but they can produce non-specific immune molecules, including AMPs, when attacked by pathogens.⁸⁷ Thus, plant genomes are very likely to be an untapped source of bioactive peptides that could be useful in many different settings.

CONCLUSIONS AND OUTLOOK

Tracing the evolutionary history of land plants with whole-genome sequences can be likened to illuminating the life of plants with the “floodlight” of genomics. By revealing both shared and unique evolutionary trajectories, we hope to begin to understand the diversity of ways in which plants can adapt to a changing environment. Much of plant biology has emphasized similarities across species. Instead, we aim to learn from the diversity of solutions that land plants have found to the same challenges. A comprehensive evolutionary perspective will contribute to a better understanding of how land plants have become so successful and help us predict the fate of natural plant populations in the face of global change. With an expansive digital library of plant genomes at our disposal, we anticipate the training and emergence of a new generation of plant genome explorers, the Darwins and Wallaces of our age.

Considering how plant diversity is distributed across the globe, PLANET will make a key contribution to uniting plant scientists from the Global South and Global North in a common

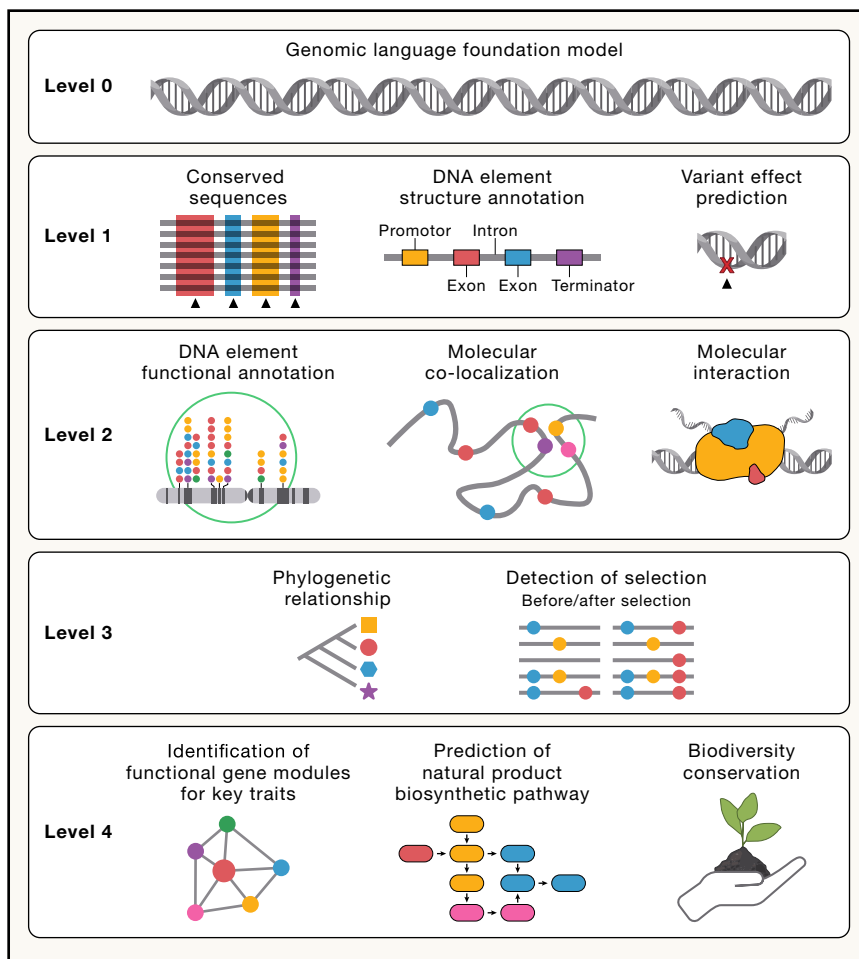


Figure 5. A diagram illustrating the implementation of AI-driven approaches in PLANeT

A genomic language foundation model will serve as the computational core, and specialized machine-learning pipelines will be developed for model fine-tuning, designed to investigate biological questions across four hierarchical scales.

cause. Without a unified international effort, research in low- and lower middle-income countries will struggle to keep pace with global advances, which, in turn, would affect the conservation and exploitation of local plant diversity and the breeding of locally adapted crops. Conversely, breeding of major crops would not live up to its full potential, as important parts of the evolutionary telescope would be missing. The integration of knowledge from wild plants, crops, and weeds will promote innovation in fundamental plant research, biodiversity conservation, agriculture in all countries, and plant-inspired drug discovery, thereby better supporting the sustainable future of our planet.

In 20 or 30 years, when we step into a forest, a desert, a tundra, or a grassland, we foresee that we will be able to use the equivalent of a tricorder, the fictional handheld sensor from Star Trek, to instantly read the genome of any species of interest. When the data are probed with sophisticated AI-based algorithms, the mysteries hidden in each species will be unlocked.

CONSORTIA

The initial members of the PLANeT consortium include Mohamed Amar, Carlos I. Arbizu, John L. Bowman, Charles H. Can-

non, David J. Cantrill, Xiaofeng Cao, Shilin Chen, Xiaoya Chen, Xuemei Chen, Kang Chong, Xingwang Deng, Weinan E, Hongwei Guo, Chuan He, Peter Hollingsworth, Qing Hu, Sanwen Huang, Hongzhi Kong, Ilia J. Leitch, Dezhu Li, Jiayang Li, Hongtao Liu, William Lucas, Eshchar Mizrahi, Mijoro Rakotoarinivo, Hai Ren, Loren Rieseberg, Jue Ruan, Rosa Scherson, Harald Schneider, Marcio C. Silva-Filho, Stephen Smith, Douglas E. Soltis, Pamela S. Soltis, Hang Sun, Alex Twyford, Björn Usadel, Yves Van de Peer, Cássio Van den Berg, Rajeev K. Varshney, Lyderson Facio Viccini, Li Wang, Qingfeng Wang, Xiao-Quan Wang, Yadong Wang, Detlef Weigel, Jonathan Wendel, Rod Wing, Gane Ka-Shu Wong, Yaowu Xing, Xun Xu, Nieng Yan, Huanming Yang, Shouzhou Zhang, and Yuxian Zhu (in alphabetical order).

ACKNOWLEDGMENTS

The concept of this consortium was initiated by the Botanical Society of China. We thank the funding support from the Shenzhen Science and Technology Program (grant no. JCYJ20241202130723030), the Agricultural Science and Technology Innovation Program (grant no. CAAS-ZDRW202601), and the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (grant no. JYB2025XDXM909). We thank Haodong

Chen, Qipian Chen, Ran Du, Qiaomei Fu, Li Guo, Lvjun Guo, Hang He, Bosheng Li, Jianpeng Ma, Baoxing Song, Hai Wang, Hongru Wang, Ren Wang, Sibo Wang, Wen Wang, Xuhui Wang, Tong Wei, Han Wen, Yaoyao Wu, Shi Xiao, Jianbin Yan, Hong Yu, Guojie Zhang, Shangwei Zhong, and Zhonghe Zhou (in alphabetical order) for contributing valuable ideas in the brainstorming meetings. We thank Bao Nie for the development of Figures 1 and 3, Zhuangwei Hou for Figures 2 and 4, and Cheng Li and Liqun Lin for Figure 5. We thank Wei Zheng, Neng Wei, Renbin Zhu, Xinxin Zhu, Guojun Hua, Chaner Zhang, Fanming Kong, Boka Li, Ang Liu, and Yechun Xu for providing the photos of *Gunnera manicata*, *Bersama abyssinica*, *Pittosporopsis kerrii*, *Gastrodia elata*, *Sapria himalayana*, *Balanophora fungosa*, *Dionaea muscipula*, *Drosera dielsiana*, *Sarracenia leucophylla*, and *Cephalotus follicularis*, respectively. The photos of *Berberidopsis corallina*, *Quintinia sieberi*, and *Vahlia capensis* were downloaded from iNaturalist, with photo IDs “252628159 (taken by islaym),” “444142503 (taken by prossington),” and “101475547 (taken by Robert Taylor),” and licensed under <https://creativecommons.org/licenses/by/4.0/>. We have not modified the content of the photos except for cropping.

AUTHOR CONTRIBUTIONS

S.H., D.W., P.S.S., K.C., X.D., and X.X. conceptualized ideas and designed the project; L.W., S.H., D.W., and P.S.S. wrote the paper; L.W. analyzed the data and coordinated the project; M.A., C.I.A., J.L.B., C.H.C., D.J.C., X. Cao, S.C., Xiaoya Chen, Xuemei Chen, K.C., X.D., W.E., H.G., C.H., P.H., Q.H., H.K., I.J.L., D.L., J.L., H.L., W.L., E.M., M.R., H.R., L.R., J.R., R.S., H. Schneider, D.E.S., H. Sun, M.C.S.-F., S.S., A.T., B.U., Y.V.d.P., C.V.d.B., R.K.V., L.F.V., Q.W., X.-Q.W., Y.W., J.W., R.W., G.K.-S.W., Y.X., X.X., N.Y., H.Y., S.Z., and Y.Z. participated in the discussions and revised the paper.

DECLARATION OF INTERESTS

D.W. holds equity in Computomics, which advises plant breeders, and has consulted for KWS SE, a globally active plant breeder and seed producer.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2026.01.026>.

REFERENCES

- Nic Lughadha, E., Bachman, S.P., and Govaerts, R. (2017). Plant states and fates: Response to Pimm and Raven. *Trends Ecol. Evol.* 32, 887–889. <https://doi.org/10.1016/j.tree.2017.09.005>.
- Govaerts, R., Nic Lughadha, E., Black, N., Turner, R., and Paton, A. (2021). The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Sci. Data* 8, 215. <https://doi.org/10.1038/s41597-021-00997-6>.
- Fernández, P., Amice, R., Bruy, D., Christenhusz, M.J.M., Leitch, I.J., Leitch, A.L., Pokorny, L., Hidalgo, O., and Pellicer, J. (2024). A 160 Gbp fern genome shatters size record for eukaryotes. *iScience* 27, 109889. <https://doi.org/10.1016/j.isci.2024.109889>.
- Bhadra, S., Leitch, I.J., and Onstein, R.E. (2023). From genome size to trait evolution during angiosperm radiation. *Trends Genet.* 39, 728–735. <https://doi.org/10.1016/j.tig.2023.07.006>.
- Folk, R.A., Siniscalchi, C.M., and Soltis, D.E. (2020). Angiosperms at the edge: Extremity, diversity, and phylogeny. *Plant Cell Environ.* 43, 2871–2893. <https://doi.org/10.1111/pce.13887>.
- Bachman, S.P., Brown, M.J.M., Leão, T.C.C., Nic Lughadha, E., and Walker, B.E. (2024). Extinction risk predictions for the world's flowering plants to support their conservation. *New Phytol.* 242, 797–808. <https://doi.org/10.1111/nph.19592>.
- Humphreys, A.M., Govaerts, R., Ficinski, S.Z., Nic Lughadha, E., and Vorontsova, M.S. (2019). Global dataset shows geography and life form predict modern plant extinction and rediscovery. *Nat. Ecol. Evol.* 3, 1043–1047. <https://doi.org/10.1038/s41559-019-0906-2>.
- Gomes, V.H.F., Vieira, I.C.G., Salomão, R.P., and ter Steege, H. (2019). Amazonian tree species threatened by deforestation and climate change. *Nat. Clim. Chang.* 9, 547–553. <https://doi.org/10.1038/s41558-019-0500-2>.
- Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. <https://doi.org/10.1038/35048692>.
- Lewin, H.A., Robinson, G.E., Kress, W.J., Baker, W.J., Coddington, J., Crandall, K.A., Durbin, R., Edwards, S.V., Forest, F., Gilbert, M.T.P., et al. (2018). Earth BioGenome Project: Sequencing life for the future of life. *Proc. Natl. Acad. Sci. USA* 115, 4325–4333. <https://doi.org/10.1073/pnas.1720115115>.
- Kress, W.J., Soltis, D.E., Kersey, P.J., Wegrzyn, J.L., Leebens-Mack, J.H., Gostel, M.R., Liu, X., and Soltis, P.S. (2022). Green plant genomes: What we know in an era of rapidly expanding opportunities. *Proc. Natl. Acad. Sci. USA* 119, e2115640118. <https://doi.org/10.1073/pnas.2115640118>.
- Xie, L., Gong, X., Yang, K., Huang, Y., Zhang, S., Shen, L., Sun, Y., Wu, D., Ye, C., Zhu, Q.-H., et al. (2024). Technology-enabled great leap in deciphering plant genomes. *Nat. Plants* 10, 551–566. <https://doi.org/10.1038/s41477-024-01655-6>.
- Sessa, E.B., and Der, J.P. (2016). Evolutionary genomics of ferns and lycophytes. In *Advances in Botanical Research*, S.A. Rensing, ed. (Academic Press), pp. 215–254. <https://doi.org/10.1016/bs.abr.2016.02.001>.
- Schafraan, P., Hauser, D.A., Nelson, J.M., Xu, X., Mueller, L.A., Kulshrestha, S., Smalley, I., de Vries, S., Irisarri, I., de Vries, J., et al. (2025). Pan-phylum genomes of hornworts reveal conserved autosomes but dynamic accessory and sex chromosomes. *Nat. Plants* 11, 49–62. <https://doi.org/10.1038/s41477-024-01883-w>.
- Marks, R.A., Hotaling, S., Frandsen, P.B., and VanBuren, R. (2021). Representation and participation across 20 years of plant genome sequencing. *Nat. Plants* 7, 1571–1578. <https://doi.org/10.1038/s41477-021-01031-8>.
- The Angiosperm Phylogeny Group (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* 181, 1–20. <https://doi.org/10.1111/boj.12385>.
- Li, H.-T., Luo, Y., Gan, L., Ma, P.-F., Gao, L.-M., Yang, J.-B., Cai, J., Gitzenanner, M.A., Fritsch, P.W., Zhang, T., et al. (2021). Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biol.* 19, 232. <https://doi.org/10.1186/s12915-021-01166-2>.
- Novakovsky, G., Dexter, N., Libbrecht, M.W., Wasserman, W.W., and Mostafavi, S. (2023). Obtaining genetics insights from deep learning via explainable artificial intelligence. *Nat. Rev. Genet.* 24, 125–137. <https://doi.org/10.1038/s41576-022-00532-2>.
- Avsec, Ž., Latysheva, N., Cheng, J., Novati, G., Taylor, K.R., Ward, T., Bycroft, C., Nicolaisen, L., Arvaniti, E., Pan, J., et al. (2026). Advancing regulatory variant effect prediction with AlphaGenome. *Nature* 649, 1206–1218. <https://doi.org/10.1038/s41586-025-10014-0>.
- Sasse, A., Chikina, M., and Mostafavi, S. (2024). Unlocking gene regulation with sequence-to-function models. *Nat. Methods* 21, 1374–1377. <https://doi.org/10.1038/s41592-024-02331-5>.
- Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P., Deac, A., et al. (2023). Scientific discovery in the age of artificial intelligence. *Nature* 620, 47–60. <https://doi.org/10.1038/s41586-023-06221-2>.
- Pavlopoulos, G.A., Baltoumas, F.A., Liu, S., Selvitopi, O., Camargo, A.P., Nayfach, S., Azad, A., Roux, S., Call, L., Ivanova, N.N., et al. (2023). Unraveling the functional dark matter through global metagenomics. *Nature* 622, 594–602. <https://doi.org/10.1038/s41586-023-06583-7>.
- Hou, X., He, Y., Fang, P., Mei, S.-Q., Xu, Z., Wu, W.-C., Tian, J.-H., Zhang, S., Zeng, Z.-Y., Gou, Q.-Y., et al. (2024). Using artificial intelligence to

- document the hidden RNA virosphere. *Cell* 187, 6929–6942.e16. <https://doi.org/10.1016/j.cell.2024.09.027>.
24. de Boer, C.G., and Taipale, J. (2024). Hold out the genome: A roadmap to solving the cis-regulatory code. *Nature* 625, 41–50. <https://doi.org/10.1038/s41586-023-06661-w>.
 25. Richards, S. (2015). It's more than stamp collecting: how genome sequencing can unify biological research. *Trends Genet.* 31, 411–421. <https://doi.org/10.1016/j.tig.2015.04.007>.
 26. Zoonomia Consortium (2020). A comparative genomics multitool for scientific discovery and conservation. *Nature* 587, 240–245. <https://doi.org/10.1038/s41586-020-2876-6>.
 27. Feng, S., Stiller, J., Deng, Y., Armstrong, J., Fang, Q., Reeve, A.H., Xie, D., Chen, G., Guo, C., Faircloth, B.C., et al. (2020). Dense sampling of bird diversity increases power of comparative genomics. *Nature* 587, 252–257. <https://doi.org/10.1038/s41586-020-2873-9>.
 28. Stiller, J., Feng, S., Chowdhury, A.-A., Rivas-González, I., Duchêne, D.A., Fang, Q., Deng, Y., Kozlov, A., Stamatakis, A., Claramunt, S., et al. (2024). Complexity of avian evolution revealed by family-level genomes. *Nature* 629, 851–860. <https://doi.org/10.1038/s41586-024-07323-1>.
 29. Rhie, A., McCarthy, S.A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., et al. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 592, 737–746. <https://doi.org/10.1038/s41586-021-03451-0>.
 30. Shao, Y., Zhou, L., Li, F., Zhao, L., Zhang, B.-L., Shao, F., Chen, J.-W., Chen, C.-Y., Bi, X., Zhuang, X.-L., et al. (2023). Phylogenomic analyses provide insights into primate evolution. *Science* 380, 913–924. <https://doi.org/10.1126/science.abn6919>.
 31. One Thousand Plant Transcriptomes Initiative (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679–685. <https://doi.org/10.1038/s41586-019-1693-2>.
 32. Darwin Tree of Life Project Consortium (2022). Sequence locally, think globally: The Darwin Tree of Life Project. *Proc. Natl. Acad. Sci. USA* 119, e2115642118. <https://doi.org/10.1073/pnas.2115642118>.
 33. Ebenezer, T.E., Muigai, A.W.T., Nouala, S., Badaoui, B., Blaxter, M., Budie, A.G., Jarvis, E.D., Korlach, J., Kuja, J.O., Lewin, H.A., et al. (2022). Africa: sequence 100,000 species to safeguard biodiversity. *Nature* 603, 388–392. <https://doi.org/10.1038/d41586-022-00712-4>.
 34. Sharaf, A., Nesengani, L.T., Hayah, I., Kuja, J.O., Mdyogolo, S., Omotoriogun, T.C., Odogwu, B.A., Beedesse, G., Smith, R.M., Barakat, A., et al. (2024). Establishing African genomics and bioinformatics programs through annual regional workshops. *Nat. Genet.* 56, 1556–1565. <https://doi.org/10.1038/s41588-024-01807-6>.
 35. Hayah, I., Ezebiro, V., Kagame, S.P., Kuja, J.O., Waruhiu, C., Nesengani, L.T., Mdyogolo, S., Molotsi, A.H., Abechi, P., Abushady, A.M., et al. (2025). Unlocking the African bioeconomy and strengthening biodiversity conservation through genomics and bioinformatics. *NPJ Biodivers.* 4, 29. <https://doi.org/10.1038/s44185-025-00102-9>.
 36. Cheng, S., Melkonian, M., Smith, S.A., Brockington, S., Archibald, J.M., Delaux, P.-M., Li, F.-W., Melkonian, B., Mavrodiev, E.V., Sun, W., et al. (2018). 10KP: A phylodiverse genome sequencing plan. *GigaScience* 7, 1–9. <https://doi.org/10.1093/gigascience/giy013>.
 37. Zhang, G., and Ma, H. (2024). Nuclear phylogenomics of angiosperms and insights into their relationships and evolution. *J. Integr. Plant Biol.* 66, 546–578. <https://doi.org/10.1111/jipb.13609>.
 38. Guo, X., Fang, D., Sahu, S.K., Yang, S., Guang, X., Folk, R., Smith, S.A., Chanderbali, A.S., Chen, S., Liu, M., et al. (2021). *Chloranthus* genome provides insights into the early diversification of angiosperms. *Nat. Commun.* 12, 6930. <https://doi.org/10.1038/s41467-021-26922-4>.
 39. Ma, J., Sun, P., Wang, D., Wang, Z., Yang, J., Li, Y., Mu, W., Xu, R., Wu, Y., Dong, C., et al. (2021). The *Chloranthus sessilifolius* genome provides insight into early diversification of angiosperms. *Nat. Commun.* 12, 6929. <https://doi.org/10.1038/s41467-021-26931-3>.
 40. Bennett, M.D. (1972). Nuclear DNA content and minimum generation time in herbaceous plants. *Proc. R. Soc. Lond., B* 181, 109–135. <https://doi.org/10.1098/rspb.1972.0042>.
 41. Doyle, J.J., and Coate, J.E. (2019). Polyploidy, the nucleotype, and novelty: The impact of genome doubling on the biology of the cell. *Int. J. Plant Sci.* 180, 1–52. <https://doi.org/10.1086/700636>.
 42. Bureš, P., Elliott, T.L., Veselý, P., Šmarda, P., Forest, F., Leitch, I.J., Nic Lughadha, E., Soto Gomez, M., Pironon, S., Brown, M.J.M., et al. (2024). The global distribution of angiosperm genome size is shaped by climate. *New Phytol.* 242, 744–759. <https://doi.org/10.1111/nph.19544>.
 43. Soto Gomez, M., Brown, M.J.M., Pironon, S., Bureš, P., Verde Arregoitia, L.D., Veselý, P., Elliott, T.L., Zedek, F., Pellicer, J., Forest, F., et al. (2024). Genome size is positively correlated with extinction risk in herbaceous angiosperms. *New Phytol.* 243, 2470–2485. <https://doi.org/10.1111/nph.19947>.
 44. Roddy, A.B., Thérout-Rancourt, G., Abbo, T., Benedetti, J.W., Brodersen, C.R., Castro, M., Castro, S., Gilbride, A.B., Jensen, B., Jiang, G.-F., et al. (2020). The scaling of genome size and cell size limits maximum rates of photosynthesis with implications for ecological strategies. *Int. J. Plant Sci.* 181, 75–87. <https://doi.org/10.1086/706186>.
 45. Thérout-Rancourt, G., Roddy, A.B., Earles, J.M., Gilbert, M.E., Zwieniecki, M.A., Boyce, C.K., Tholen, D., McElrone, A.J., Simonin, K.A., and Brodersen, C.R. (2021). Maximum CO₂ diffusion inside leaves is limited by the scaling of cell size and genome size. *Proc. Biol. Sci.* 288, 20203145. <https://doi.org/10.1098/rspb.2020.3145>.
 46. Faizullah, L., Morton, J.A., Hersch-Green, E.I., Walczyk, A.M., Leitch, A.R., and Leitch, I.J. (2021). Exploring environmental selection on genome size in angiosperms. *Trends Plant Sci.* 26, 1039–1049. <https://doi.org/10.1016/j.tplants.2021.06.001>.
 47. Knight, C.A., Molinari, N.A., and Petrov, D.A. (2005). The large genome constraint hypothesis: Evolution, ecology and phenotype. *Ann. Bot.* 95, 177–190. <https://doi.org/10.1093/aob/mci011>.
 48. Forest, F., Brown, R., Buerki, S., Colville, J.F., Moat, J., Lughadha, E.N., Owen, N.R., Raimondo, D.C., Rivers, M., Rosindell, J., et al. (2025). 300 million years of angiosperm evolution at risk of extinction. Preprint at bioRxiv. <https://doi.org/10.1101/2025.1104.1129.651190>.
 49. Finn, C., Grattarola, F., and Pincheira-Donoso, D. (2023). More losers than winners: Investigating Anthropocene defaunation through the diversity of population trends. *Biol. Rev. Camb. Philos. Soc.* 98, 1732–1748. <https://doi.org/10.1111/brv.12974>.
 50. Theissinger, K., Fernandes, C., Formenti, G., Bista, I., Berg, P.R., Bleidorn, C., Bombarely, A., Crottini, A., Gallo, G.R., Godoy, J.A., et al. (2023). How genomics can help biodiversity conservation. *Trends Genet.* 39, 545–559. <https://doi.org/10.1016/j.tig.2023.01.005>.
 51. De León, L.F., Silva, B., Avilés-Rodríguez, K.J., and Buitrago-Rosas, D. (2023). Harnessing the omics revolution to address the global biodiversity crisis. *Curr. Opin. Biotechnol.* 80, 102901. <https://doi.org/10.1016/j.copbio.2023.102901>.
 52. Pfenninger, M., Reuss, F., Kiebler, A., Schönnenbeck, P., Caliendo, C., Gerber, S., Cocchiara, B., Reuter, S., Blüthgen, N., Mody, K., et al. (2021). Genomic basis for drought resistance in European beech forests threatened by climate change. *eLife* 10, e65532. <https://doi.org/10.7554/eLife.65532>.
 53. Morrell, P.L., Buckler, E.S., and Ross-Ibarra, J. (2011). Crop genomics: advances and applications. *Nat. Rev. Genet.* 13, 85–96. <https://doi.org/10.1038/nrg3097>.
 54. Meyer, R.S., and Purugganan, M.D. (2013). Evolution of crop species: Genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840–852. <https://doi.org/10.1038/nrg3605>.
 55. Lye, Z., Choi, J.Y., and Purugganan, M.D. (2022). Deleterious mutations and the rare allele burden on rice gene expression. *Mol. Biol. Evol.* 39, msac193. <https://doi.org/10.1093/molbev/msac193>.

56. Wang, L., Beissinger, T.M., Lorient, A., Ross-Ibarra, C., Ross-Ibarra, J., and Hufford, M.B. (2017). The interplay of demography and selection during maize domestication and expansion. *Genome Biol.* 18, 215. <https://doi.org/10.1186/s13059-017-1346-4>.
57. Zhang, C., Wang, P., Tang, D., Yang, Z., Lu, F., Qi, J., Tawari, N.R., Shang, Y., Li, C., and Huang, S. (2019). The genetic basis of inbreeding depression in potato. *Nat. Genet.* 51, 374–378. <https://doi.org/10.1038/s41588-018-0319-1>.
58. Lozano, R., Gazave, E., Dos Santos, J.P.R., Stetter, M.G., Valluru, R., Bandidillo, N., Fernandes, S.B., Brown, P.J., Shakoor, N., Mockler, T.C., et al. (2021). Comparative evolutionary genetics of deleterious load in sorghum and maize. *Nat. Plants* 7, 17–24. <https://doi.org/10.1038/s41477-020-00834-5>.
59. Bohra, A., Kilian, B., Sivasankar, S., Caccamo, M., Mba, C., McCouch, S.R., and Varshney, R.K. (2022). Reap the crop wild relatives for breeding future crops. *Trends Biotechnol.* 40, 412–431. <https://doi.org/10.1016/j.tibtech.2021.08.009>.
60. Zhang, F., and Batley, J. (2020). Exploring the application of wild species for crop improvement in a changing climate. *Curr. Opin. Plant Biol.* 56, 218–222. <https://doi.org/10.1016/j.cpb.2019.12.013>.
61. Sullivan, P.F., Meadows, J.R.S., Gazal, S., Phan, B.N., Li, X., Genereux, D.P., Dong, M.X., Bianchi, M., Andrews, G., Sakthikumar, S., et al. (2023). Leveraging base-pair mammalian constraint to understand genetic variation and human disease. *Science* 380, eabn2937. <https://doi.org/10.1126/science.abn2937>.
62. Christmas, M.J., Kaplow, I.M., Genereux, D.P., Dong, M.X., Hughes, G.M., Li, X., Sullivan, P.F., Hindle, A.G., Andrews, G., Armstrong, J.C., et al. (2023). Evolutionary constraint and innovation across hundreds of placental mammals. *Science* 380, eabn3943. <https://doi.org/10.1126/science.abn3943>.
63. Wu, Y., Li, D., Hu, Y., Li, H., Ramstein, G.P., Zhou, S., Zhang, X., Bao, Z., Zhang, Y., Song, B., et al. (2023). Phylogenomic discovery of deleterious mutations facilitates hybrid potato breeding. *Cell* 186, 2313–2328.e15. <https://doi.org/10.1016/j.cell.2023.04.008>.
64. Atanasov, A.G., Zotchev, S.B., Dirsch, V.M., and International Natural Product Sciences Taskforce, and Supuran, C.T. (2021). Natural products in drug discovery: advances and opportunities. *Nat. Rev. Drug Discov.* 20, 200–216. <https://doi.org/10.1038/s41573-020-00114-z>.
65. Tu, Y. (2016). Artemisinin—A gift from traditional chinese medicine to the world (Nobel Lecture). *Angew. Chem. Int. Ed.* 55, 10210–10226. <https://doi.org/10.1002/anie.201601967>.
66. Wall, M.E., and Wani, M.C. (1995). Camptothecin and taxol: Discovery to clinic—thirteenth Bruce F. Cain Memorial Award Lecture. *Cancer Res.* 55, 753–760.
67. Tzvetkov, N.T., Kirilov, K., Matin, M., and Atanasov, A.G. (2024). Natural product drug discovery and drug design: two approaches shaping new pharmaceutical development. *Nephrol. Dial. Transplant.* 39, 375–378. <https://doi.org/10.1093/ndt/gfad208>.
68. Jiang, B., Gao, L., Wang, H., Sun, Y., Zhang, X., Ke, H., Liu, S., Ma, P., Liao, Q., Wang, Y., et al. (2024). Characterization and heterologous reconstitution of *Taxus* biosynthetic enzymes leading to baccatin III. *Science* 383, 622–629. <https://doi.org/10.1126/science.adj3484>.
69. Jeon, J.E., Kim, J.G., Fischer, C.R., Mehta, N., Dufour-Schroif, C., Wemmer, K., Mudgett, M.B., and Sattely, E. (2020). A pathogen-responsive gene cluster for highly modified fatty acids in tomato. *Cell* 180, 176–187.e19. <https://doi.org/10.1016/j.cell.2019.11.037>.
70. Lichman, B.R., Godden, G.T., Hamilton, J.P., Palmer, L., Kamileen, M.O., Zhao, D.Y., Vaillancourt, B., Wood, J.C., Sun, M., Kinser, T.J., et al. (2020). The evolutionary origins of the cat attractant nepetalactone in catnip. *Sci. Adv.* 6, eaba0721. <https://doi.org/10.1126/sciadv.aba0721>.
71. Hu, J., Qiu, S., Wang, F., Li, Q., Xiang, C.-L., Di, P., Wu, Z., Jiang, R., Li, J., Zeng, Z., et al. (2023). Functional divergence of CYP76AKs shapes the chemodiversity of abietane-type diterpenoids in genus *Salvia*. *Nat. Commun.* 14, 4696. <https://doi.org/10.1038/s41467-023-40401-y>.
72. Jacobowitz, J.R., and Weng, J.K. (2020). Exploring uncharted territories of plant specialized metabolism in the postgenomic era. *Annu. Rev. Plant Biol.* 71, 631–658. <https://doi.org/10.1146/annurev-arplant-081519-035634>.
73. Yu, H., Yang, H., Sun, W., Yan, Z., Yang, X., Zhang, H., Ding, Y., and Li, K. (2024). An interpretable RNA foundation model for exploring functional RNA motifs in plants. *Nat. Mach. Intell.* 6, 1616–1625. <https://doi.org/10.1038/s42256-024-00946-z>.
74. Mendoza-Revilla, J., Trop, E., Gonzalez, L., Roller, M., Dalla-Torre, H., de Almeida, B.P., Richard, G., Caton, J., Lopez Carranza, N., Skwark, M., et al. (2024). A foundational large language model for edible plant genomes. *Commun. Biol.* 7, 835. <https://doi.org/10.1038/s42003-024-06465-2>.
75. Zhai, J., Gokaslan, A., Schiff, Y., Berthel, A., Liu, Z.-Y., Lai, W.-Y., Miller, Z.R., Scheben, A., Stitzer, M.C., Romay, M.C., et al. (2025). Cross-species modeling of plant genomes at single-nucleotide resolution using a pretrained DNA language model. *Proc. Natl. Acad. Sci. USA* 122, e2421738122. <https://doi.org/10.1073/pnas.2421738122>.
76. Brix, G., Durrant, M.G., Ku, J., Naghipourfar, M., Poli, M., Sun, G., Brockman, G., Chang, D., Fanton, A., Gonzalez, G.A., et al. (2026). Genome modeling and design across all domains of life with Evo2. *Nature*. <https://doi.org/10.1038/s41586-026-10176-5>.
77. Stiehler, F., Steinborn, M., Scholz, S., Dey, D., Weber, A.P.M., and Denton, A.K. (2021). Helix: Cross-species gene annotation of large eukaryotic genomes using deep learning. *Bioinformatics* 36, 5291–5298. <https://doi.org/10.1093/bioinformatics/btaa1044>.
78. Peleke, F.F., Zumkeller, S.M., Gültas, M., Schmitt, A., and Szymański, J. (2024). Deep learning the cis-regulatory code for gene expression in selected model plants. *Nat. Commun.* 15, 3488. <https://doi.org/10.1038/s41467-024-47744-0>.
79. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249. <https://doi.org/10.1038/nmeth0410-248>.
80. Tang, Z., Toneyan, S., and Koo, P.K. (2023). Current approaches to genomic deep learning struggle to fully capture human genetic variation. *Nat. Genet.* 55, 2021–2022. <https://doi.org/10.1038/s41588-023-01517-5>.
81. Chen, J., Brunner, A.-D., Cogan, J.Z., Nuñez, J.K., Fields, A.P., Adamson, B., Itzhak, D.N., Li, J.Y., Mann, M., Leonetti, M.D., et al. (2020). Pervasive functional translation of noncanonical human open reading frames. *Science* 367, 1140–1146. <https://doi.org/10.1126/science.aay0262>.
82. Miller, B., de Souza, E.V., Pai, V.J., Kim, H., Vaughan, J.M., Lau, C.J., Dieckrich, J.K., and Saghatelian, A. (2025). ShortStop: A machine learning framework for microprotein discovery. *BMC Methods* 2, 16. <https://doi.org/10.1186/s44330-025-00037-4>.
83. Santos-Júnior, C.D., Torres, M.D.T., Duan, Y., Rodríguez del Río, Á., Schmidt, T.S.B., Chong, H., Fullam, A., Kuhn, M., Zhu, C., Houseman, A., et al. (2024). Discovery of antimicrobial peptides in the global microbiome with machine learning. *Cell* 187, 3761–3778.e16. <https://doi.org/10.1016/j.cell.2024.05.013>.
84. Torres, M.D.T., Brooks, E.F., Cesaro, A., Sberro, H., Gill, M.O., Nicolaou, C., Bhatt, A.S., and de la Fuente-Nunez, C. (2024). Mining human microbiomes reveals an untapped source of peptide antibiotics. *Cell* 187, 5453–5467.e15. <https://doi.org/10.1016/j.cell.2024.07.027>.
85. Wan, F., Torres, M.D.T., Peng, J., and de la Fuente-Nunez, C. (2024). Deep-learning-enabled antibiotic discovery through molecular de-extinction. *Nat. Biomed. Eng.* 8, 854–871. <https://doi.org/10.1038/s41551-024-01201-x>.

86. Klapper, M., Hübner, A., Ibrahim, A., Wasmuth, I., Borry, M., Haensch, V.G., Zhang, S., Al-Jammal, W.K., Suma, H., Fellows Yates, J.A., et al. (2023). Natural products from reconstructed bacterial genomes of the Middle and Upper Paleolithic. *Science* 380, 619–624. <https://doi.org/10.1126/science.adf5300>.
87. Zou, F., Tan, C., Shinali, T.S., Zhang, B., Zhang, L., Han, Z., and Shang, N. (2023). Plant antimicrobial peptides: a comprehensive review of their classification, production, mode of action, functions, applications, and challenges. *Food Funct.* 14, 5492–5515. <https://doi.org/10.1039/D3FO01119D>.