Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Plant Biology

# Editorial overview: Technology development as a driver of biological discovery

Josh T Cuperus and Christine Queitsch

**Current Opinion in Plant Biology** 2020, **54**:A1–A4

For a complete overview see the Issue

https://doi.org/10.1016/j.pbi.2020.06.001



## Josh T Cuperus and Christine Queitsch

Josh T Cuperus conducted his doctoral work with James Carrington on small RNAs at Oregon State University. He joined Stanley Fields at the University of Washington for his postdoctoral work, developing synthetic biology assays and massively parallel reporter assays to determine sequence function. As a faculty member in the Department of Genome Sciences at the University of Washington, he focuses with Dr. Queitsch on the relationship between genome architecture, chromatin accessibility, and gene expression. Dr. Cuperus aims to manipulate gene expression at single-cell resolution in order to create crops that are able to thrive in the marginal environments of the near future.

Christine Queitsch grew up in the former East Germany and obtained a Fulbright Fellowship to study the function of the chaperone Hsp90 in plant development and evolution with Susan Lindquist at the University of Chicago. After a Bauer Fellowship at Harvard University, she joined the faculty of the Department of Genome Sciences at the University of Washington, where, jointly with Dr. Cuperus, she continues to pursue her life-long fascination with the translation of genotype into phenotype.

Here, we present a series of reviews that are connected by a common thread: technology development enables biological discovery. Affordable short-read sequencing has revolutionized plant genomics, giving access to diverse genomes, including those of most major crops and their most commonly grown cultivars, and evolutionarily distant non-model species. We highlight in this issue the ways in which short-read sequencing has driven biological discovery. Nevertheless, the technology falls literally short when it comes to assembling the repetitive, transposon-rich genome sequences that dominate large crop genomes. Thus, we start with reviews that discuss long-read sequencing and its applications in plants.

Michael VanBuren describe how long-read technologies, such as from PacBio and Oxford Nanopore, combine with physical mapping approaches and computational advances to allow chromosome-scale assemblies. Haplotype phasing and resolution of structural variants, in particular in polyploid and heterozygous species, as well as *de novo* pangenomics, are described as emerging frontiers. The authors posit that we have entered the golden age of genome assembly for crops and non-model species alike, as the cost of long-read sequencing drops precipitously.

Danilevicz and co-authors take up Michael and VanBuren's call for pan-genome assembly with long-read sequencing. They further discuss the need for deep learning approaches in stitching these genomes together and interpreting their content. Pangenomes with their broad delineation of a crop's genomic landscape will enable increased use of natural variation in designer crop development, providing a data-driven template for genome editing.

Shahid and Slotkin propose that the field most radically altered by long-read sequencing is the biology of transposable elements. These elements can occupy up to 90% of plant genomes. They have the capacity to dramatically expand or restructure genomes, in addition to remodeling gene expression and chromatin landscapes. The authors describe how using long-read sequencing will reveal the effects of transposons on local gene expression, chromosomal rearrangements and the epigenome, through the use of carefully selected plant populations, including those bred by Barbara McClintock.

Transposons also play a starring role in the review by Alger and Edger, which focuses on subgenome dominance. In alloploid species, one of the parental genomes typically shows higher levels of gene expression and, ultimately, greater gene retention. The authors use a beautiful and simple cartoon to illustrate the interplay of transposon abundance and their epigenetic

modifications with their ensuing effect on gene expression in subgenome dominance. They highlight future areas of research, such as the impact of nuclear organization — subgenomes are likely organized separately — and the impact of environmental change that may render submissive subgenomes dominant, presumably by altering transposon silencing or subgenome nuclear organization.

With Pontvianne and Liu's review on chromatin domains, the discussion moves from genome assembly to genome organization and function. Mammalian genomes form self-organizing, largely insulated chromatin domains known as topologically associated domains; they also form domains associated with lamin fibers at the nuclear periphery and domains associated with the nucleolar periphery. Although chromatin domains are found in plants, the authors point out that these seem not fully equivalent to those found in animals. In addition to reviewing commonly used technologies to assess 3D genome organization, including those relying on short-read sequencing, the authors discuss the functional implications of spatial chromatin domain organization for gene regulation and replication timing. They identify liquid–liquid phase separation as a likely crucial process in spatially arranging chromatin domains and speculate about the role of intrinsically disordered proteins in this process.

In contrast to our relative naïveté about the impact of the 3D genome on gene regulation, our understanding of gene regulation in the context of the linear genome sequence, that is, chromatin accessibility and transcription factor binding sites, has become quite sophisticated in the past decade. As Bubb and Deal describe, this understanding has come in large part from robust genome-scale methods to assess chromatin accessibility, enabled by affordable short-read sequencing. As these methods are increasingly applied to diverse crops to identify tissue-specific and condition-specific regulatory elements, the authors provide carefully considered guidelines on sample preparation, sequencing, read mapping, and analysis, focusing in particular on methods for peak calling and motif analysis. This review is essential and timely reading; as the field moves toward assessing single cells, we need to maximize the power of these approaches and minimize their pitfalls.

Diving head-first into plant single-cell genomics, McFaline-Figuerao and co-authors take on the task of summarizing recent studies of single-cell transcriptomes of *A. thaliana* roots. By combining the root data that have been published, the authors characterize rare cell types and intermediate cell states, identifying insufficient cell numbers as a limitation. Another limitation in applying this approach to crops is the scarcity of cell type annotations. The authors review existing technologies to capture single cells, including recent advances that allow for a far greater number of cells to be profiled. Their detailed discussion of workflow and existing computational analysis pipelines is of particular value, as plant single-cell genomics is still an emerging field. The authors clarify the enormous promise single cell genomics holds for understanding plant development, tissue-specific responses to stress, and, ultimately, targeted plant engineering and breeding. They point to the need for future technology advances to enable co-assays for different classes of RNAs or regulatory landscapes and gene expression. These co-assays promise to shed light on the interplay of small RNAs and mRNAs, the transcriptional responses to endoreplication, and the surprising stasis of plant regulatory landscapes in bulk studies.

Endoreplication, the process in which a genome is repeatedly replicated in the absence of mitosis, leading to polyploidy, is pervasive in plants and animals. Lang and Schnittger put to rest the assumption that endoreplication directly promotes growth; rather, it appears to be a self-enhancement program that facilitates the pre-programmed developmental fate of cells in which it occurs. The authors discuss endoreplication as a possible strategy to overcome stress such as drought and DNA damage. At the molecular level, endoreplication appears to increase transcription, in particular of cell wall and ribosomal RNA genes, the latter hinting at increased translation. At first glance, these findings appear to contradict recent single-cell genomics findings in several systems, including *A. thaliana* root epidermal cells, which show decreased overall transcription in older cells, which are more likely to be endoreplicated. Combining the assessment of endoreplication state with single-cell measures of chromatin accessibility and gene expression will be informative for resolving this seeming contradiction and further exploring endoreplication's impact on cell states.

Jones and Vandepoele review our community's still insufficient efforts to integrate chromatin accessibility, gene expression, and transcription factor binding sites into robust gene regulatory networks. Comparing gene regulatory networks across tissues, in development, in response to stress and among species has the potential to identify important network nodes for future manipulation. The authors discuss the challenges inherent in applying this promising strategy, namely, the fast divergence of regulatory elements, the scarcity of unique motifs for non-model transcription factors or factors belonging to large families, and the notoriously weak correlation of chromatin accessibility and gene expression. In addition, they offer solutions such as the need to annotate enhancers and the generation of single-cell data for chromatin accessibility and gene expression. Their discussion of transcription factor evolution highlights gene duplication coupled with changes in expression timing and location (i.e. neofunctionalization and

subfunctionalization) as a major factor in diversifying and expanding gene regulatory networks.

Parcy and colleagues take up the thread of transcription factor evolution by reviewing how massive short-read sequencing and comparative genomics have allowed for a rigorous investigation of how transcription factors families have driven plant diversification from charophyte algae to angiosperms. The authors discuss progress on transcription factor family reconstruction and identification of distantly related transcription factors, contrasting evolutionary trajectories of different transcription factor families and illustrating how conserved transcription factors can adopt diverse roles. Provocatively, the authors identify changes in transcription factor oligomerization state and protein–protein interaction specificity as possibly crucial events in the neofunctionalization and subfunctionalization of transcription factors after duplication.

Turner-Hissong and colleagues take us from transcription factor evolution to crop domestication, describing how insights from evolutionary biology will allow for deeper understanding of the genetic architecture and short-term evolution of complex traits in crops. The authors argue convincingly how accounting for the diverse life histories of crops and their ancient and recent polyploidy will inform our understanding of crop variation and their potential for improvement through breeding. Breeding occurs on a relatively short time scale, typically drawing on standing variation rather than *de novo* mutations, necessitating bottlenecks that impose 'domestication costs' by reducing genetic variation. Another consideration is the nature of the selective sweeps involved in fixing traits. The authors suggest soft sweeps are likely more important when breeding for complex traits such as yield, albeit they are harder to detect with common genotype-phenotype association approaches. According to the authors, the future of breeding lies in *de novo* editing of crop genomes as well as in precision breeding, requiring significant advances in population and complex trait genetics.

Wang and colleagues take on the central challenge posed by the explosion of available crop genomes: multi-dimensional genome-wide molecular phenotypes and organismal phenotypes. They discuss how this information can be interpreted systematically to improve crops. Their answer lies in the application of deep learning approaches in two key areas, the first focusing on modeling information flow from genome sequence to phenotype, and the second on identifying functional (i.e. beneficial) variants in natural populations. Beyond natural variations, the authors outline the potential of deep learning methods to design synthetic genomic elements with beneficial functions for editing-based improvement of future crops. For the uninitiated, computationally naive among us, a particular value of this review lies in its meticulous review of concepts, tools and limitations in deep learning approaches.

Genome editing has become the magic bullet in the tool box of plant geneticists and crop breeders. In their comprehensive and thoughtful piece, Atkins and Voytas deliver another must read by outlining the substantial obstacles for precise and efficient genome editing in plants. They identify inefficiency in creating the desired DNA modification and ineffective delivery of gene editing reagents as the most serious bottlenecks and discuss recent advances in plants and other systems to resolve them.

With readily available, efficient genome editing in crops not yet here, Baxter is concerned about our inability to pick good candidate genes to edit in the future. He writes that while we have amassed vast collections of genomic and phenotype data, we miss the tools to organize, integrate, and translate this knowledge into causal genes. He sees solutions in field-based phenotyping combined with association and linkage studies, and he calls for genomic selections aided by artificial intelligence. He urges us to avoid confirmation bias at all levels, asking that we improve annotations across all plant species in order to understand and manipulate crops.

Kliebenstein answers Baxter's challenge by describing how biological networks can reduce the dimensionality inherent in today's imFengmensely complex data sets. Combining biological networks with natural variation data can determine network nodes that are present or absent across phenotyped cultivars or species, thereby facilitating the identification of candidate genes for editing. Kliebenstein's review is a fitting capstone to the technology-heavy arc of this series.

Plants are wondrous creatures to those who study them and to the many who grow them for food or pleasure. Among the many unique features of plants compared to their animal brethren, the diversity of plant reproductive strategies may be the most stunning one, which is amply demonstrated in the last two reviews.

Guanqiao Feng and co-authors dig deeply into the evolution of dioecy and sex determination in plants. With increased access to genomes and transcriptomes, the authors find evidence for conservation in sex-biased gene expression across evolutionarily distant plants. In animals, sex determination pathways show conservation in a bottom-up manner. In plants, as the authors argue, there is evidence for both bottom-up and top-down conservation, which may have contributed to the stunning diversity in their reproductive strategies.

Chow, Chakraborty and Mosher discuss our changing understanding of RNA-directed DNA (RdDM) in

reproduction. In the model plant *A. thaliana*, RdDM mediates the balance between maternal and paternal contributions to the endosperm. However, studies in non-flowering plants that lack endosperm reveal an ancestral pathway with a broad role in sexual reproduction.

We hope that these reviews illustrate the promise of technology development to drive biological discovery and ensure future food and energy security. By 2050, the demand for agricultural products will double due to our exploding needs for food, animal feed, and environmentally sustainable biofuels. This increased demand, coupled with rapid loss of arable land and unpredictable weather patterns, calls for vast investments in plant research. As outlined by several authors here, precision breeding and genome editing of crops, enabled by cell-type specific knowledge of transcription and regulatory elements, are promising paths.

The past decade has seen plant research fall behind the development and application of technology in human and animal research. This slippage is particularly regrettable, as some of the 'hot' and tech-heavy fields such as epigenetics and transposon biology were pioneered in plants. A major and obvious factor in this decline has been the lack of funding. Another, more recent factor has been the reluctance to fund advanced genomics research in simple non-crop model plants. The value of piloting new technologies in simple well-annotated models has become clear once again in the single-cell genomics era, which has drawn heavily on animal cell fate maps derived decades earlier. These pilot studies provided the blueprint for translational single-cell genomics applied to human disorders, including most recently COVD-19 infections.

Tackling the vexing question of how genotype translates into phenotype in different environments requires all hands on deck and all tools in the toolbox. We have made a strong argument for advancing the toolbox of plant research, but the 'hands on deck' are probably even more important for future success. Of the Ph.D. students who graduated from our lab since 2014, fewer than half continued to pursue plant research. This loss of talent to other scientific disciplines is alarming. Young successful researchers will flock into fields that offer opportunities for impactful discoveries and solid career prospects. As a community, we need to raise public awareness that our shared future critically depends on recruiting the best young scientists as much as it does on developing and applying advanced technology.