

## Editorial

# Next-generation sequencing technologies: opportunities and obligations in plant genomics

The year 2003 marked the completion of the Human Genome Project. In the ~9 years since then, genomics has become a vital tool for biomedical research and a driver for improved human health. An often ignored component of human health is plant-derived human nutrition. Plant and agricultural genomics have benefited from many of the same drivers leading technical advances in the development and application in human genomics. The most disruptive technological advance has been a doubling of sequencing data output on an average of every 5 months and has resulted in a free-fall in cost per DNA base sequenced [1]. One recalls when it was acceptable to submit, review and publish RNA-sequencing manuscripts in prestigious scientific journals with zero biological or technical replicates because the cost was prohibitive. We soon arrive at the point where it requires less resource to re-sequence the genome or repeat the sequence data generation or expression profiling experiment than to simply store and manage the data. The articles assembled for this Special Issue address current trends and applications along with opportunities and challenges in the use of next-generation DNA sequencing (NGS) technologies in plant genomics research.

Established NGS technology platforms continue to evolve and have enabled the current state of plant functional genomics. In this issue, Thudi *et al.*, reviews the status of existing next- or second-generation sequencing technologies as well as the development and potential plant genomics applications of third- and fourth-generation sequencing technologies. Commensurate with the evolution of data generation has been the rapid advancement of database and informatics and assembly tools to manage and analyze NGS data. Reviews by both Lee *et al.* and Li *et al.*, describe the development of computational tools designed to address the

challenges of rapidly analyzing and assembling massive sequence data sets. NGS technologies are now widely used to characterize plant genomes and transcriptomes. Genome analysis is often directed toward determining species- or population-level genetic diversity, and in many cases, toward leveraging this diversity for crop improvement. Benjamin Kilian and Andreas Graner provide a synopsis of recent advancements and challenges in the application of NGS technologies to the analysis of germplasm diversity in plant genetic resource collections. The article by Robert Henry describes the use of resequencing to better understand sequence diversity and complexity associated with the domestication of crop species. Genetic variation at the DNA sequence level may evoke changes in gene expression resulting in phenotypic variation. Transcriptome analysis continues to benefit from NGS broad adoption and technical advance. Unfathomable sequence data costs can also be a curse to scientists as thoughts toward experimental design possibly begin to diminish with little experiment cost. Auer *et al.*, describe the complexities of and review many of the key experimental design challenges common to RNA sequencing experiments, while, Mukesh Jain reviews the current use of NGS technologies for gene expression profiling and the informatics challenges inherent in data analysis. Sequence analysis is not limited to the RNA species destined for translation. Small RNAs play critical roles in plant gene regulation, development and response to the environment. In this issue, David Studholme describes the challenges and opportunities in the informatics analysis of massive, relatively short sequence read data sets inherent to small RNA experiments.

According to the Food and Agriculture Organization of the United Nations, the demand for plant-based products will double by the first half of this century. Food and fiber production will

increasingly compete with bioenergy crops for resources. In addition, the world will face environmental and social challenges by way of reductions in arable land and water resources along with the trend toward fewer farmers.

It has been stated that the world's current capacity to generate DNA sequence data is 13 terabases annually. But to what end? With the addition of each genome to our current stamp collection in this age of the '\$1,000 genome', we should ensure that we direct the majority of this capacity toward hypothesis-driven research.

The biology of the plant genome is multi-dimensional, but has until now been interrogated using two-dimensional approaches. The advent, rapid evolution and adoption of genomics technologies enable researchers to address fundamental questions about genome biology, heterosis and the heritability of complex agronomic traits [2].

As scientists in the field of plant and agricultural genomics, we have a moral obligation to identify keys to unlocking increased productivity in

crop species. Let's trust that we will fulfill this obligation. In 2050, 9 billion people will be counting on it.

Rajeev K. Varshney

*International Crops Research Institute for the Semi-Arid*

*Topics (ICRISAT)*

*Patancheru, AP, India*

*CGIAR Generation Challenge Program*

*c/o CIMMYT, Mexico DF, Mexico*

Gregory D. May

*National Center for Genome Resources*

*Santa Fe, NM, USA*

## References

1. Stein LD. The case for cloud computing in genome informatics. *Genome Biol* 2010;**11**:207.
2. Cook DR, Varshney RK. From genome studies to agricultural biotechnology: closing the gap between basic plant science and applied agriculture. *Curr Opin Plant Biol* 2010; **13**:115–8.