Current Plant Biology 6 (2016) 2–10

Contents lists available at ScienceDirect

Current Plant Biology

journal homepage: www.elsevier.com/locate/cpb

Advances in genomics for adapting crops to climate change☆

Armin Scheben, Yuxuan Yuan, David Edwards*

University of Western Australia, Australia

ARTICLE INFO

Article history:
Received 27 June 2016
Received in revised form 9 September 2016
Accepted 9 September 2016

Keywords:
Breeding
Climate change
Crop improvement
Genomics

ABSTRACT

Climate change is a major threat to food security in a world of rising crop demand. Although increases in crop production have previously been achieved through the use of fertilisers and chemicals for better control of weeds and pests, these methods rely on finite resources and are often unsustainable. Recent advances in genomics are laying the foundations for sustainable intensification of agriculture and heightened resilience of crops to climate change. The number of available high-quality reference genomes has been constantly growing due to the widespread application of genome sequencing technology. Advances in population-level genotyping have further contributed to a more comprehensive understanding of genomic variation. These increasing volumes of genomic data facilitate the move towards plant pangenomics, providing deeper insights into the diversity available for crop improvement and breeding of new cultivars. Genomics-assisted breeding is benefiting from these advances, allowing rapid identification of genes implicated in climate related agronomic traits, for breeding of crops adapted to a changing climate.

Contents

1. Introduction

Producing sufficient food to feed the rising global population is a huge challenge for agriculture, especially under the threat of unpredictable consequences of climate change [1,2]. Climate change may alter weather patterns, rainfall regimes, temperature and carbon dioxide concentrations in particular regions [3,4]. These changes can lead to increased abiotic stress in crops, increased incidence of pests and pathogens, and an overall reduction in crop yield. During recent decades, increased crop production has been mainly achieved through refining agronomic management and breeding improved crop varieties [5]. However, maintaining a continued increase of crop yield using these methods to ensure food secu-

☆ This article is part of a special issue entitled "Plants and global climate change: a need for sustainable agriculture", published in the journal Current Plant Biology 6, 2016.

* Corresponding author.
E-mail address: dave.edwards@uwa.edu.au (D. Edwards).

http://dx.doi.org/10.1016/j.cpb.2016.09.001

rity is unsustainable, as most of them rely on finite resource such as phosphorus or nitrogenous fertiliser and there is little room for further optimisation [1,5]. Genomics-assisted breeding is considered to have the greatest potential for overcoming these challenges and ensuring a sustainable increase of food production by adapting available crops to biotic and abiotic stresses and breeding novel crop varieties [1,4].

Reference crop genome sequences are the basis of crop genetic and genomic studies, as they provide insights into gene content, genomic variation and the genetic basis for agronomic traits [5,6]. Since use of genome sequencing technologies has become more widespread, an increasing number of plant genomes have been assembled, including crops and wild crop relatives [7]. This has shown that unlike most animal genomes, plant genomes are often large, highly repetitive and polyploid [8]. A major challenge in genome assembly using the prevailing short read sequencing methods is the difficulty of reconstructing repetitive regions in the plant genome [9]. The increasing use of long read sequencing and optical mapping aims to overcome this issue and improve plant genome assemblies.

As more genome sequence information becomes available, an emerging consensus is that the genomic information contained in a single crop individual does not accurately represent the diversity of the species [4]. Population-level genotyping has provided opportunities to identify the widespread genomic variation within species [6]. The study of crop pangenomes, which aim to accurately represent the genomic diversity within a species, has also contributed to greater knowledge of within-species diversity in crops [10]. With high quality genome assemblies, accurate characterisation of genomic diversity, and precise association of heritable agronomic traits and genotypes, crop yield stability and environmental resilience will be improved [1]. Furthermore, genome editing approaches hold great promise for engineering climate-adapted crops and accelerating breeding [11]. Building on the increasing amount of genomic data and advances in genome editing, genomics-assisted breeding will play an important role in ensuring food security in a changing climate.

2. Genome sequencing and assemblies

Since the completion of the first human draft genome in 2001, the study of other species using genome sequencing technologies has been growing rapidly. Sanger sequencing, the first generation of sequencing technology, has been used to assemble several plant genomes [12]. Despite the long read length and high assembly accuracy, the low throughput and high cost have limited the widespread adoption of Sanger sequencing for genome assembly [13]. Second generation sequencing (SGS) technologies such as Illumina are faster, with higher throughput and lower cost, and have become dominant [14]. According to the National Center for Biotechnology Information (NCBI), there are currently over 100 plant reference genome sequences publicly available, the majority of which were assembled using SGS data. However, due to the short read length produced by SGS, misassembles in the long repetitive regions and gaps in the assemblies are common [9]. Depending on genome complexity and sequencing depth and quality, SGS can also lead to short contig length and thus low N50. This can compromise the quality of gene predictions, as genes may be split across contigs causing inflation of gene numbers [15]. Misassembles and split genes in assemblies are an important consideration for downstream analyses such as pangenomics and genome diversity analysis. Recently, long read sequencing and optical mapping have provided new approaches to increase contig length, reconstruct repetitive regions and fill the gaps in genome assemblies.

2.1. Long read sequencing

In contrast to short read sequencing, the reads produced by long read sequencing can be several thousand bases long, and can thus span complex and repetitive regions. The use of long sequence reads in transcriptomic studies can facilitate identification of the connectivity of exons and discern gene isoforms by spanning entire mRNA transcripts [14]. Currently, the available long read sequencing methods are single molecule based and short read synthetized long read sequencing technologies.

Pacific Biosciences (PacBio) single molecule real time sequencing and Oxford Nanopore MinION sequencing are the major single molecule based long read sequencing technologies, producing long sequencing reads in real-time. PacBio and Oxford Nanopore MinION sequencing steps are PCR-free, eliminating PCR amplification biases [16]. First commercially used in 2011, the PacBio RS II platform can now produce single molecule reads up to 60 Kb, with an average read length over 10 Kb [17,18]. However, error rates are high (13%–18%), particularly due to many indel errors [19,20]. Formation of recombinant, or chimeric, reads during library preparation may also be a pitfall of PacBio sequencing, though increasing coverage or applying appropriate quality control algorithms can decrease chimera frequency [21]. To lower error rates, different algorithms have been developed, for instance PacBio Corrected Reads [22], the hierarchical genome-assembly process [23] and the MinHash Alignment Process [19]. After read correction, the accuracy can be increased up to ~99.99% [19]. The Oxford Nanopore MinION was first made available in 2014 [24]. It can sequence DNA fragments longer than 100 Kb [25]. However, high indel error rates (~15%) also occur in Oxford Nanopore reads [19,24]. To algorithmically address this error rate, different methods have also been developed for nanopore data [24], including NanoCorr [26], NanoPolish [27], PoreSeq [28] and marginAlign [29].

Illumina synthetic long read sequencing and 10X Genomics GemCode technology are short read synthetized long read sequencing technologies. Illumina synthetic long read sequencing relies on TruSeq library preparation to construct synthetic long reads from short read sequencing reads generated by its HiSeq platform [30–32]. 10X Genomice GemCode technology uses microfluidic techniques to partition long DNA molecules into oil-encased droplets that are then barcoded [33]. Using Illumina short-read sequencing, a novel algorithm is applied to link the sequenced reads to their original molecules and construct contiguous DNA fragments [33,34].

2.2. Optical mapping

Optical mapping is a type physical mapping, which uses the physical locations of restriction enzyme sites to produce maps that can improve genome assemblies. First reported in the early 1990s [35], optical mapping is currently dominated by the BioNano Irys and OpGen Argus platforms. The average length of the BioNano single molecule maps produced is around 225 Kb [36], while the optical maps generated by OpGen span 200 Kb on average.

Using the overlap-layout-consensus paradigm, *de novo* assemblies are implemented to construct consensus optical maps [37]. By aligning consensus maps to the digested reference sequence assemblies, optical mapping identifies assembly errors including false joins, false inversions, and translocation errors. The results are then visualised using analysis tools such as BioNano IrysView and OpGen MapSolver. In addition, optical mapping can efficiently correct the gap size in the assemblies [38] and anchor scaffolds in the assembly to form super scaffolds [39]. Optical mapping has been applied to assist the genome assembly of the plants *Amborella tri-*