A chromosomal genomics approach to assess and validate the *desi* and *kabuli* draft chickpea genome assemblies

Pradeep Ruperao[1,2,3], Philipp E. Bayer[1,2,], Chon-Kit Kenneth Chan[1], Sarwar Azam[3], Miroslava Karafiátová[4], Satomi Hayashi[1], Jana Čížková[4], Rachit K Saxena[3], Hana Šimková[4], Chi Song[5], Jan Vrána[4], Annapurna Chitikineni[3], Paul Visendi[1,2], Pooran M Gaur[3], Teresa Millán[6], Karam B. Singh[7,8], Bunyamin Taran[9], Jun Wang[5], Jacqueline Batley[1], Jaroslav Doležel[4] , Rajeev K. Varshney[3] , David Edwards[1,2]

[1]University of Queensland, St. Lucia, Queensland, Australia. [2]Australian Centre for Plant Functional Genomics, University of Queensland, St. Lucia, Queensland, Australia. [3]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, Andhra Pradesh, India. [4]Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelů 31, 783 71 Olomouc-Holice, Czech Republic. [5]Beijing Genomics Institute (BGI) - Shenzhen, China. [6]Department of Genetics, University of Cordoba, Cordoba, Spain. [7]The University of Western Australia Institute of Agriculture, The University of Western Australia, Crawley, Australia. [8]CSIRO Plant Industry, Private Bag 5, Wembley, WA 6913, Australia. [9]Crop Development Centre, Plant Sciences, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

Abstract

The ability to isolate individual chromosomes combined with next generation sequencing permits the validation of genome assemblies at the chromosome level. We demonstrate this approach by the assessment of the recently published chickpea *kabuli* (Varshney et al 2013) and *desi* (Jain et al 2013) genomes. While previous genetic analysis suggests that these genomes should be very similar, a comparison of their published assemblies highlights significant differences. Our chromosomal genomics analysis highlights regions that appear to have been mis-assembled in the *kabuli* genome and identifies large scale mis-assembly in the draft *desi* genome. The integration of chromosomal genomics with skimGBS based genome assembly tools has the potential to significantly improve genome assemblies. The approach could be applied both for new genome assemblies as well as published assemblies, and complements currently applied genome assembly strategies.

Aims

▪To assess and validate the chickpea *desi* and *kabuli* reference genome assemblies using chromosomal genomics and skim based genotyping by sequencing

Results

The *desi* genome assembly is more fragmented and the pseudomolecules are much smaller than for the *kabuli* assembly, and pairwise comparison of each of the pseudomolecules from the two assemblies revealed numerous structural variations (Figure 1). Illumina DNA sequencing and mapping of reads from isolated *kabuli* chromosomes identified 46 mis-assembled regions in *kabuli* reference genome (Figure 2). Mapping *kabuli* and *desi* isolated chromosome sequence reads, together with whole genome *desi* Illumina data to the *desi* draft assembly highlighted major errors in the assembly (Figure 3), including whole misplaced pseudomolecules (*desi* Ca3 includes both Ca3 and Ca8), and regions of the *desi* reference pseudomolecules that appear not to be composed of sequence from the chickpea genome (purple colour on Figure 3).

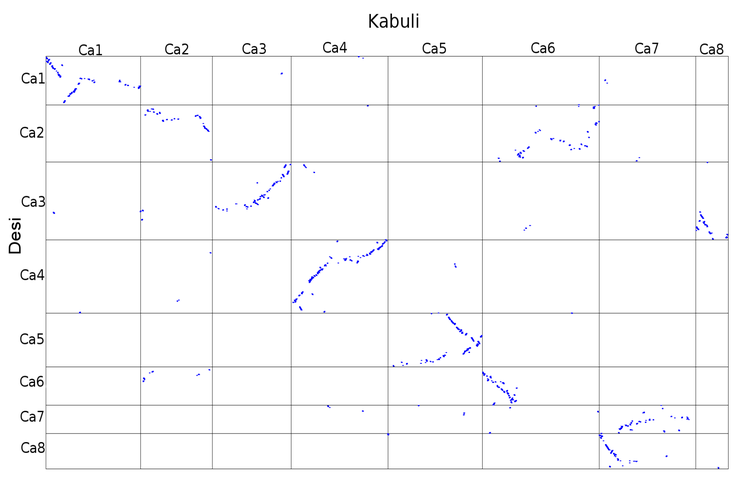

Figure 1. Comparison of the *desi* and *kabuli* reference genomes.

Using SGSautoSNP (Lorenc et al. 2012) we predicted 84,963 single nucleotide polymorphic positions between two chickpea parental varieties ICC 1882 and ICC 4958. SkimGBS was performed to genotype 46 F2 individuals to produce an imputed genotype map (Figure 4). The haplotype map highlights the mis-assembly regions of the *kabuli* genome and assists in locating these regions in their proper pseudomolecule positions. This GBS data also assists in characterising recombination and gene conversion frequencies across the genome and for precise trait mapping.

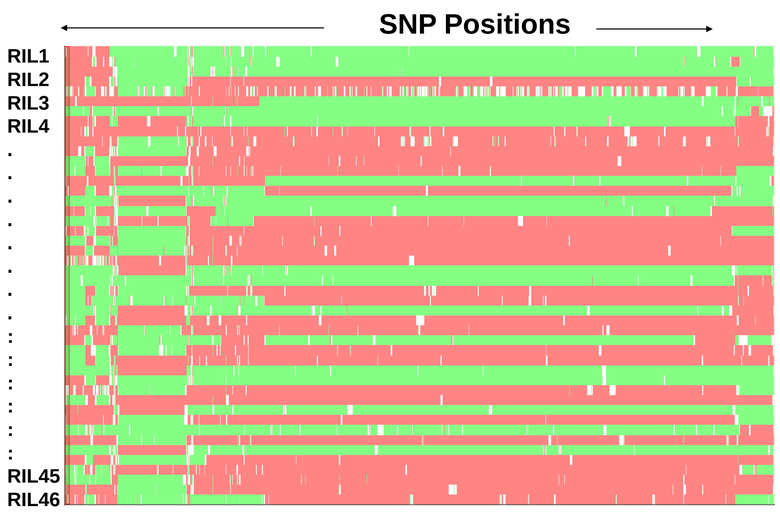

Figure 4. Imputed haplotype blocks from the ICC1882, ICC4958 F2 population showing mis-assembled regions.

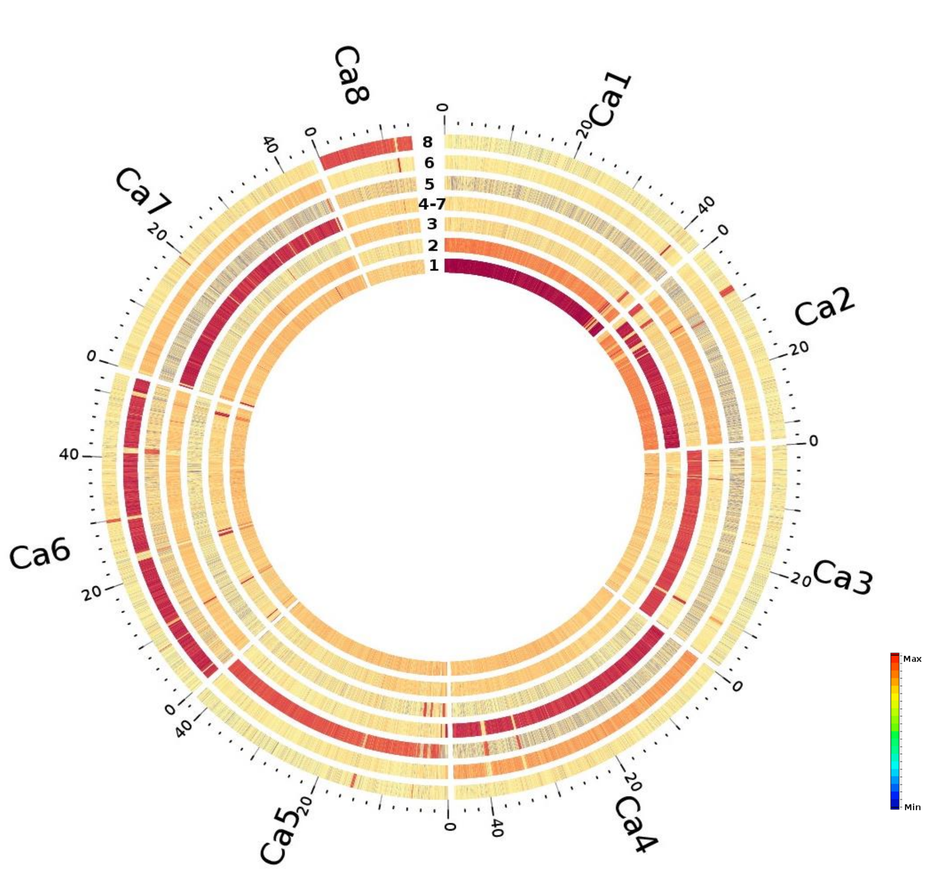

Figure 2. Map of the *kabuli* reference genome demonstrating the density of Illumina paired sequence reads (red colour) from isolated *kabuli* chromosomes 1,2,3, (4,7), 5, 6 and 8

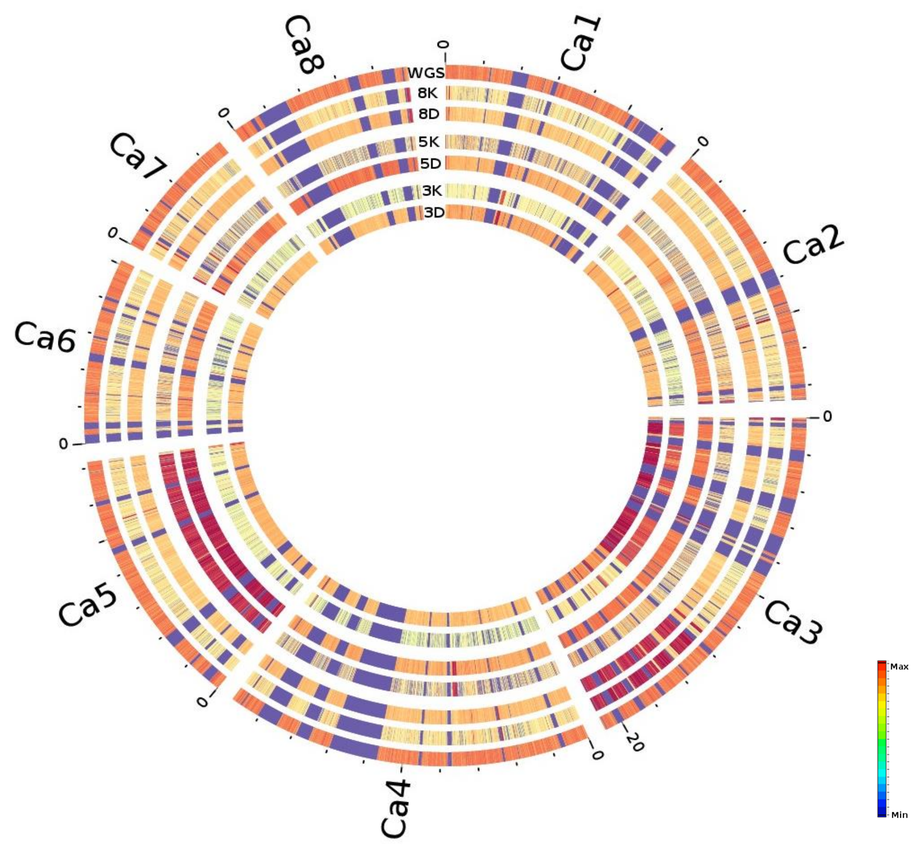

Figure 3. Map of the *desi* reference genome demonstrating the density of sequence reads (red colour) from isolated *kabuli* (K) and *desi* (D) chromosomes 3, 5 and 8, as well as whole genome data (WGS). Purple regions indicate no specific read mapping and hence regions in the assembly which might not reflect chickpea genome sequence.

Conclusion

We have established a chromosomal genomics approach to validate and compare reference genome assemblies. Overall, the assembly quality of the *kabuli* genome is high, with relatively few regions in the reference pseudomolecules which appear to have been misassembled into the wrong pseudomolecule. In addition to chromosomal genomics, skim based GBS can be applied to validate and improve the accurate assembly of the chickpea genomes.

References and acknowledgements

1. Varshney et al: Draft genome sequence of kabuli chickpea (*Cicer arietinum*): genetic structure and breeding constraints for crop improvement. Nature Biotechnology (2013) 31: 240-246
2. Jain et al: A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). The Plant Journal (2013) 74: 715-729
3. Lorenc et al: Discovery of single nucleotide polymorphisms in complex genomes using SGSautoSNP. Biology (2012) 1: 370-382

The authors would like to acknowledge funding support from the Australian Research Council (Projects LP0882095, LP0883462, LP110100200 and DP0985953) and the AISRF grand challenge fund GCF010013. This work was supported AGRF, QCIF, APAC.