Assessing and validating the amphidiploid genome of *Brassica napus* using Genotyping By Sequencing

Philipp Bayer
Applied Bioinformatics Group
University of Western Australia &
University of Queensland

Outline

- Background: *Brassica*
- Method: Skim-based genotyping by sequencing
- *B. napus* cv. Darmor genome assembly improvement using SkimGBS
- Takeaways

Skim GBS

- Determine SNPs by sequencing parents and running SGSautoSNP
- Low coverage skim sequence segregating population
- Map reads to the reference genome
- Call genotype where reads cover previously defined SNP
- Impute and clean to define haplotype blocks

Genotype calling

- We use SGSautoSNP (Lorenc *et al.* 2012), calls SNPs using read alignments (doesn’t allow heterozygous SNPs in cultivars)
- Script compares called genotypes by comparing SNPs in parents to positions in offspring
- Cleaning-step removes individuals with >90% empty alleles, removes SNPs with > 70% empty alleles
Skim-based GbyS II

- "Sideways" imputation fills up genotype between adjacent SNPs, simple rules:
  - T, Missing, Missing, T -> T, t, t, T
  - T, Missing, Missing, N -> T, Missing, Missing, N
- Some individuals exhibit large amount of recombinations, probably unwarranted gene-flow, removed these

Problem: Quality of reference

- Two possibilities:
  - Translocations between reference cultivar and Tapidor x Ningyou
  - Misplaced or misassembled contigs
- Can now identify misassembled contigs!

GBS in Darmor

- Both parental individuals high coverage (~50x)
- 92 double haploid Tapidor x Ningyou individuals, average coverage 1.6x
- Aligned using SOAPaligner (unique only)
- Called SNPs using SGSautoSNP (Lorenc et al. 2012)
Called alleles in the DH population

contigPlacer
- Novel algorithm: contigPlacer
- Uses tagging SNPs per contig and compares genotype patterns with placed contigs, places unplaced contigs

Improvement of Darmor genome

Darmor genome
- Chalhoub et al.: “Early allopolyploid evolution in the post-neolithic Brassica napus oilseed genome”, Science 2014
- Expected size ~ 1,100.0 Mbp
- 10 A-chromosomes, 9 C-chromosomes, 20 sets of unplaced contigs
  - Assembled size: 850.29 Mbp
  - 645.95 Mbp placed (75.8%), 204.33 Mbp unplaced contigs

Darmor improvement
- GBS: 1,006,985 Tapidor/Ningyou SNPs, using Darmor genome as reference
- After contigPlacer: 798.95 Mbp placed, 51.33 Mbp unplaced
- From 75.8% placed contigs to 93.9%
  - 98.5% of unplaced contigs with initial chromosome assignment on same chromosome!

Takeaways
- SkimGBS fast and cheap genotyping method
- Resulting data starting point to find misplaced/misassembled contigs in genome assembly
- Improved B. napus genomes about to be published
- Using SkimGBS it is possible to place unplaced contigs using contigPlacer thereby improving already assembled genomes
Acknowledgements

Pradeep Ruperao
Kenneth Chan
Paula Martinez
Michal Lorenc
Kaitao Lai
Agneszka Golic
Paul Vaand
Jenny Huay-Ting Lee
Juan D. Montenegro
Bhavna Hurgobin
Dave Edwards

Contact:
Philipp.bayer@uq.edu.au

Jacqueline Batley
Annalessa S. Mason
Alice Hayward
Emma Campbell
Reece Tollenaere
Salman Alamery
Jessica Dalton-Morgan
Satomi Hayashi
Harsh Raman
Yan Long
Jingjing Meng
Isobel Parkin
Bart Lambert
Benjamin Laga

Questions?

Source: http://www.botanicalgarden.ubc.ca/potd/brassica-napus1.jpg