Assembly and validation of a draft genome of *Brassica napus* using skim genotyping by sequencing

David Edwards

University of Queensland

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

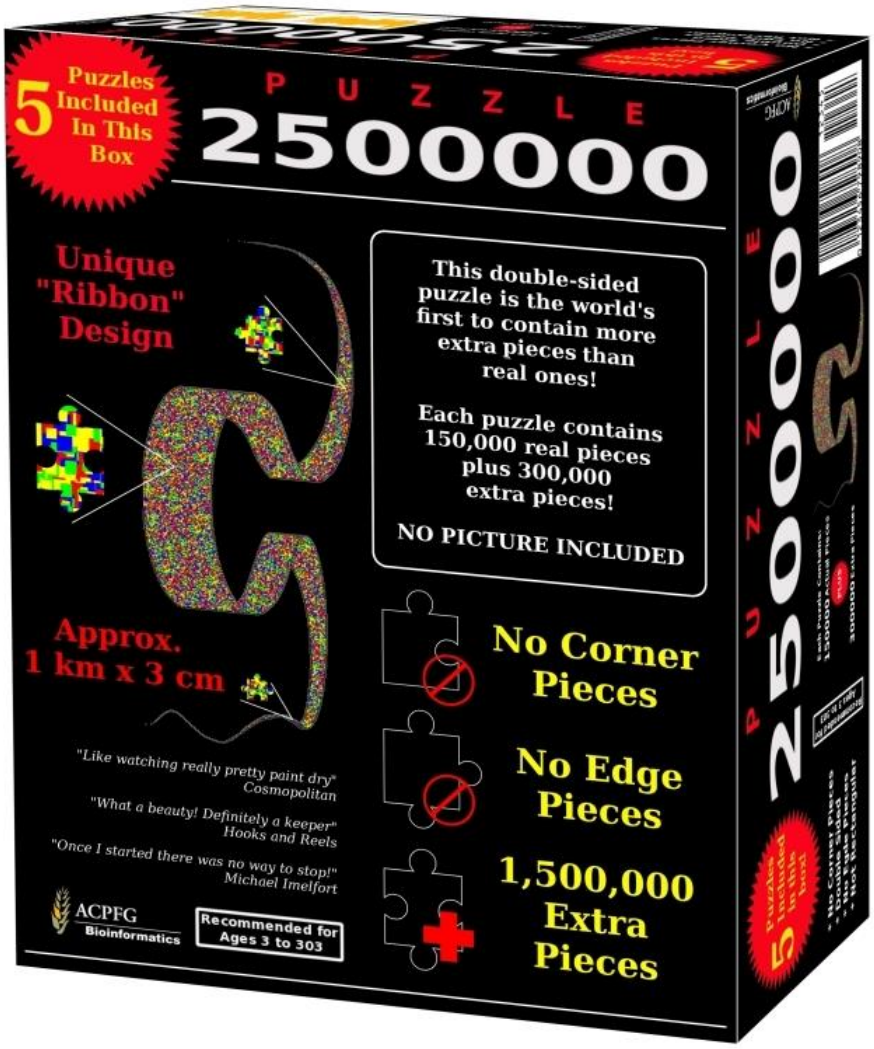
Overview

- Genome assembly challenges
- Validating genome assemblies
- Genotyping by sequencing
- Placing missing contigs
- Clustering based validation

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

The challenge of genome sequencing

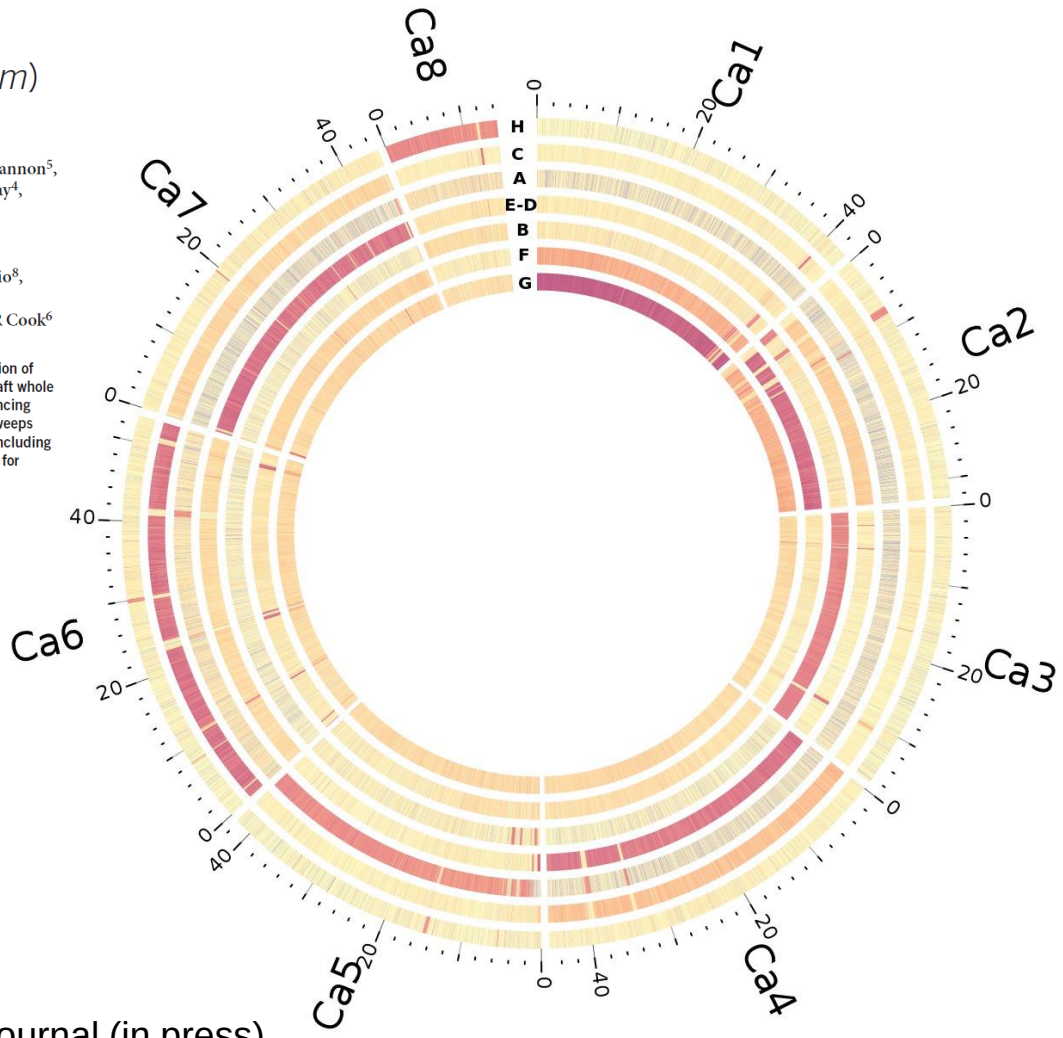


The challenge of genome sequencing

Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement

Rajeev K Varshney[1,2], Chi Song[3], Rachit K Saxena[1], Sarwar Azam[1], Sheng Yu[3], Andrew G Sharpe[4], Steven Cannon[5], Jongmin Baek[6], Benjamin D Rosen[6], Bunyamin Tar'an[7], Teresa Millan[8], Xudong Zhang[3], Larissa D Ramsay[4], Aiko Iwata[9], Ying Wang[3], William Nelson[10], Andrew D Farmer[11], Pooran M Gaur[1], Carol Soderlund[10], R Varma Penmetsa[6], Chunyan Xu[3], Arvind K Bharti[11], Weiming He[3], Peter Winter[12], Shancen Zhao[3], James K Hane[13], Noelia Carrasquilla-Garcia[6], Janet A Condie[4], Hari D Upadhyaya[1], Ming-Cheng Luo[6], Mahendar Thudi[1], C L L Gowda[1], Narendra P Singh[14], Judith Lichtenzveig[15], Krishna K Gali[4], Josefa Rubio[8], N Nadarajan[16], Jaroslav Dolezel[17], Kailash C Bansal[18], Xun Xu[3], David Edwards[19], Gengyun Zhang[3], Guenter Kahl[20], Juan Gil[8], Karam B Singh[13,21], Swapan K Datta[22], Scott A Jackson[9], Jun Wang[3,23] & Douglas R Cook[6]

Chickpea (*Cicer arietinum*) is the second most widely grown legume crop after soybean, accounting for a substantial proportion of human dietary nitrogen intake and playing a crucial role in food security in developing countries. We report the ~738-Mb draft whole genome shotgun sequence of CDC Frontier, a *kabuli* chickpea variety, which contains an estimated 28,269 genes. Resequencing and analysis of 90 cultivated and wild genotypes from ten countries identifies targets of both breeding-associated genetic sweeps and breeding-associated balancing selection. Candidate genes for disease resistance and agronomic traits are highlighted, including traits that distinguish the two main market classes of cultivated chickpea—*desi* and *kabuli*. These data comprise a resource for chickpea improvement through molecular breeding and provide insights into both genome diversity and domestication.

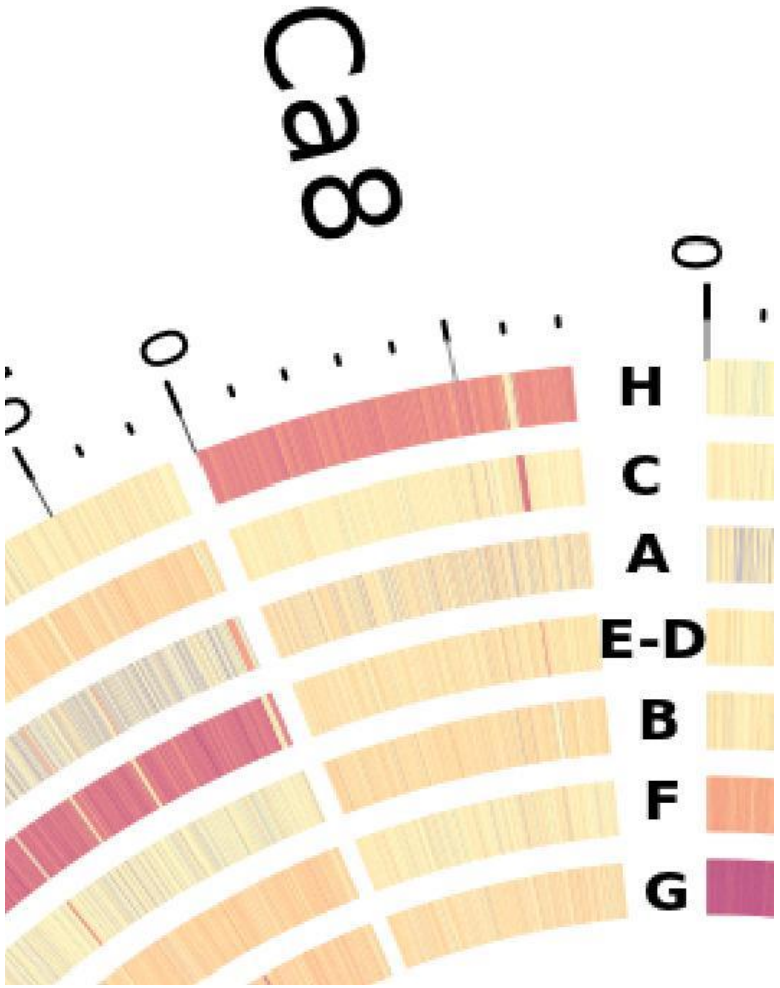


Ruperao et al. The Plant Biotechnology Journal (in press)

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

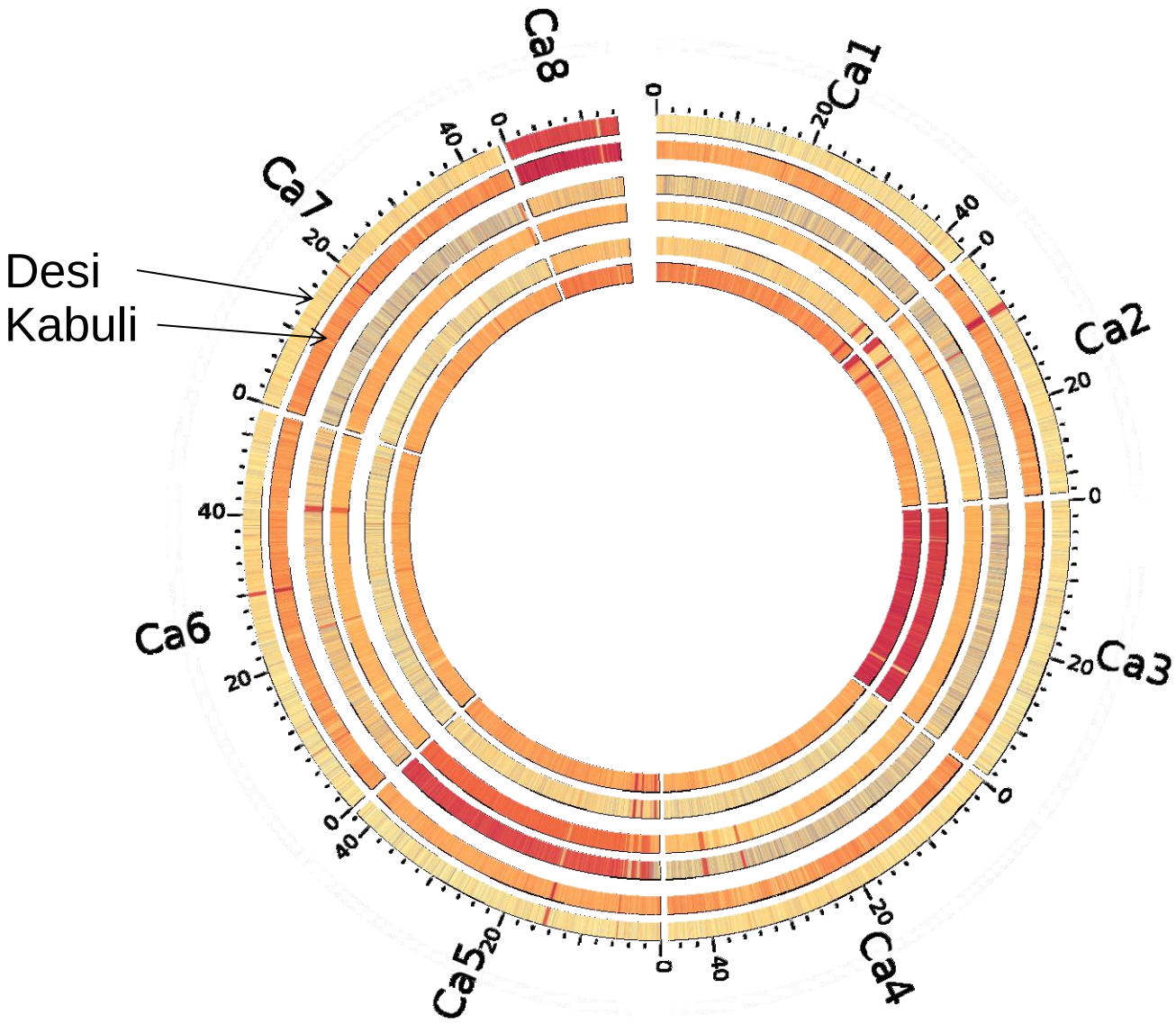
Kabuli chickpea reference



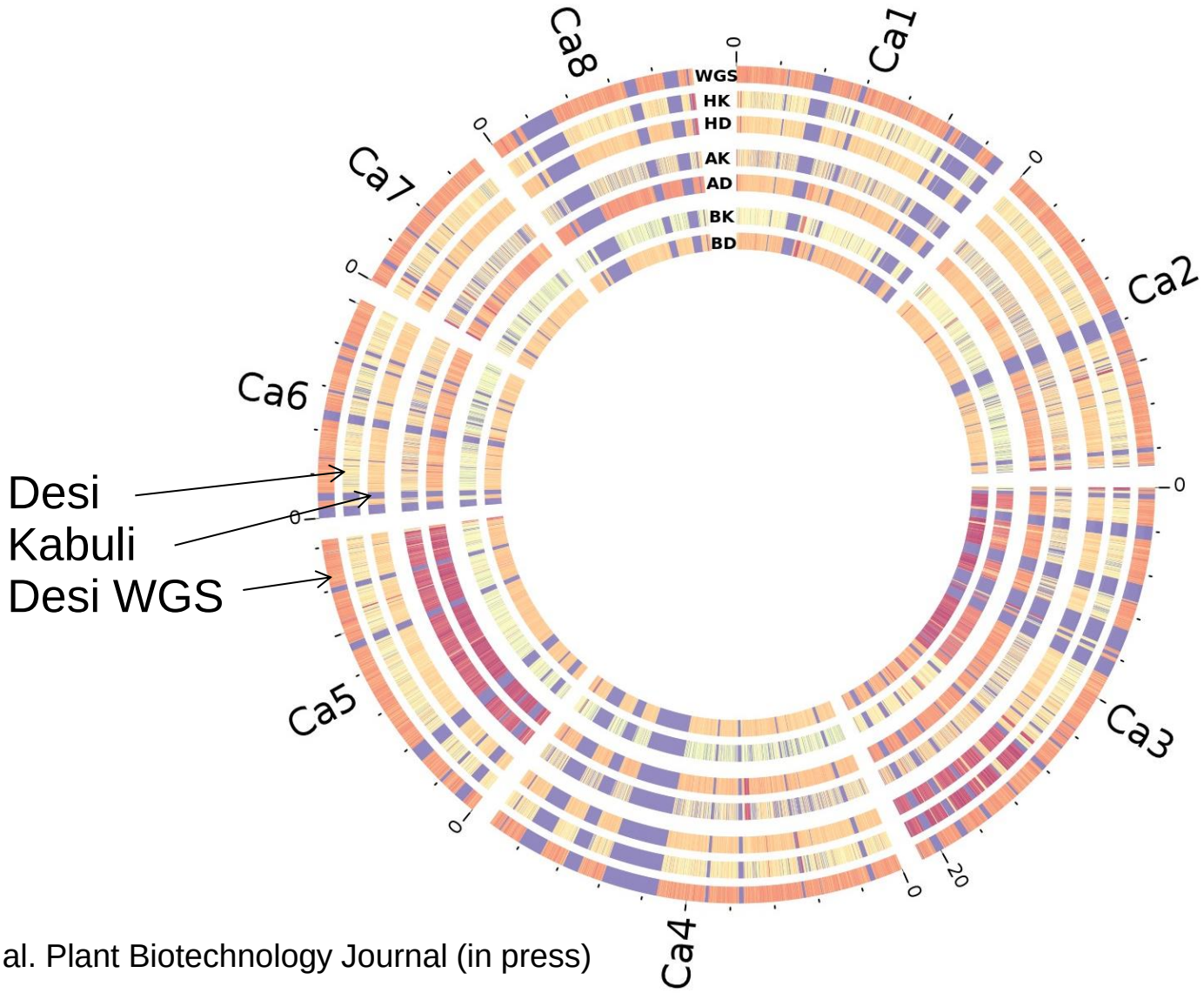
THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

Kabuli chickpea reference



Desi chickpea reference

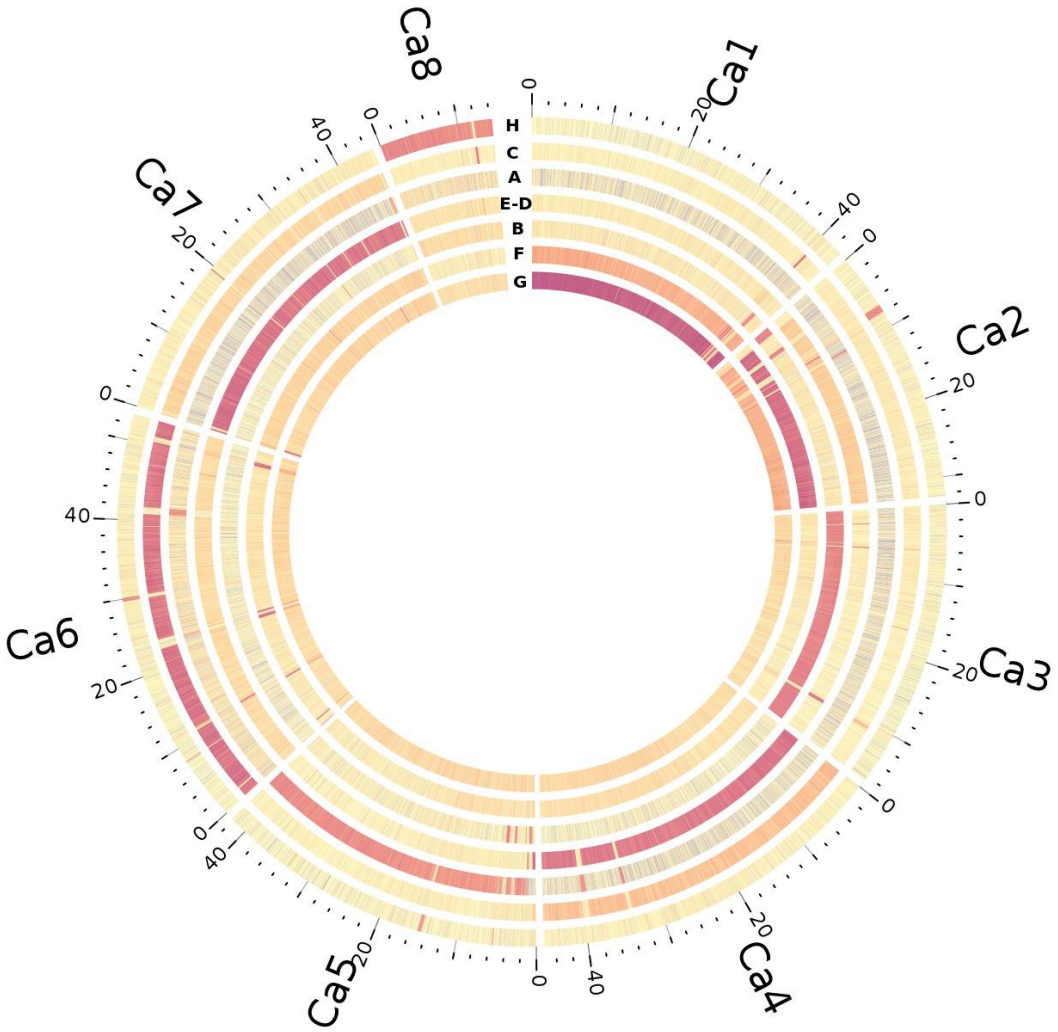


Ruperao et al. Plant Biotechnology Journal (in press)

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

How do we validate and fix a reference genome



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

Call genotype of previously predicted SNPs

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

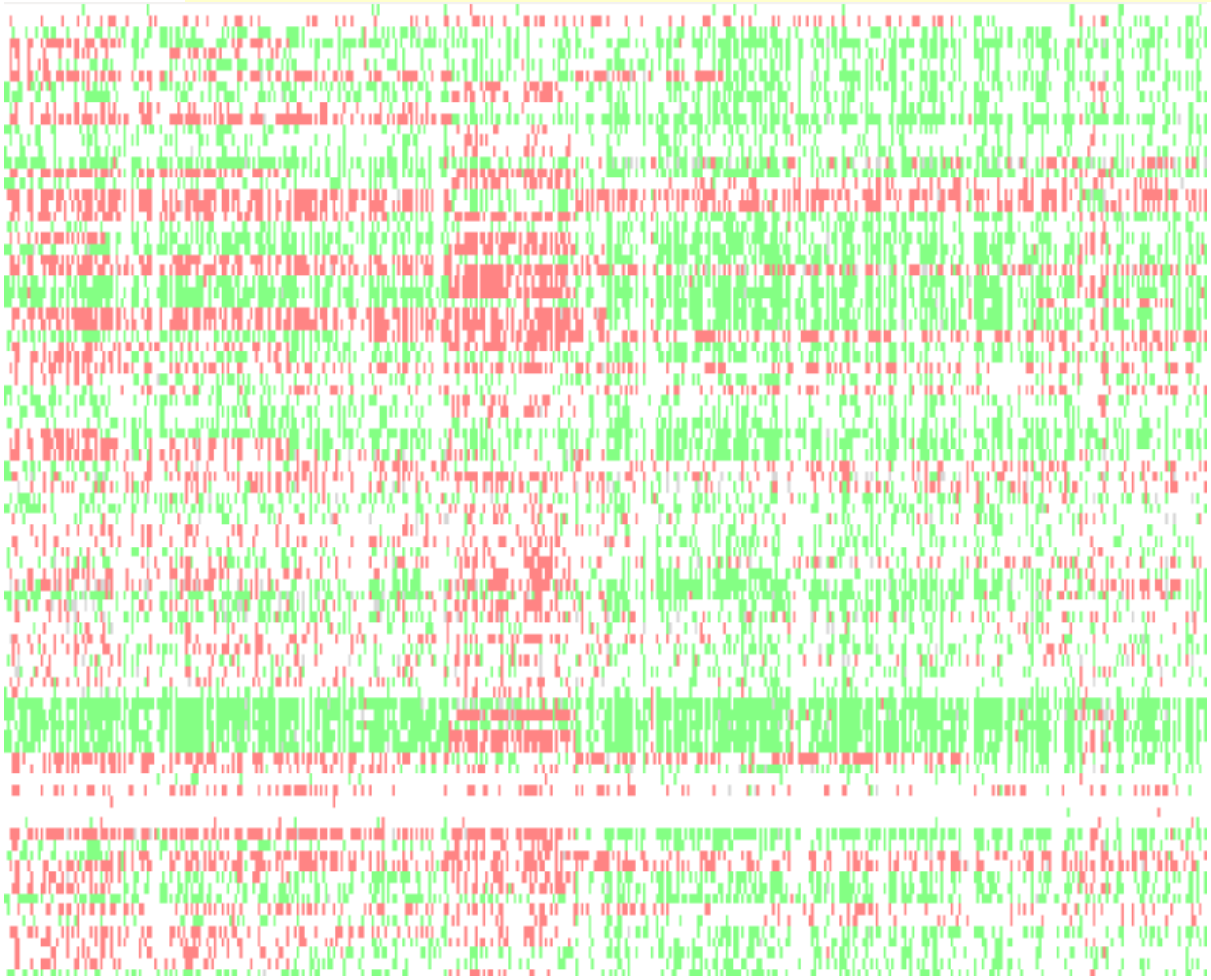
Haplotype blocks

T	A	G	G	T	C	C	A	G	G	A	T	A	A	T
N	T	C	C	A	G	G	C	T	C	G	C	G	G	C
TN1	A	G	G	T	C	C	A	G	G	A	T	A	A	T
TN2	A	G	G	T	C	C	A	G	G	A	T	A	A	T
TN3	T	C	C	A	G	G	C	G	G	A	T	A	A	T
TN4	A	G	G	T	C	C	A	G	G	A	T	A	A	T
TN5	T	C	C	A	G	G	C	T	C	G	C	G	G	C
TN6	A	G	G	T	C	C	A	G	G	A	T	A	A	T
TN7	T	C	C	A	G	G	C	T	C	G	C	G	G	C

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

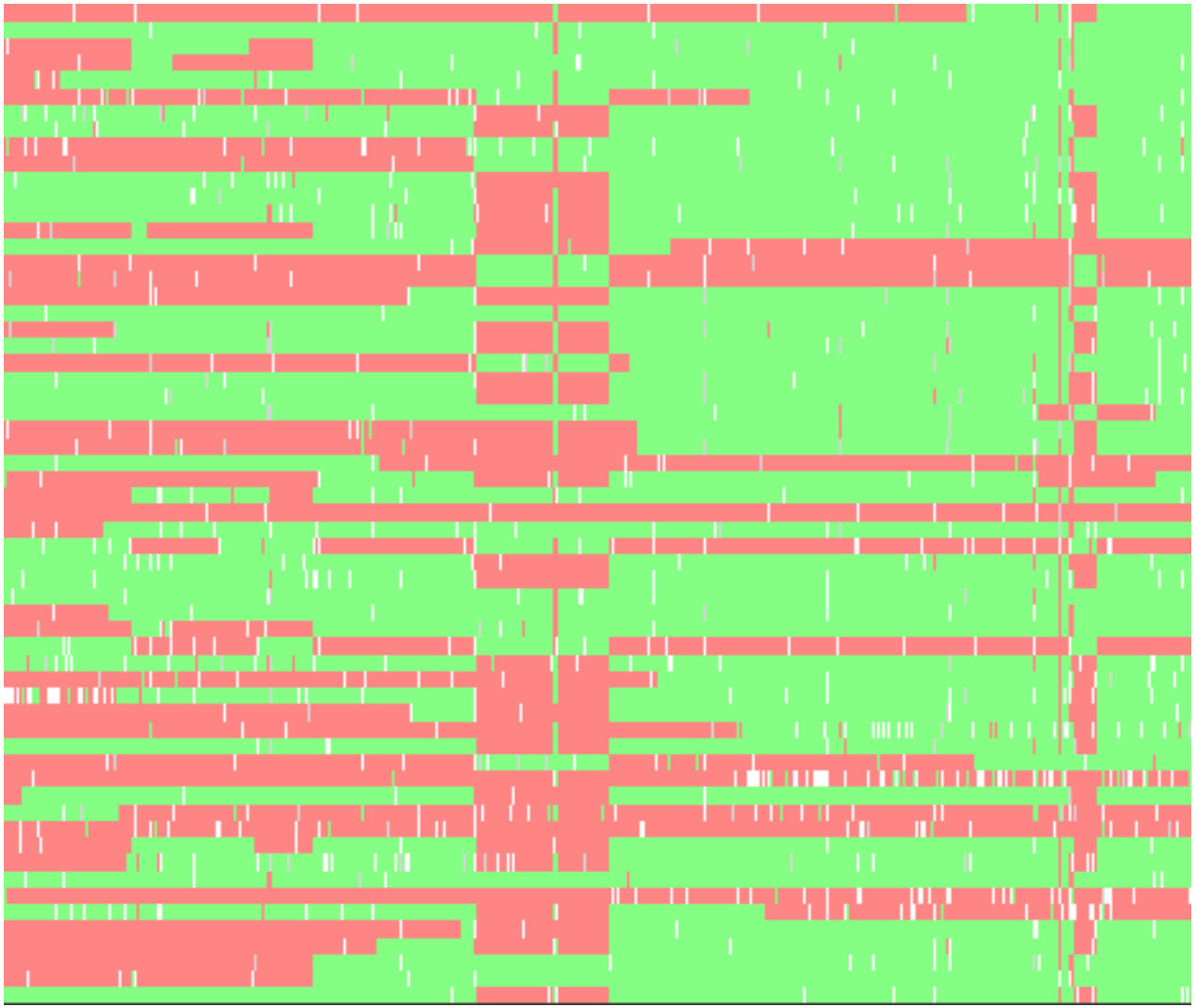
Pre imputation



THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

After imputation and cleaning



THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

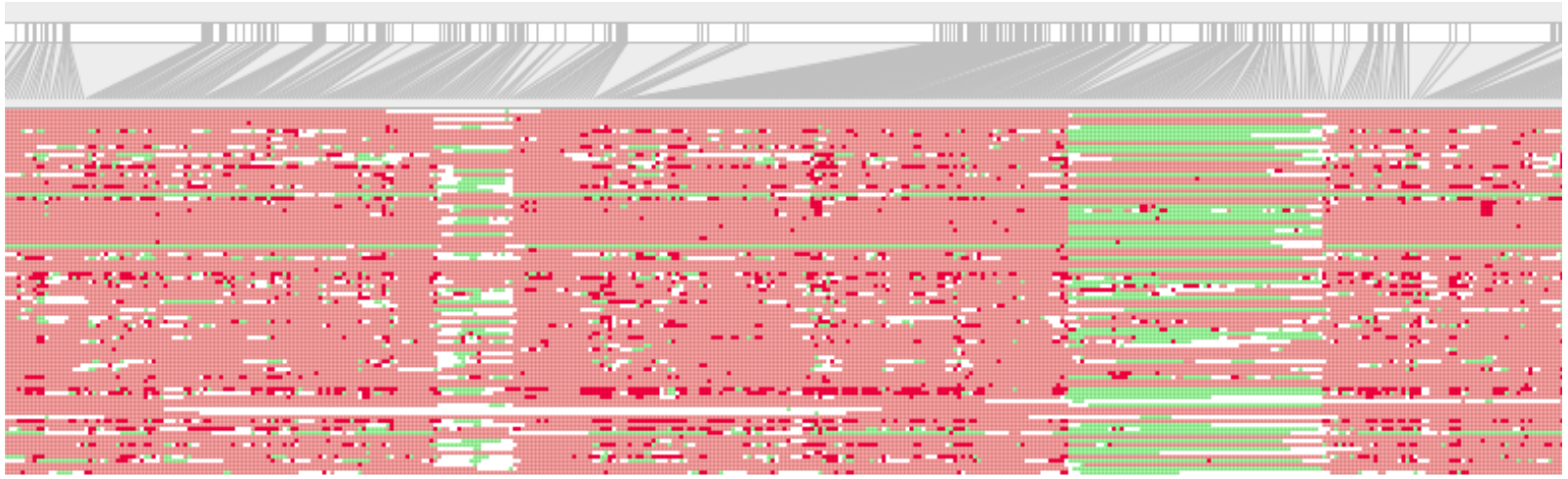
ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

Genotyping by sequencing

- 92 double haploid Tapidor x Ningyou individuals
- Called SNPs for parents, assigned parental genotypes to all alleles in the population
- Used these alleles to
  - create genetic maps
  - place unplaced contigs
  - identify misplaced and chimeric contigs

GBS applications

- Check for misplaced contigs based on recombination events shared between individuals



- Red:Tapidor, green: Ningyou, dark red: heterozygous allele, white: missing

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

Placing unplaced contigs: contigPlacer

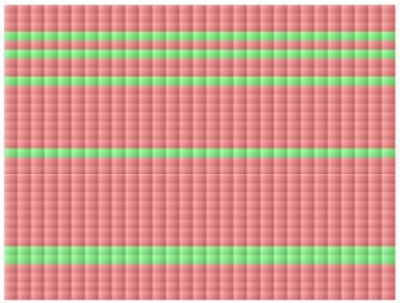
- Compare unplaced contigs with all placed contigs
- Use metaSNPs
- Penalized Hamming distance to compare alleles between two SNPs
- Places unplaced contig next to the best possible placed contig
- Possible to reverse contig if latter half of contig fits better to best partner than first half

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

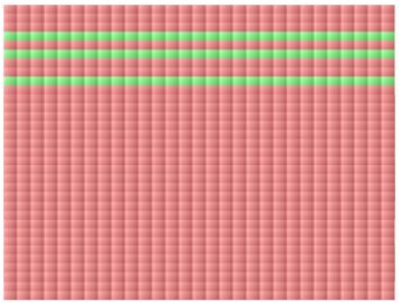
ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

contigPlacer 1

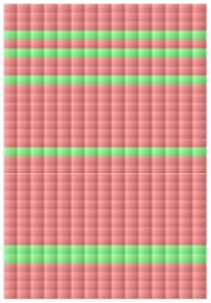
Placed contig 1



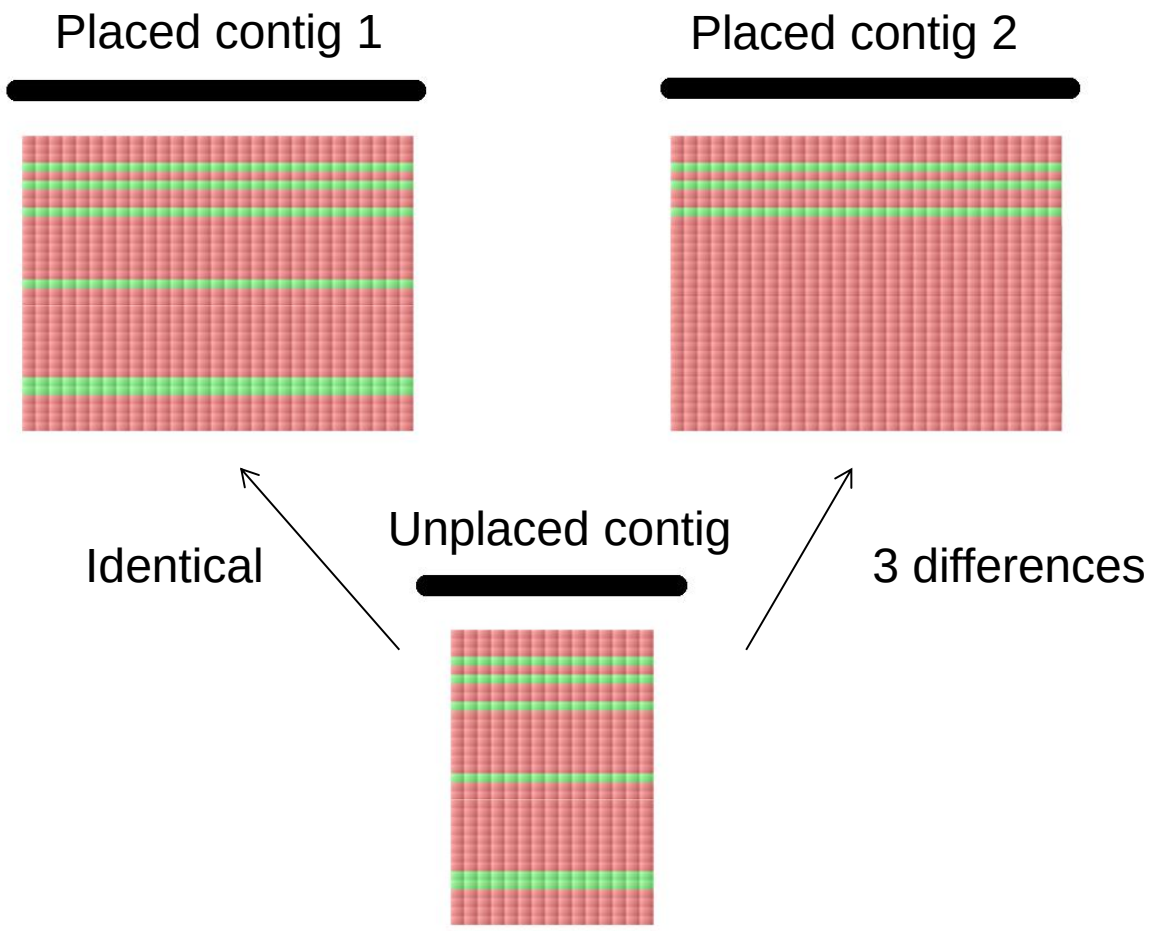
Placed contig 2



Unplaced contig



contigPlacer 2

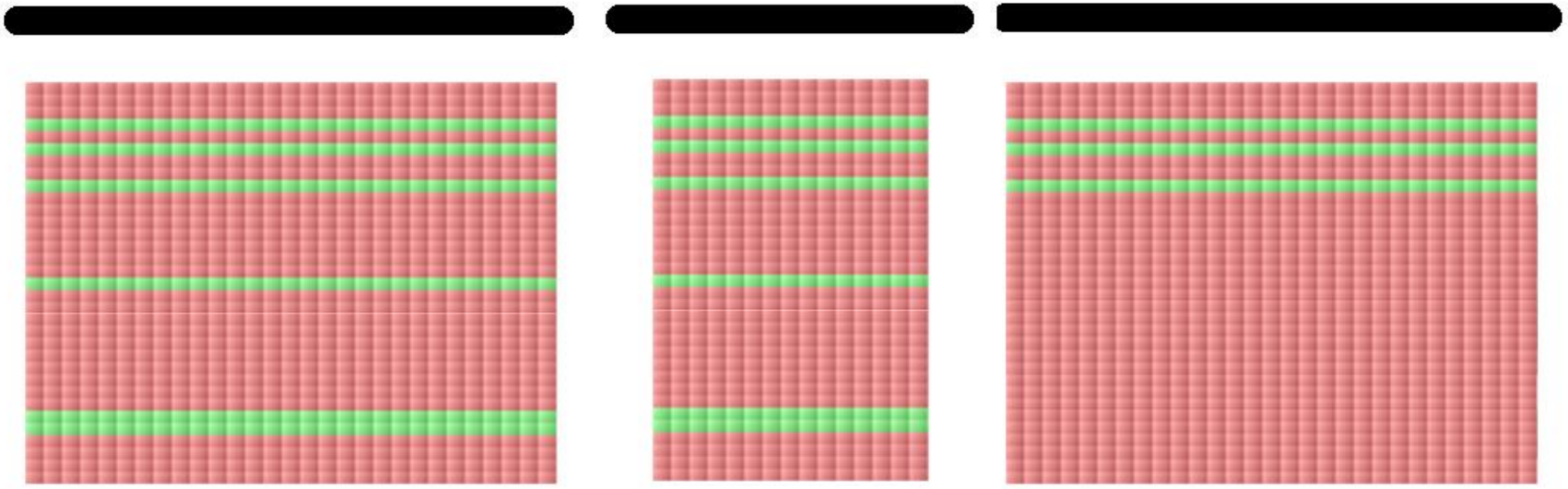


contigPlacer 3

Placed contig 1

Now placed contig

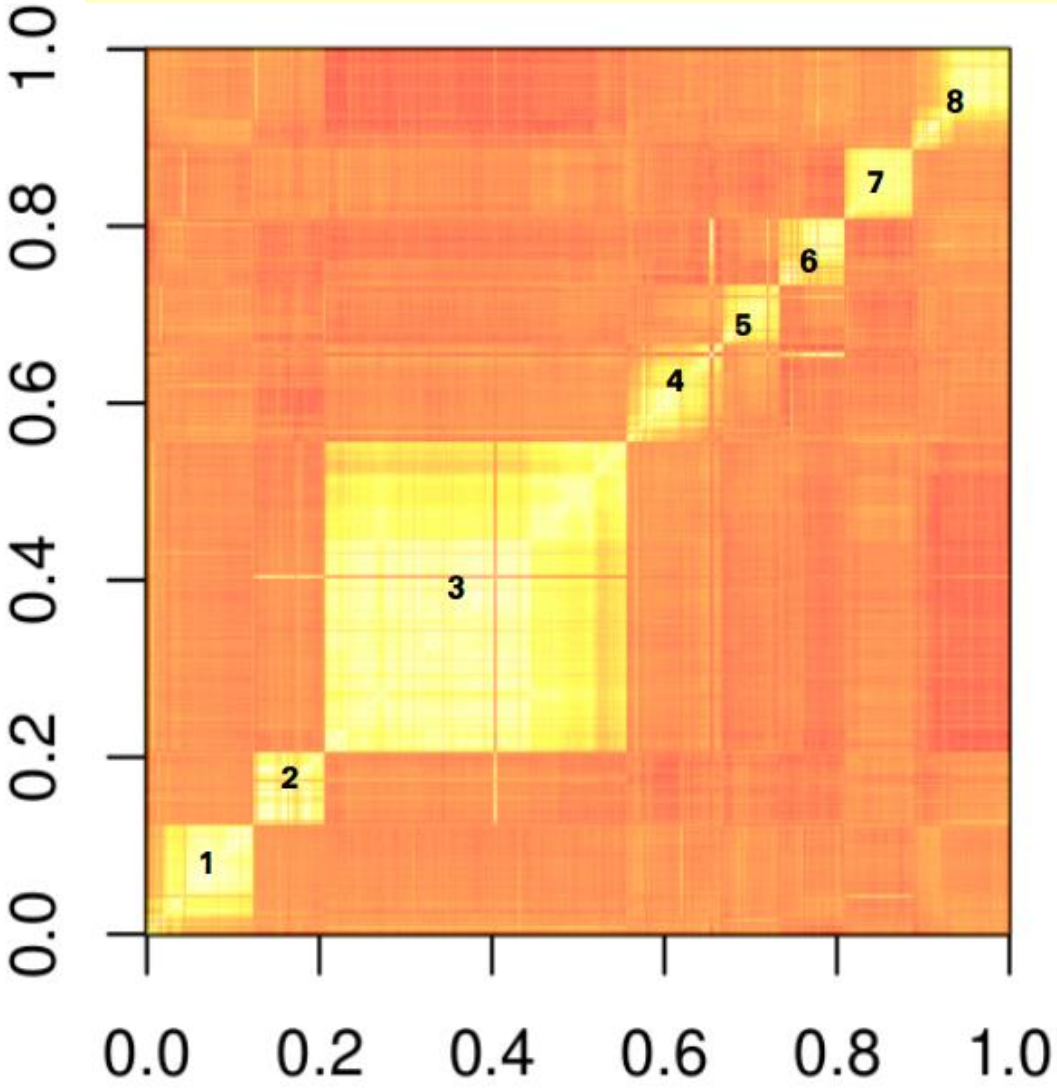
Placed contig 2



THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

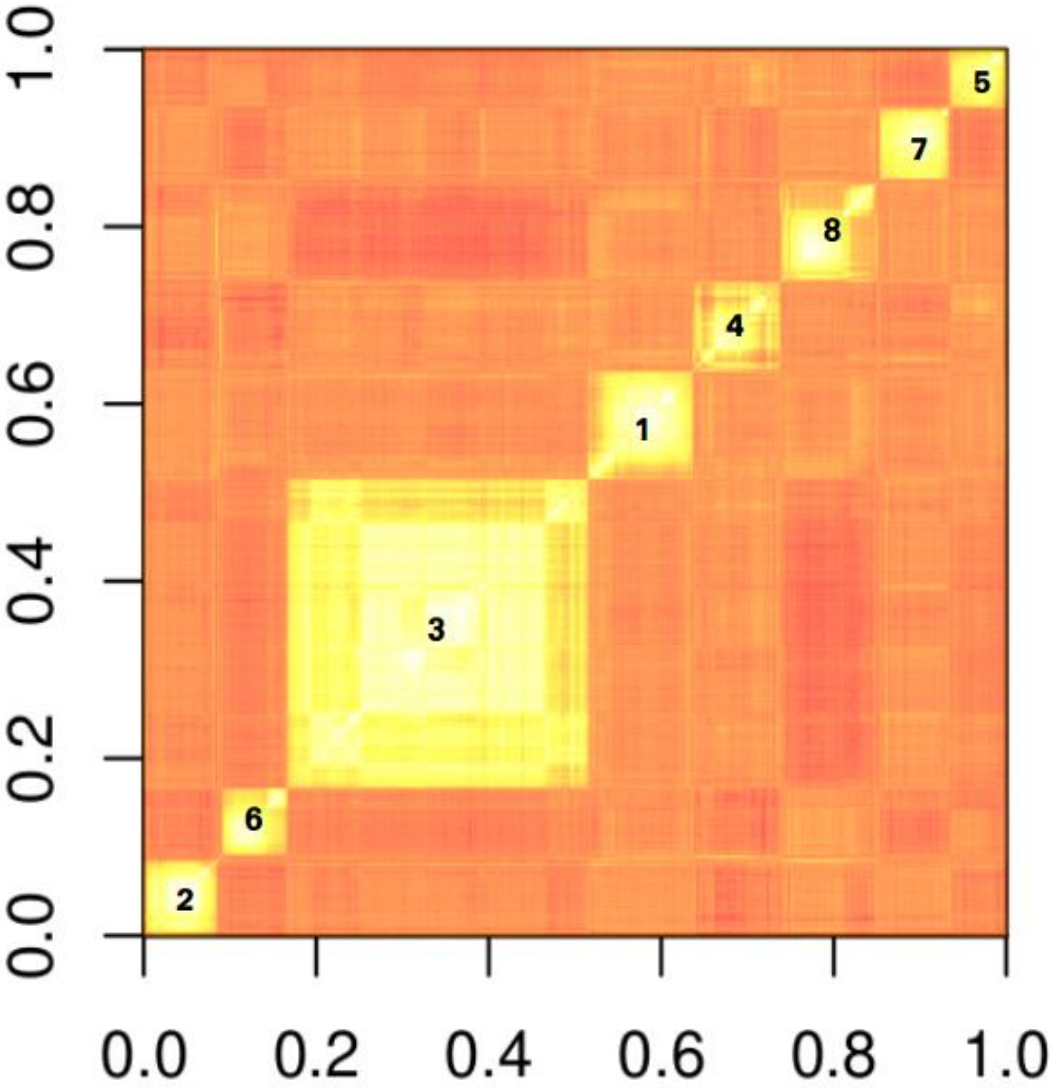
Clustering



THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

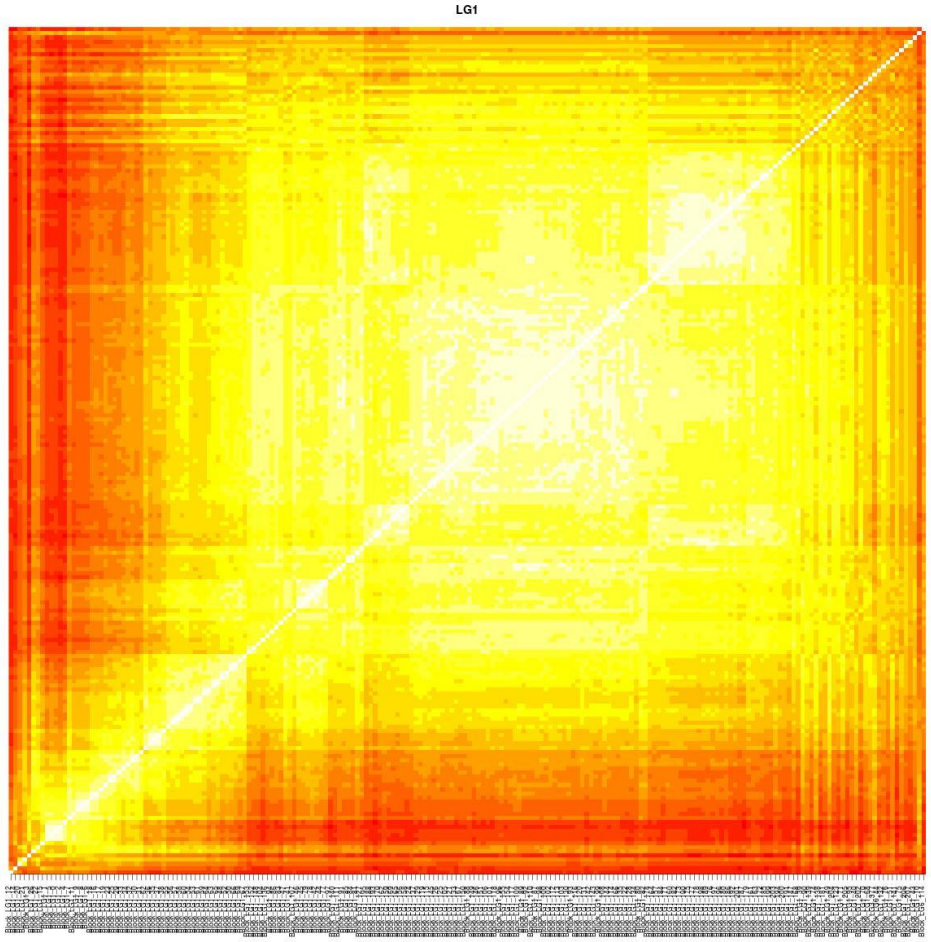
Clustering



THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

Clustering



THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

Placing contigs in Darmor

- Total size: 850 Mbp
- Was 645 Mbp placed in pseudomolecules
- Now 800 Mbp placed in pseudomolecules,
- 50 Mbp unplaced (contigs with no SNPs or chimeric)
- Genes on pseudomolecules from 63,904 to 75,955
- Only 3,528 genes remain unplaced

Acknowledgements

Philipp Bayer
Kenneth Chan
Michal Lorenc
Pradeep Ruperao
Kaitao Lai
Agnieszka Golicz
Paul Visendi
Paula Martinez
Bhavna Hurgobin
Huey Tyng Lee
Juan Montenegro

Jacqueline Batley
Alice Hayward
Emma Campbell
Reece Tollenaere
Salman Alamery
Jessica Dalton-Morgan
Satomi Hayashi

Jinling Meng
Yan Long
Boulos Chalhoub
Harsh Raman
Isobel Parkin
Shengyi Liu
Bart Lambert
Benjamin Laga

Contact:
Dave.Edwards@uq.edu.au

ELSEVIER

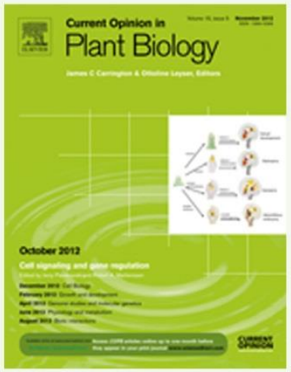
Current Opinion
Conferences

Agriculture and Climate Change

Adapting crops to increased uncertainty

Amsterdam, The Netherlands | 15-17 February, 2015

Supporting publication



Conference Chairs

David Edwards
Australian Centre for Plant Functional Genomics
School of Agriculture and Food Sciences
University of Queensland, Australia

Giles E.D. Oldroyd
Cell & Developmental Biology
John Innes Centre, United Kingdom

Advisory Board
Jeff Bennetzen
Jose Crossa
Robert Henry
Rodomiro Ortiz
Andrew Paterson
Kadambot Siddique
Mark Sorrells
Mark Tester
Michael Udvardi
Senthold Asseng
Robin Buell

www.agricultureandclimatechange.com