



THE UNIVERSITY OF  
WESTERN AUSTRALIA



# Genomics

Dave Edwards

[Dave.Edwards@uwa.edu.au](mailto:Dave.Edwards@uwa.edu.au)

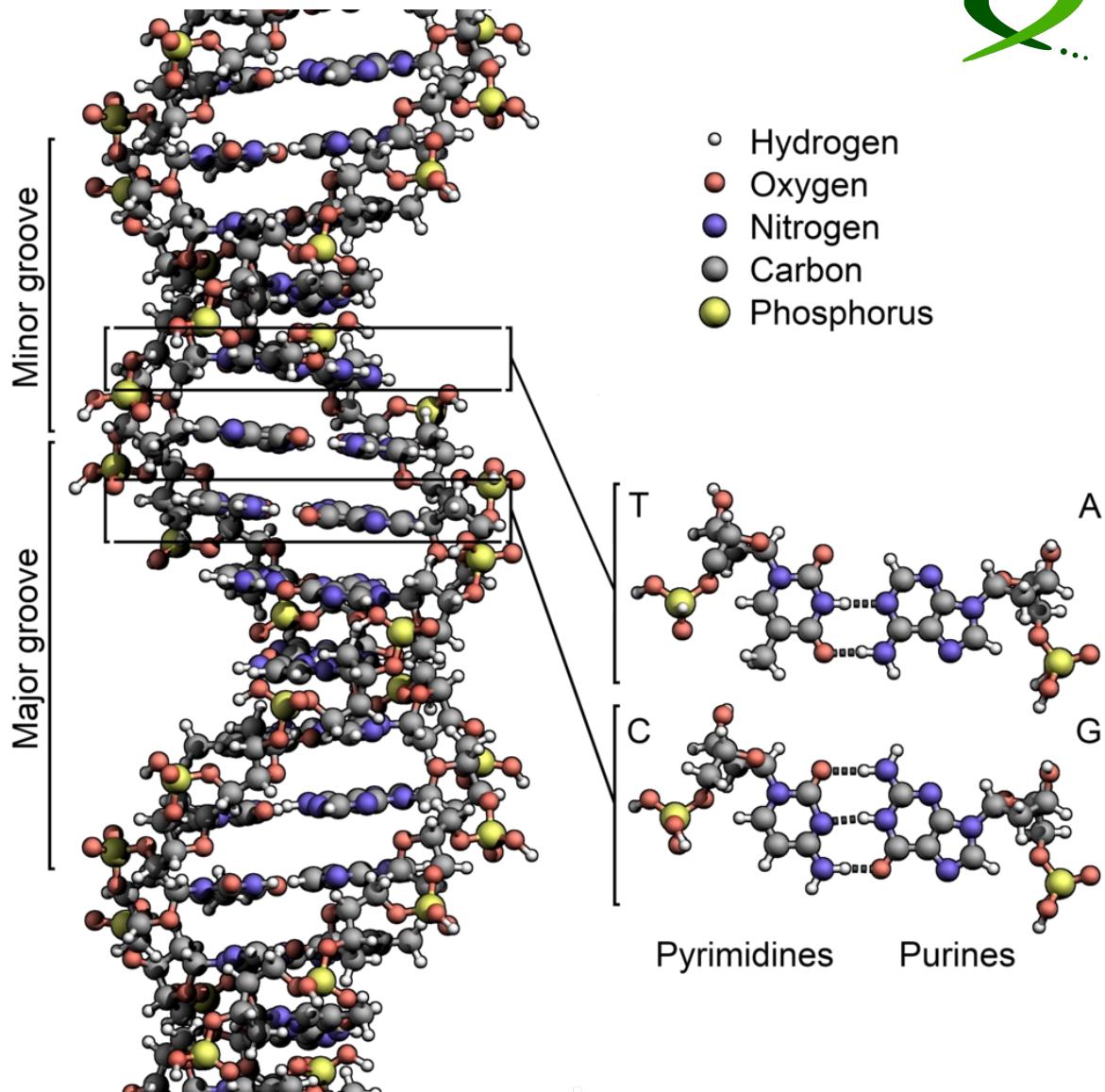


# About me

- Applied Bioinformatics Group
- Work on genomics, diversity and deep learning
- Work on *Brassica*, wheat, chickpea, pigeonpea, soybean, lupin, clover, pea, lentil etc.
- Also work on *Hakea*, seagrasses, Amborella, humans, chicken, fish, coral and more
- Teach practical bioinformatics in SCIE4002 and SCIE5003
- Established microcredentials in Applied Bioinformatics
- Established the Master of Bioinformatics
- Established the UWA Centre for Applied Bioinformatics
- Support student projects in applied bioinformatics



THE UNIVERSITY OF  
WESTERN AUSTRALIA





THE UNIVERSITY OF  
WESTERN AUSTRALIA



Centre for  
Applied  
Bioinformatics



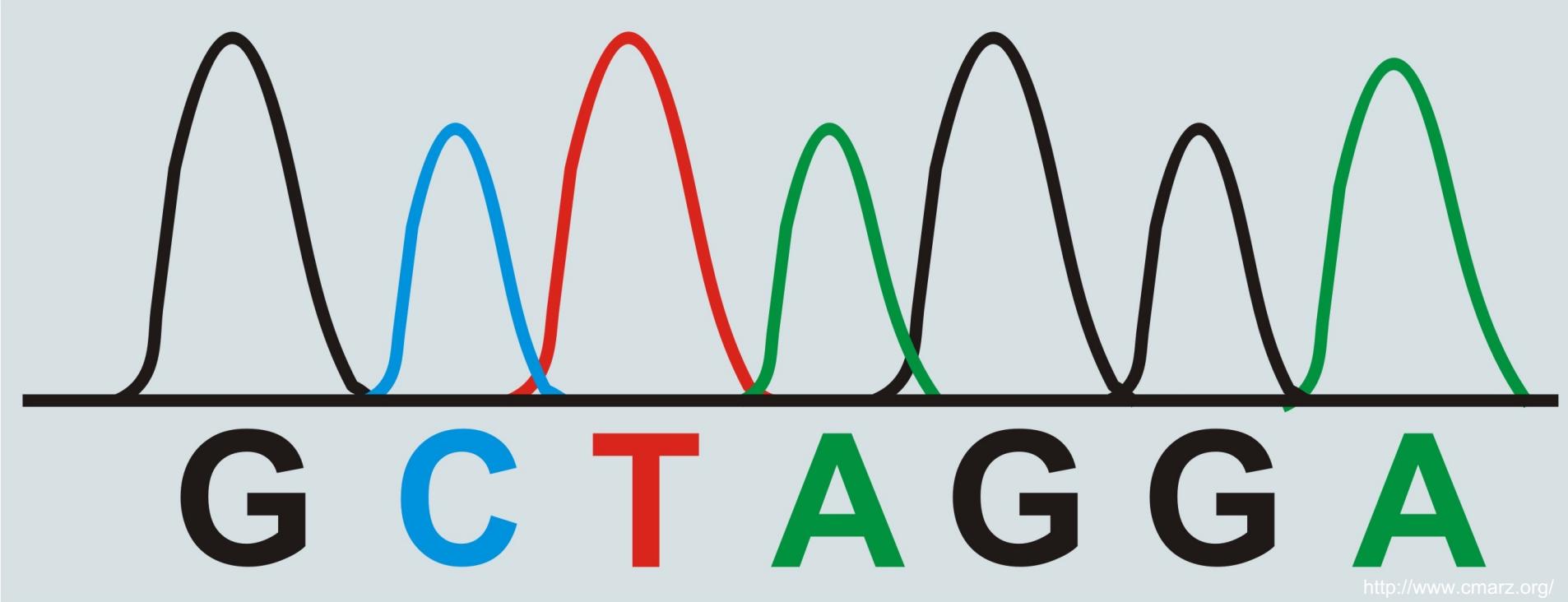
<http://www.geneticliteracyproject.org/>



THE UNIVERSITY OF  
WESTERN AUSTRALIA

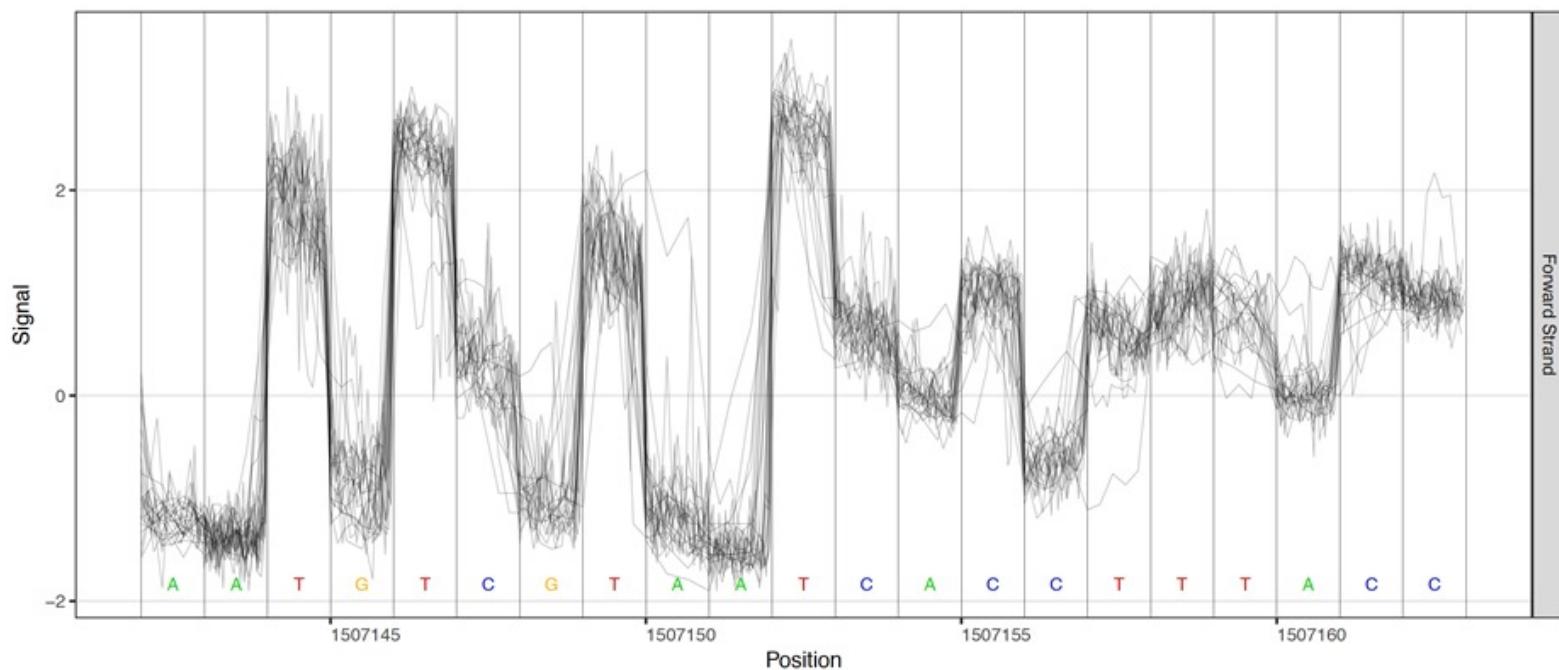


Centre for  
Applied  
Bioinformatics





THE UNIVERSITY OF  
WESTERN AUSTRALIA



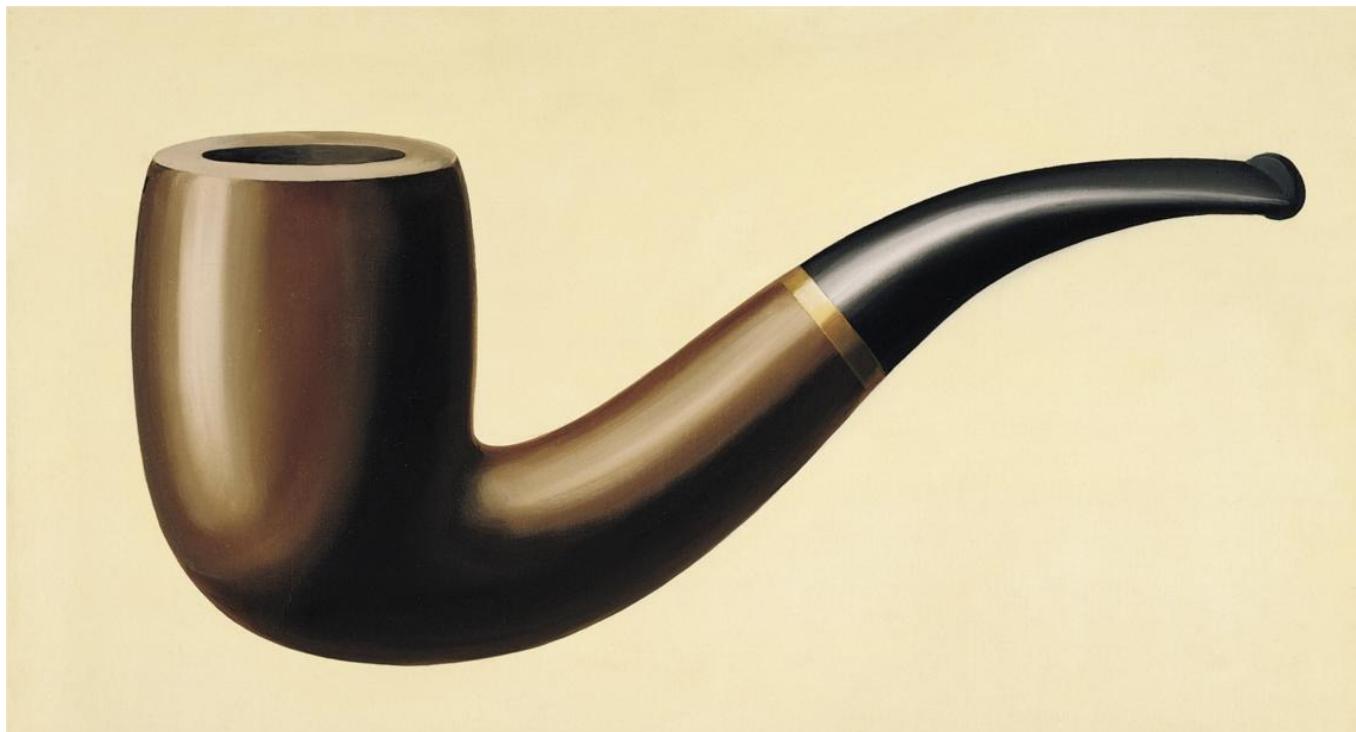


>gi|126513242|gb|EF428113.1| Triticum aestivum  
cultivar Chinese Spring FT protein (FTD) gene,  
FTD-h1 allele, partial cds

TCAGGGTGACCTCGGGAACAGGACCGTGTCCAACGGCTGCGAGCTCAAG  
CCGTCCATGGTCGCCAGCAGCCCAGGGTTGAGGTGGCGGCAATGAGAT  
GAGGACCTTCTACACACTCGTACGTACACAGTCACTATCTAATGCCTATA  
TGTTTAGCTCTGAAAGTGCTGCCACACGCACATGATCGATCGAGCTCTA  
TATATATAGTACGTGGGAAGATGATTCTCGATGCTTCTGTTCACAGAA  
TGTTGTCTTGGCAGGCACATGACTAATGCTCCATCTTGCATATGGCTCT  
GTGCTAGCTCTGGTGTGATCATGATTCTATGCTTGTTCATTTTC  
TTTCGGGAAACATTGATTTGATGCTTCTATTCACATGTTGTTCAT  
GTTTGTCTAGCAAGCACACGACTAATTAAAGCTCGGTCTAAATACATA  
TGCTTATTACGGTAGTACTCTACTTCTCTAGTATTGATCATGATGTG  
CACCGTGTACTGCCCGCAGGTGATGGTAGACCCAGATGCTCCAAGTCCA  
AGCGATCCAACCTTAGGGAGTATCTCCACTGGTAAGTACTAAATTGTA  
ACTCAGTTGAATAATTCTCTGTCCCTAGATATACACAGTAGCTCATGTG  
TGTGTGTCTACATGTGTGCAGGCTTGTGACAGATATCCCCGGTACAAC  
TGGTGCATCCTCGGGCAGGAGGTGATGTGCTACGAGAGCCCTCGTCCGA  
CCATGGGATCCCATCGCTCGTGCTCTCCAGCAGCTCGGCCGG  
CAGACCGTGTACGCTCCGGTGGCGCC



THE UNIVERSITY OF  
WESTERN AUSTRALIA





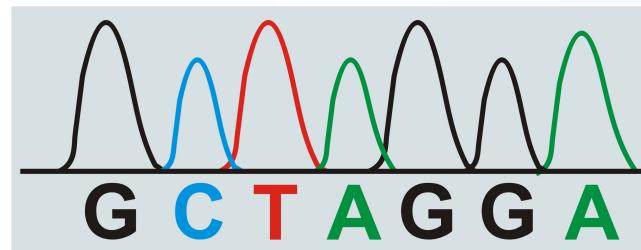
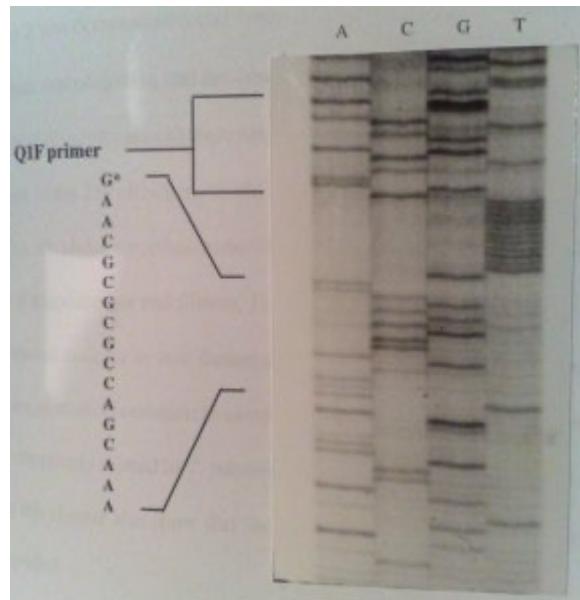
THE UNIVERSITY OF  
WESTERN AUSTRALIA





THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Sanger sequencing





THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Back in 2007



- Genome sequencing using BACs
- Roche 454 sequencing
- Illumina Solexa 35 bp reads





THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Illumina NovaSeq



- Draft genome assemblies
- Illumina 150 bp reads
- 3 Tbp per run
- 300 Gbp <\$2000





THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Oxford Nanopore



Centre for  
Applied  
Bioinformatics

- Oxford Nanopore
- MinION
- Up to 50 Gb
- PromethION
- Up to 290 Gb per cell





THE UNIVERSITY OF  
WESTERN AUSTRALIA

# PacBio



- Pacific Biosciences Revio (after RSI, RSII, Sequel I, Sequel II )
- HiFi reads highly accurate





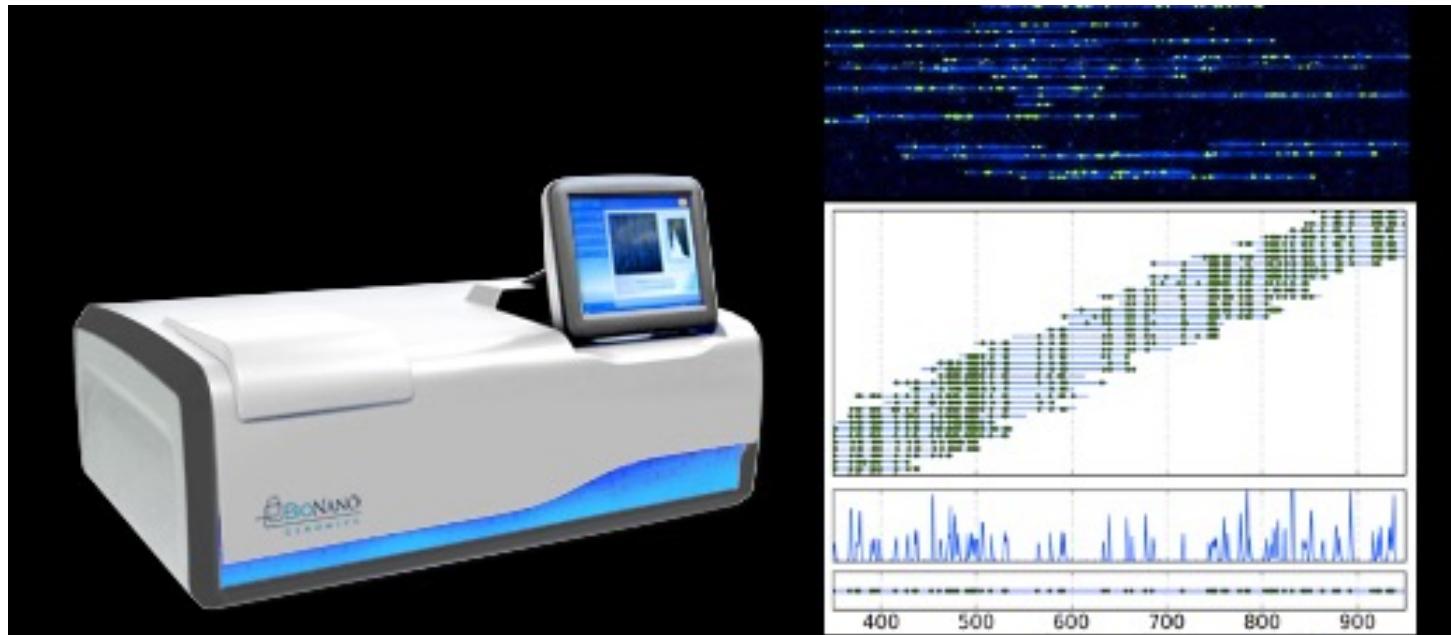
THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Bionano



Centre for  
Applied  
Bioinformatics

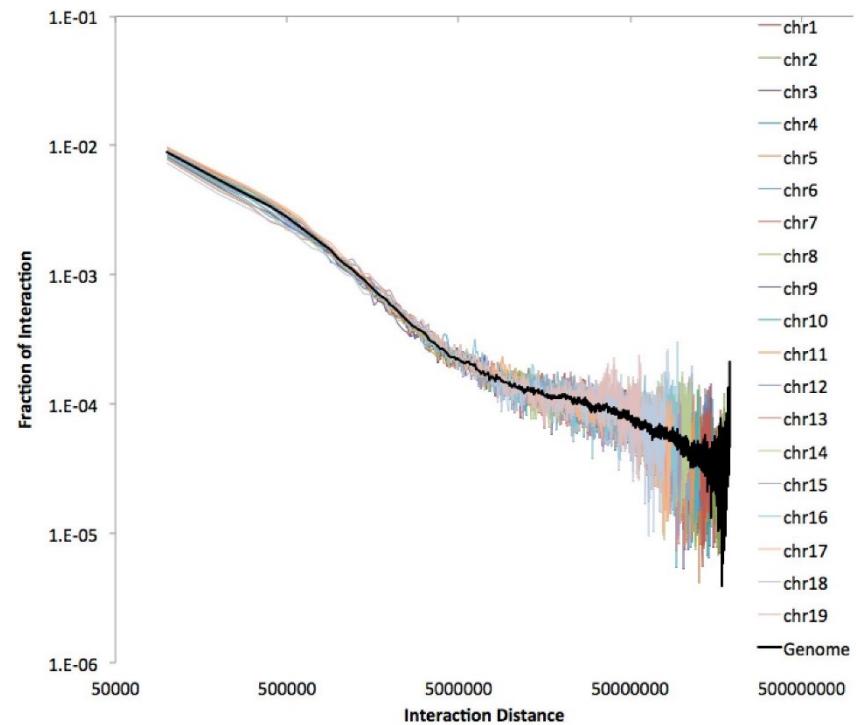
- Bionano physical mapping
- Large scale scaffolding or structural variations





# HiC

- Chromosome conformation capture
- HiC/Dovetail
- Long range reads





# What can we do with this data?



- Sequence genomes
- Characterise genome variation
  - Presence/absence variants (PAVs)
  - Single nucleotide polymorphisms (SNPs)
- Gene expression
- Associate genome variation with traits
- Apply this knowledge for breeding
  - Marker assisted selection
  - Genomic selection
  - Genome editing

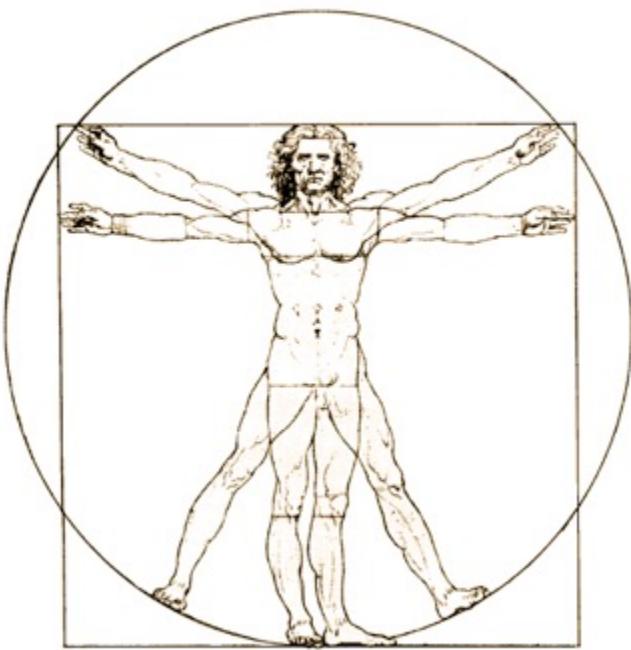


THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Complexity



Centre for  
Applied  
Bioinformatics





THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Sequence reads are small



Centre for  
Applied  
Bioinformatics

	bp	distance	
Illumina read	100	1.88 m	me
Sanger read	900	17 m	telegraph pole
Nanopore read	120,000	2.3 km	short walk to Kings Park



# Genomes are big

	bp	distance	
Illumina read	100	1.88 m	me
Sanger read	900	17 m	telegraph pole
Nanopore read	120,000	2.3 km	short walk to Kings Park
Canola genome	1,300,000,000	24,700 km	Perth to Moscow (and back)
Human genome	3,000,000,000	56,400 km	~3 times the Great Wall
Wheat genome	17,000,000,000	323,000 km	around the world ~4 times
Loblolly pine	22,000,000,000	413,600 km	Earth-moon distance
Axolotl genome	32,000,000,000	601,600 km	2 light seconds
<i>Paris japonica</i>	150,000,000,000	2,820,000 km	twice the diameter of the sun



THE UNIVERSITY OF  
WESTERN AUSTRALIA

# The challenge of genome sequencing



Centre for  
Applied  
Bioinformatics





THE UNIVERSITY OF  
WESTERN AUSTRALIA

# The challenge of genome sequencing



Centre for  
Applied  
Bioinformatics



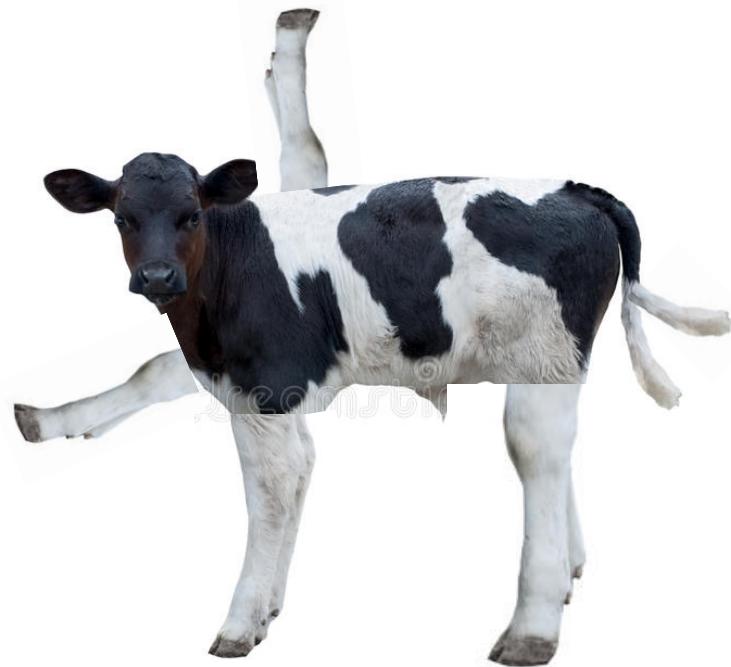
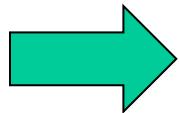


THE UNIVERSITY OF  
WESTERN AUSTRALIA

# The challenge of genome sequencing



Centre for  
Applied  
Bioinformatics



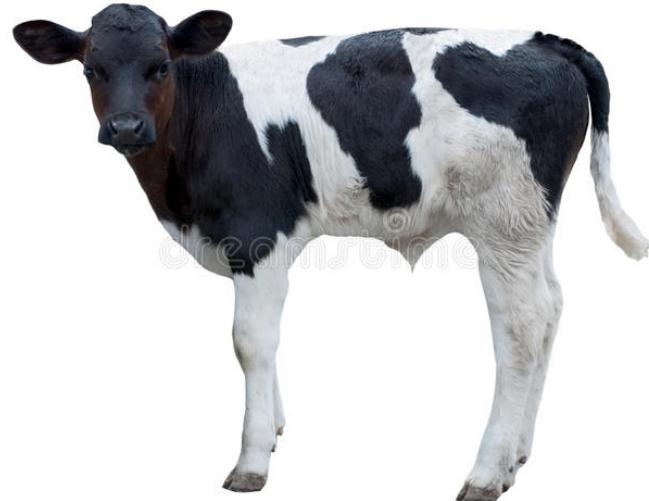


THE UNIVERSITY OF  
WESTERN AUSTRALIA

# The challenge of genome sequencing



Centre for  
Applied  
Bioinformatics





# Sequencing a genome



- Generate data
  - Reduce complexity?
    - BACs
    - Isolated chromosomes
  - Whole genome shotgun
  - Data types
    - Sanger
    - Illumina
      - Paired end
      - Mate paired
      - Long mate paired
  - Pacific Biosciences
  - Oxford Nanopore
  - Bionano
  - Population sequencing/genetic maps



# Sequencing a genome

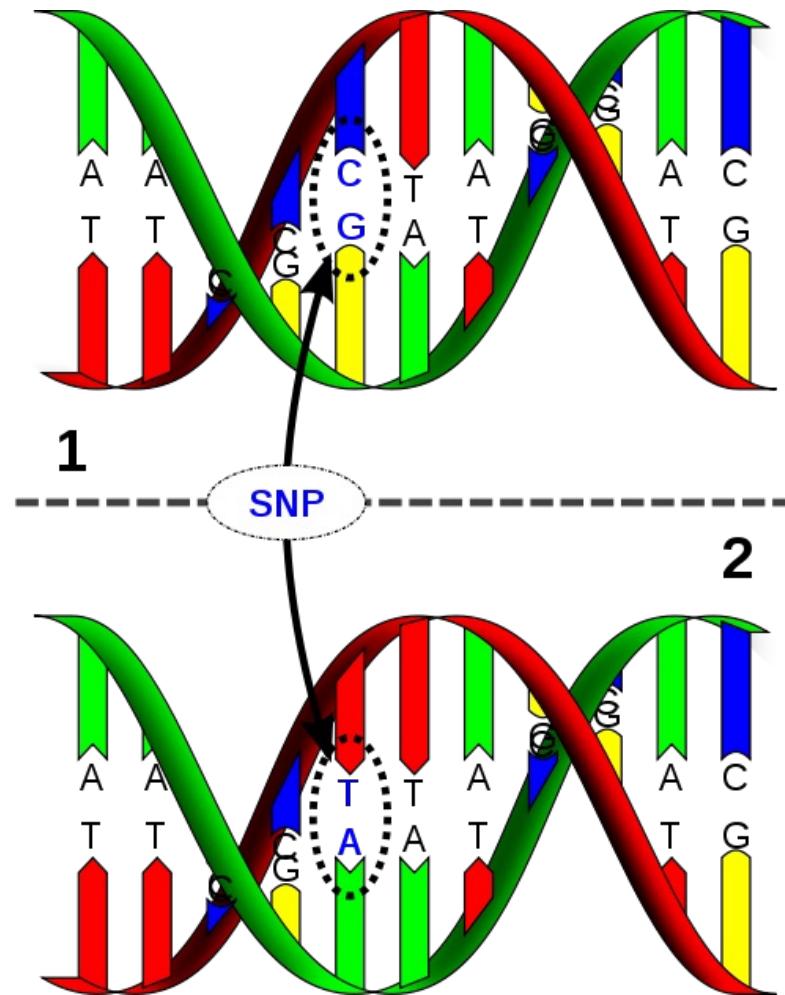


- Assemble the data
  - Which algorithm?
  - Stepwise or all together?
- Validation – is it any good?
  - N50
  - BUSCO
  - %Ns
  - % chimeric contigs
  - True pseudomolecules



THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Single Nucleotide Polymorphisms (SNPs)





THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Single Nucleotide Polymorphisms (SNPs)



- Single nucleotide differences are linked to phenotype differences, for example, different human eye colors
- More often just linked to a phenotype, not causing it





# Single Nucleotide Polymorphisms (SNPs)

- Discover SNPs
  - Sequencing
- Genotype SNPs
  - Sequencing
  - Hybridisation arrays (Affymetrix)
  - Ligation assays (infinium/goldengate)
  - KASP (allele specific PCR)
- Use SNPs for
  - Genetic mapping
  - Trait association
  - Crop and animal breeding
  - Population analysis



# Other marker types



- Copy number variations
  - Extra copy number of amylase is linked to starch diet in humans and dogs
- Insertions, deletions
  - Insertion in *FMR1* causes fragile X syndrome
- Presence/absence of genes
  - Very common in plants and bacteria, rare in humans
- SSRs/microsatellites, AFLPs, RFLPs



THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Single Nucleotide Polymorphisms (SNPs)



- Identify SNPs
  - Map reads to a genome
  - Look for differences
  - Synonymous? Non-synonymous?  
If non-synonymous, how severe is the change?



THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Single Nucleotide Polymorphisms (SNPs)



Centre for  
Applied  
Bioinformatics

- Sources of error
  - Read sequence errors
  - Mismapping errors

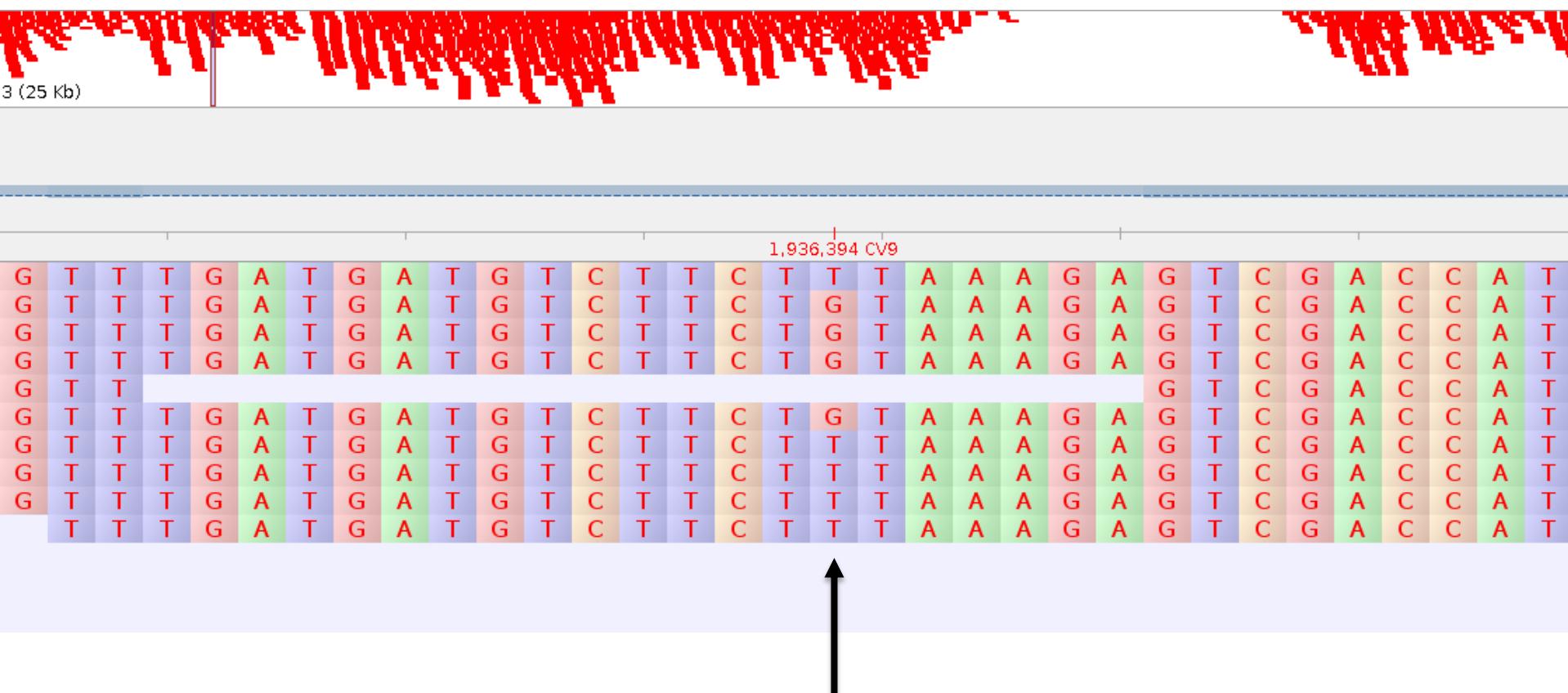


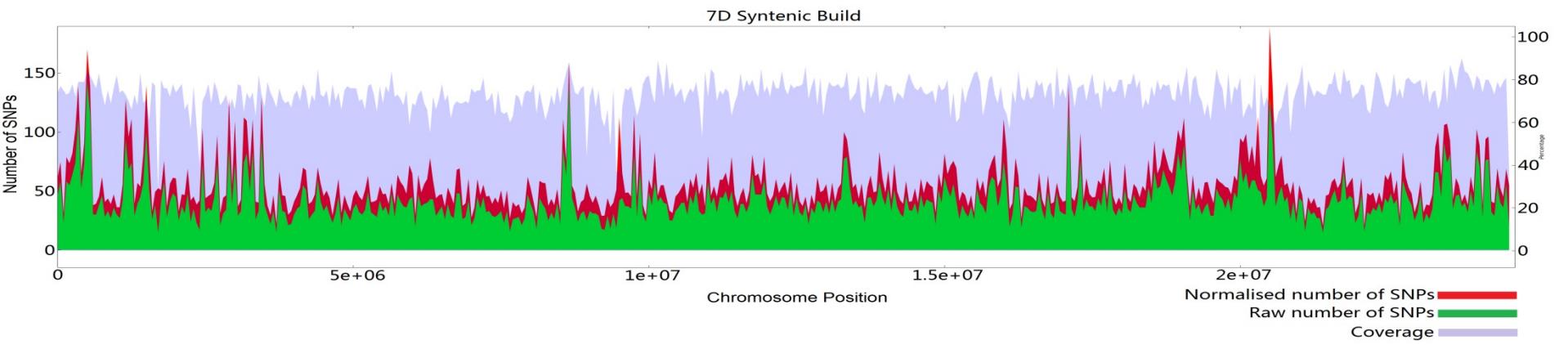
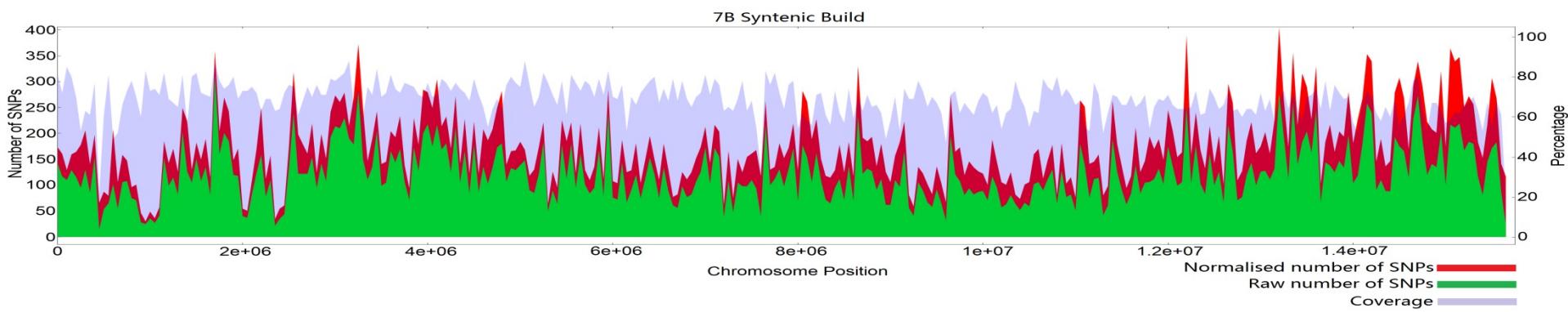
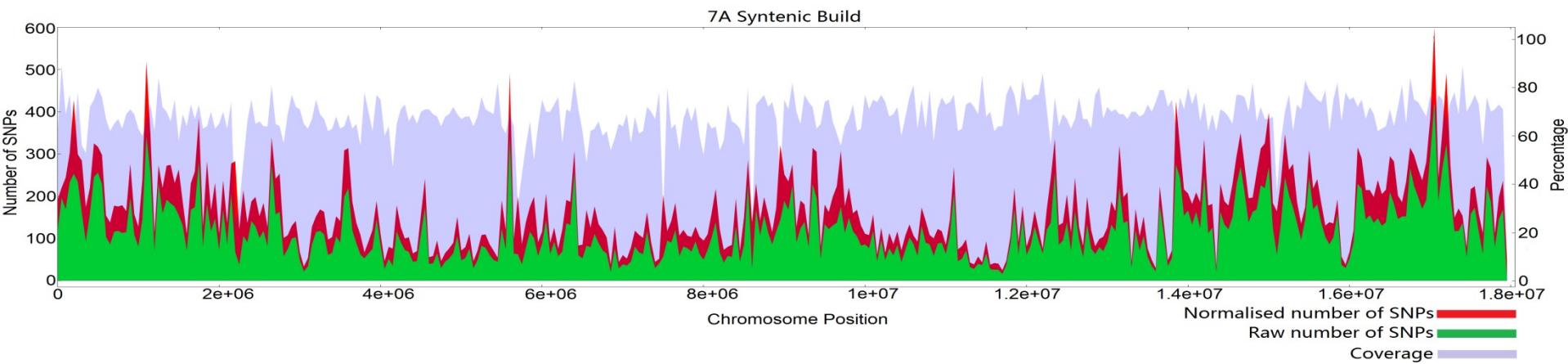
THE UNIVERSITY OF  
WESTERN AUSTRALIA

# SNP calling



Centre for  
Applied  
Bioinformatics





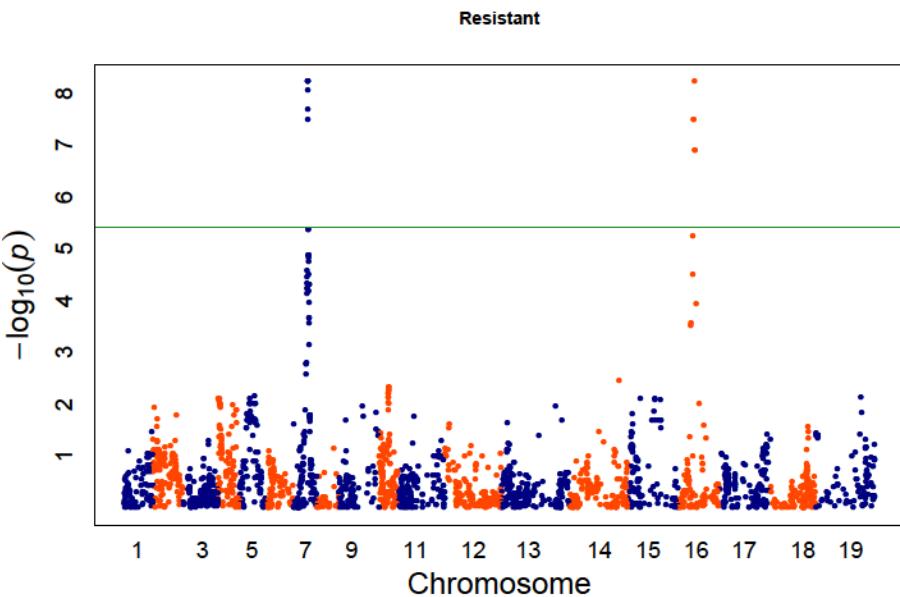


THE UNIVERSITY OF  
WESTERN AUSTRALIA

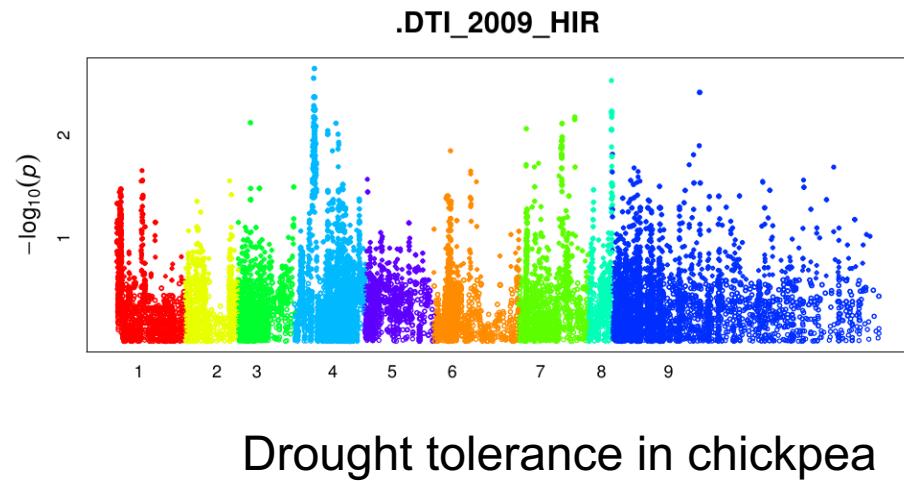
# Trait association



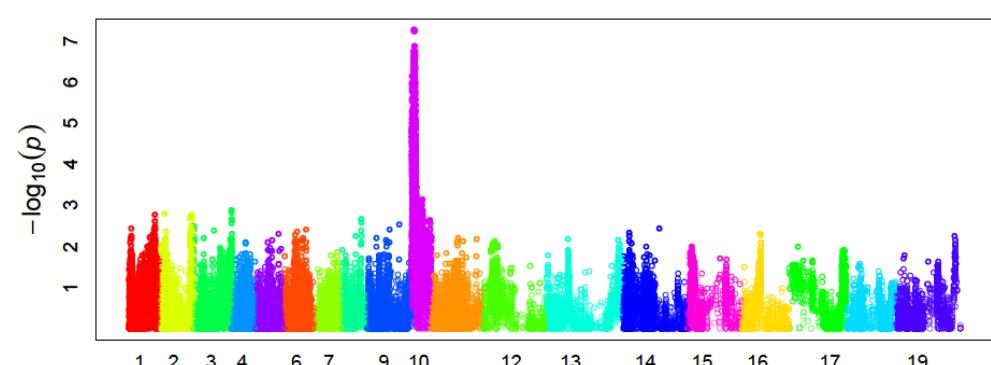
Centre for  
Applied  
Bioinformatics



Disease resistance in canola



Drought tolerance in chickpea





# Detour: Problems

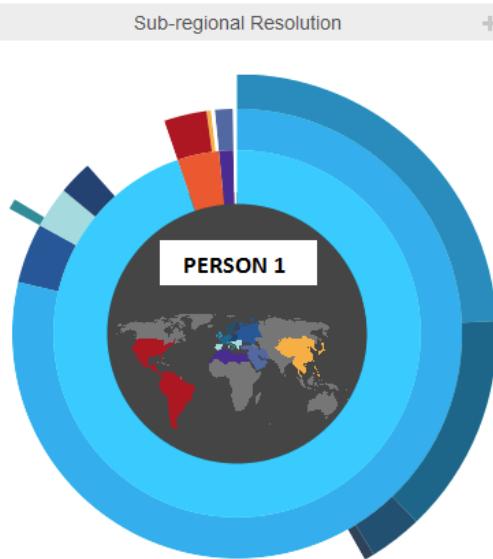
- In every species and fields, the reference data is biased based on research history, researchers' location, funding, etc.
- In human disease, lots of data is based on Europeans/US citizens
- In wheat research, lots of data is based on Chinese Spring reference cultivar



# Detour: Problems

## ancestry composition

Speculative Estimate ▾

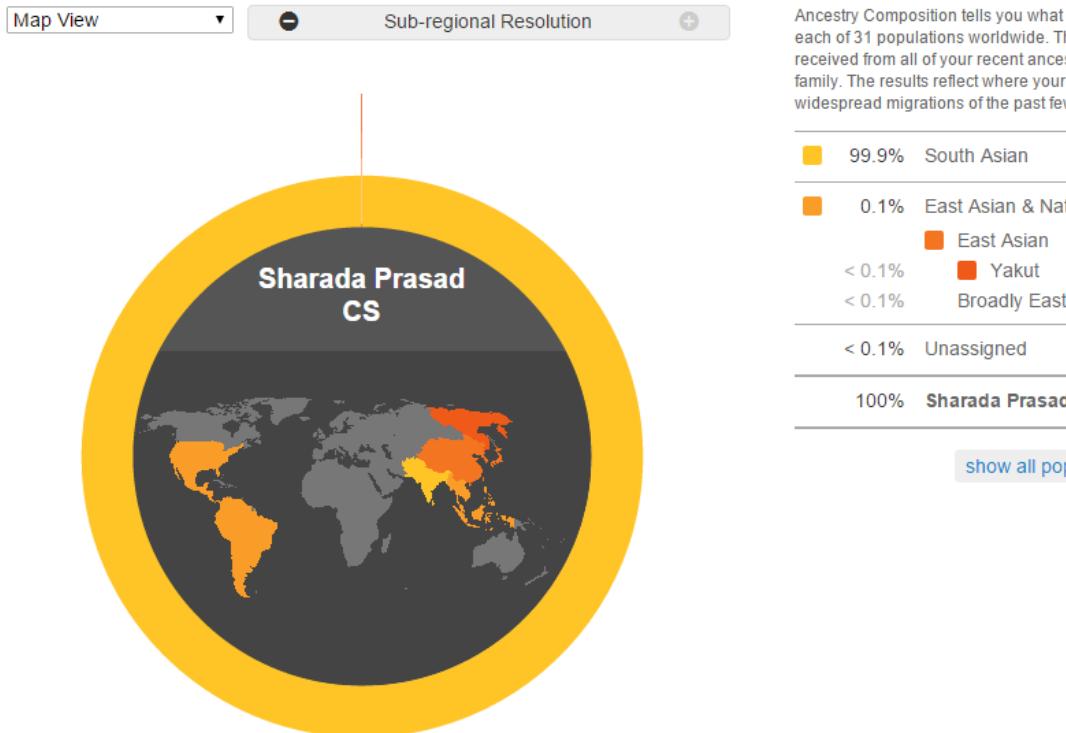


Ancestry Composition tells you what percent of your DNA comes from each of 22 populations worldwide. The analysis includes DNA you received from all of your ancestors, on both sides of your family. The results reflect where your ancestors lived 500 years ago, before ocean-crossing ships and airplanes came on the scene.

■ 94.7% European
24.2%      ■ Northern European
13.7%      ■ British and Irish
3.3%      ■ French and German
0.6%      ■ Scandinavian
0.6%      ■ Finnish
36.9%      ■ Nonspecific Northern Euro...
4.3%      ■ Eastern European
0.6%      ■ Southern European
0.6%      ■ Italian
2.3%      ■ Nonspecific Southern Eur...
2.4%      ■ Ashkenazi
6.3%      ■ Nonspecific European
■ 3.7% East Asian & Native American
3.1%      ■ Native American
0.3%      ■ East Asian
0.3%      ■ Nonspecific East Asian & Nativ...
■ 1.3% Arab
1.3%      ■ Middle Eastern
0.3% Unassigned



# Detour: Problems



<http://sharadaprasad.com/post/478/>



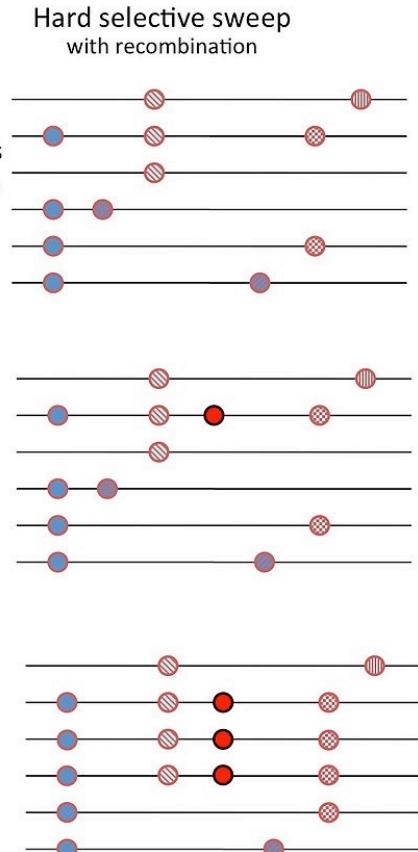
# Evolution

- Non-synonymous vs synonymous SNP changes
- Usually abbreviated as Ka/Ks
- Genes under positive selection  $Ka/Ks > 1$ 
  - Examples: HIV env
- Genes under negative selection  $Ka/Ks < 1$ 
  - Examples: house keeping genes



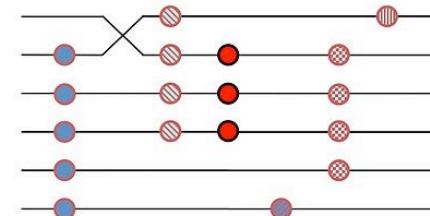
# Selective sweeps

1. Six sequences from a population. Each line represents a DNA sequence and each dot is a neutral mutation, present in one or more sequences.



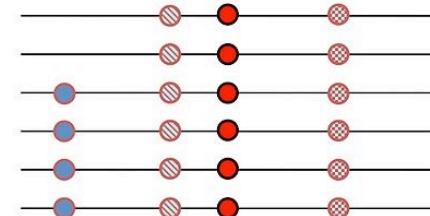
2. A beneficial mutation occurs (bright red dot on the second sequence).

4. A recombination event creates a new combination (the beneficial mutation is no longer always associated with the blue neutral mutation).



3. The beneficial mutation increases in frequency in the population, and so does the genomic background it is associated with.

5. The beneficial mutation is fixed in the population. Close to the beneficial mutation there is now no genetic variation left, all sequences look the same. Three neutral mutations have hitchhiked along with the beneficial mutation and reached high frequency.





# Selective sweeps



- ‘Hard’ vs ‘Soft’
- Hard sweeps: novel beneficial allele is introduced, spreads rapidly (0% to ~100%)
- Soft sweeps: existing neutral allele becomes beneficial, is already spread (~50% to ~100%)
- Can be hard to distinguish from random drift
- Useful to search for artificial selection



# Discussion



- In how many different ways can we describe DNA?
- What can we do with a genome assembly?
- When is a genome assembly finished?
- Do we want to know the impact of human genome variation?



THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Lecture 2



Centre for  
Applied  
Bioinformatics



# Why a pangenome?

- A reference genome does not represent the diversity of a species
- PAV genes are responsible for important traits
- Need to know gene content for genome editing



# Why a pangenome?

Previous model:

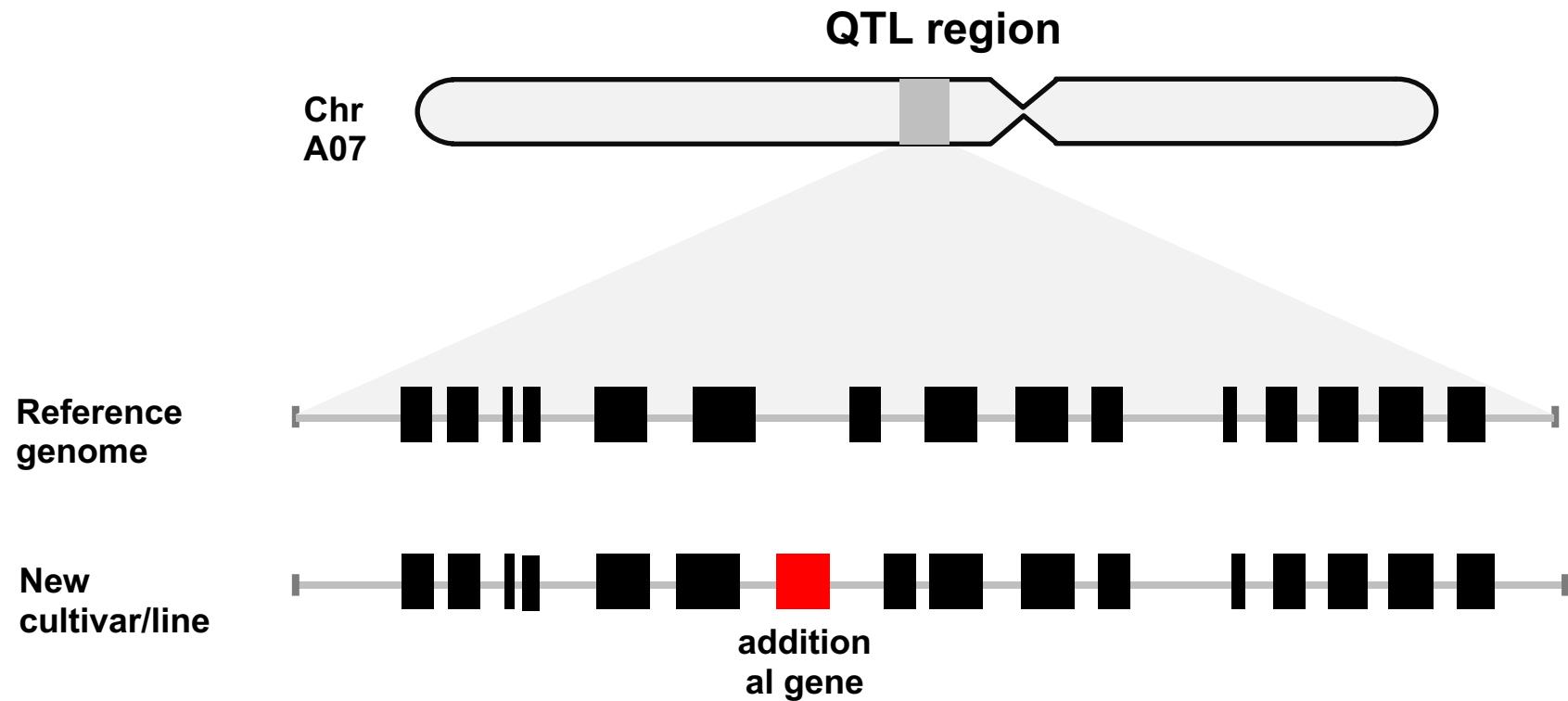
- Phenotypic variation is due to small changes (SNPs) in the sequence of genes (allelic variation)

Current model:

- Gene presence/absence variation is also a significant cause of phenotypic variation (PAV)



# Why a pangenome?





Bayer PE, Peterit J, Durant E, Monat C, Rouard M, Hu H, Chapman B, Li C, Cheng S, Batley J, Edwards D. (2022) Wheat Panache: A pangenome graph database representing presence-absence variation across sixteen bread wheat genomes. *The Plant Genome*. 15 (3): e20221

## THE UNIVERSITY OF WESTERN AUSTRALIA

Garg G, Kamphuis L, Bayer P, Kaur P, Dudchenko O, Taylor C, Frick K, Foley R, Gao LL, Lieberman Aiden E, Edwards D, Singh K. (2022) A pan-genome and chromosome-length reference genome of narrow-leaved lupin (*Lupinus angustifolius*) reveals genomic diversity and insights into key industry and biological traits. *The Plant Journal*. 111 (5): 1252-1266



Centre for  
Applied  
Bioinformatics

Varshney et al. (2021) A chickpea genetic variation map based on the sequencing of 3,366 genomes. *Nature*. 599: 622-627

Hurgobin et al. (2018) Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnology Journal*. 16 (7), 1265-1274

Khan et al. (2020) Super-pangenome by integrating the wild side of a species for accelerated crop improvement. *Trends in Plant Science*. 25 (2): 148-158

Ruperao et al. (2021) Sorghum pangenome explores the functional utility for genomic-assisted breeding to accelerate the genetic gain. *Frontiers in Plant Science*. 12:963

Zhao et al. (2020) Trait associations in the pangenome of pigeon pea (*Cajanus cajan*) *Plant Biotechnology Journal*. 18: 1946-1954

Bayer et al. (2017) Assembly and comparison of two closely related *Brassica napus* genomes. *Plant Biotechnology Journal*. 15 (12):1602-1610

Bayer et al. (2020) Plant pan-genomes are the new reference. *Nature Plants*. 6 (8): 1-7

Golicz A, Batley J and Edwards D. (2016) Towards plant pangenomics. *Plant Biotechnology Journal*. 14 (4):1099-105

Montenegro JDM, Golicz AA, Bayer PE, Hurgobin B, Lee HT, Chan CKK, Visendi P, Lai K, Doležel J, Batley J, Edwards D. (2017) The pangenome of modern hexaploid bread wheat. *Plant Journal*. 90 (5): 1007-1013

Hurgobin H and Edwards D. (2017) SNP discovery using a pangenome: has the single reference approach become obsolete? *Biology* 6 (1): E21

Khan et al. (2024) Cicer super-pangenome provides insights into species evolution and agronomic trait loci for crop improvement in chickpea (*Cicer arietinum*). *Nature Genetics* 56 (6):1225-1234

Zanini et al. (2021) Pangenomics in crop improvement – from coding SVs to finding regulatory variants with pangenome graphs. *The Plant Genome*. 15 (1): e20177

Edwards D and Batley J. (2022) Graph pangenomes find missing heritability. *Nature Genetics*. 54 (7): 919-920

Golicz et al. (2020) Pangenomics comes of age: From bacteria to plant and animal applications. *Trends in Genetics* 36(2): 132-145

Hu H, Li R, Zhao J, Batley J, Edwards D. (2024) Technological development and advances for constructing and analysing plant pangenomes. *Genome Biology and Evolution* 16(4): evae081

Wang J, Yang W, Zhang S, Hu H, Yuan Y, Dong J, Chen L, Ma Y, Yang T, Zhou L, Chen J, Liu B, Li C, Edwards D and Zhao J. (2023) A pangenome analysis pipeline provides insights into functional gene identification in rice. *Genome Biology*. 24(1):19

Yu et al. (2019) Insight into the evolution and functional characteristics of the pan-genome assembly from sesame landraces and modern cultivars. *Plant Biotechnology Journal*. 17 (5): 881-892

Bayer et al. (2021) The application of pangenomics and machine learning in genomic selection. *Plant Genome*. e20112.

Song B, Ning W, Wei D, Jiang M, Zhu K, Wang X, Edwards D, Odeny DA, Cheng S. (2023) Plant genome resequencing and population genomics: current status and future prospects. *Molecular Plant*. 16 (8): 1252-1268

Hu H, Wang J, Nie S, Zhao J, Batley J and Edwards D (2024) Plant pangenomics, current practice and future direction. *Agriculture Communications* (accepted March 2024)

Tay Fernandez et al. (2022) Expanding gene-editing potential in crop improvement with pangenomes. *International Journal of Molecular Sciences* 23(4): 2276.

Hu H, Scheben A, Verpaalen B, Tirnaz S, Bayer PE, Hodel R, Batley J, Soltis D, Soltis P, Edwards D. (2022) Amborella gene presence/absence variation is associated with abiotic stress responses that may contribute to environmental adaptation. *New Phytologist*. 233 (4): 1548-1555

Bayer et al. (2021) Modelling of gene loss propensity in the pangenomes of three *Brassica* species suggests different mechanisms between polyploids and diploids. *Plant Biotechnology Journal*. 19 (12): 2488-2500

Bayer et al. (2022) Sequencing the USDA core soybean collection reveals gene loss during domestication and breeding. *The Plant Genome*. 15: e20109

Peterit J, Bayer PE, Thomas WJW, Tay Fernandez C, Amas J, Zhang Y, Batley J, Edwards D. (2022) Pangenomics and crop genome adaptation in a changing climate. *Plants*. 11 (15): 1949

Dolatabadian A, Bayer P, Tirnaz S, Hurgobin B, Edwards D, Batley J. (2020) Characterisation of disease resistance genes in the *Brassica napus* pangenome reveals significant structural variation. *Plant Biotechnology Journal*. 18 (4): 969-982

Golicz et al. (2016) The pangenome of an agriculturally important crop *Brassica oleracea*. *Nature Communications* 7:13390

Rijaani H, Bayer PE, Rouard M, Doležel J, Batley J, Edwards D. (2022) The pangenome of banana highlights differences between genera and genomes. *Plant Genome*. 15: e20100

Wang et al. (2021). The chicken pan-genome reveals gene content variation and a regulatory region deletion in IGF2BP1 affecting body size. *Molecular Biology and Evolution*. 38 (11): 5066–5081

Danilevitz et al. (2020) Plant Pangenomics: Approaches, Applications and Advancements. *Current Opinion in Plant Biology*. 54: 15-25

Amas J, Bayer P, Hong Tan W, Tirnaz S, Thomas W, Edwards D, Batley J, (2023) Comparative pangenome analyses provide insights into the evolution of *Brassica rapa* resistance gene analogs (RGAs). *Plant Biotechnology Journal*. 21(10) 2100-2112

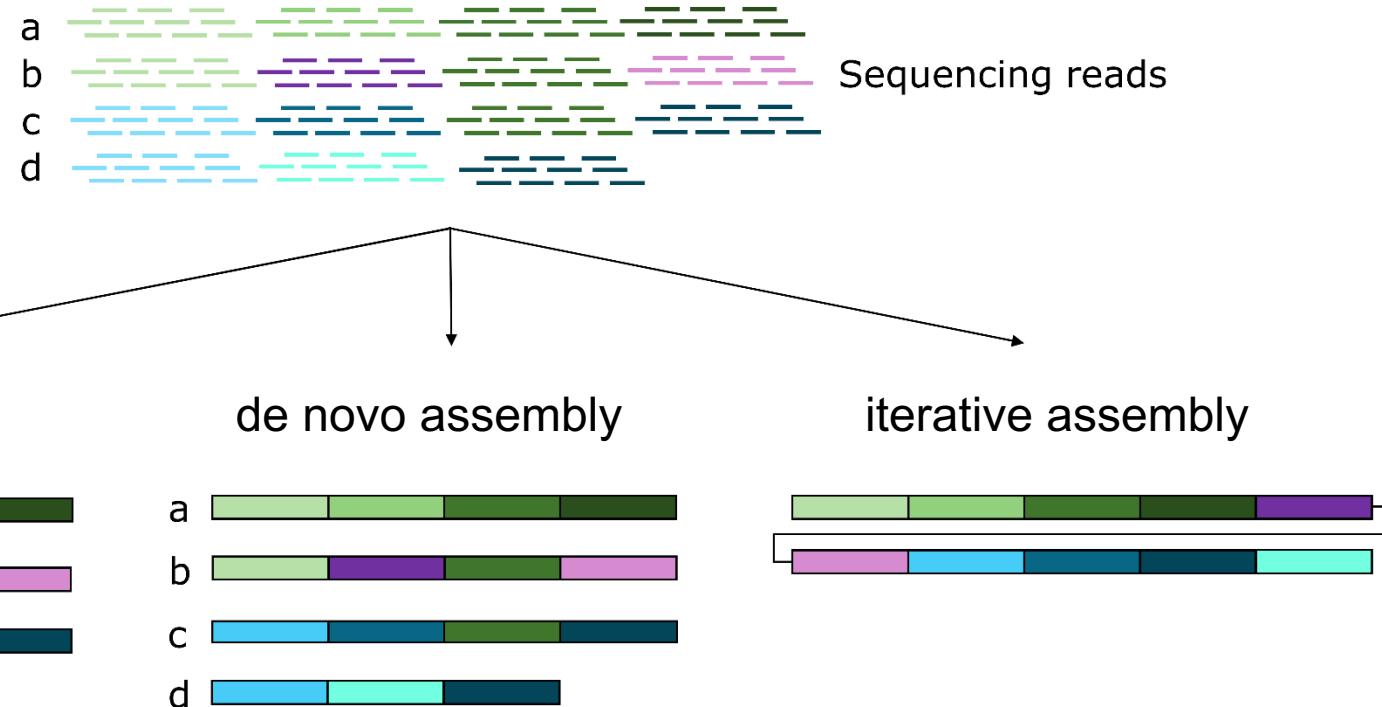
Edwards D and Batley J. (2023) Teatime for pangenomics. *Nature Plants* 9(12): 1944-1945

Tay Fernandez et al. (2022) Pangenomes as a resource to accelerate breeding of under-utilised crop species. *International Journal of Molecular Sciences* 23 (5): 2671



THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Building a pangenome





# *De novo assembly*



1. Generate high coverage sequence data for each individual
2. Assemble each independently
3. Annotate each assembly
4. Compare gene content
  - Issues:
    - Very similar genomes can have quite different annotations
    - Differences in assembly or annotation may mask real differences



# Iterative assembly



1. Start with a reference
2. Map reads to the reference
3. Assemble reads which don't map, add new contigs to the assembly
4. Repeat 2 and 3 with additional individuals (single or pooled)
5. Place new contigs where possible using PE and MP data, or LD, or genetic map
6. For each variety, perform presence/absence assessment on the reference

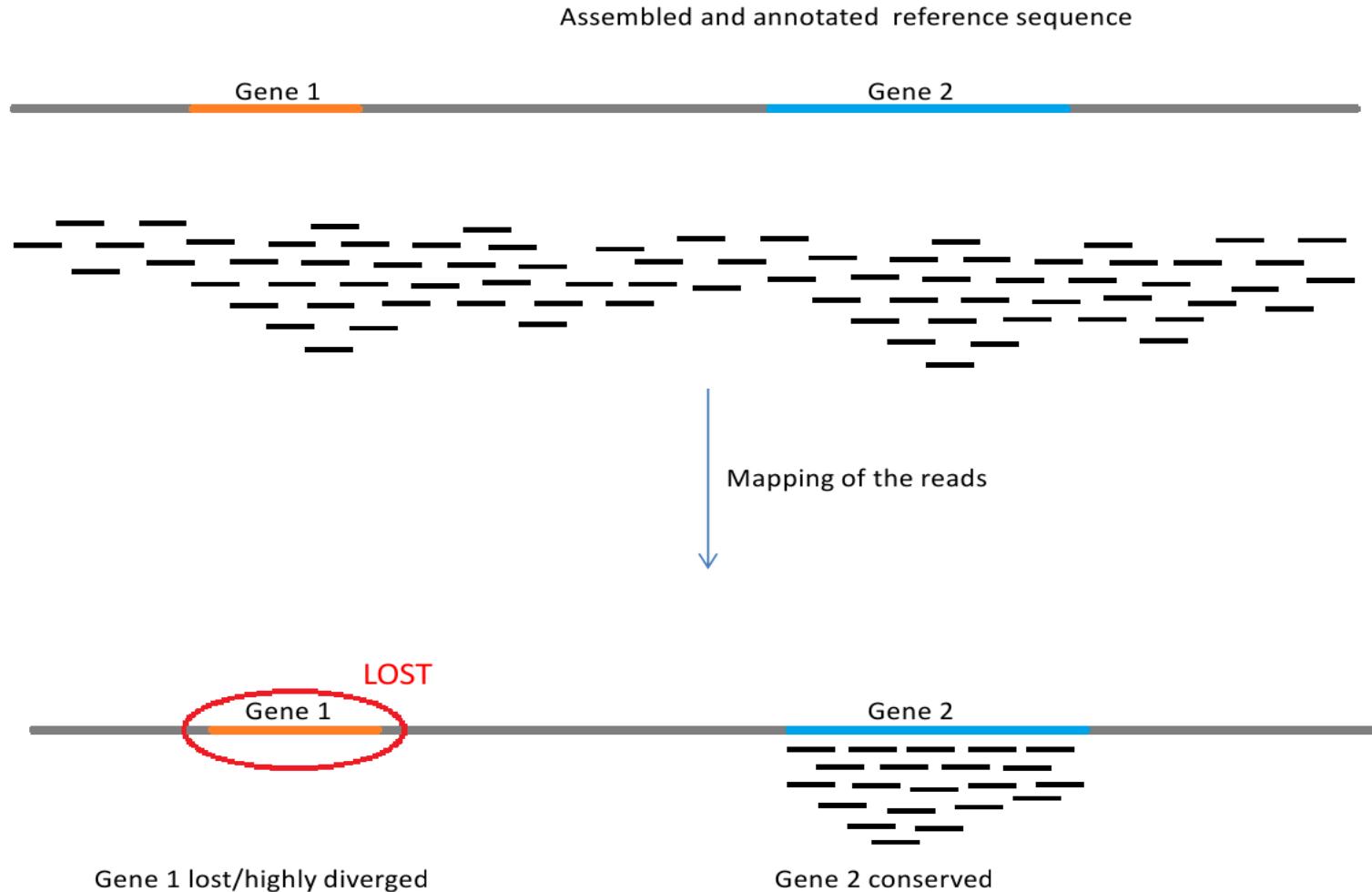


THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Presence/absence (SGSgeneloss)

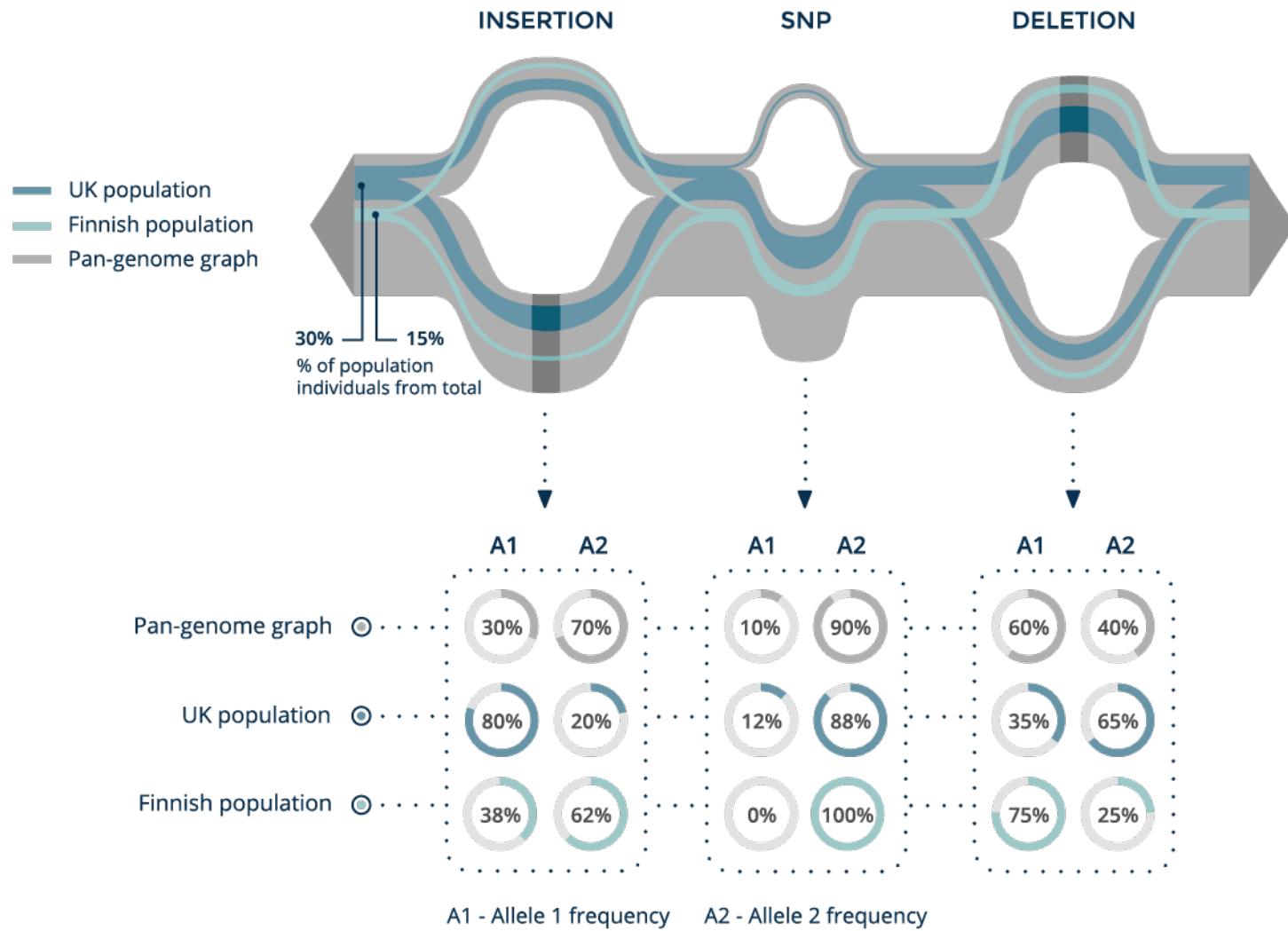


Centre for  
Applied  
Bioinformatics





# Population graphs





THE UNIVERSITY OF  
WESTERN AUSTRALIA

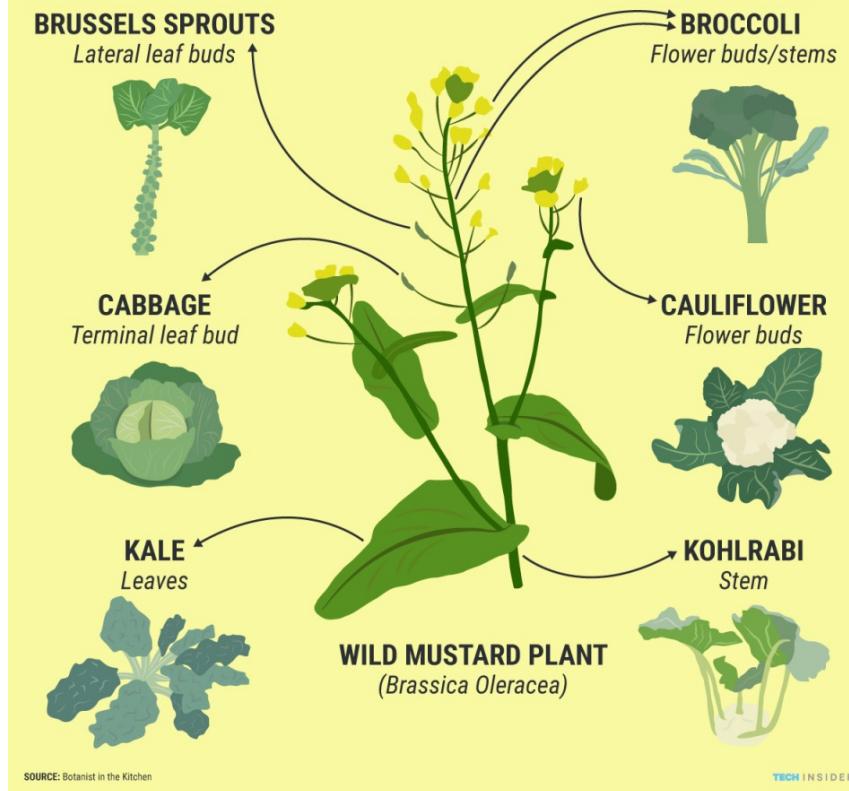
# Presence/absence in *B. oleracea*



Centre for  
Applied  
Bioinformatics

## 6 vegetables that are the same plant

Over hundreds of years farmers have been breeding one plant – called *Brassica Oleracea* – into dozens of different varieties. These six vegetables you can find in the grocery store are actually all the same plant.



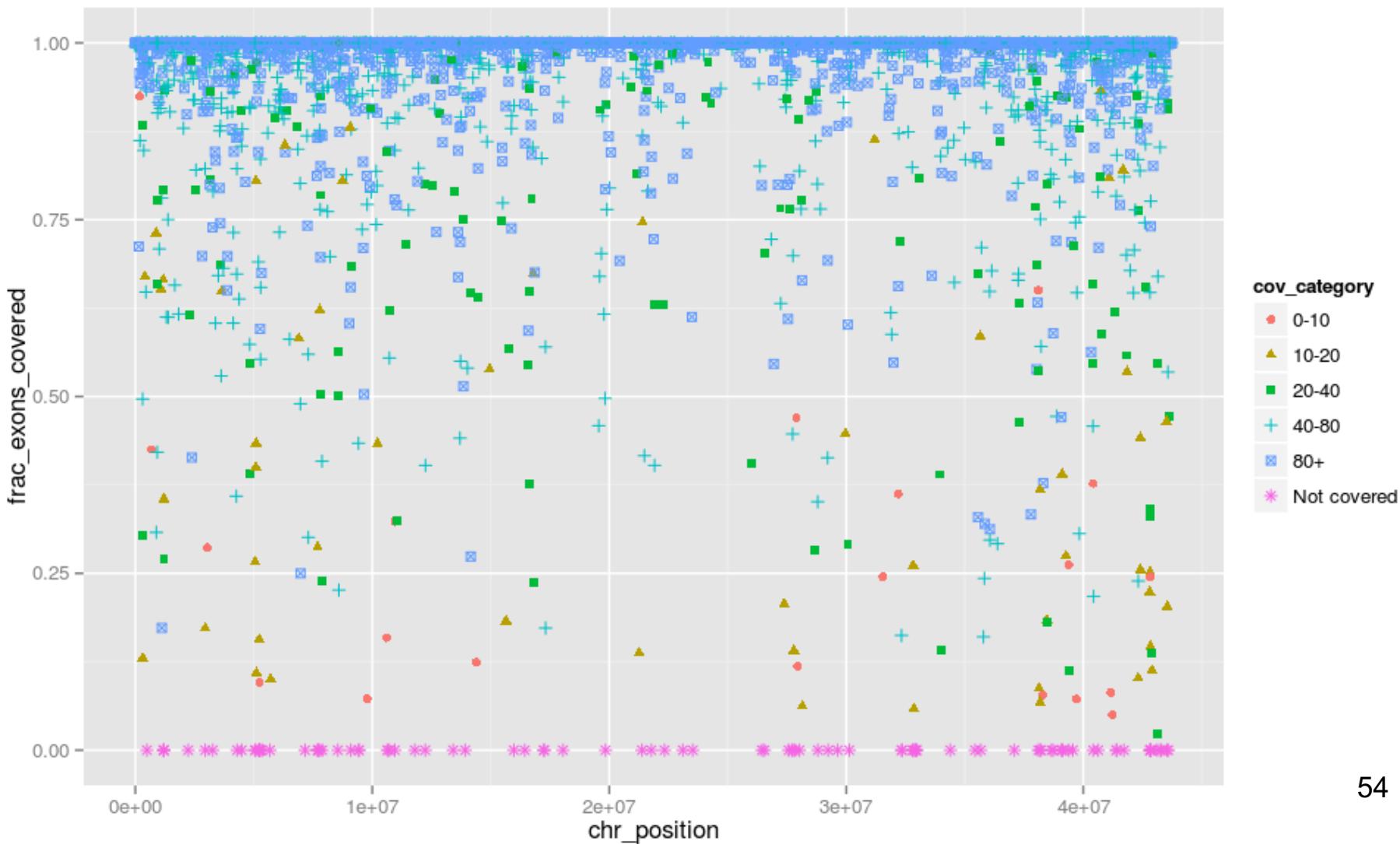


THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Cabbage



Centre for  
Applied  
Bioinformatics





THE UNIVERSITY OF  
WESTERN AUSTRALIA

# *Brassica oleracea* pangenome



Centre for  
Applied  
Bioinformatics

## ARTICLE

Received 22 Aug 2016 | Accepted 28 Sep 2016 | Published 11 Nov 2016

DOI: 10.1038/ncomms13390

OPEN

## The pangenome of an agronomically important crop plant *Brassica oleracea*

Agnieszka A. Golicz<sup>1</sup>, Philipp E. Bayer<sup>2</sup>, Guy C. Barker<sup>3</sup>, Patrick P. Edger<sup>4</sup>, HyeRan Kim<sup>5</sup>, Paula A. Martinez<sup>1</sup>, Chon Kit Kenneth Chan<sup>2</sup>, Anita Severn-Ellis<sup>2</sup>, W. Richard McCombie<sup>6</sup>, Isobel A.P. Parkin<sup>7</sup>, Andrew H. Paterson<sup>8</sup>, J. Chris Pires<sup>9</sup>, Andrew G. Sharpe<sup>10</sup>, Haibao Tang<sup>11</sup>, Graham R. Teakle<sup>3</sup>, Christopher D. Town<sup>12</sup>, Jacqueline Batley<sup>2</sup> & David Edwards<sup>2</sup>

## Diverse morphotypes

- Cabbage (2)
- Cauliflower (2)
- Broccoli
- Kale
- Brussels sprout
- Rapid cybler TO1000
- *B. macrocarpa*



THE UNIVERSITY OF  
WESTERN AUSTRALIA

# *Brassica oleracea* pangenome



Centre for  
Applied  
Bioinformatics

Previous reference genomes:

*B. oleracea* TO1000      54,458 genes, 488 Mbp  
*B. oleracea* var. *capitata*    45,758 genes, 535 Mbp

Pangenome      61,379 genes, 587 Mbp,

18.7% of genes are variable

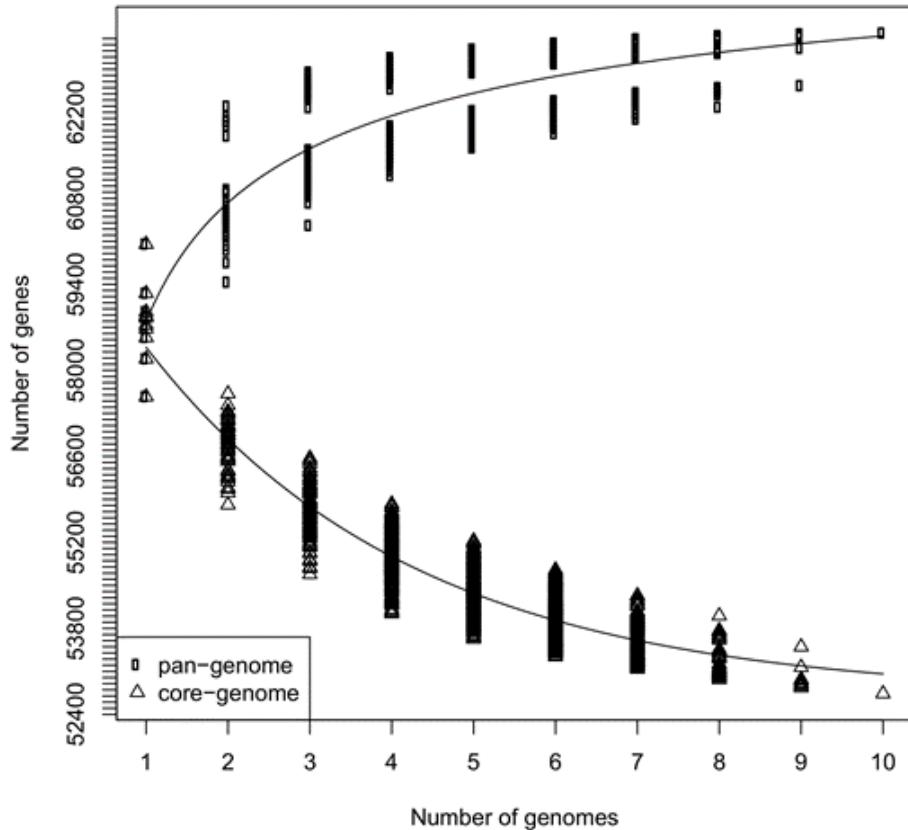


THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Brassica pangenome



Adding more genomes captures more genes and defined the core and variable genome



*Brassica oleracea*

Predicted pangenome of  $61,198 \pm 394$  genes ( $35,462 \pm 250$  gene families)

Predicted core genome size of  $49,676 \pm 96$  genes ( $28,489 \pm 51$  gene families)



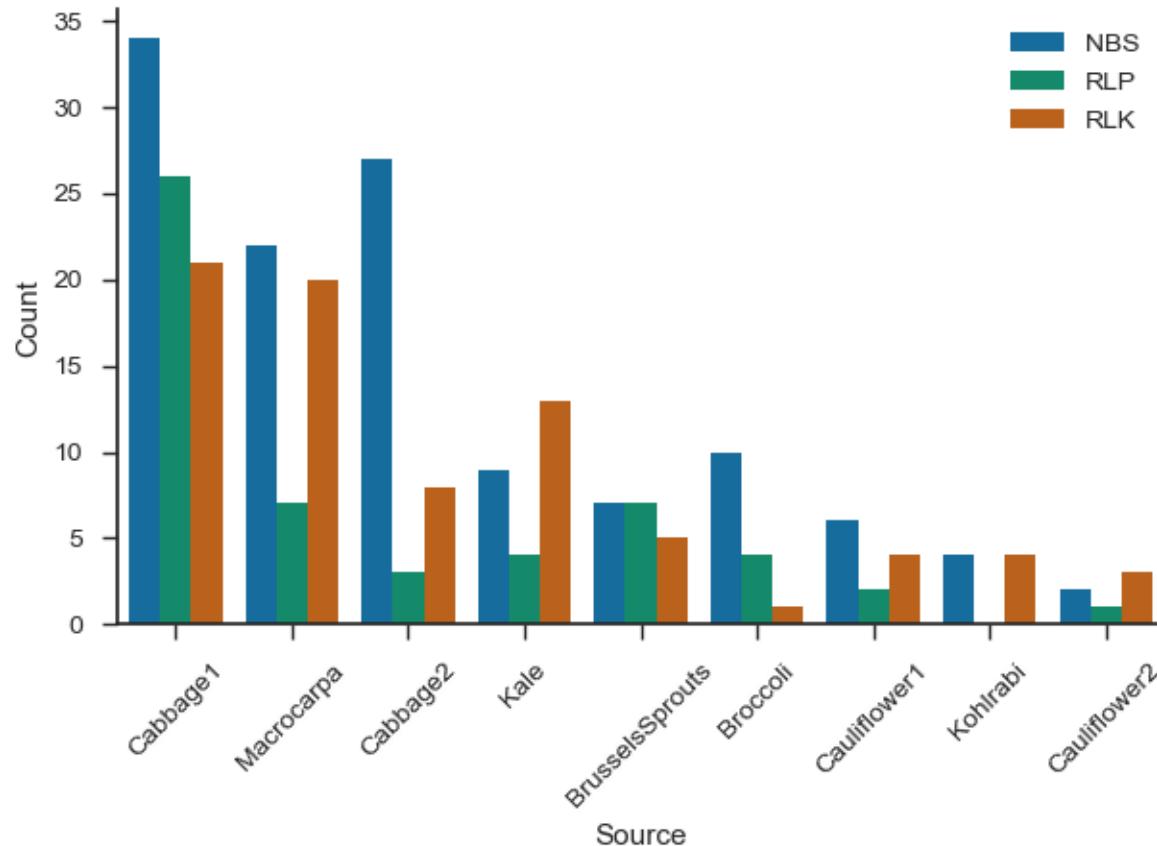
# Brassica pangenome



SCF-dependent proteasomal ubiquitin-dependent protein catabolic process  
defense response to oomycetes  
systemic acquired resistance, salicylic acid mediated signaling pathway  
detection of external stimulus  
defense response signaling pathway, resistance gene-dependent  
regulation of hydrogen peroxide metabolic process  
RNA 5'-end processing response to virus  
response to molecule of bacterial origin  
response to bacterium cellular water homeostasis  
sesquiterpene biosynthetic process  
defense response to bacterium  
defense response to bacterium, incompatible interaction  
defense response signaling pathway, resistance gene-independent  
positive regulation of defense response to virus by host  
cGMP biosynthetic process wax biosynthetic process  
cAMP biosynthetic process MAPK cascade  
transport of virus in host, tissue to tissue  
regulation of protein dephosphorylation  
transmembrane receptor protein tyrosine kinase signaling pathway  
response to nickel cation



# Different resistance genes

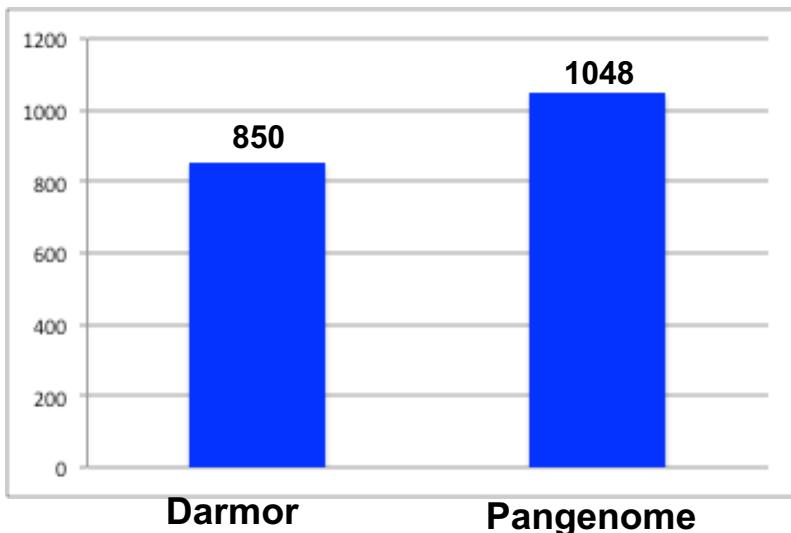




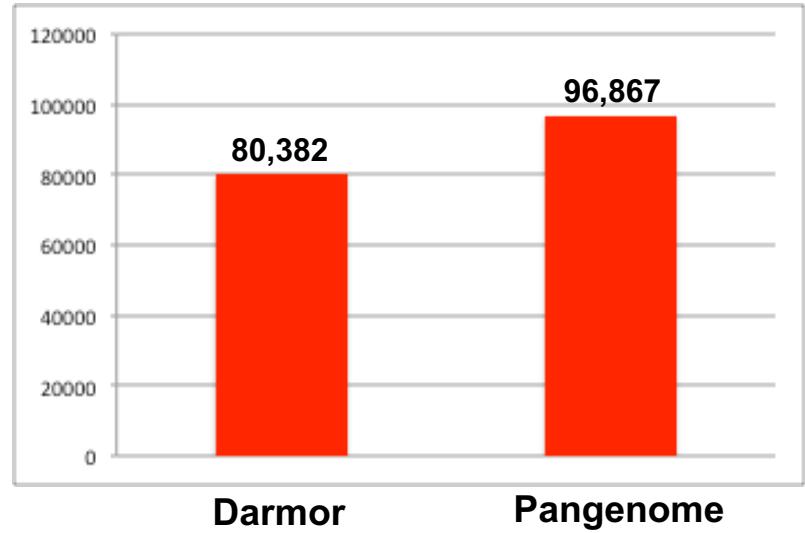
# Results in *Brassica napus*

- 33 *B. napus*
- 20 resynthesised *B. napus*

Assembly size (Mb)

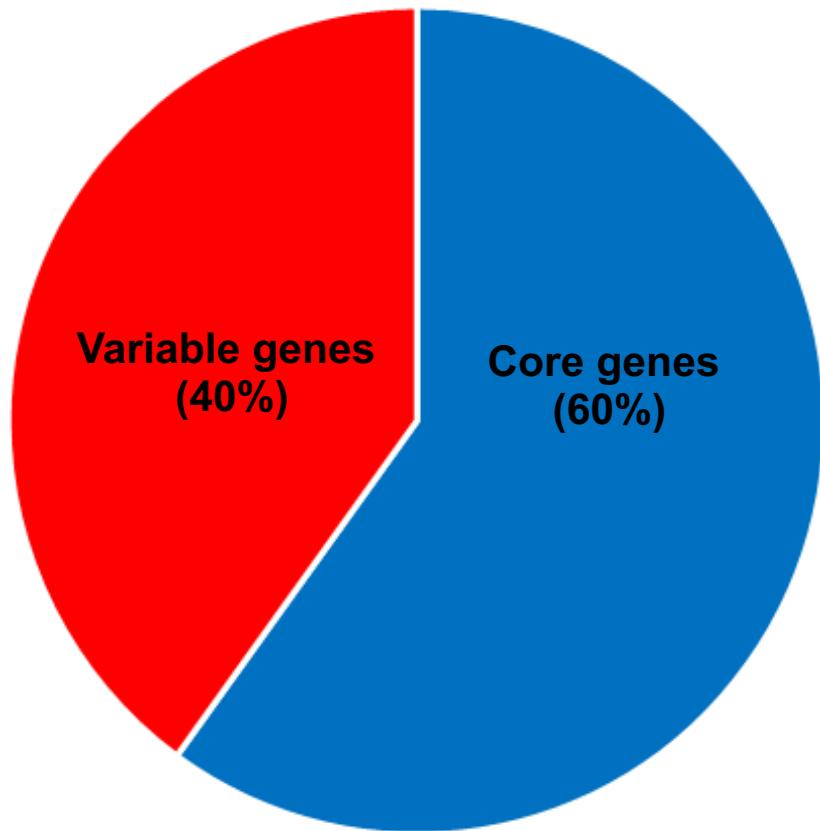


No. of genes





# Results in *Brassica napus*



- Variable genes are shorter and have fewer exons than core genes
- Number of variable genes indicates how much diversity is present



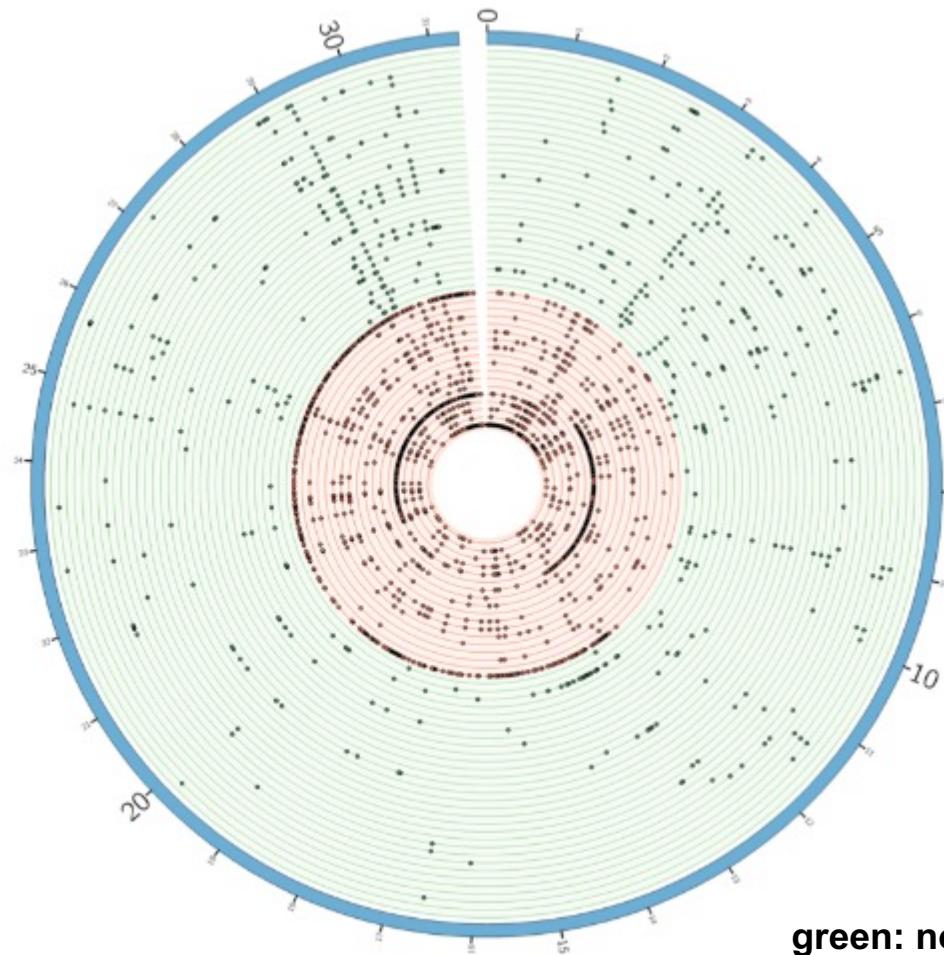
THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Results in *Brassica napus*

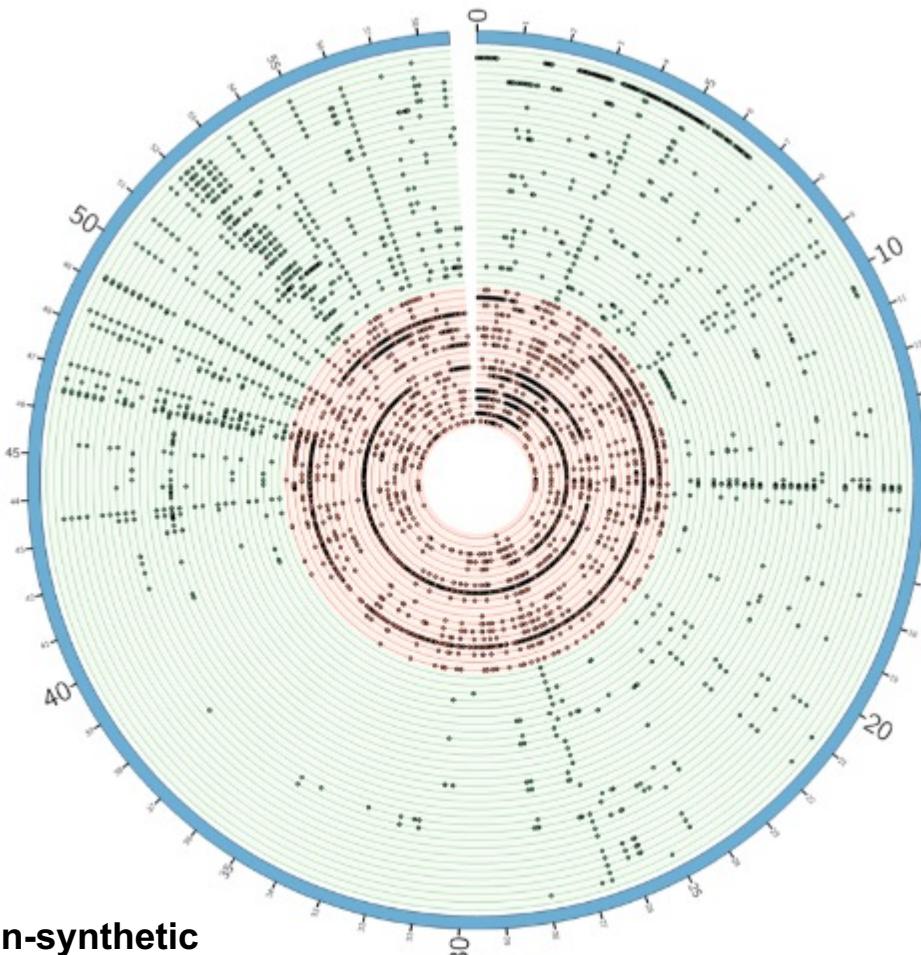


Centre for  
Applied  
Bioinformatics

A02



C02



green: non-synthetic  
red: synthetic



# Functional annotation of variable genes



regulation of cellular metabolic process  
monocarboxylic acid transport  
transport of virus in host, tissue to tissue  
regulation of defense response by callose deposition  
positive regulation of defense response  
DNA topological changes single-organism process  
cGMP biosynthetic process translation  
sesquiterpenoid biosynthetic process  
immune response-regulating signaling pathway  
**regulation of innate immune response**  
nucleic acid metabolic process

## plant-type hypersensitive response

defense response to virus  
defense response signaling pathway  
sesquiterpene biosynthetic process  
RNA 5'-end processing acetyl-CoA biosynthetic process  
cAMP biosynthetic process response to virus  
defense response to bacterium  
protein deglycosylation  
cellular macromolecule metabolic process  
positive regulation of transport  
meiotic DNA double-strand break formation



# Variable *R*-genes are located in known in disease QTL

*B. rapa*



Amas et al. 2023 (Plant Biotechnology Journal)

Downy mildew (7)  
Blackleg (4)  
Clubroot (10)  
TuMV (4)  
Fusarium wilt (2)

*B. napus*



Dolatabadian et al., 2020 (Plant Biotechnology Journal)

Blackleg (22)

*B. oleracea*



Bayer et al., 2018 (Plant Biotechnology Journal)

Black rot (7)  
*Sclerotinia* (20)  
Clubroot (14)

CORE

VARIABLE

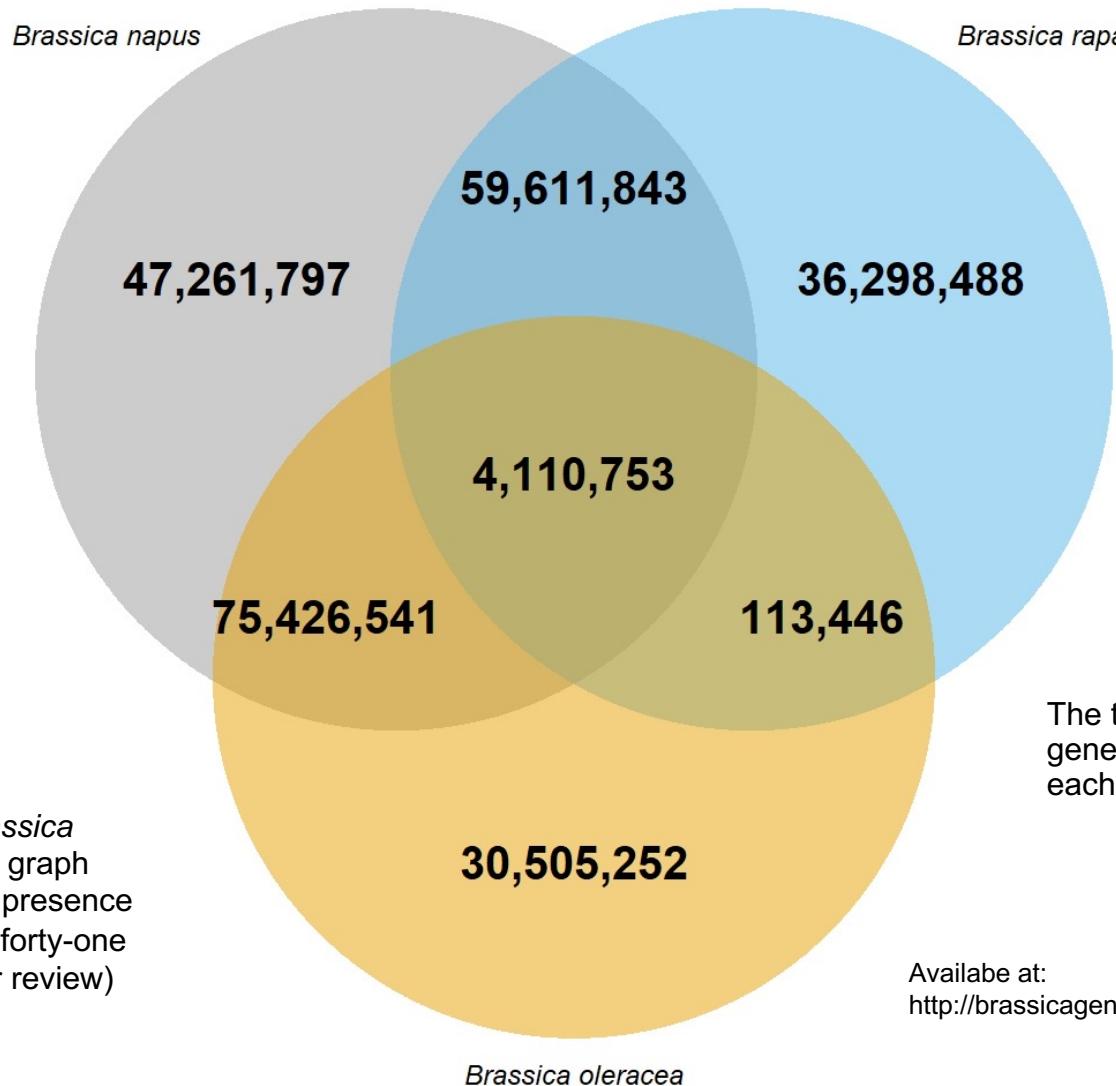


THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Brassica graph pangenome



Centre for  
Applied  
Bioinformatics

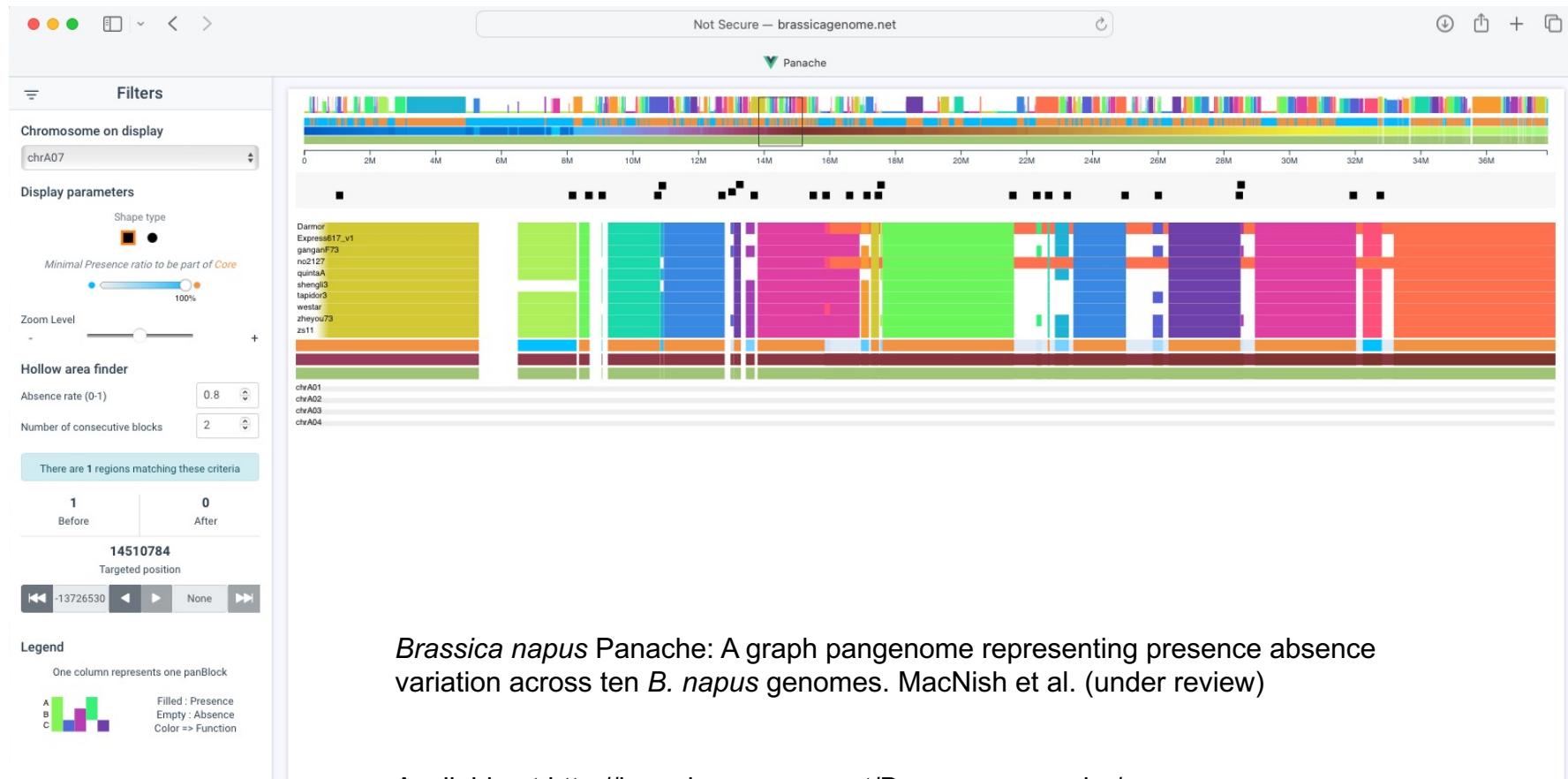


MacNish et al. (2024) *Brassica Panache*: A multi-species graph pangenome representing presence absence variation across forty-one *Brassica* genomes (under review)

Available at:  
[http://brassicagenome.net/brassica\\_panache](http://brassicagenome.net/brassica_panache)



# *B. napus* graph pangenome



Available at [http://brassicagenome.net/B\\_napus\\_panache/](http://brassicagenome.net/B_napus_panache/)

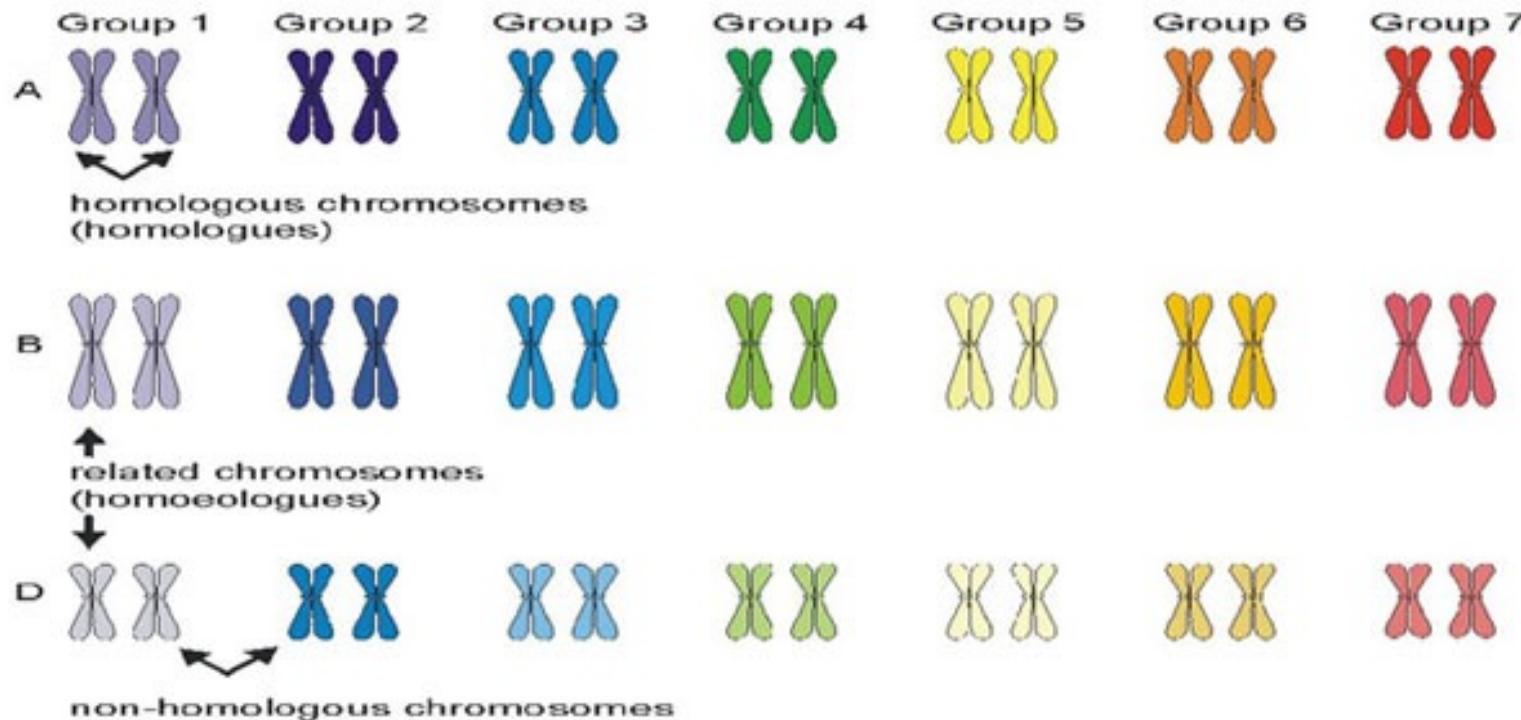


THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Wheat genome



Centre for  
Applied  
Bioinformatics



17 billion bases

[http://www.jic.ac.uk/staff/graham-moore/wheat\\_meiosis.htm](http://www.jic.ac.uk/staff/graham-moore/wheat_meiosis.htm)



THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Bread wheat pangenome

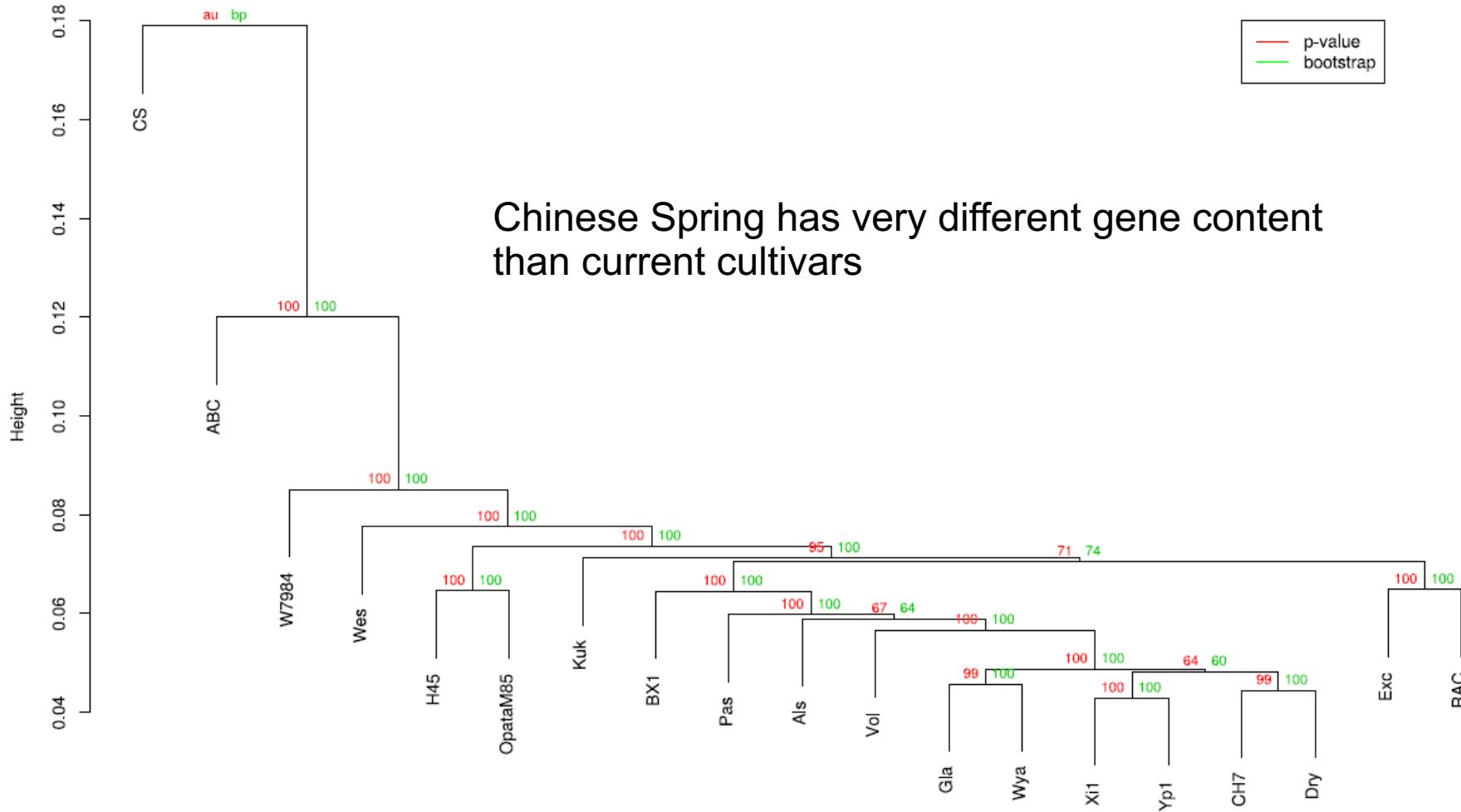


Centre for  
Applied  
Bioinformatics

- Pangenome of 18 elite cultivars plus Chinese Spring
- 350 Mb additional to Chinese Spring
- Average of 128,656 genes in each cultivar
- Pangenome size of 140,500 +/- 102 genes
- Core genome of 81,070 +/- 1,631 genes
- Chinese Spring very different from elite varieties
  - 245 genes in Chinese Spring which are absent from the 18 cultivars
  - 12,150 genes were identified in all 18 cultivars but are not found in Chinese Spring



# Bread wheat pangenome





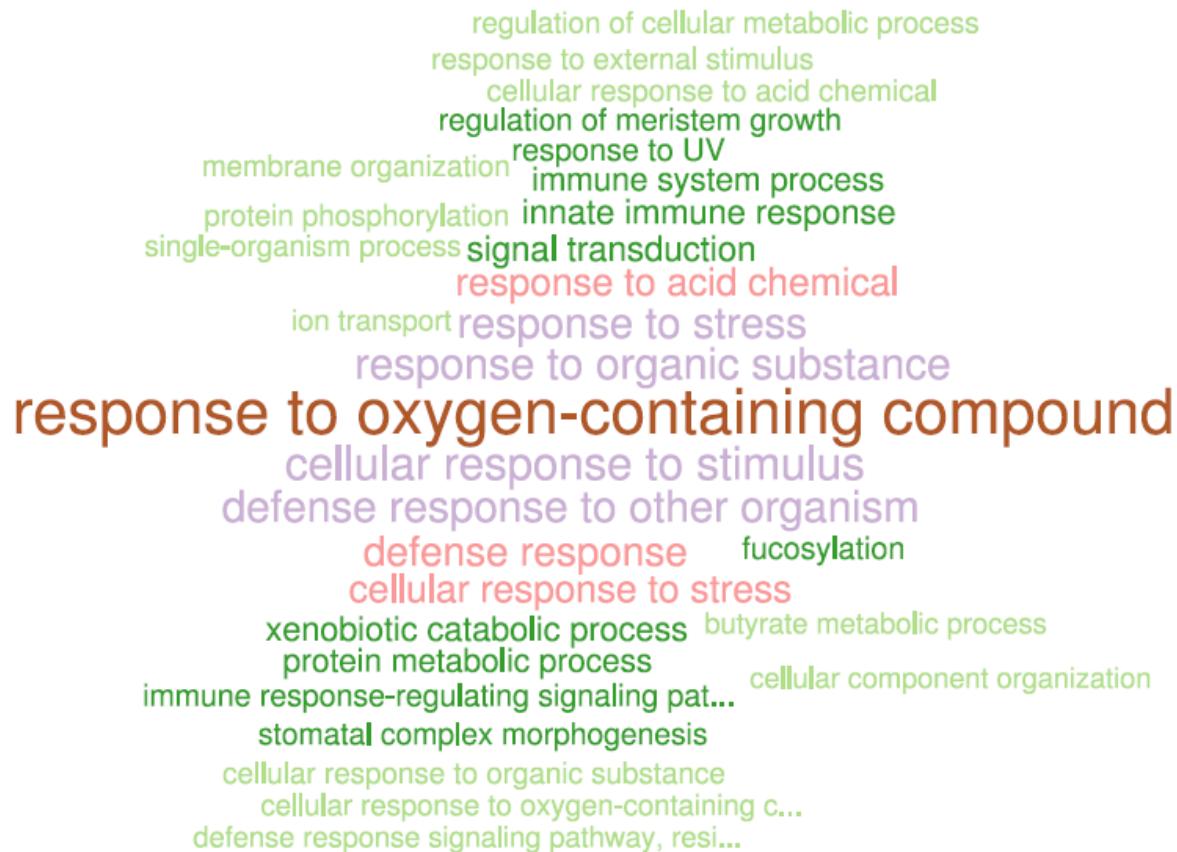
THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Bread wheat pangenome



Centre for  
Applied  
Bioinformatics

GO terms  
enriched in  
variable genes





THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Bread wheat pangenome



Centre for  
Applied  
Bioinformatics

**Wheat Panache: A pangenome graph database representing presence-absence variation across sixteen bread wheat genomes**

(2022)

Philipp E. Bayer<sup>1</sup> | Jakob Peterer<sup>1</sup> | Éloi Durant<sup>2,3,4,5</sup> | Cécile Monat<sup>3</sup> |  
Mathieu Rouard<sup>4,5</sup> | Haifei Hu<sup>6</sup> | Brett Chapman<sup>6</sup> | Chengdao Li<sup>6</sup> |  
Shifeng Cheng<sup>7</sup> | Jacqueline Batley<sup>1</sup> | David Edwards<sup>1</sup>



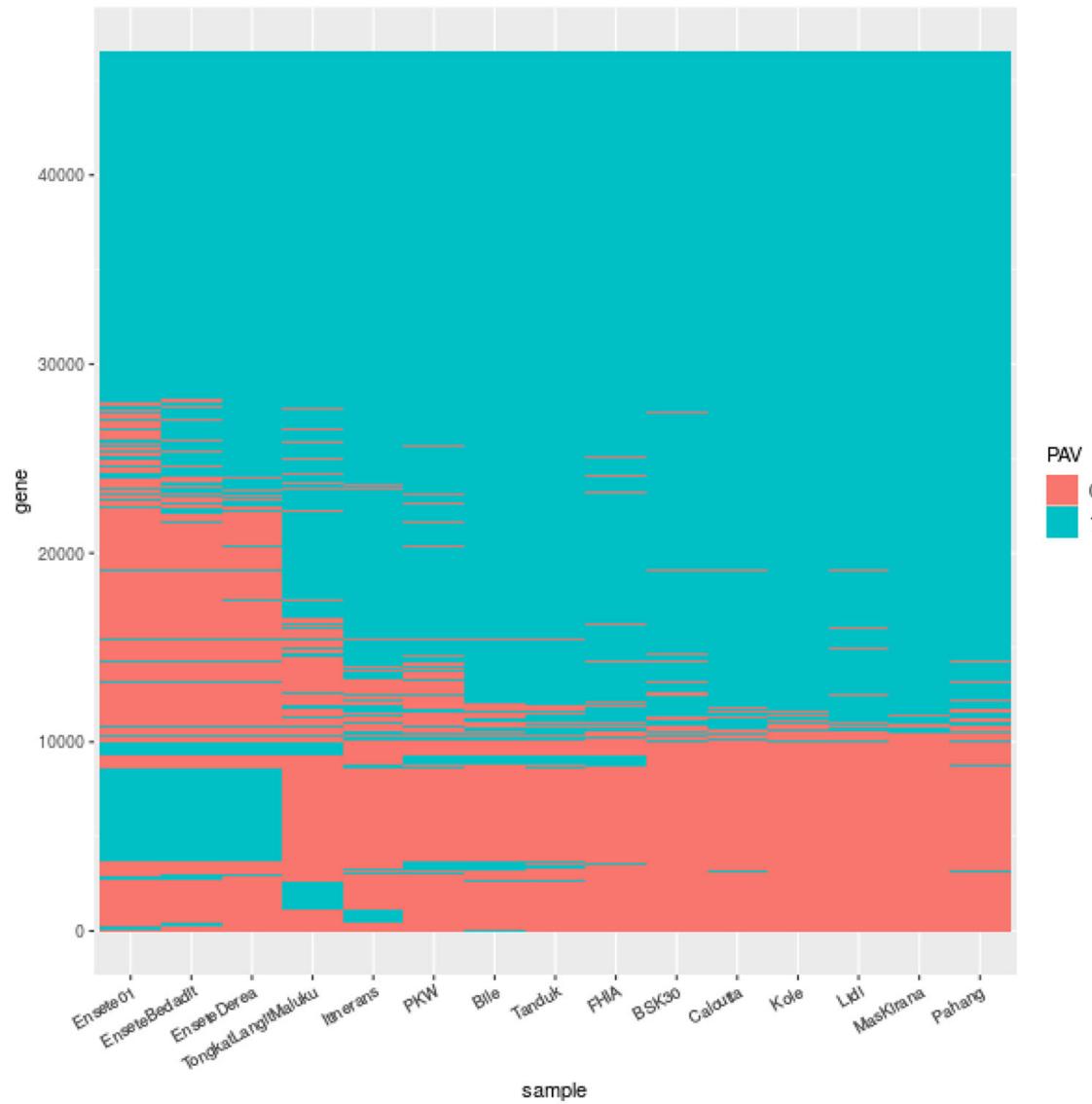


THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Banana pangenome



Rijzaani et al. (2022)  
Plant Genome



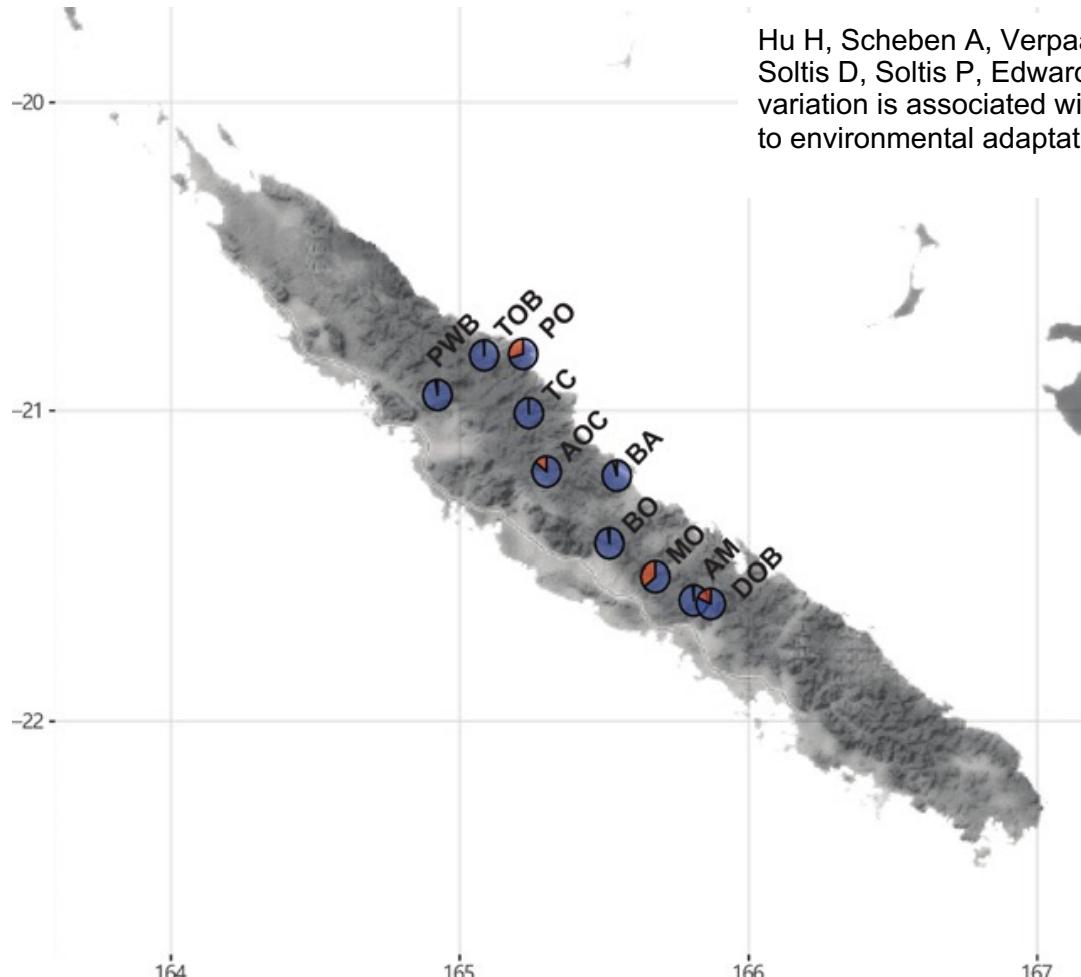


THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Amborella pangenome



Centre for  
Applied  
Bioinformatics



Hu H, Scheben A, Verpaalen B, Tirnaz S, Bayer PE, Hodel R, Batley J, Soltis D, Soltis P, Edwards D. (2022) Amborella gene presence/absence variation is associated with abiotic stress responses that may contribute to environmental adaptation. *New Phytologist*. 233 (4): 1548-1555



THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Soybean pangenome



Centre for  
Applied  
Bioinformatics

- > 1000 individuals from US national collection
- Wild, landrace, old and new cultivars
- Showed selection and loss of genes during domestication and breeding

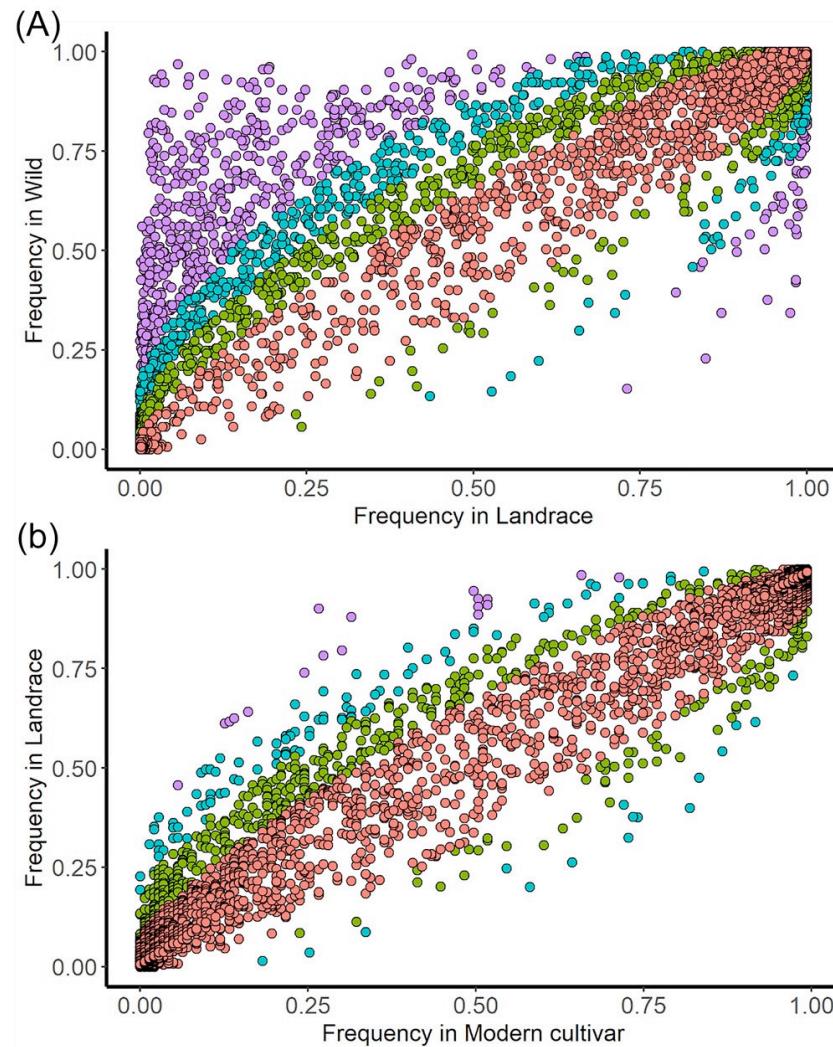


THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Soybean pangenome



Centre for  
Applied  
Bioinformatics



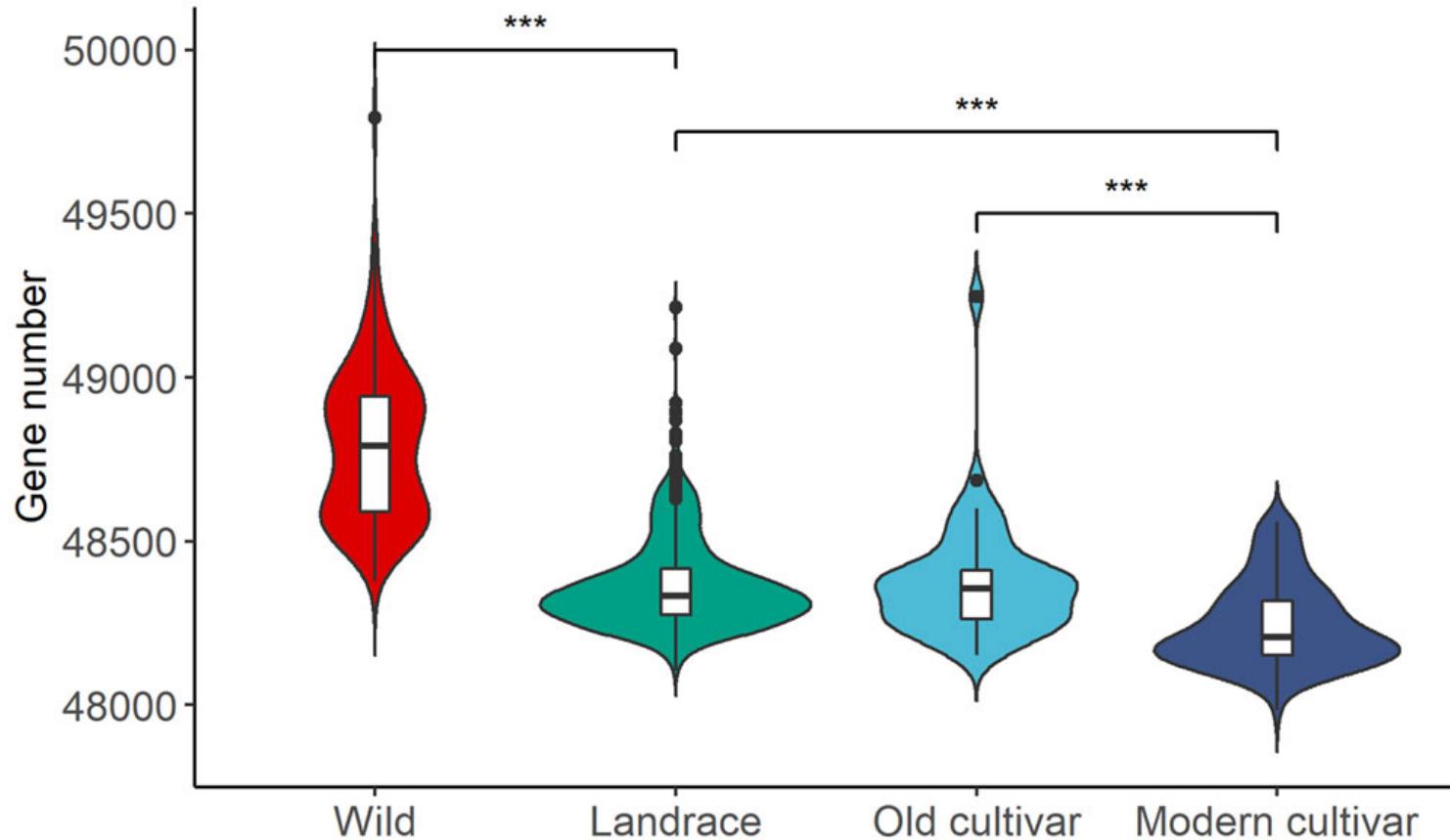


THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Soybean pangenome



Centre for  
Applied  
Bioinformatics





# The future of graph pangenomes



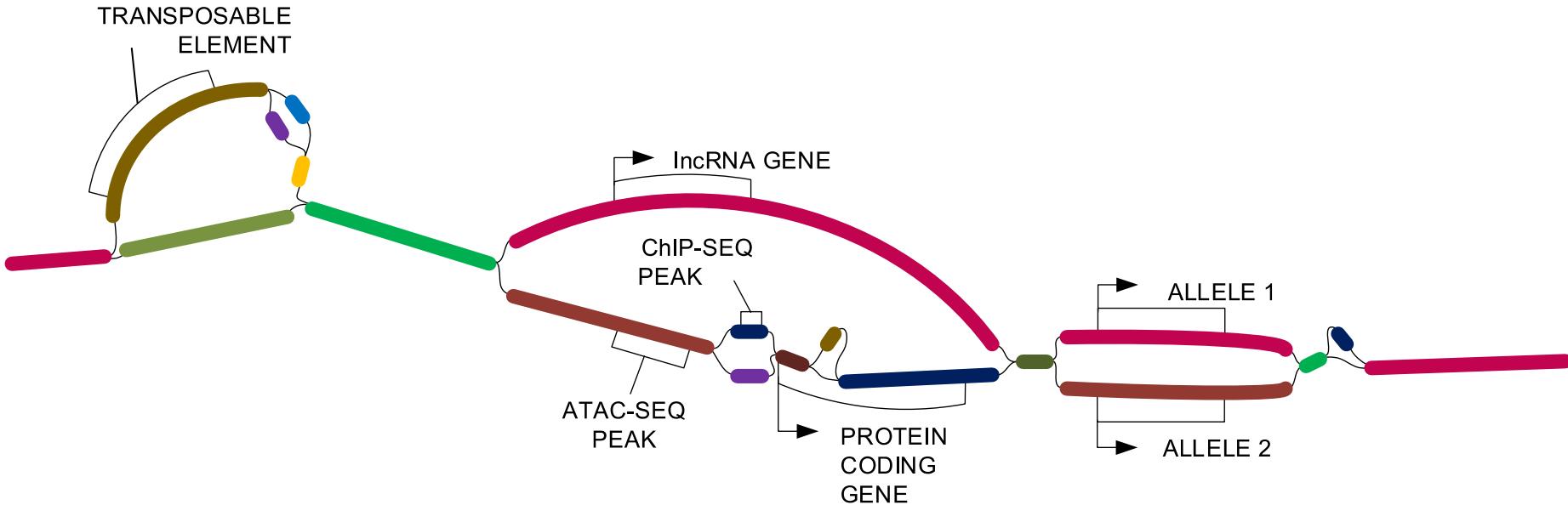
PLANT GENOMICS

## Graph pangenomes find missing heritability

The use of association studies to identify candidate genes for complex biological traits in plants has been challenging due to a reliance on single reference genomes, leading to missing heritability. Graphical pangenomes and the identification of causal variants help overcome this and provide an important advance for crop breeding.

David Edwards and Jacqueline Batley

Edwards and Batley (2022) Nature Genetics 54: 919-920





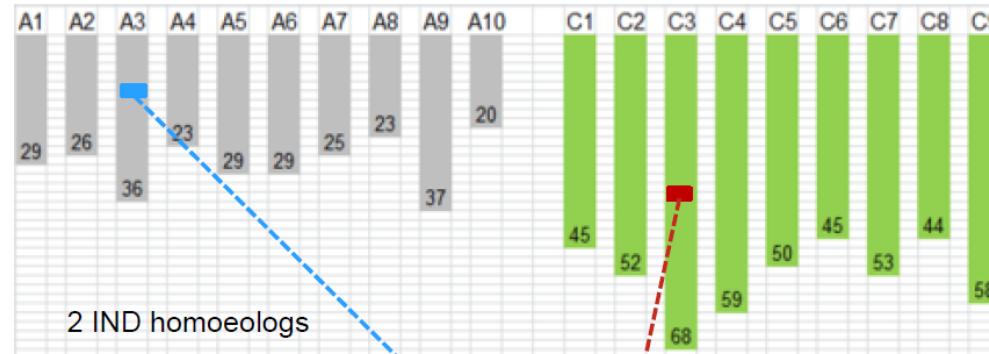
THE UNIVERSITY OF  
WESTERN AUSTRALIA

# PodGuard – a candidate gene approach



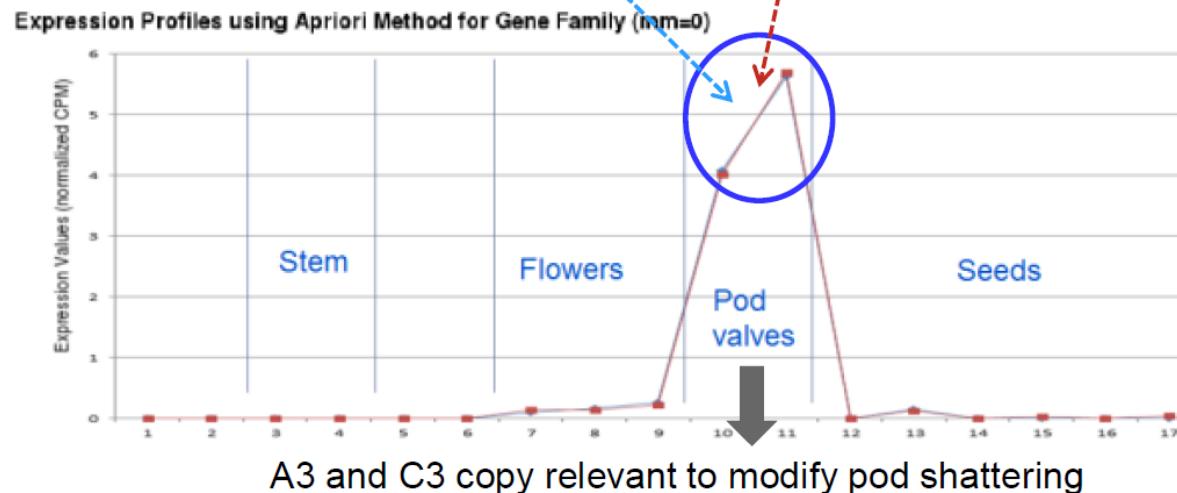
Centre for  
Applied  
Bioinformatics

*B.napus*  
genome sequence



*B.napus*  
transcript atlas

- 17 tissue-time points



Bart Lambert





THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Crispr genome editing



Centre for  
Applied  
Bioinformatics

Gene knockout – targets gene, cuts DNA and repair causes deletions

Gene editing – specifically modify bases using template DNA

Extremely powerful way to generate novel and specific variation



THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Potential of Crispr crop production



Centre for  
Applied  
Bioinformatics

- Accelerate crop production
- Domesticate new crops
- Rapidly adapt elite varieties to new/changing environments
- Add novel diversity and traits that are unavailable in wild or domesticated species
- More acceptable than transgenic crops



# Potential of Crispr crop production



## Crop improvement schemes

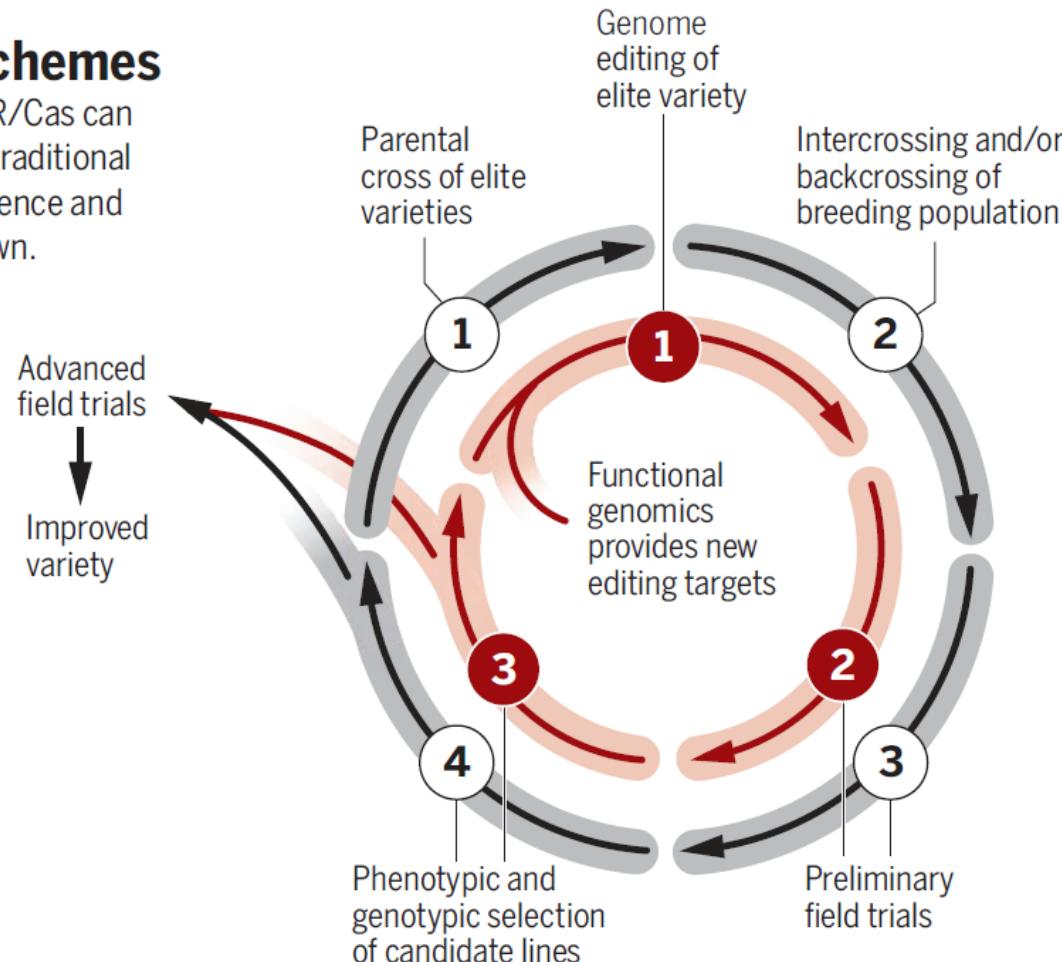
Gene editing tools such as CRISPR/Cas can improve crops more quickly than traditional approaches if the nucleotide sequence and function of the target site are known.

### Conventional crop breeding cycle

Crop traits are combined via recombination over multiple generations to produce improved varieties.

### CRISPR/Cas-assisted crop breeding cycle

Crops with different edits of known targets are produced in a single step, and selected for advanced trials based on phenotypic traits.





THE UNIVERSITY OF  
WESTERN AUSTRALIA

# CRISPR crops?



Genome editing based breeding

need to identify gene content

need to be able to predict the impact of modifying the sequence of genes

continuous editing and evaluation cycles

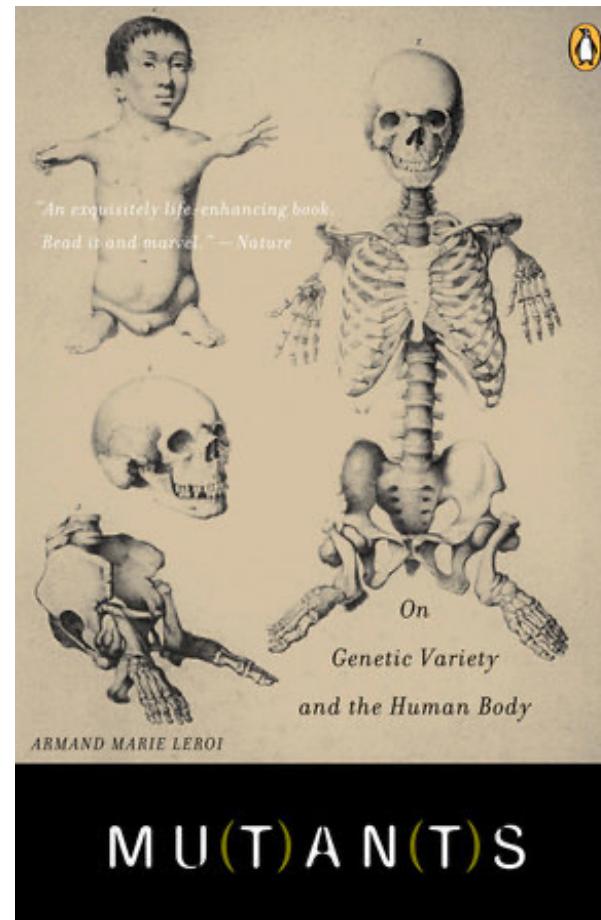
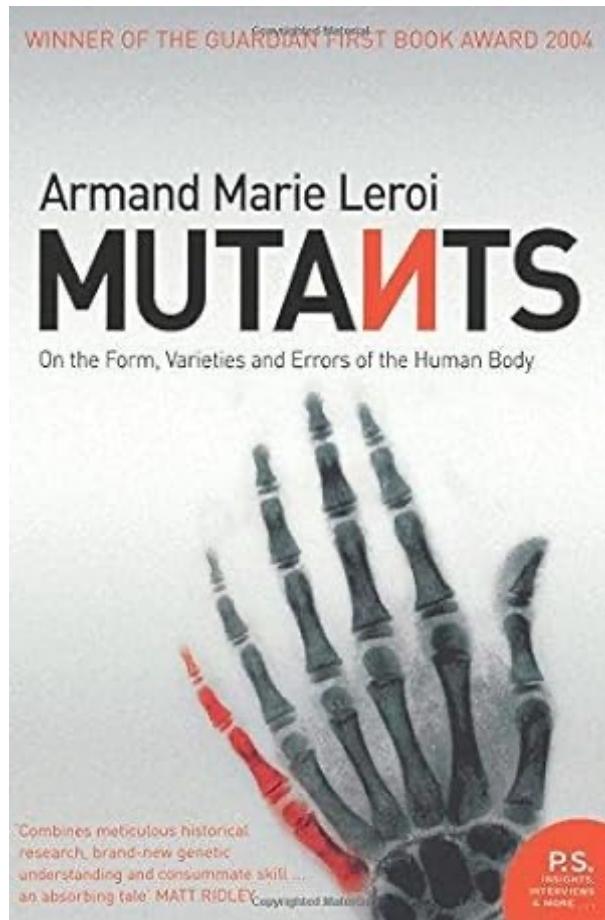


THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Mutant people



Centre for  
Applied  
Bioinformatics





THE UNIVERSITY OF  
WESTERN AUSTRALIA

# CRISPR people?



We can now change the genomics of the human population by

- Selecting embryos (designer babies)
- Editing the genomes of human germ cells
- Editing the tissues of adult humans

Designed to achieve more desired looks, skills, talent, health

Societal views change over time



THE UNIVERSITY OF  
WESTERN AUSTRALIA

# Designer babies



Centre for  
Applied  
Bioinformatics





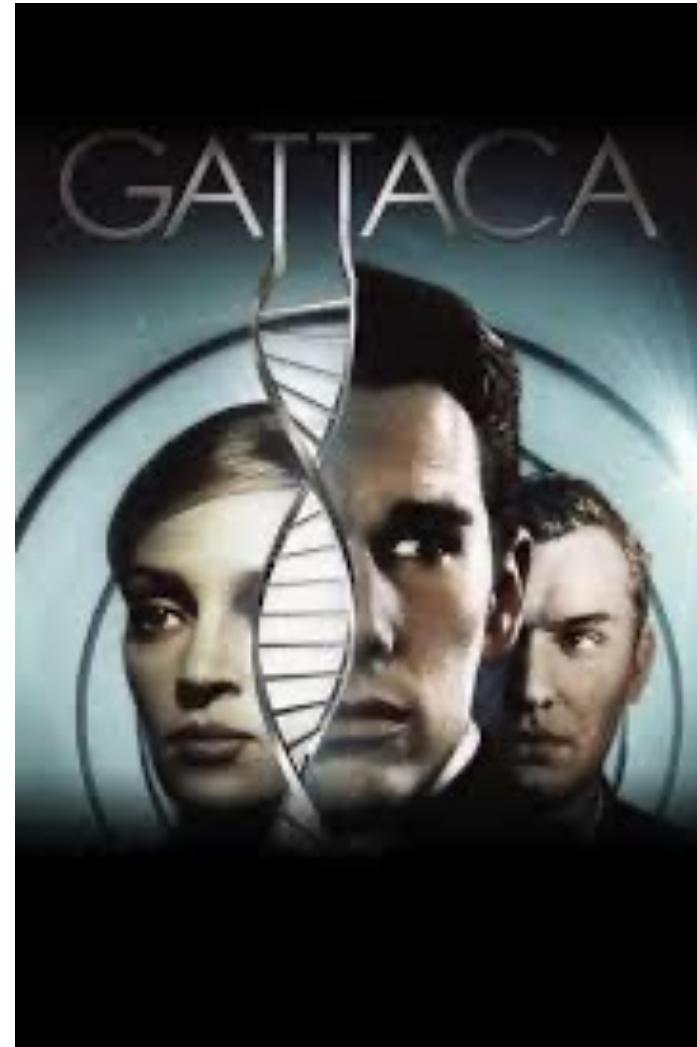
THE UNIVERSITY OF  
WESTERN AUSTRALIA

# CRISPR people?



Centre for  
Applied  
Bioinformatics

Designer society





# Discussion



- What is the value of a genome?
- What is the value of a pangenome?
- Why is variation in a genome important?
- How can a genome sequence be used to improve crops?
- Should we apply genome editing in humans?