

A chromosome level avocado genome

Stephen Fletcher

Disease resistance



A high quality avocado genome



Drought tolerance



Dwarfing rootstock



Clonal propagation

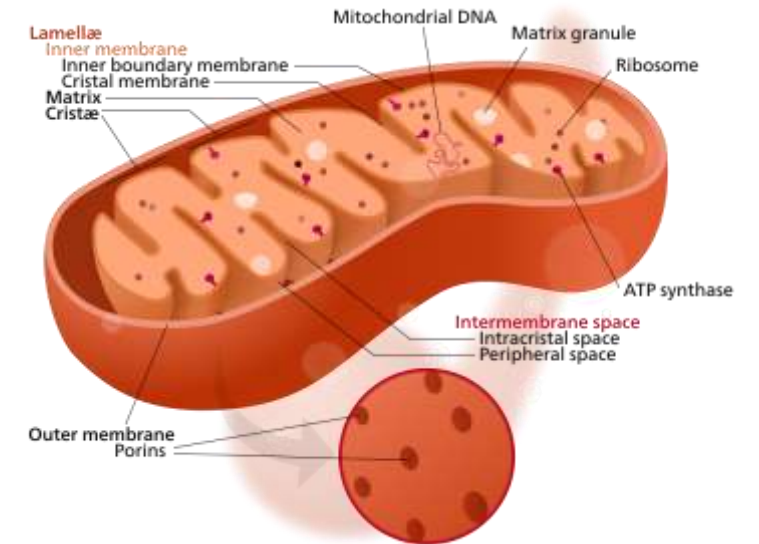
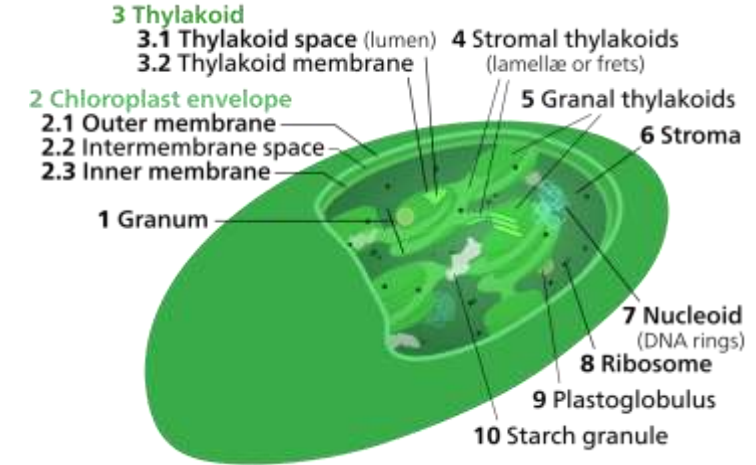
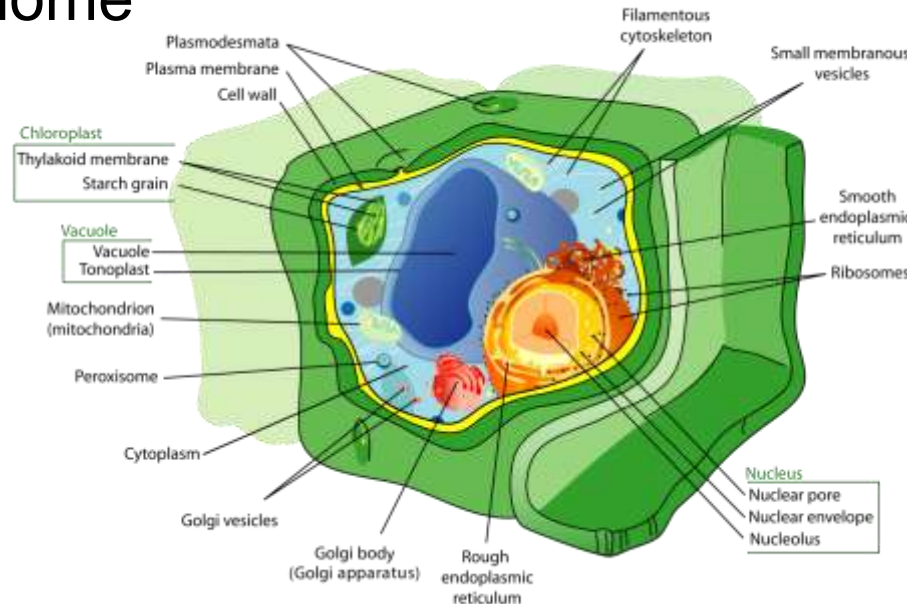


Breeding



The avocado genome

- The nuclear genome consists of 12 pairs of chromosomes
- Mitochondrial genome
- Chloroplast genome



Images obtained from <https://www.peirsoncenter.com/articles/mitochondria-why-theyre-important-and-what-they-need-to-function>, https://en.wikipedia.org/wiki/Chloroplast_membrane,

What is a genome assembly?

- A collection of sequence records that represent a genome
 - Reads from genome sequencing are assembled using De Bruijn graphs
 - Reads are assembled in contigs (contiguous sequences)
 - Contigs → scaffolds (supercontigs)
- BUT:
 - Which chromosome does a contig come from?
 - Alleles are collapsed (consensus assembly)

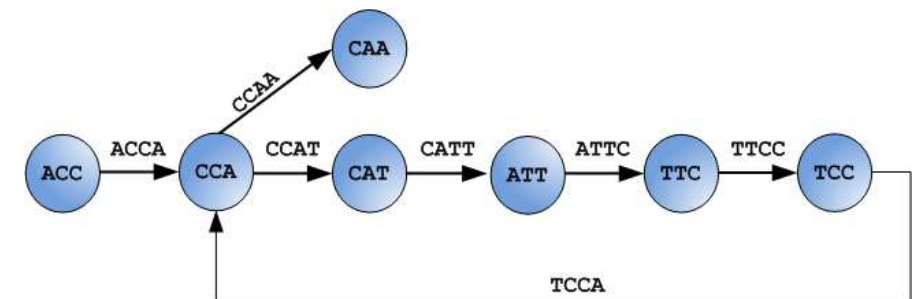
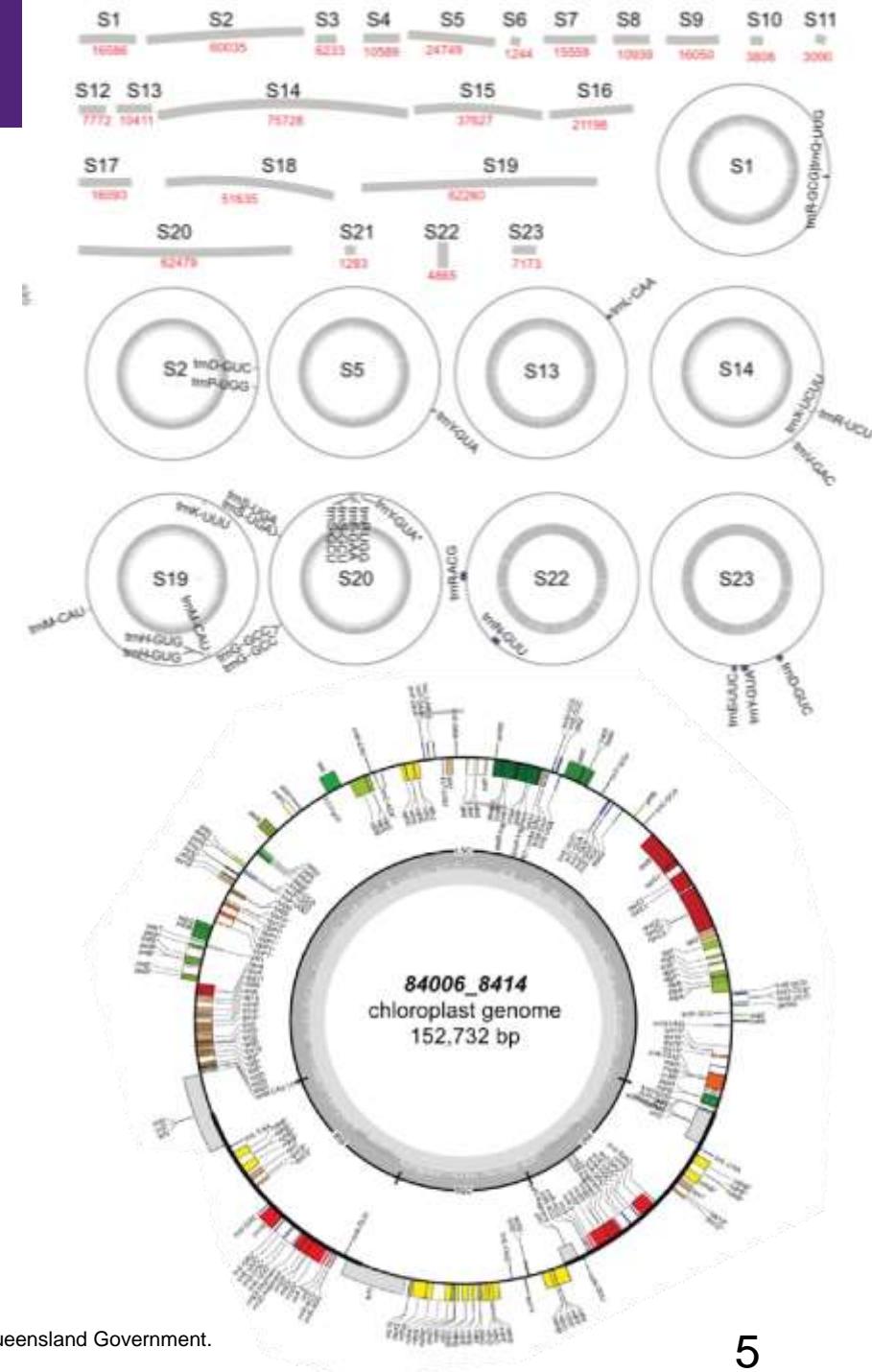


Image obtained from https://www.researchgate.net/publication/261424721_Approaches_to_DNA_de_novo_assembly

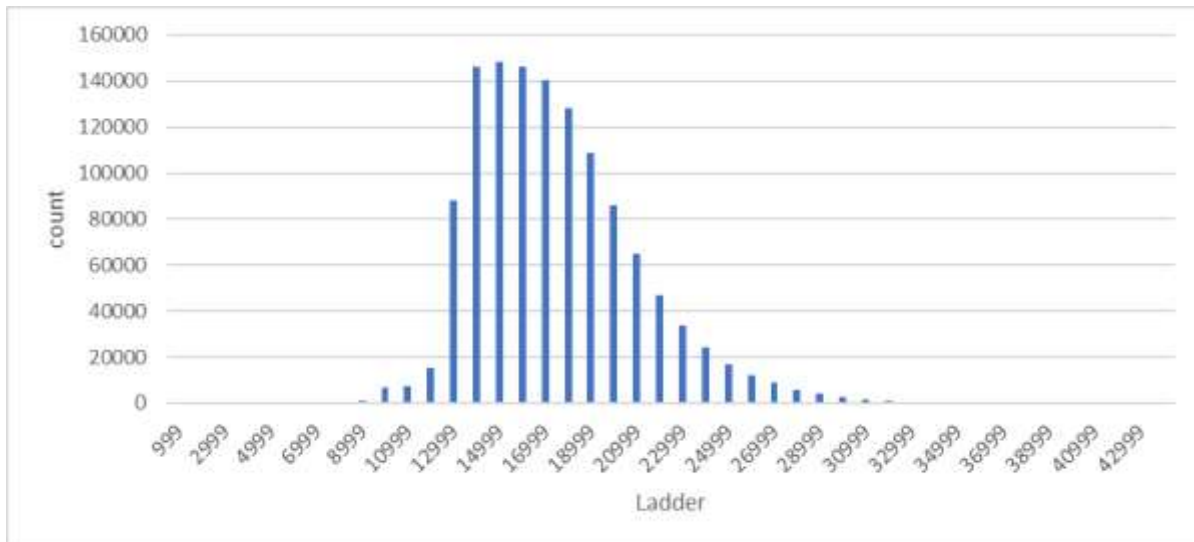
Organelle genome assembly

- Assembled using Illumina short reads
- Chloroplast genome → ~152 kb
 - 2 versions with SSC in different orientations
- Mitochondrial genome → ~527 kb
 - 23 contigs
 - Mitochondrial genomes have complex structures



Nuclear genome assembly

- DNA extracted with custom high quality protocol (Nath *et al.*, 2022)
- PacBio Hi-Fi reads → long reads with the base accuracy of Illumina short reads
- Hass sequenced 48x coverage

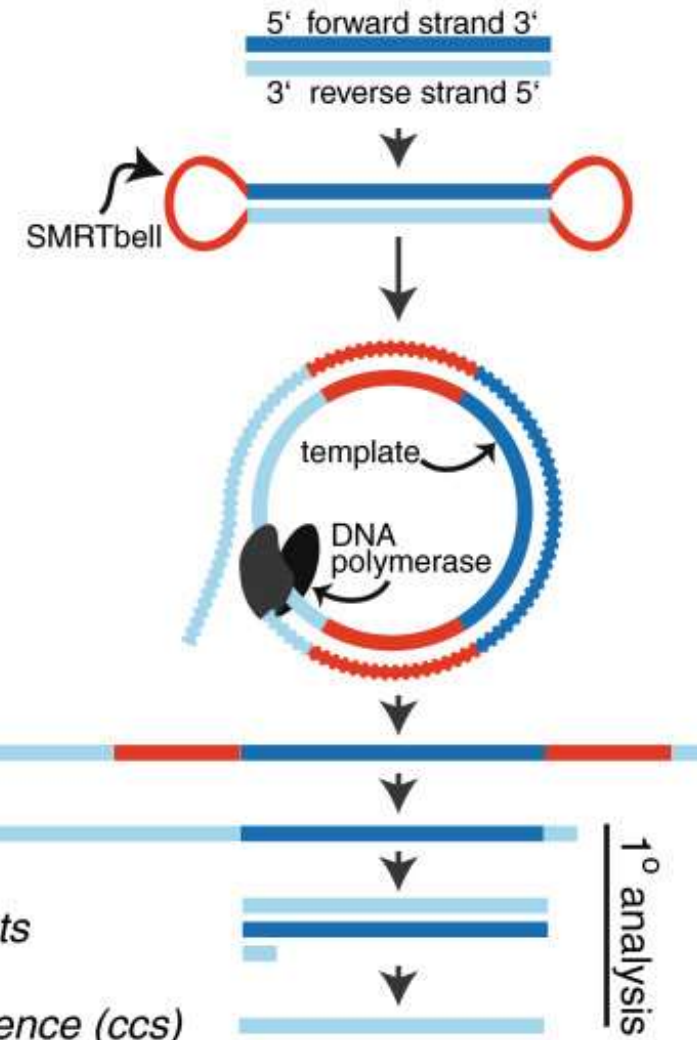
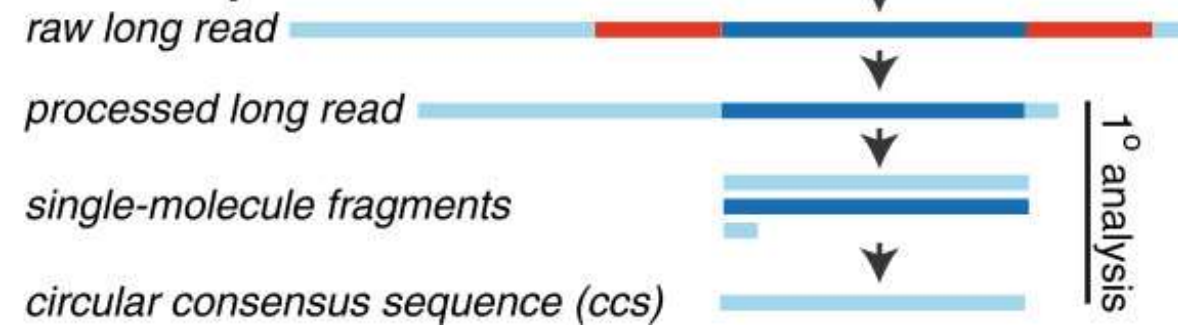


1. generate amplicon

2. ligate adaptors

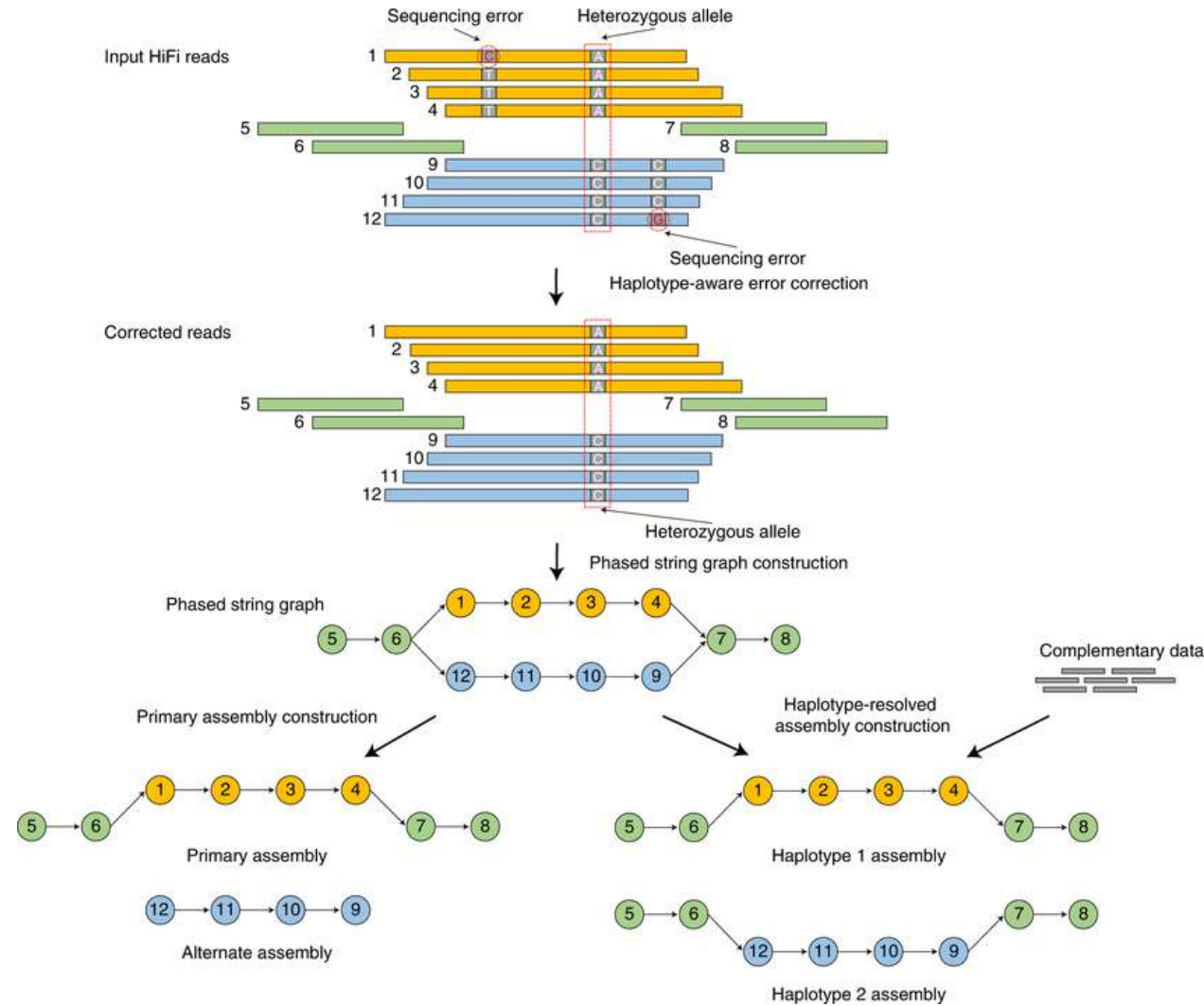
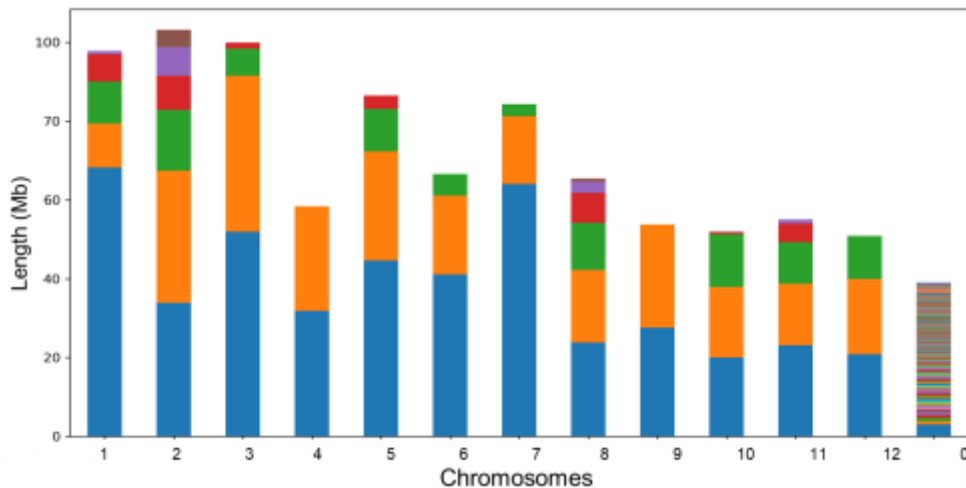
3. sequence

4. data analysis



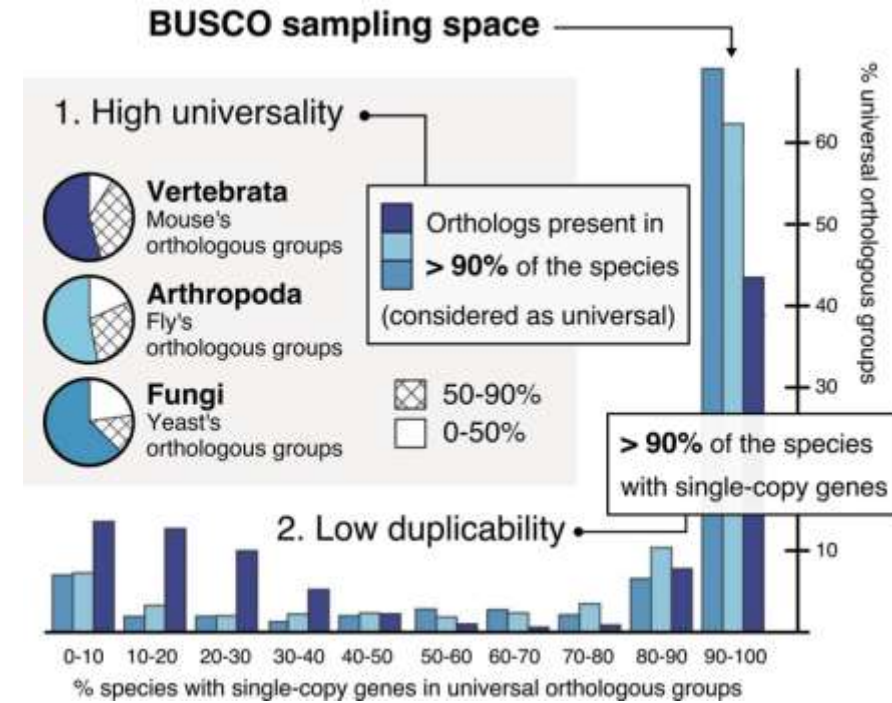
Nuclear genome assembly

- Phased assembly and primary/alternate assembly using HiFiasm
- 913 Mb primary genome in 516 contigs following organellar screening (N50 = 26 Mb)
- SynMap (CoGe) placed and orientate contigs on the Rendon-Anaya et al genome
- Gaps between placed contigs filled by Ns



BUSCO

- 98.8% completeness using the embryophyte (land plant) lineage
- Benchmarking Universal Single-Copy Orthologs
- Can be used for assembly optimisation



Results from dataset embryophyta_odb10

C:98.8%[S:93.7%,D:5.1%],F:0.5%,M:0.7%,n:1614
 1595 Complete BUSCOs (C)
 1512 Complete and single-copy BUSCOs (S)
 83 Complete and duplicated BUSCOs (D)
 8 Fragmented BUSCOs (F)
 11 Missing BUSCOs (M)
 1614 Total BUSCO groups searched

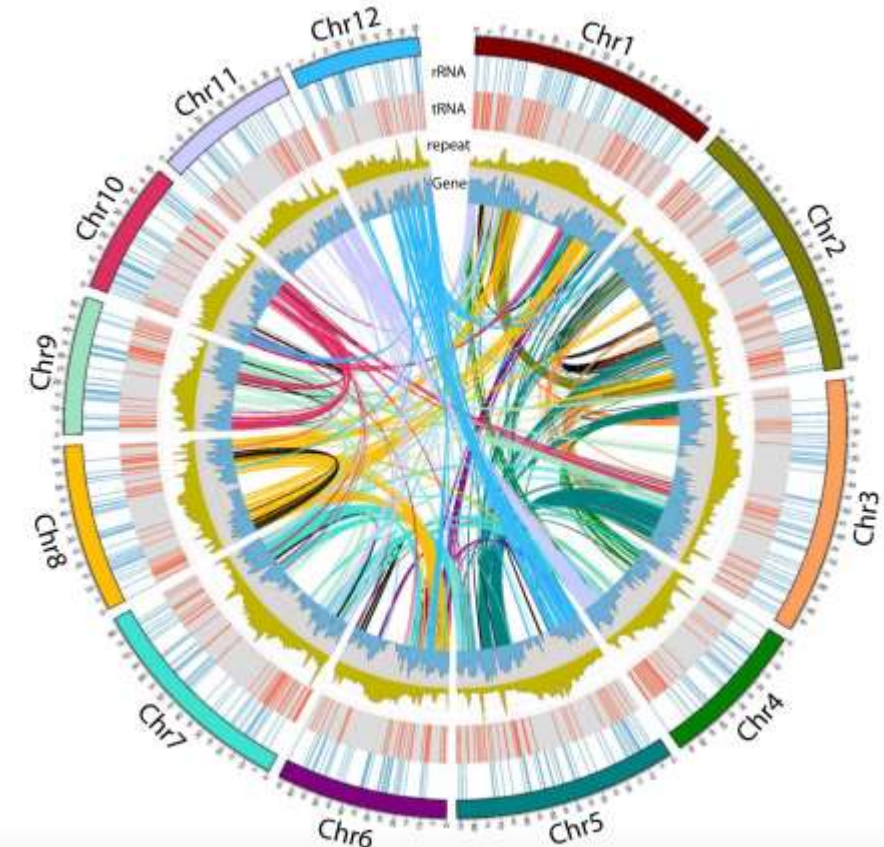
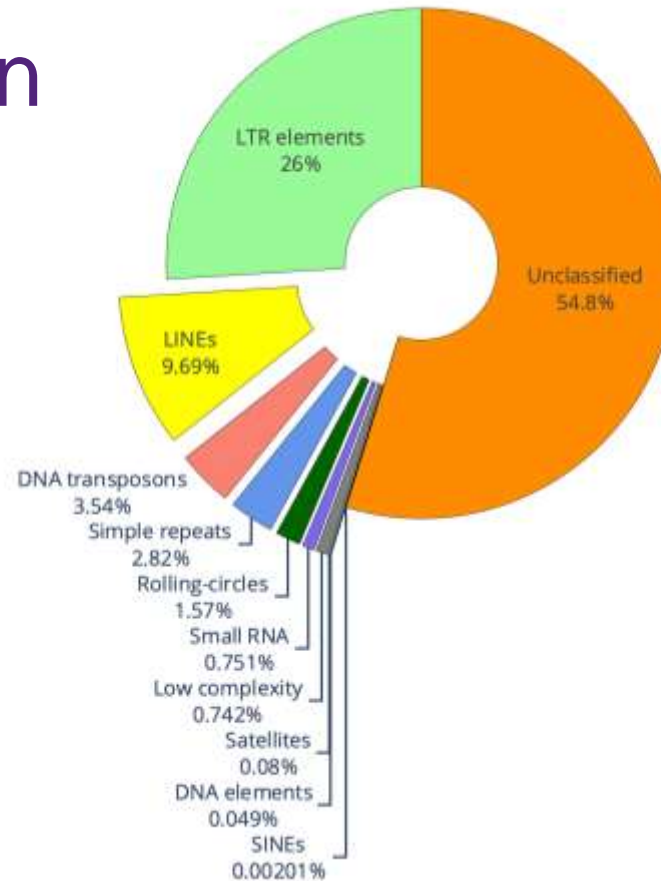


Results from dataset embryophyta_odb10

C:99.0%[S:94.4%,D:4.6%],F:0.5%,M:0.5%,n:1614
 1597 Complete BUSCOs (C)
 1523 Complete and single-copy BUSCOs (S)
 74 Complete and duplicated BUSCOs (D)
 8 Fragmented BUSCOs (F)
 9 Missing BUSCOs (M)
 1614 Total BUSCO groups searched

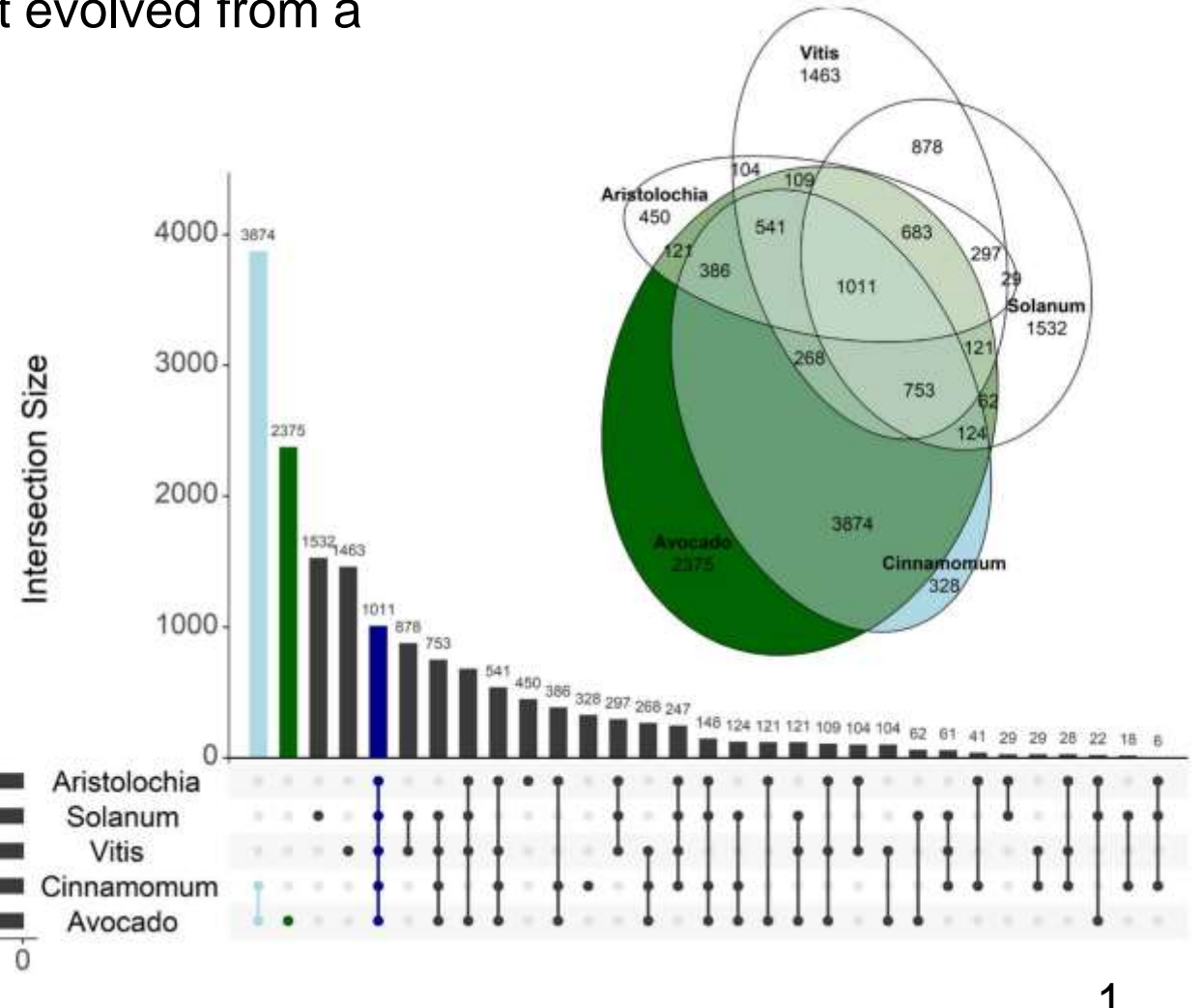
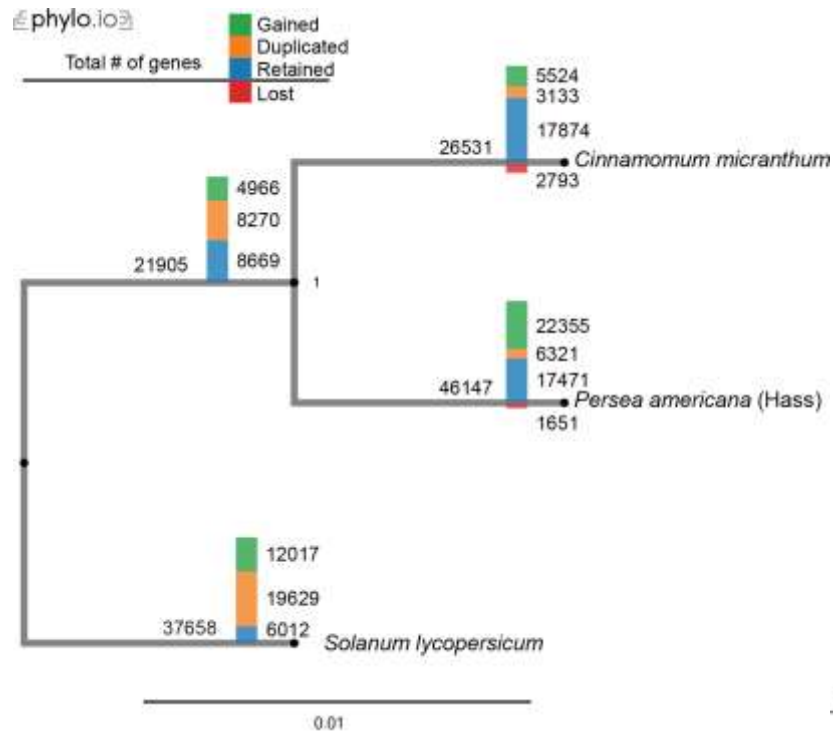
Structural annotation

- Repeat modeller and masker
- *ab initio* gene prediction using Braker2
- Splice aligned RNA-seq data used to self-train models
 - Multiple tissue types and developmental stages
- 42,442 genes
- More genes predicted than Rendon-Anaya et al
 - 98.8% v. 85% BUSCO
 - Purely *ab initio* not previously reported



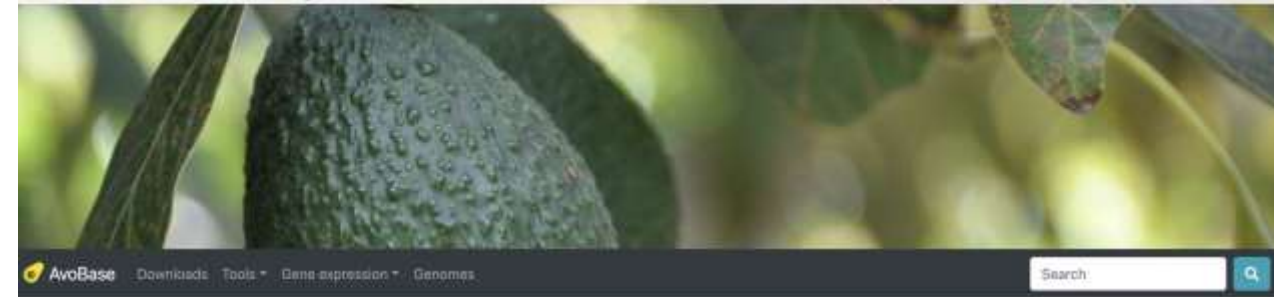
Ortholog analysis

- Orthologs are genes in different species that evolved from a common ancestral gene
- Often retain the same function



Functional annotation

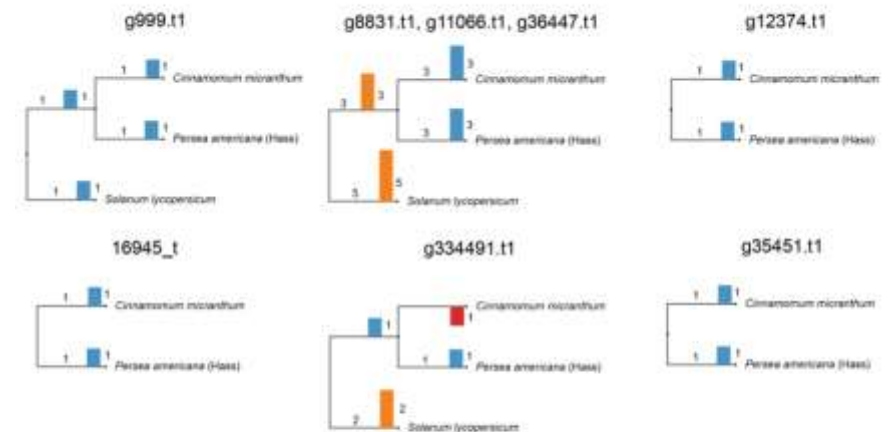
- Functional annotation using OmicsBox
- 33,663 transcripts had protein family classification
- Heptose sugar biosynthesis, cellulase degradation, and potassium biosynthesis and transport were analysed
- Data available at AvoBase: https://www.avocado.uma.es/easy_gdb/index.php (developed by Noé Fernandez @ CSIC)



Welcome to AvoBase

This site is dedicated to avocado genomics, and provides multiple bioinformatics tools to explore and download omics data related to *Persea americana* and related species.

Avocado (*Persea americana* Mill.) (2n = 24) is a member of the flowering plant family Lauraceae, which includes at least 2500 pantropical to temperate species, most of which are woody trees. This family contains a few edible fruit trees, and includes cinnamon, bay laurel and saffron. Avocado has been divided into three landraces: Mexican, Guatemalan and West Indian. 'Hass', a hybrid of the Guatemalan and Mexican races (without a known breeding history), is the principal cultivar, originating from a chance seedling that has been clonally propagated ever since.



Hexokinase gained genes



Where to next?

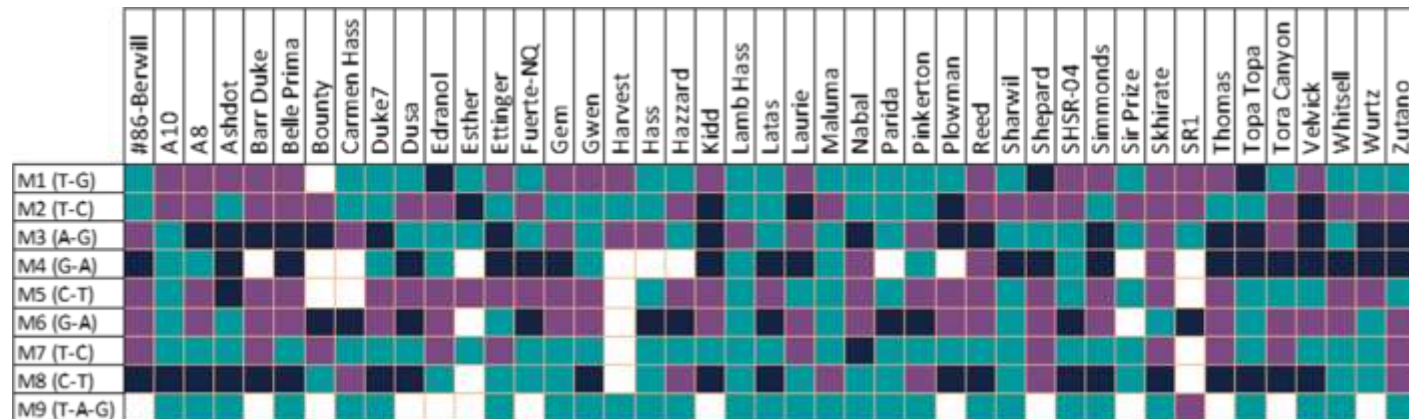
- Telomere-to-telomere, gap free, haplotype resolved genome
 - E.g. achieved for kiwifruit in 2022
 - Combination of HiFi reads (29.7Gb), nanopore sequencing (20.6Gb), Hi-C (105Gb) + short reads
- Doesn't involved starting over
- Pan-genome sequencing → obtaining the full repertoire of genetic variation (inc. structural variation) with the species



Additional work

- 70 individuals re-sequenced to ~20x coverage
 - Represent 50 cultivars
 - 15 samples from the USA and NZ
 - 5 TC plants to assess somaclonal variation
- Variant calling using 3 tools
- Minimal marker set development – 9 markers for 43 CVs

Sample No.	Cultivar name	Rootstock/ Scion	Clonal/ Seedling
S1	A8	R	Un
S2	A10	R	Un
S3	Ashdot	R	Un
S4	Ashdot	R	Un
S5	Barr Duke	R	Un
S6	Belle Prima	S	Un
S7	Bounty	R	Un
S8	Duke7	R	C
S9	Duke7	R	Un
S10	Dusa	R	C
S11	Edranol	S	Un
S12	Esther	S	Un
S13	Fuerte (Cole)	S	Un
S14	Fuerte (N. Qld)	S	Un
S15	Gem	S	C
S16	Gwen	S	Un
S17	Hass	S	Un
S18	Hazzard	S	Un
S19	Kidd	R	S/C
S20	Lamb Hass	S	Un
S21	Latas	R	Un
S22	Maluma	S	C
S23	Nabal	R	Un
S24	Parida	R & S	Un
S25	Pinkerton	S	Un
S26	Plowman	R	Un
S27	Reed	R & S	Un
S28	Reed	R & S	Un
S29	Reed	R & S	Un
S30	Sharwil	S	Un
S31	Shepard	S	Un
S32	SHSR-04	R	C
S33	Simmonds	S	Un
S34	Skhirate	S	Un
S35	Thomas	R	Un
S36	Tora Canyon	R	Un
S37	Velvick	R	S/C
S38	Velvick	R	Un
S39	Velvick	R	Un
S40	Whitsell	S	Un
S41	Wurtz	S	Un
S42	Zutano	R & S	Un
S43	Zutano	R & S	Un
S44	Zutano	R & S	Un
S45	Reed (TC)	R	Un
S46	Reed (TC)	R	Un
S47	Zutano (TC)	R	Un
S48	Zutano (TC)	R	Un
S49	Kidd (TC)	R	Un
S50	Velvick (TC)	R	Un
S51	Velvick (TC)	R	Un
S52	Topa Topa	R	Un
S53	Dusa (TC)	R	Un
S54	Dusa (TC)	R	Un
S55	Mother Hass	S	Un
S56	Sir Prize	Un	Un
S57	442712-05	Un	Un
S58	Harvest	Un	Un
S59	465518-99	Un	Un
S60	#86-Berwill	Un	Un
S61	BL-667	Un	Un
S62	Laurie	Un	Un
S63	Ettinger	Un	Un
S64	BL-516	Un	Un
S65	465418-99	Un	Un
S66	467352-00	Un	Un
S67	GEM	Un	Un
S68	464918-99	Un	Un
S69	SR1	Un	Un
S70	Carmen Hass	Un	Un



Acknowledgments

- Onkar Nath



- Alice Hayward
- Lindsay Shaw
- Ardashir Kharabian Masouleh
- Agnelo Furtado
- Robert Henry
- Neena Mitter

This project was jointly supported by the Department of Agriculture and Fisheries, University of Queensland, and Horticulture Innovation Australia Limited

