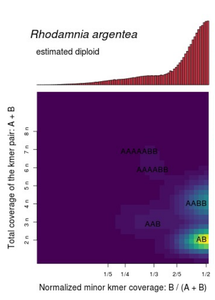
Small but mitey: high-quality long-read assembly of a streamlined mite genome from contaminated sequencing data

Richard J Edwards, Stephanie H Chen & Jason G Bragg

Evolution & Ecology Research Centre, School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, Australia
Australian Institute of Botanical Science, Royal Botanic Garden, Sydney, Australia

The problem: a contaminated genome?

- Sequenced and assembled two Myrtaceae genomes in 2018 using 10x Genomics linked reads
  - *Rhodamnia argentea* (Malletwood) & *Syzygium oleosum* (Blue lilly pilly)
  - Pilot data for an ARC Linkage Project with the Royal Botanic Gardens in Sydney
  - Develop conservation populations of Myrtaceae in face of threats from Myrtle Rust
    - ➢ Invasive fungal pathogen (Arrived 2010)
    - ➢ Affects >390 native species



385 Mbp diploid. 2n=22

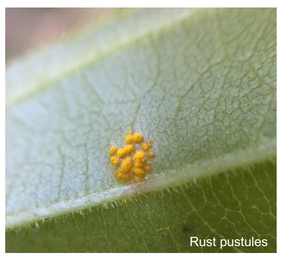


Photo by R Edwards

Photo by S Chen

Photo by S Chen

Photo by S Chen

- Possible mite or insect contamination in Malletwood v1.0 assembly?
  - ~1700 scaffolds (754 kb) with detectable homology to mite genome
- The plan:
  - Improved chromosome-level assembly (+ONT & HiC)
  - Stringent contamination identification (*R. rubescens* homology filter)
  - Improved taxonomic assessment of assembly contigs

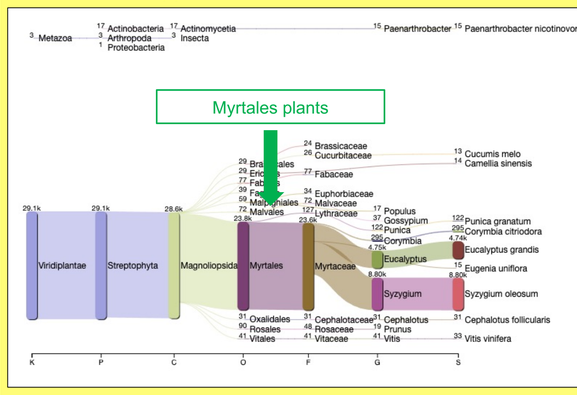
Taxolotl: assembly taxonomy summary and assessment

https://github.com/slimsuite/taxolotl

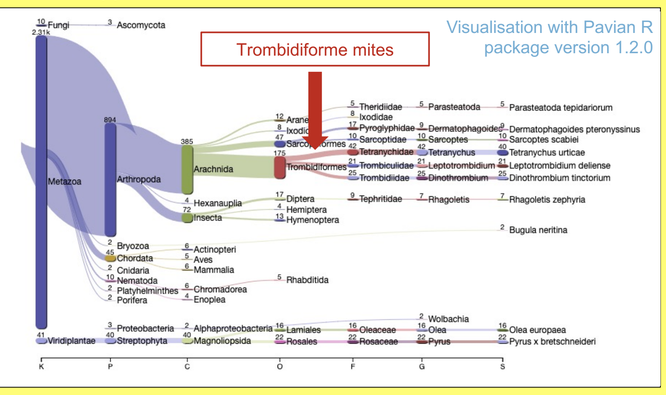
1. Annotate protein-coding genes in assembly (GeMoMa)
2. Use MMseqs2 "easy-taxonomy" to assign proteins to taxonomic groups
3. Remove redundancy and report taxonomy rankings for assembly/scaffolds/contigs
4. Flag sequences and contigs that appear to contradict consensus

Malletwood v2.0 annotation

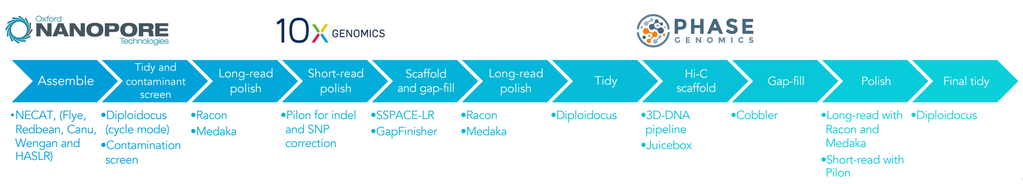
74 scaffolds mapped to *R. rubescens*



12 unmapped scaffolds

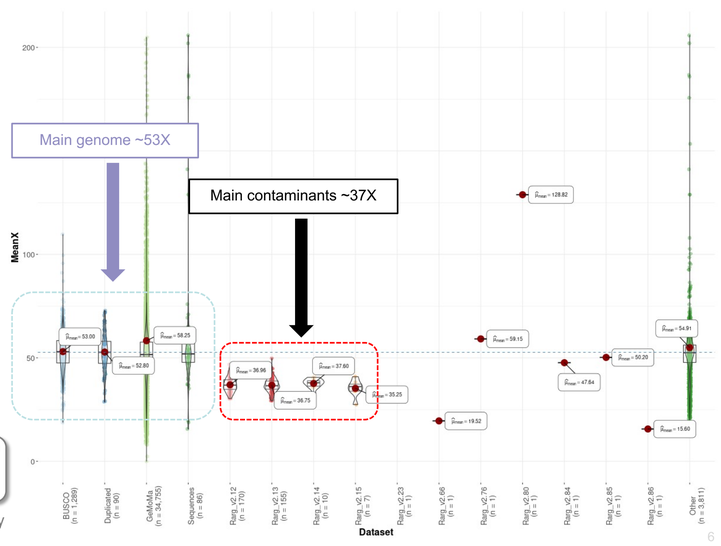


The solution: Malletwood genome v2.0 & re-assembly of contamination

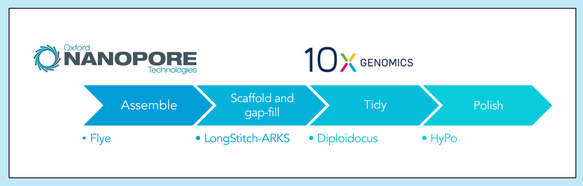


- Chromosome-level assembly (ONT + 10x & Hi-C)
  - 384 Mb genome on 86 scaffolds
  - 11 chromosomes (343 Mb)
- v2.0 scaffolds either pure tree or pure contaminant (see Taxolotl box)
  - 74 scaffolds (347 Mb; 99.93% plant) mapped to pure (tissue culture) *R. rubescens* genome
    - ➢ ~53X coverage
  - 12 unmapped scaffolds (34.6 Mb; 97.8% animal)
    - ➢ ~37X coverage = surprisingly deep!
    - ➢ BUSCO v5 metazoa only 58% Complete!
    - ➢ 35 Mb too small for a metazoan genome?

https://github.com/slimsuite/depthkopy



Genome assembly of contaminating reads

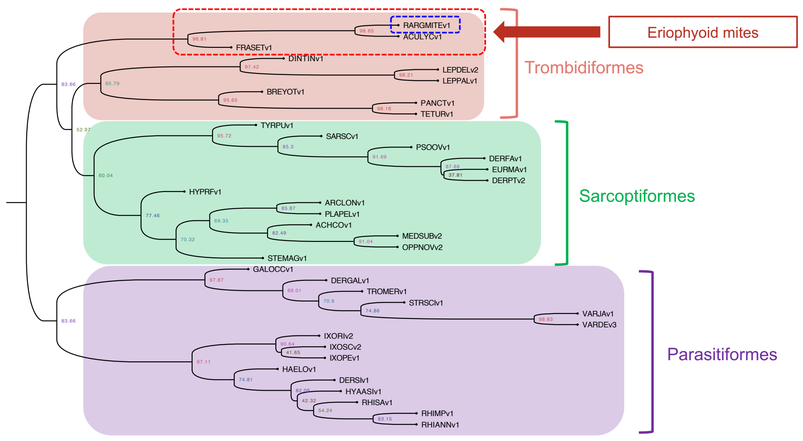


- Partitioned out the long reads and re-assembled
  - 34.6 Mb with two main contigs (17.4 Mb and 17.0 Mb)
  - 67.7% BUSCO Complete (Metazoa)
  - 87.4% BUSCO Complete (Eukaryota)

Telomere-to-telomere assembly of a small mite with a very small genome?

Phylogenomics: ASTRAL species tree (37 Acari genomes)

- 1,877 BUSCO v5 arachnida Complete » HAQESAC (ClustalO » FastTree) » ASTRAL



RARGMITEv1

Unknown mite (different mite shown)

ACULYCv1

Aculops lycopersici

FRASETv1

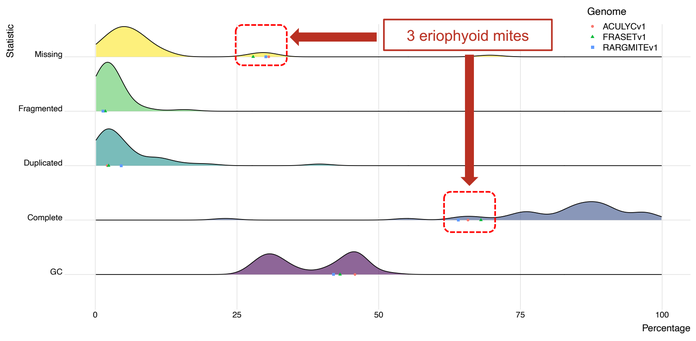
Fragariocoptes setiger

- Novel mite identified as an unknown eriophyoid mite
  - Closest relative with available genome is the Tomato russet mite (*Aculops lycopersici*)

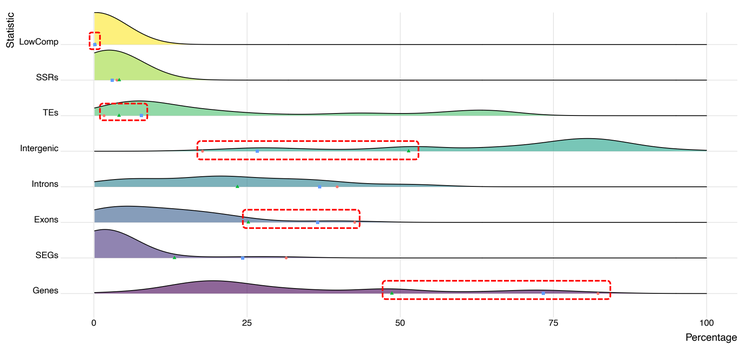
Genome reduction in eriophyoid mites (statistics for 37 Acari genomes)

- Tomato russet mite (*Aculops lycopersici*) has a "streamlined" genome:
  - 32.5 Mb (8 scaffolds)
  - 67.0% BUSCO Complete (Metazoa)
  - 86.3% BUSCO Complete (Eukaryota)
  - 0.2 mm long!

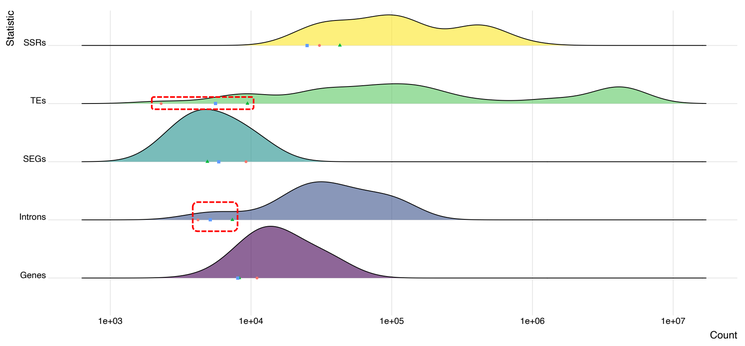
- Low BUSCO Complete and high Missing values show genome reduction
  - Comparable/superior assembly quality



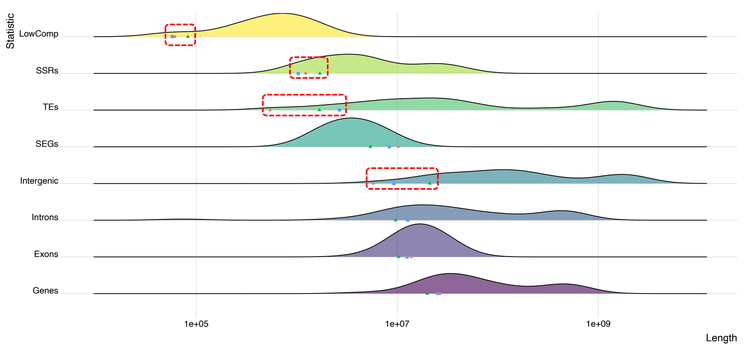
- Reduced repetitive and intergenic content (high % exons/genes):



- Reduced numbers of introns and transposable elements:

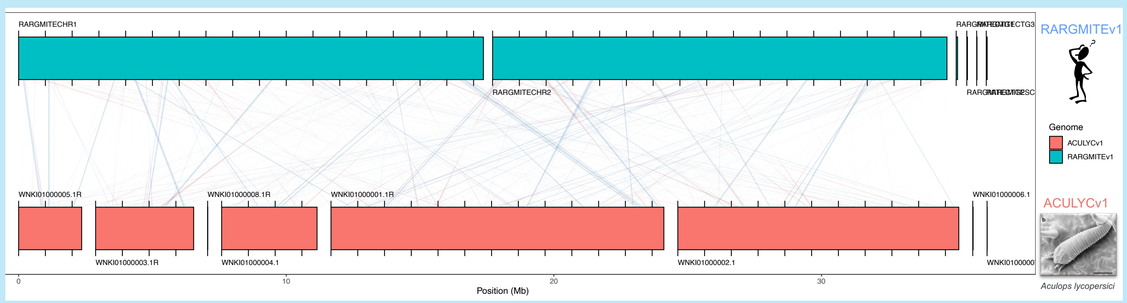


- Reduced repetitive and intergenic content:

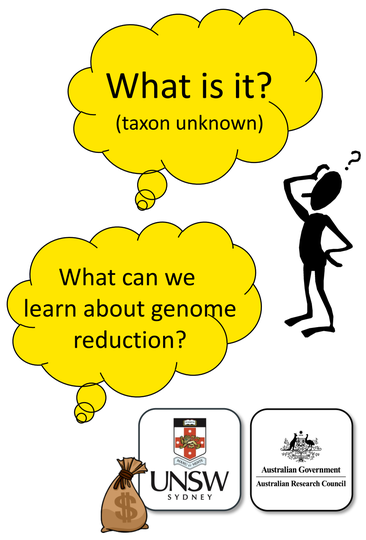


The first telomere-to-telomere assembly of an eriophyoid mite!

- Two gapless chromosomes, capped with telomeres (novel TTTGG sequence)



- Mapped synteny to Tomato russet mite via Complete BUSCO genes
  - Extensive rearrangements (~71% protein sequence identity)



https://www.slimsuite.unsw.edu.au
https://edwardslab.blogspot.com.au
https://github.com/slimsuite
@slimsuite
@cabbagesofdoom
E-mail: richard.edwards@unsw.edu.au