

Chromosome-level assembly of allotetraploid *Coffea arabica* reveals the complex history of a recent allopolyploid

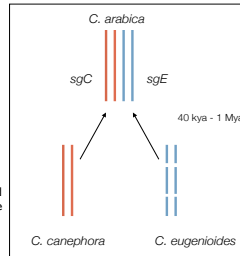
On behalf of Arabica Coffee Genome Sequencing Consortium, presented by Jarkko Salojärvi!

¹School of Biological Sciences, Nanyang Technological University, Singapore

S1-PO-19

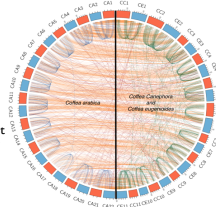
Chromosome-level assembly

- C. arabica* is a recent allopolyploid of *C. canephora* and *C. eugenioides* progenitors.
- Here, we carried out chromosome-level assemblies of *C. arabica* and *C. canephora*, as well as long read assembly of *C. eugenioides*.
- For *C. canephora*, the assembly was verified by establishing an ultra-high density linkage map from whole genome sequencing of parent-progeny population.



Conserved genome structure

- Overall, the sub-genomes were very conserved when compared to diploid progenitors.
- Genome fractionation had progressed in excisions of blocks of genes rather than pseudogenization.
- Preferential retention of genes associated with regulatory functions.
- No dominance of either of sub-genomes, but we saw neofunctionalization when inspecting certain gene families.



ABSTRACT

RATIONALE

Coffea arabica formed as an allotetraploid hybrid of *C. eugenioides* and *C. canephora* approximately 2000-4000 generations ago. The quality of arabica coffee surpasses the other coffee species, but the species is susceptible to many plant pathogens. To increase pathogen resistance, currently bred *C. arabica* cultivars are introgressed with diploid *C. canephora*. Unfortunately this breeding strategy, which is commonly pursued for many monocultured crop plants, introduces unwanted side-effects to the novel hybrids, such as decreased gustatory quality of the coffee beverage. To address this problem and facilitate enhanced breeding and bioengineering strategies, modern genomic tools are needed.

METHODS

We, the Arabica Coffee Genome Consortium, sequenced the genomes of a di-haploid *C. arabica* accession as well as modern representatives of its diploid progenitors *C. eugenioides* and *C. canephora* using a combination of Pacbio long read sequencing and chromosome conformation capture technology. Genome annotation was carried out using a comprehensive library of RNAseq evidence from different tissues and protein homology.

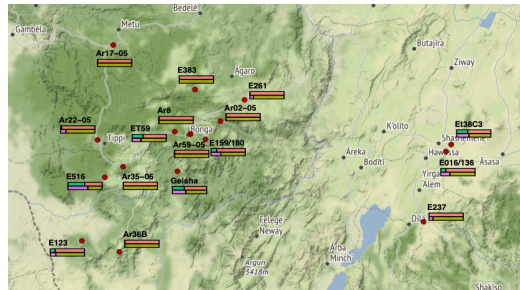
We further resequenced the genomes of 18 wild accessions, including the lectotype used by Carl von Linné, 15 commonly cultivated accessions as well as six modern cultivars from a spontaneous hybrid originating from Timor containing introgression from *C. canephora*.

RESULTS

The assembly yielded 2 x 11 high quality pseudo-chromosomes for *C. arabica*, with 89% of the total of 54,562 predicted genes placed on the chromosomes. Since *C. arabica* hybridization is a recent event, the chromosome-level assemblies made it possible to study the processes of genome evolution in a newly formed tetraploid hybrid. Genomic excisions were found to be the dominant process. Even though the diversity of subgenome *C. eugenioides* was higher, we did not detect subgenome dominance within the species. Using molecular data, we followed the breeding history and resolved the relationships between coffee cultivars. We were able to shed light to the geographic origins of domesticated *C. arabica*. Furthermore, the study of Timor hybrid lines illustrated how introgression shaped the genomes of these cultivars and suggested regions underlying their resistance to coffee leaf rust disease.

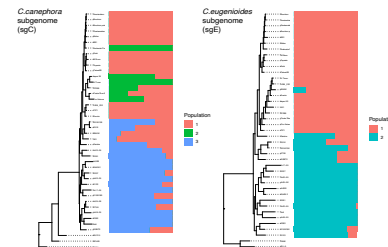
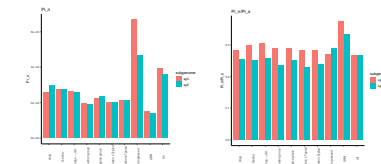
Whole genome sequencing of 39 coffee accessions

- Representative set of *C. arabica* diversity:
 - 13 Bourbon and Typica group cultivars, 5 Timor hybrid accessions, an Indian variant Jackson 1 and Geisha, a newly established cultivar.
 - 17 wild individuals from Ethiopia.
 - Including lectotype herbarium sample (Linnean society) from 1700's.



- ADMIXTURE shows grouping into wild vs. cultivars. Introgression in Timor hybrid lines (green) is visible only in *C. canephora* subgenome.
- Population split due to modern cultivation bottleneck or the original establishment of cultivated coffee in Yemen ca. 1600?

Effect of cultivation is visible in reduced nucleotide diversity



Conclusions & perspectives

- The chromosome level assembly of *C. arabica* reveals the evolution and cultivation history of the species and provides molecular tools for accelerating coffee breeding in future.
- More results in an upcoming paper.

ACGC consortium:

