

28 June - 1 July

www.alphavisa.com/asic/2021



28th Conference **Asic 2021**

BOOK OF ABSTRACTS



- Agronomy
- Chemistry
- Technology
- Physiological effects
- Sustainability, climate changes

A single polyploidization event at the origin of the tetraploid genome of *Coffea arabica* is responsible for extremely low genetic variation in wild & cultivated germplasm

Scalabrin Simone¹ (sscalabrin@igatechnology.com), Toniutti Lucile², Di Gaspero Gabriele³, Scaglione Davide⁴, Magris Gabriele⁵, Vidotto Michele⁴, Pinosio Sara⁶, Cattonaro Federica⁴, Magni Federica⁴, Jurman Irena⁴, Cerutti Mario⁷, Suggi Liverani Furio⁸, Navarini Luciano⁸, Del Terra Lorenzo⁸, Pellegrino Gloria⁷, Ruosi Manuela⁷, Vitulo Nicola⁹, Valle Giorgio¹⁰, Pallavicini Alberto¹¹, Graziosi Giorgio¹², Klein Patricia¹³, Bentley Nolan¹³, Murray Seth¹³, Solano William¹³, Al Hakimi Amin¹⁴, Schilling Timothy¹⁵, Montagnon Christophe¹⁶, Kotch George¹⁵, Bertrand Benoit², Morgante Michele³

¹IGA Technology Services, Udine, UD, Italy ; ²CIRAD, Montpellier, France ; ³Istituto di Genomica Applicata, Udine, Italy ; ⁴IGA Technology Services, Udine, Italy ; ⁵Università degli Studi di Udine, Udine, Italy ; ⁶CNR, Sesto Fiorentino, Italy ; ⁷Lavazza, Torino, Italy ; ⁸illycaffè, Trieste, Italy ; ⁹Università degli studi di Verona, Verona, Italy ; ¹⁰Università di Padova, Padova, Italy ; ¹¹Università degli Studi di Trieste, Trieste, Italy ; ¹²DNA Analytica, Trieste, Italy ; ¹³Texas A&M University, Bryan, Texas, United States ; ¹⁴Sana'a University, Sana'a, Yemen ; ¹⁵World Coffee Research, Portland, Oregon, United States ; ¹⁶RD2vision, Montpellier, France

RATIONALE

The genome of the allotetraploid species *Coffea arabica* L. was sequenced to assemble independently the two component subgenomes (putatively deriving from *C. canephora* and *C. eugenioides*) and to perform a genome-wide analysis of the genetic diversity in cultivated coffee germplasm and in wild populations growing in the center of origin of the species.

METHODS

We studied an individual of *C. arabica* 'Bourbon Vermelho'. A BAC library of 175,872 BAC clones was constructed and sequenced using an Illumina HiSeq2000. Each BAC pool was assembled independently with the tool ABySS and scaffolded with SSPACE. Genotyping by sequencing (GBS) was conducted using the restriction enzyme *Pst*I followed by single-end sequencing on an Illumina HiSeq2000. SNP calling was performed using Stacks. Principal Component Analysis was performed using the R package ade4. A hierarchical study of the diversity has been conducted using a model-based clustering procedure with admixture as implemented in STRUCTURE.

RESULTS

We assembled a total length of 1.536 Gbp, 444 Mb and 527 Mb of which were assigned to the *canephora* and *eugenioides* subgenomes, respectively, and predicted 46,562 gene models, 21,254 and 22,888 of which were assigned to the *canephora* and to the *eugenioides* subgenome, respectively. Through a genome-wide SNP genotyping of 736 *C. arabica* accessions, we analyzed the genetic diversity in the species and its relationship with geographic distribution and historical records.

CONCLUSIONS & PERSPECTIVES

We observed a weak population structure due to low-frequency derived alleles and highly negative values of Tajima's *D*, suggesting a recent and severe bottleneck, most likely resulting from a single event of polyploidization, not only for the cultivated germplasm but also for the entire species. This conclusion is strongly supported by forward simulations of mutation accumulation. However, PCA revealed a cline of genetic diversity reflecting a west-to-east geographical distribution from the center of origin in East Africa to the Arabian Peninsula. The extremely low levels of variation observed in the species, as a consequence of the polyploidization event, make the exploitation of diversity within the species for breeding purposes less interesting than in most crop species and stress the need for introgression of new variability from the diploid progenitors.