

S1-34

SEQUENCE ORGANISATION AND CONSERVATION AT HOMEOLOGOUS REGIONS IN THE RECENT ALLOTETRAPLOID COFFEE (*COFFEA ARABICA* L.)

Cenci A., Combes M.C., Ribas A., Etienne H., Lashermes P.

UMR RPB (CIRAD, IRD, UM2), Centre IRD de Montpellier, BP 64501, F-34394, Montpellier, France

Coffee is one of the world's largest traded commodities produced in more than 60 countries. Coffee species belong to the Rubiaceae family and commercial coffee production relies mainly on two closely related species: *Coffea arabica* and *C. canephora*, which account respectively for 65 and 35% of the coffee production. All coffee species are diploid ($2n=2x=22$) and generally self-incompatible, except for *C. arabica* which is the only tetraploid ($2n=4x=44$) and self-fertile. Molecular phylogenies have indicated that *C. arabica* is a recent allotetraploid (CaEa genome) formed by hybridisation between two related diploid species: *C. canephora* (C genome) and *C. eugenioides* (E genome) or ecotypes related to those diploid species. In spite of the close relationship between the two constitutive sub-genomes, *C. arabica* displays diploid-like meiotic behaviour with bivalent formation.

In order to estimate the sub-genome divergence at fine scale, a sequence comparison was performed between orthologous regions from both Ca and Ea sub-genomes of *C. arabica*. In particular two homeologous BAC clones, named clones 45 and 52, and sharing around 180 Kb, were analyzed for gene composition, sequence divergence and for presence/absence of repetitive elements. The analyzed regions include 17 putatively functional genes. All predicted genes were strictly conserved in the same order and orientation on both sub-genomes. On more than 31 Kb of coding sequence, around 1% of substitution rate was observed, corresponding to 1.5% of amino acid substitutions. In intronic regions the substitution rate increased to around 2%. Some repetitive elements were shared by the two sub-genomes, indicating that their insertion predated the divergence between *C. canephora* and *C. eugenioides*. Based on the comparison of the predicted coding sequences of both clones with the publicly available *C. canephora* ESTs, the clones 45 and 52 were attributed to the Ca and Ea sub-genomes, respectively.

Our data point out the high sequence and structure conservation between the Ca and Ea sub-genomes of *C. arabica*. This result is consistent with the evolutionary history of Arabica progenitors, where diploid C- and E-genome species involved in the allopolyploidization event derived from the same ancestor approximately 150,000-350,000 yr BP. Moreover, limited changes in genome structure of the analysed regions appear to be associated with polyploidy in *C. arabica*. Although partial, this study provides us with the first view of Arabica genome organisation and will be very useful when defining a whole genome sequencing strategy.

Keywords: Allopolyploidy, Genome evolution, subgenome, Microcolinearity, Coffee