

P-1032

S-42 From theory to genome-wide data: inferring selection, demography, gene flow and admixture

 21/08/18

 Level 3

Local Ancestry Inference approaches to unravel plant genome mosaic: a simulation-based evaluation

COTTIN A. ^(1, 2), PENAUD B. ^(1, 2), MARTIN G. ^(1, 2), D. SANTOS J. ^(1, 2), CURK F. ^(2, 3), D'HONT A. ^(1, 2), GLASZMANN J. C. ^(1, 2), YAHIAOUI N. ^(1, 2), GAUTIER M. ⁽⁴⁾

(1) CIRAD, UMR AGAP, F-34398, Montpellier FRANCE

(2) AGAP, Univ. Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier FRANCE

(3) INRA, UMR AGAP, F-20230, San Giuliano FRANCE

(4) CBGP, Univ. Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier FRANCE

AIMS:

Hybridization events between species and subspecies are considered as major evolutionary steps, possibly contributing to the advent of new phenotypes. These events are widespread in several crop species and are expected to produce genomes with a mosaic structure of sequence blocks originating from different ancestry. With the development of NGS genotyping technologies, several population genomics approaches have been proposed to infer the ancestry of genome segments, by comparing polymorphism patterns across individuals along chromosomes. However, these Local Ancestry Inference (LAI) methods have mainly been developed for applications in animal models, and human most particularly. They are based on assumptions which do not always fit plant models due to more complex genome structures (e.g. different ploidy levels, variable heterozygosity levels within species) or different reproductive systems (e.g. vegetative propagation, selfing). In this context, there is a need to evaluate available methods on plant models. To that end, we developed a small and flexible R tool to simulate mosaic genome data under a wide variety of scenarios representative of plant model characteristics. We evaluated two main types of LAI methods i) exploratory approaches (based on multivariate analysis), and ii) full probabilistic approaches (based on Hidden Markov Models) that are able to use unphased genotypic data and handle more than two ancestries. The results will be presented and discussed.