

[POSTER] Fractionation patterns in subgenomes

Zhe Yu¹, Chunfang Zheng¹, and David Sankoff¹

University of Ottawa, Canada sankoff@uottawa.ca

The evolutionary process of gene loss, through DNA excision, pseudogenization or other mechanism, is the obverse of novel gene acquisition by a genome through processes such as tandem duplication, gene family expansion, whole genome doubling, neo- and subfunctionalization and horizontal transfer. Loss serves a number of functional and structural roles, mainly compensating for the energetic, material and structural costs of gene complement expansion.

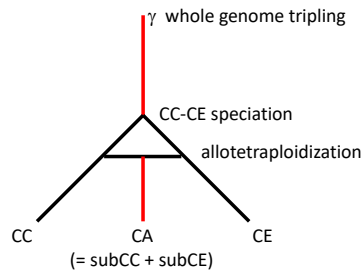


Fig. 1. *Coffea* phylogeny. Fractionation operates in lineages coloured red.

in Figure 1. (Genome sequence and annotation produced by the Arabica Coffee Genome Consortium [1].) We survey gene loss in three periods. These are: loss from the ancestral lineage leading from the γ whole genome tripling event [2] 120 million years ago due at least partly to fractionation, independent losses from the CC and CE genomes after speciation around 10 million years ago [3], and loss from the CC and CE, and the subCC and subCE subgenomes of CA, following the allotetraploidization event.

Our main analytical construct is the frequency distribution of gap lengths within syntenic blocks calculated during the comparison of chromosomes from two genomes or subgenomes.

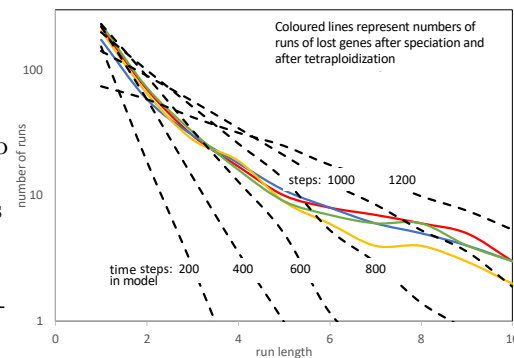


Fig. 2. Inability of one-at-a-time model to fit *Coffea* data on runs of zeros at various time intervals (steps)

In the simplest model, proposed over ten years ago [4–6], at each step a random gene pair is selected to lose one member. In a new version of this model that takes into account chromosome length, we develop an exact recurrence to calculate the expected number of gaps of each length after a given number of steps. The evidence from the *Coffea* data demonstrates a systematic departure from this model, as in Figure 2.

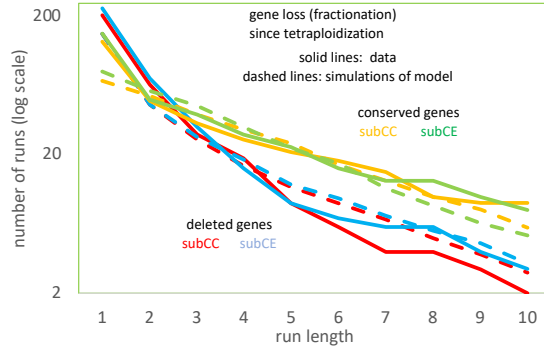


Fig. 3. Comparisons of the distributions of run sizes (0's and 1's) with simulations of combined model in the tetraploidization data).

In a competing class of models [7], gene loss is effected by excision of a variable length fragment of a chromosome, often formulated in terms of a gamma distribution. In the *Coffea* data, there are far too many single-gene deletions for this solution, but a mixture of the two models, where the gamma is actually a single-parameter geometric distribution, fits well (cf Figure 3).

References

1. Aims and goals of the Arabica Coffee Genome Consortium (ACGC) (A De Kochko, D Crouzillat, Arabica Coffee Genome Consortium) 12th Solanaceae Conference (2015)
2. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla (O Jaillon, JM Aury, B Noel, A Policriti, C Clepet, *et al.*) *Nature* 449, 463–467 (2007)
3. Genotyping-by-sequencing provides the first well-resolved phylogeny for coffee (*Coffea*) and insights into the evolution of caffeine content in its species (P Hamon, CE Grover *et al.*) *Molecular Phylogenetics and Evolution* 109, 351–361 (2017)
4. The role of mutational dynamics in genome shrinkage (MJ van Hoek, P Hogeweg) *Molecular Biology and Evolution* 24, 2485–2494 (2007)
5. Reorganization of adjacent gene relationships in yeast genomes by whole-genome duplication and gene deletion (JK Byrnes, GP Morris, WH Li) *Molecular Biology and Evolution* 23, 1136–1143 (2006)
6. Gene loss under neighbourhood selection following whole genome duplication and the reconstruction of the ancestral *Populus* diploid (C Zheng, PK Wall, J Leebens-Mack, C dePamphilis, VA Albert, D Sankoff) *Journal of Bioinformatics and Computational Biology* 7, 499–520 (2009)
7. Structural *vs.* functional mechanisms of duplicate gene loss following whole genome doubling (D Sankoff, C Zheng, B Wang, C Fernando Buen Abad Najar) *BMC Genomics* 15, DOI: 10.1109/ICCABS.2014.6863915 (2015)