

On the origin and structure of haplotype blocks

Daria Shipilina¹, Sean Stankowski², Arka Pal², Yingguang Frank Chan³, and Nick Barton²

¹Uppsala University

²IST Austria

³Friedrich Miescher Laboratory of the Max Planck Society

February 7, 2022

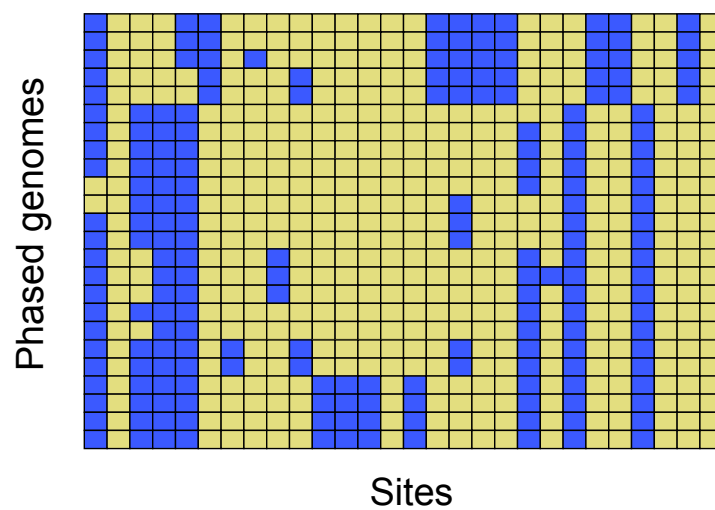
Abstract

The term “haplotype block” is commonly used in the developing field of haplotype-based inference methods. We argue that the term should be defined based on the structure of the Ancestral Recombination Graph (ARG), which contains complete information on the ancestry of a sample. We use simulated examples to demonstrate key features of the relation between haplotype blocks and ancestral structure, emphasising the stochasticity of the processes that generate them. Even the simplest cases of neutrality or of a “hard” selective sweep produce a rich structure, which is missed by commonly used statistics. We highlight a number of novel methods for inferring haplotype structure as full ARG, or as a sequence of trees. While some of these new methods are computationally efficient, they still lack features to aid exploration of the haplotype blocks, as we define them, thus calling for the development of new methods. Understanding and applying the concept of the haplotype block will be essential to fully exploit long and linked-read sequencing technologies.

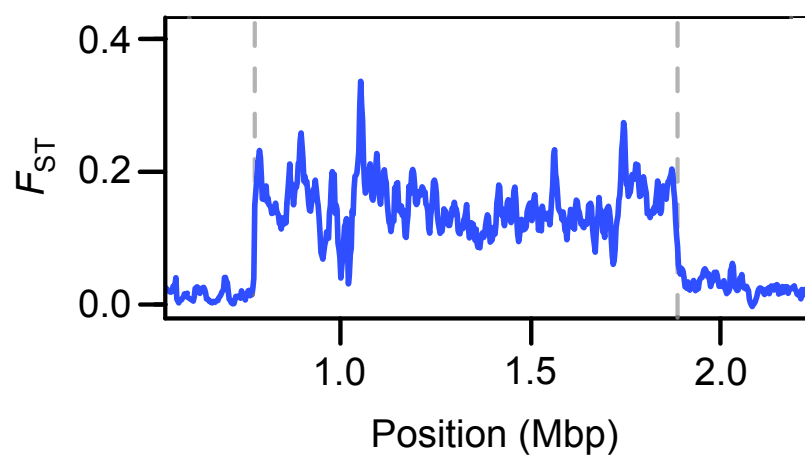
Hosted file

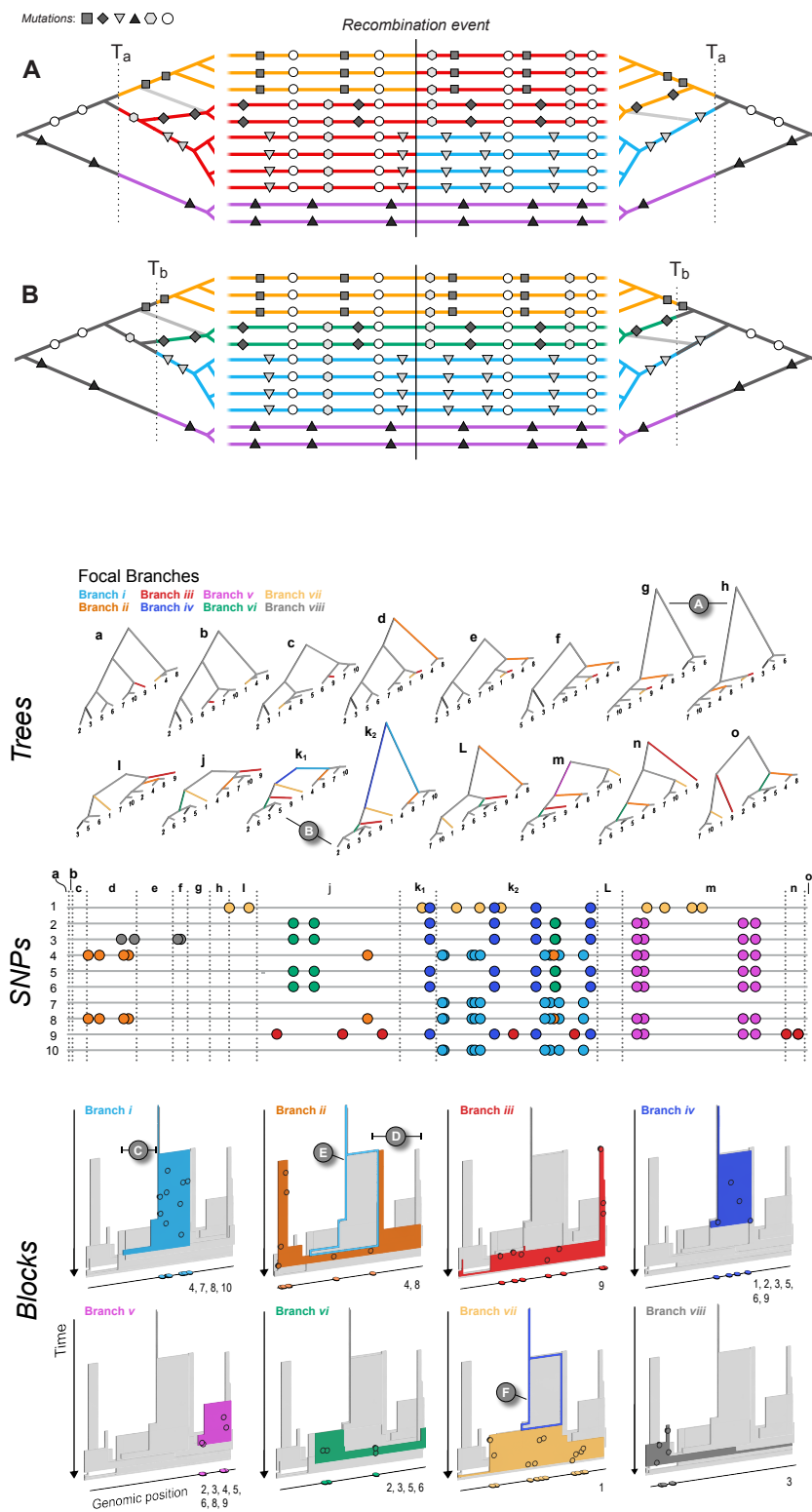
Shiplina_et_al_MolEcol.docx available at <https://authorea.com/users/459348/articles/555633-on-the-origin-and-structure-of-haplotype-blocks>

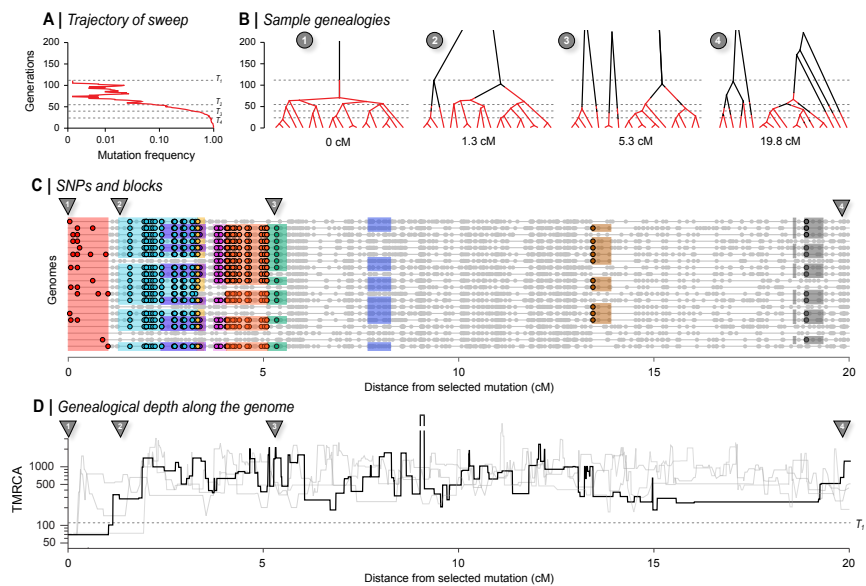
A | *Phased sequence data*



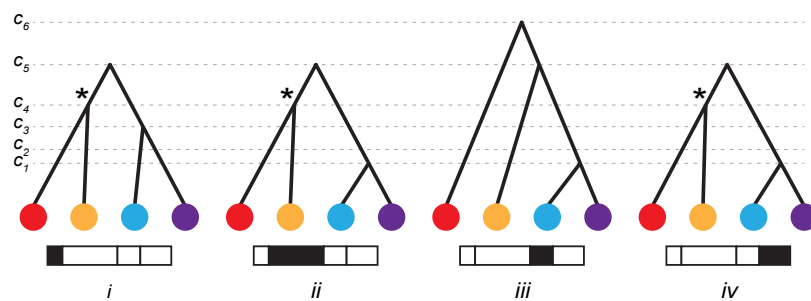
B | *Scan of chromosome-wide variation*







A | Genealogies along the genome



B | Ancestral Recombination Graph

