

# On the origin and structure of haplotype blocks

Shipilina, Daria <sup>1,2</sup>, Pal, Arka <sup>2</sup>, Stankowski, Sean<sup>2</sup>, Chan, Yingguang Frank <sup>2,3</sup>, and Barton, Nicholas<sup>4</sup>

<sup>1</sup>Evolutionary Biology Program, Department of Ecology and Genetics (IEG)

<sup>2</sup>Institute of Science and Technology Austria

<sup>3</sup>Friedrich Miescher Laboratory, Max Planck Society

<sup>4</sup>Affiliation not available

January 27, 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.pdf available at <https://authorea.com/users/456606/articles/554177-on-the-origin-and-structure-of-haplotype-blocks>