## 

[**Understanding haplotypes**](#_ygqz5buswerh) **1**

[Introduction](#_htx9fdgd44jd) 1

[Defining haplotype block](#_mdf4s0pnl2sz) 1

[Practical definition of the haplotype block](#_x0b7uoyue9w2) 1

[Length of the haplotype: expectation from simulation](#_b2tb8gamqt6c) 2

[Future directions](#_8p73tzm2kfb9) 2

[References](#_40wrs0l54gi9) 2

## 

# Understanding haplotypes

### Introduction

1. Side-based vs. haplotype-based approach
2. Motivation to use haplotype structure
   1. New inference methods (tsinfer, Relate)
   2. New methods for molecular phasing (linked reads, haplotagging)

“Haplotype”” and “Haplotype block” are widely used terms, and have increased in importance for several reasons:

- phasing and imputation depend on assumptions about haplotype structure

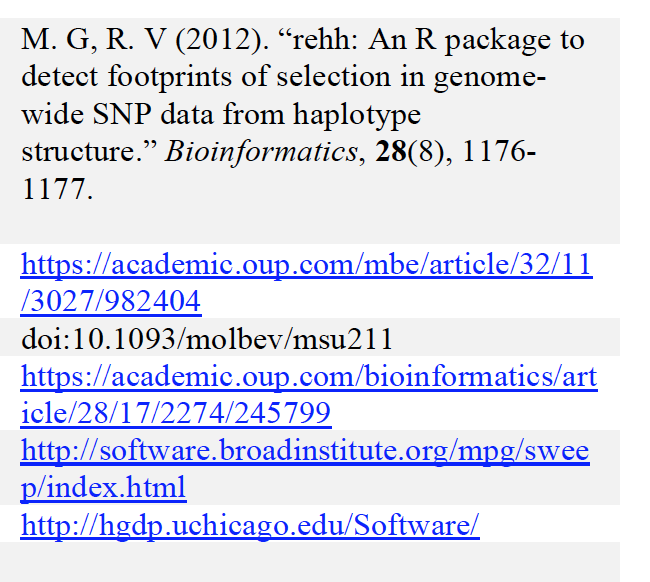
- haplotypes are increasingly used in inference, of selective sweeps, introgression, and population structure.

### Defining haplotype block

1. Relation to the concept of identity by descent

### Practical definition of the haplotype block

1. Before considering haplotype block: what limitations can be added by phasing
2. Definition of the haplotype block in analysis of the population structure:
   1. iHS (Integrated Haplotype Score): length of derived haplotype is longer then length of ancestral one, based on LD.

Limitation: point estimate for one focal gene

iHS is implemented in

1. Integrated analysis: tsinfer + tsdate and Relate
   1. Ancestral Haplotype reconstruction serves as a basis for block length defifnition

### Length of the haplotype: expectation from simulation

1. Neutral scenario
2. Selective sweep
3. Island model: population structure
4. Special case: balancing selection

### Future directions

1. What type of analysis can be improved while using haplotype length and frequency as a statistics?
2. What is the new information we gain?

### References

<https://www.frontiersin.org/research-topics/8750/haplotype-analysis-applied-to-livestock-genomics>

2. Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, et al. 2002 Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832-837.

**Introduction**

“Haplotype”” and “Haplotype block” are widely used terms, and have increased in importance for several reasons:

- phasing and imputation depend on assumptions about haplotype structure

- haplotypes are increasingly used in inference, of selective sweeps, introgression, and population structure.

Definition & theory

- original and simple meaning of “haplotype”

- “haplotype blocks” can be defined through identity by descent from a reference population

- in some cases, there is an obvious reference population (eg in a selection experiment), but in general, it is arbitrary - we should think in terms of the ARG; then, we see that haplotype blocks correspond to sets of genomes that depend from particular branches.

- using this definition, in which blocks descend from some branch, they can be detected through carrying mutations with a certain configuration;

- the number of such mutations corresponds to the length of the branch.

- close relation with Konrad’s blockwise SFS, which also looks at the numbers of each mutational configuration in a window.

- what extra information comes from knowing the block length (and how exactly do we define that?)

Usage

- What definitions are used in the literature?

- Verbally, and implemented in software ?

- how accurate are the algorithms in inferring the “true” blocks? (hard to determine without a lot of simulation, but maybe one can make some comment)

- sequential Markov coalescent (SMC) is assumed as an approximation, whenever a Hidden Markov Model is used.

Example/Simulation/Cartoon

- This should, if possible, represent a real simulated example, even if we are aiming just for an illustration

- Illustrate typical patterns:

- neutral

- hard sweep

- false positive - a region with (by chance) an unusually short genealogy

- island model

- balancing selection