perm.pub/dsi:0iV47kWzQAuyONrIDG538k3x3Qc/0.1
Additional formats and editions available online.

Edition 0.1

Obsolete

Author date: 2022-06-16

Citation:
E. Castedo Ellerman (2022) "Gametic lineage space" perm.pub https://perm.pub/dsi:0iV47kWzQAuyONrIDG538k3x3Qc/0.1

Gametic lineage space

E. Castedo Ellerman (castedo@castedo.com)

Abstract

STAGE: EARLY DRAFT

DOCUMENT TYPE: Definition Document

OBJECTIVES

- Formal mathematical definition to be used in stochastic process model under development.
- Precise mathematical definition for clarification of certain concepts discussed in the population genetics literature.

Gametic genealogy

A *gametic genealogy* is a convenient mathematical formalism of the genealogy of a population from the perspective of gametes. Mathematically, it is a quadruple

$$(\mathsf{Gam}, \mathsf{Mate}, \mathsf{Par}, \mathsf{Fert})$$

with components

- Gam, the set of underlying gametes,
- Mate, the set of zygotes formed by the fusion of egg gametes and sperm gametes,
- Par, a mapping from child gametes to parent zygotes, and
- Fert, a mapping from zygotes to fertilization time.

For convenience, given a *gametic genealogy*,

- Gam_0 denotes the set of egg gametes,
- Gam_1 denotes the set of sperm gametes, and
- Mate_* denotes the mapping from gametes to the zygotes they formed during fertilization.

Formally, a *gametic genealogy* must satisfy the following conditions.

$\mathsf{Mate} \subset \mathsf{Gam}_0 \times \mathsf{Gam}_1$ where $\mathsf{Gam}_0 \cap \mathsf{Gam}_1 = \emptyset$, $\mathsf{Gam}_0 \cup \mathsf{Gam}_1 = \mathsf{Gam}$ and Mate forms a one-to-one mapping between Gam_0 and Gam_1.

Par is a function $C \mapsto \mathsf{Mate}$, where C is a subset of Gam representing child gametes.

Fert is a function Mate $\mapsto \mathbb{R}$ such that for all child gametes $g \in \operatorname{dom}$ Par,

$$\mathsf{Fert}(\mathsf{Mate}_*(g)) > \mathsf{Fert}(\mathsf{Par}(g))$$

Note that dom Par denotes the domain of Par, that is, the set of child gametes.

Gametic lineage space

A *gametic lineage space* is a mathematical formalism representing the lines of transmission of genetic information via gametes of a population over time. It is a triplet

$$(\mathsf{Loc}, G, \mathsf{Lin})$$

where

- Loc is the set of all genomic locations,
- G is a gametic genealogy (Gam, Mate, Par, Fert), and
- Lin is a function Loc $\times$ Gam $\mapsto 2^{\mathsf{Gam}}$ mapping a genomic position in a gamete to the set of gametes that transmitted genetic information to that position.

For every location $\ell \in$ Loc and gamete $g \in$ Gam, Lin(ℓ, g) is the lineage ending at gamete g via locus ℓ and it must satisfy the condition

$$\mathsf{Lin}(\ell, g) = \{g\} \cup \mathsf{Lin}(\ell, \mathsf{Par}(g)_i) \text{ for either } i = 0 \text{ or } i = 1$$

when $g \in \operatorname{dom}$ Par, otherwise Lin$(\ell, g) = \{g\}$.

Example mathematical application

Given a sample of gametes S, define the genomic locations reached by an ancestral gamete as

$$R_S(g) := \{\ell \in \mathsf{Loc} : \exists g' \in S\,(g \in \mathsf{Lin}(\ell, g'))\}$$

We conjecture that the set

$$\{R_S(g) : g \in \mathsf{Gam}\}$$

is the set of haplotype blocks defined in [1].

References

1. Shipilina D, Stankowski S, Pal A, Chan YF, Barton N. On the origin and structure of haplotype blocks. Preprints; 2022 Feb. doi:10.22541/au.164425910.09070763/v1