



Edition 0.1

⚠ Obsolete

Author date: 2022-06-16

Citation:

E. Castedo Ellerman (2022) "Gametic lineage space" perm.pub
https://perm.pub/dsi:Fs4CjTor07Ssbb69Qcggp5yiBA0/0.1

Copyright:

creativecommons.org/licenses/by/4.0/
2022 © The Authors. This document is distributed under a Creative Commons Attribution 4.0 International license.

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

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

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.

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

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

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

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

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

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

where

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

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

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

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

Example mathematical application

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

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

We conjecture that the set

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

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](https://doi.org/10.22541/au.164425910.09070763/v1)