

Population Genetics

Peerteaching KW24

Samuel Hehn
Swastik Kashyap

Universität Tübingen

11. Juni 2024

Inhalt

Introduction

Basic models of population growth

Alleles and Ploidy

Calculating Populations

The coalescent process

The standard coalescent model

Most recent common ancestor

Introduction

Population
Genetics

Samuel Hehn
Swastik Kashyap

Population genetics aims to infer details of evolutionary processes based on the current population's genetic composition.

Subfield of Genetics

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

Single Nucleotide Variant(SNV)

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

A single-nucleotide variant (SNV) is a variation in a single nucleotide that occurs at a specific position in the genome. This may be rare or common in a population.

Single Nucleotide Polymorphism(SNP)

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

A single-nucleotide polymorphism (SNP) is a SNV that occurs in a significant proportion (typically $> 1\%$) of a population.

Polymorphisms describe sites (nucleotide positions, etc.) that are variable within a species; divergence describes sites variable between species.

Basic models of population growth

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

$$r = \frac{s(T) - s(T_0)}{s(T_0)}$$

where $s(T)$ is the population size at time T and $s(T_0)$ the size of the initial population at time T_0 .

Alleles and ploidy

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

An allele is the variant form of a given gene found at the same chromosomal location.

Ploidy is the number of sets of chromosomes in a cell and hence the number of possible alleles for genes.

Haploid and Diploid

Haploid Cells: Haploid cells have half the number of chromosomes (n) as diploid - germ cells Result of meiosis

Example: sperm and egg cells

Diploid Cells: Diploid cells contain two copies of each chromosome ($2n$) - somatic cells Result of mitosis

Example: skin cells

Haploid Reproduction Model

- Assume constant population size
- Each gene i in generation $t + 1$ is found by randomly choosing a predecessor gene g in generation t . This implies, that it is possible that one gene can have more than one offspring.
- Any gene in generation t not chosen “dies out”.

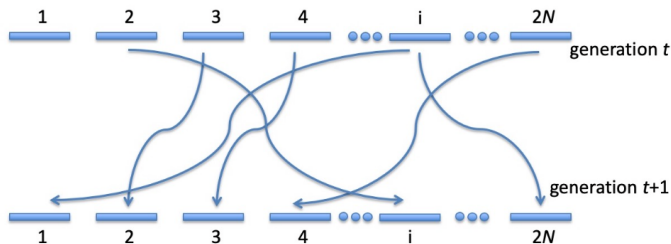


Abbildung: Haploid reproduction model

Diploid populations

Assumptions - Species has 2 Sexes, Male and Female.

Each individual has 2 copies of each gene, a gene is chosen for each parent with equal odds

In this model, each gene has one parent gene and each individual has two parents.

For large values of N , N_f and N_m , one can approximate the diploid model by the haploid model. Thus, we will only consider the haploid model.

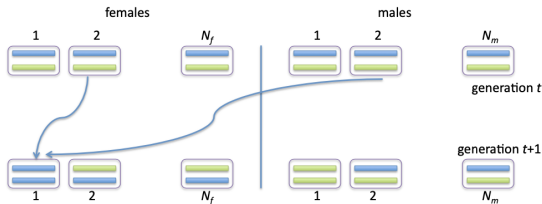


Abbildung: Diploid reproduction model

Haplotypes and Genotypes

Population
Genetics

Samuel Hehn
Swastik Kashyap

Haplotype: A haplotype is a sequence of an individual's genome that occurs together on one or two homologous copies of a chromosome.

Genotype: The two alleles at the same site on two homologous chromosomes form the genotype at that site. We denote that by 'P|Q', if 'P' and 'Q' are the alleles, respectively. Example

An individual has the two haplotypes

A T T G A C A T C

A C T G A C A C T

Then the genotype at the first site is A|A, at the second site T|C, and so on.

Task- Show the full genotype of the individual.

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

Homozygous and Heterozygous

A site c is called homozygous if the genotype consists of two identical alleles at c . The site c is called heterozygous if the genotype consists of two different alleles at c .

Task- for an individual that has the two haplotypes

A T T G A C A T C

A C T G A C A C T

give the set of homozygous and heterozygous sites.

Calculating decendants

What is the probability of any gene i to have at least one decendant (It does not “die out”)?

Calculating decendants

What is the probability of any gene i to have at least one decendant (It does not “die out”)?

We look at the inverse probability, a gene has no decendants.

Calculating decendants

What is the probability of any gene i to have at least one decendant (It does not “die out”)?

We look at the inverse probability, a gene has no decendants.

$$\implies \mathbb{P}_{\text{Poisson}}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Calculating decendants

What is the probability of any gene i to have at least one decendant (It does not “die out”)?

We look at the inverse probability, a gene has no decendants.

$$\implies \mathbb{P}_{\text{Poisson}}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Since the population keeps it size from generation to generation, the average number of decendants per gene is 1.

Calculating decendants

What is the probability of any gene i to have at least one decendant (It does not “die out”)?

We look at the inverse probability, a gene has no decendants.

$$\implies \mathbb{P}_{\text{Poisson}}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Since the population keeps it size from generation to generation, the average number of decendants per gene is 1.

$$\implies \mathbb{P}_{\text{Poisson}}(X = k) = \frac{1}{k!} e^{-1}$$

Calculating decendants

What is the probability of any gene i to have at least one decendant (It does not “die out”)?

We look at the inverse probability, a gene has no decendants.

$$\implies \mathbb{P}_{\text{Poisson}}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Since the population keeps it size from generation to generation, the average number of decendants per gene is 1.

$$\implies \mathbb{P}_{\text{Poisson}}(X = k) = \frac{1}{k!} e^{-1}$$

No decendants: ($k = 0$)

$$\mathbb{P}_{\text{Poisson}}(X = 0) = e^{-1} \approx 0.37$$

At least one decendant:

$$1 - 0.37 \approx 0.63$$

Calculating decendants

Why is this useful?

What was the size of a population of 10 000 genes $t = 15$ generations ago?

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

Calculating decendants

Why is this useful?

What was the size of a population of 10 000 genes $t = 15$ generations ago?

$$10\,000 \cdot 0.63^{15}$$

Calculating decendants

Why is this useful?

What was the size of a population of 10 000 genes $t = 15$ generations ago?

$$10\,000 \cdot 0.63^{15} \approx 10$$

Calculating decendants

Why is this useful?

What was the size of a population of 10 000 genes $t = 15$ generations ago?

$$10\,000 \cdot 0.63^{15} \approx 10$$

So a population of 10 000 genes would have most likely evolved from just 10 genes over a course of 15 generations.

Calculating decendants

Why is this useful?

What was the size of a population of 10 000 genes $t = 15$ generations ago?

$$10\,000 \cdot 0.63^{15} \approx 10$$

So a population of 10 000 genes would have most likely evolved from just 10 genes over a course of 15 generations.

This means we can now find out how many generations back there was just one individual (the most recent common ancestor of our whole population).

The coalescent process

Population
Genetics

Samuel Hehn
Swastik Kashyap

“*to coalesce*”: grow together, to join, to fuse

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

**The coalescent
process**

The standard
coalescent model

Most recent
common ancestor

The coalescent process

Population
Genetics

Samuel Hehn
Swastik Kashyap

“to coalesce”: grow together, to join, to fuse

Definition (coalescent event)

If traversing the sequence-transmission paths backward in time, two sequence transmission paths intersect at some sequence, the paths coalesce at that intersection point. This is called a coalescent event.

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

The coalescent process

Population
Genetics

Samuel Hehn
Swastik Kashyap

“to coalesce”: grow together, to join, to fuse

Definition (coalescent event)

If traversing the sequence-transmission paths backward in time, two sequence transmission paths intersect at some sequence, the paths coalesce at that intersection point. This is called a coalescent event.

Basic idea

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

The coalescent process

Population
Genetics

Samuel Hehn
Swastik Kashyap

“to coalesce”: grow together, to join, to fuse

Definition (coalescent event)

If traversing the sequence-transmission paths backward in time, two sequence transmission paths intersect at some sequence, the paths coalesce at that intersection point. This is called a coalescent event.

Basic idea

- ▶ Start with present-day generation

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

The coalescent process

“to coalesce”: grow together, to join, to fuse

Definition (coalescent event)

If traversing the sequence-transmission paths backward in time, two sequence transmission paths intersect at some sequence, the paths coalesce at that intersection point. This is called a coalescent event.

Basic idea

- ▶ Start with present-day generation
- ▶ Construct previous generations

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

The coalescent process

“to coalesce”: grow together, to join, to fuse

Definition (coalescent event)

If traversing the sequence-transmission paths backward in time, two sequence transmission paths intersect at some sequence, the paths coalesce at that intersection point. This is called a coalescent event.

Basic idea

- ▶ Start with present-day generation
- ▶ Construct previous generations
- ▶ By randomly choosing parents in the previous generation

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

The coalescent process

Example

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

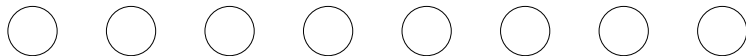
**The coalescent
process**

The standard
coalescent model

Most recent
common ancestor

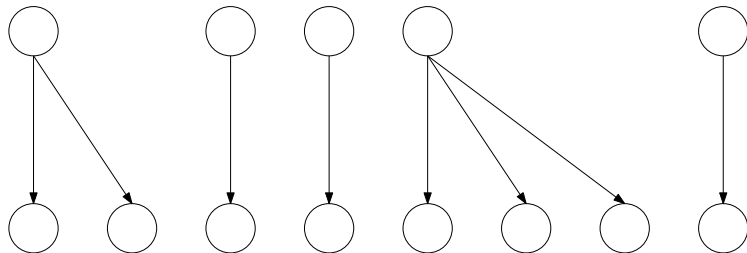
The coalescent process

Example



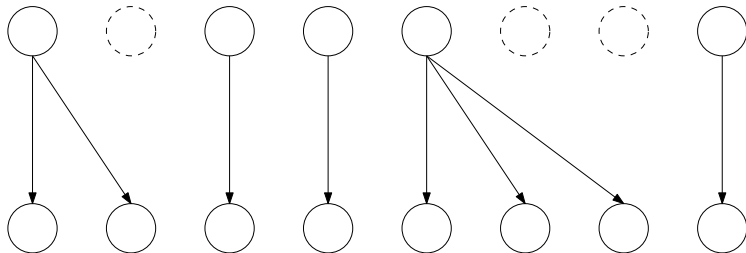
The coalescent process

Example



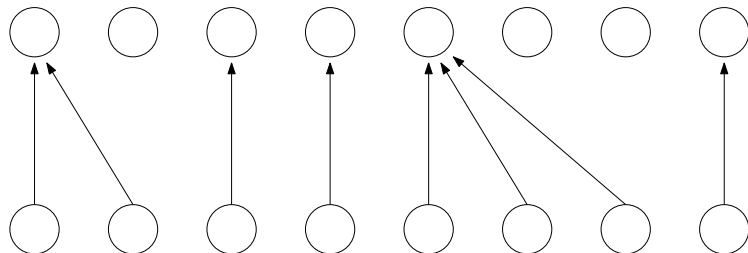
The coalescent process

Example



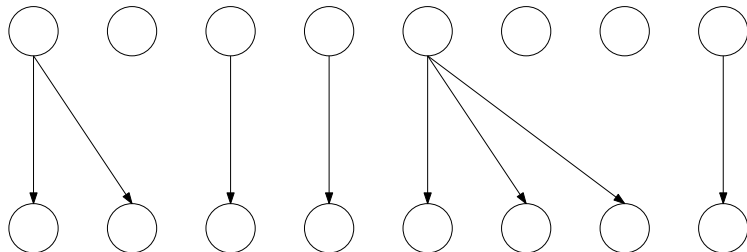
The coalescent process

Example



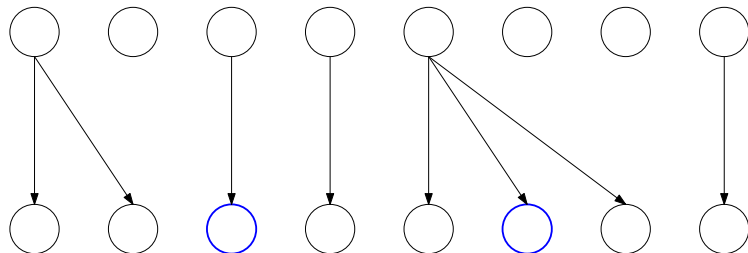
The coalescent process

Example



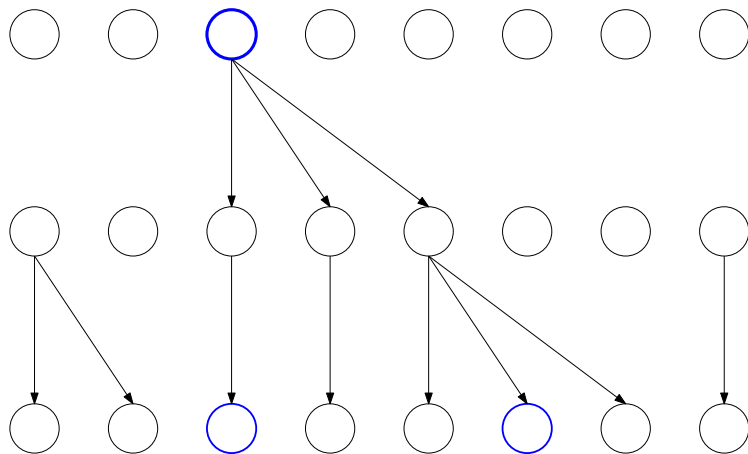
The coalescent process

Example



The coalescent process

Example



Coalescence of 2 genes

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

Considering a haploid model with n genes.

For two present day genes i and j , when did they coalesce?

Coalescence of 2 genes

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

Considering a haploid model with n genes.

For two present day genes i and j , when did they coalesce?

We want to know two things:

Coalescence of 2 genes

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

Considering a haploid model with n genes.

For two present day genes i and j , when did they coalesce?

We want to know two things:

1. When did the two genes coalesce?

Coalescence of 2 genes

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

Considering a haploid model with n genes.

For two present day genes i and j , when did they coalesce?

We want to know two things:

1. When did the two genes coalesce?
→ Who is their common ancestor

Coalescence of 2 genes

Considering a haploid model with n genes.

For two present day genes i and j , when did they coalesce?

We want to know two things:

1. When did the two genes coalesce?
→ Who is their common ancestor
2. How long is the waiting time until the two genes coalesced?

Coalescence of 2 genes

Considering a haploid model with n genes.

For two present day genes i and j , when did they coalesce?

We want to know two things:

1. When did the two genes coalesce?
→ Who is their common ancestor
2. How long is the waiting time until the two genes coalesced?
→ How many generations back is their common ancestor?

Coalescence of 2 genes

When did the two genes coalesce?

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

**The standard
coalescent model**

Most recent
common ancestor

Coalescence of 2 genes

When did the two genes coalesce?

We select a random ancestor for each individual:

Coalescence of 2 genes

When did the two genes coalesce?

We select a random ancestor for each individual:
Probability to select the right ancestor of i is 1, since there are no requirements.

Coalescence of 2 genes

When did the two genes coalesce?

We select a random ancestor for each individual:
Probability to select the right ancestor of i is 1, since there are no requirements.
Probability to select the right ancestor of j is

Coalescence of 2 genes

When did the two genes coalesce?

We select a random ancestor for each individual:

Probability to select the right ancestor of i is $\frac{1}{n}$, since there are no requirements.

Probability to select the right ancestor of j is

$$P(T_2 = 1) = \frac{1}{n}$$

since we need to "hit" the ancestor we've chosen for i .

Coalescence of 2 genes

How long is the waiting time until the two genes coalesced?

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

**The standard
coalescent model**

Most recent
common ancestor

Coalescence of 2 genes

How long is the waiting time until the two genes coalesced?

What is the Probability that the common ancestor is in Generation t ?

Coalescence of 2 genes

How long is the waiting time until the two genes coalesced?

What is the Probability that the common ancestor is in
Generation t ?
($n - 1$ failures following one success)

$$P(T_2 = t) =$$

Coalescence of 2 genes

How long is the waiting time until the two genes coalesced?

What is the Probability that the common ancestor is in Generation t ?
($n - 1$ failures following one success)

$$P(T_2 = t) = \left(1 - \frac{1}{n}\right)^{n-1}$$

Coalescence of 2 genes

How long is the waiting time until the two genes coalesced?

What is the Probability that the common ancestor is in Generation t ?
($n - 1$ failures following one success)

$$P(T_2 = t) = \left(1 - \frac{1}{n}\right)^{n-1} \cdot \left(\frac{1}{n}\right)$$

Coalescence of 2 genes

Exercise

Let's assume $n = 50$

What is the probability that two genes coalesce exactly 5 generations in the past?

$$P(T_2 = 5) =$$

Coalescence of 2 genes

Exercise

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

Let's assume $n = 50$

What is the probability that two genes coalesce exactly 5 generations in the past?

$$P(T_2 = 5) = \left(1 - \frac{1}{50}\right)^{5-1} \cdot \frac{1}{50}$$

Coalescence of 2 genes

Exercise

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

Let's assume $n = 50$

What is the probability that two genes coalesce exactly 5 generations in the past?

$$P(T_2 = 5) = \left(1 - \frac{1}{50}\right)^{5-1} \cdot \frac{1}{50} \approx 0.01845 = 1.845\%$$

Coalescence of 2 genes

Exercise

Let's assume $n = 50$

What is the probability that it takes them at least 5 generations to coalesce?

$$P(T_2 > 5) =$$

Coalescence of 2 genes

Exercise

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

Let's assume $n = 50$

What is the probability that it takes them at least 5 generations to coalesce?

$$P(T_2 > 5) = \left(1 - \frac{1}{50}\right)^{5-1}$$

Coalescence of 2 genes

Exercise

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

Let's assume $n = 50$

What is the probability that it takes them at least 5 generations to coalesce?

$$P(T_2 > 5) = \left(1 - \frac{1}{50}\right)^{5-1} \approx 0.92237 = 92.237\%$$

Most recent common ancestor(MRCA)

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

**Most recent
common ancestor**

How many generations back is the most recent common ancestor (MRCA) of two present-day genes?

Wright Fisher Model

Population
Genetics

Samuel Hehn
Swastik Kashyap

Introduction

Basic models of
population growth

Alleles and Ploidy

Calculating
Populations

The coalescent
process

The standard
coalescent model

Most recent
common ancestor

Key assumptions of the Wright-Fisher model:

1. Discrete and non overlapping generations.
2. Constant population size.
3. Haploid individuals
4. All individuals are equally fit.
5. The population has no geographic or social structure
6. No recombinations of genes (or sequences)