A genomicist introduces quantitative genetics

Pradipta Ray,

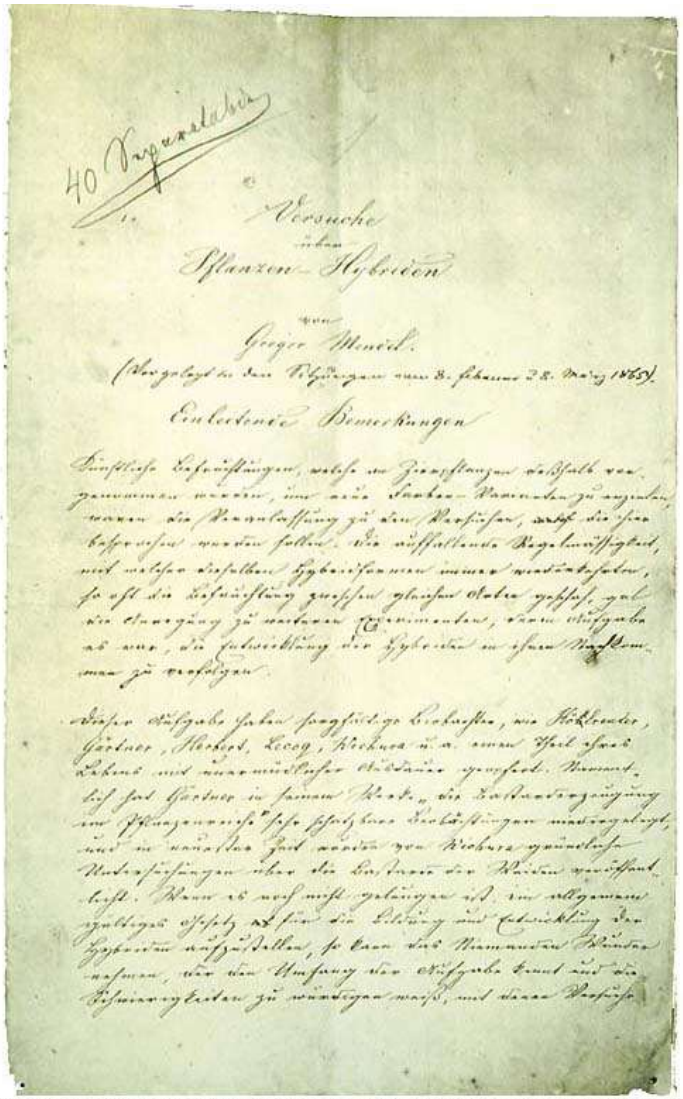
BIOL 6385 / BMEN 6389,

The University of Texas at Dallas

(some material based on content by PR in Eric Xing's 10-810 Carnegie Mellon class)

Wikipedia

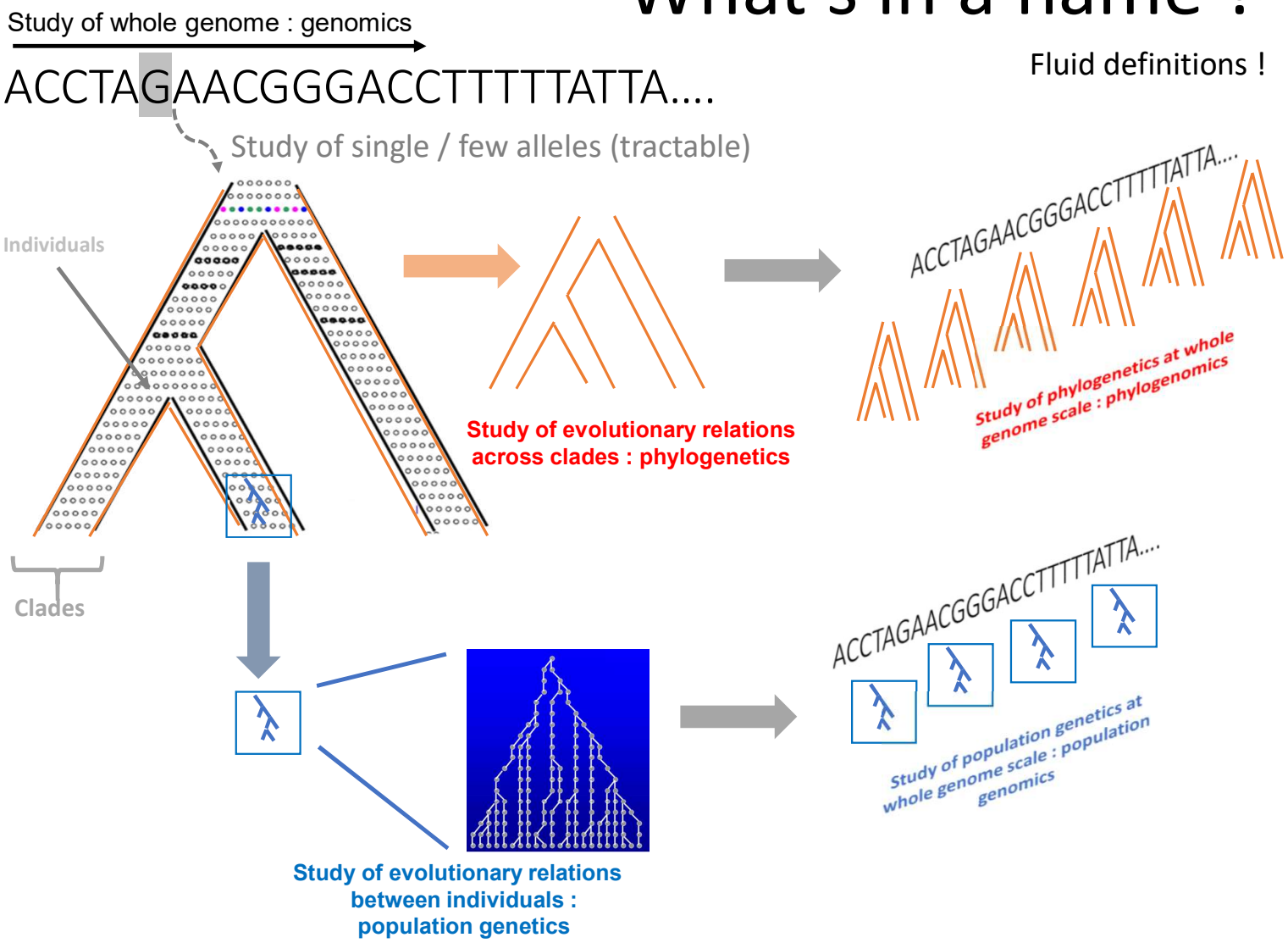
Father of modern genetics



Courtesy of American Philosophical Society, Curt Stern Papers. Noncommercial, educational use only.

What's in a name ?

Fluid definitions !



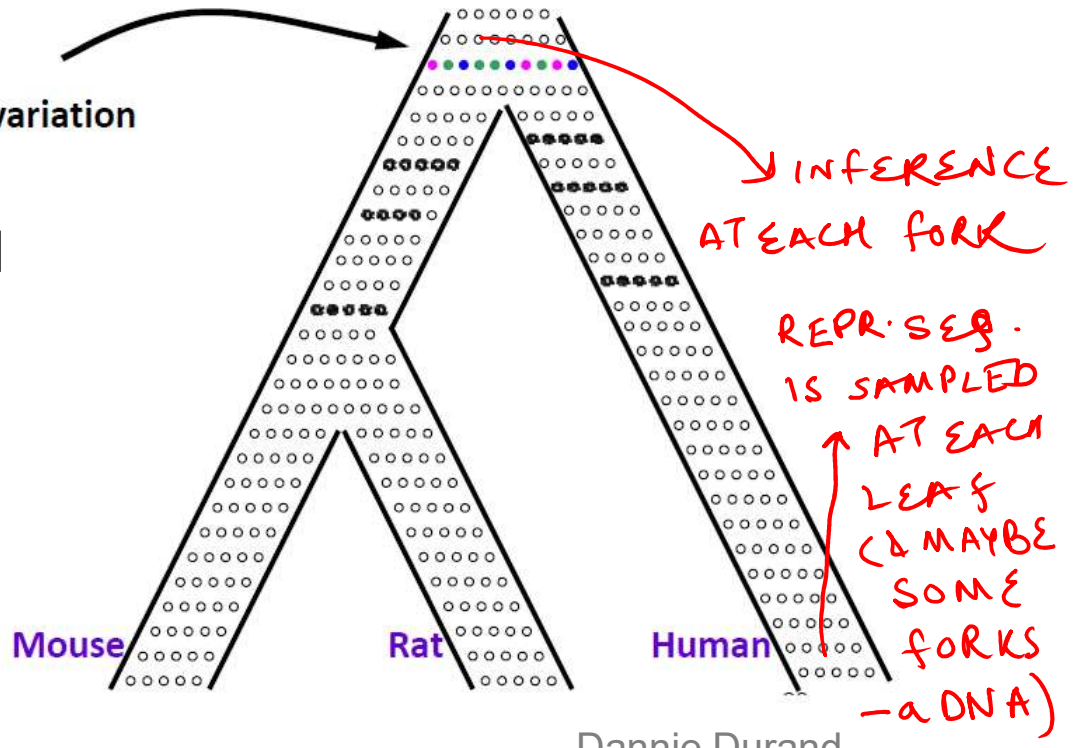
Interconnected disciplines

- Evolutionary process : cooking pot
- Alleles : ingredients
- Drift, mutation, selection, recombination, population structure and migration, **stochasticity** : recipe
- Changes in allele frequencies : outcome of the process !
- Often, the goal is to observe the outcome and make evidence-driven guesses about **missing pieces** of the recipe
  - **CLASSICAL GENETICS : genotype – phenotype relationships**: phenotype-associated loci, epistasis model, quantitative trait models, **pedigree based inference**
  - **POPULATION GENETICS: evolutionary forces**: mutation rates, selectional model, recombination rate, **demography:** migratory model, population size, **population structure:** coalescents

What is a species phylogeny, really ?

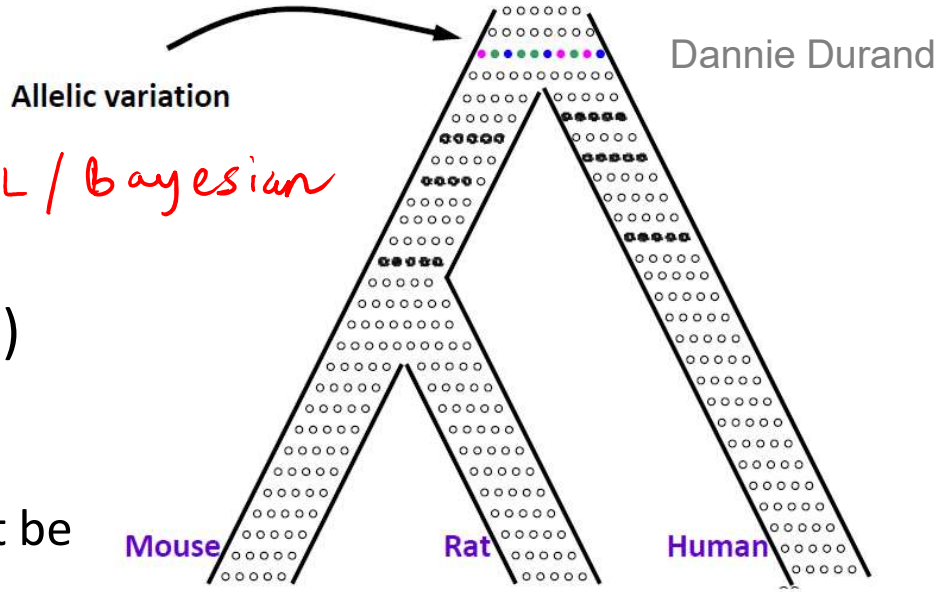
- It's a jungle out there !

Pop gen variation should be modelled for a clearer insight into evolutionary dynamics



Problems with phylogenetic inference

- ML inference: may not be representative



ML / bayesian

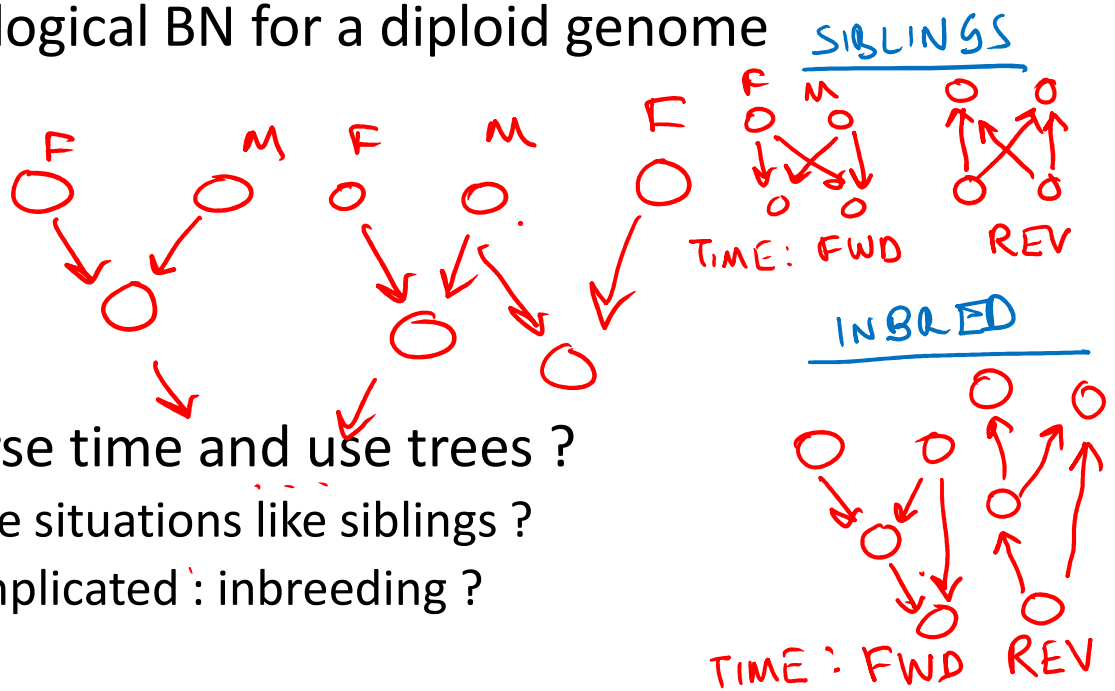
- Multi dimensionality (multiple genomic loci)
  - full joint : may not be tractable
  - full marginal : may not be consistent

TRADE OFF

REMEMBER POSTERIOR DECODING?

Why not use phylo methods for pop genetics ?

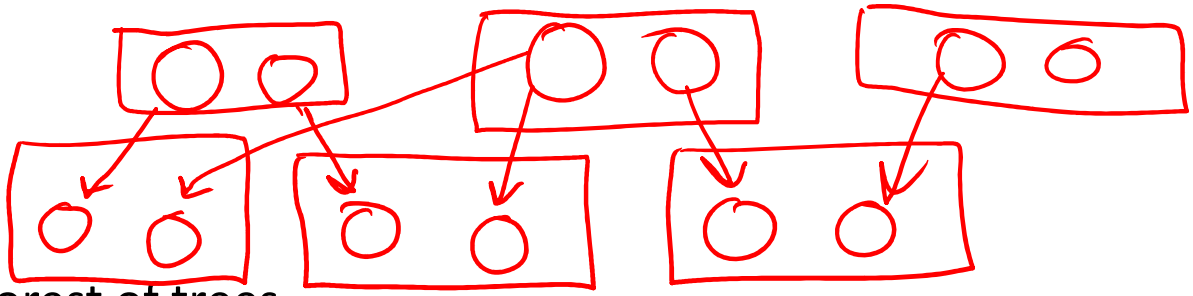
- Because populations are not well modelled by trees
- Typical genealogical BN for a diploid genome population



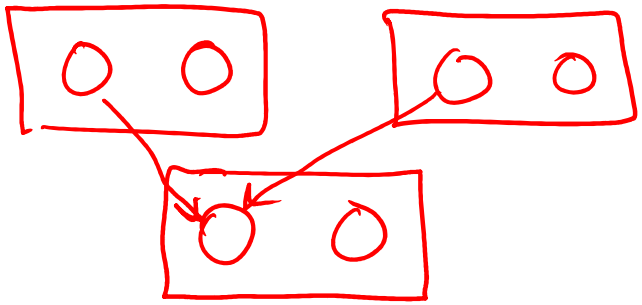
- Cant we reverse time and use trees ?
  - how to tackle situations like siblings ?
  - or more complicated : inbreeding ?

Why not use phylogenetic methods ?

- Maybe we can just model chromosomes …



  - a forest of trees
- But recombination breaks the tree structure

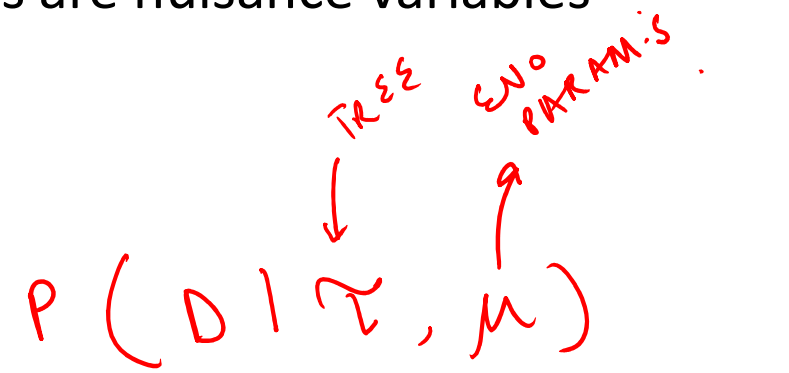


Why not use phylogenetic methods ?

- How about just modelling single alleles ? (chances are recombination inside allele would be low)
  - even then tree models are not appropriate
    - migration (lateral movement of genes)
    - longer alleles or multi locus models : recombinants cannot be ignored

Why not use phylogenetic methods ?

- In pop genetics, trees or genealogies are nuisance variables (marginalized out)
- Phylogenetics :



- Pop genetics:

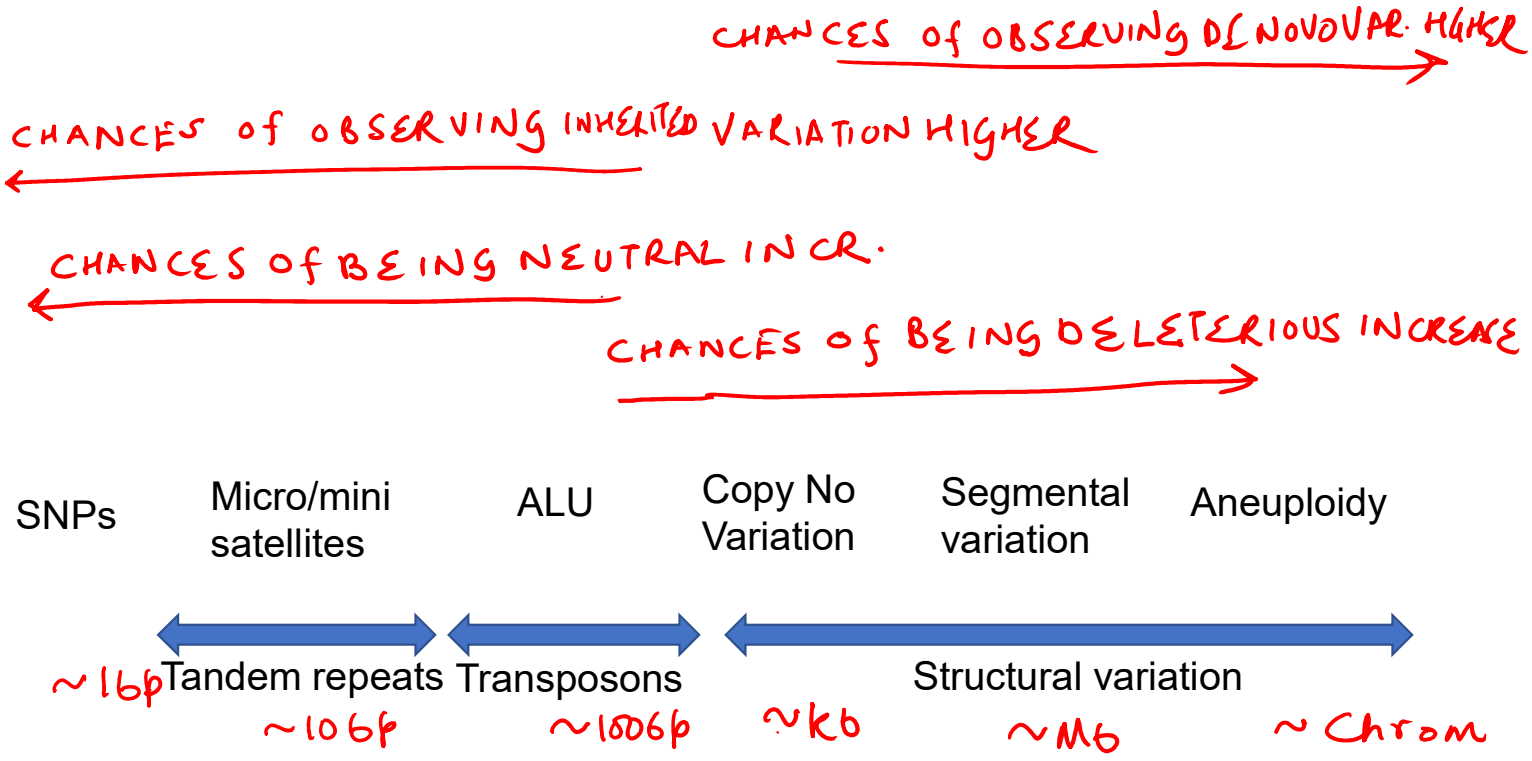
$$\sum_{\tau} P(D \mid \tau, \mu) P(\tau, \mu) = P(D, \mu)$$

[SOMETIMES P(D | μ) MODEL]

→ TYPICALLY, ALLELE FREQ. EVOLUTION IS MODELLED AS STOCH PROCESS EVOLVING OVER TIME (GENERATIONS)

Genetic variability of population

Due to polymorphisms at various genetic loci :

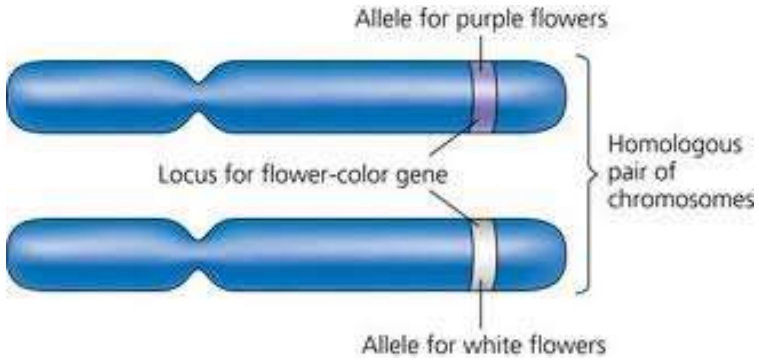


Nomenclature and brief explanations

- “Traits”
- Alleles, identity by state and by descent
- Phenotype, genotype, haplotype

Unit of pop gen evolution : alleles

- Mutually exclusive variants at a genetic locus
- Organism, wrt an allele :
  - hemizygous : only one copy of chromosome
  - homozygous : both copies have same allele
  - heterozygous : copies have different alleles



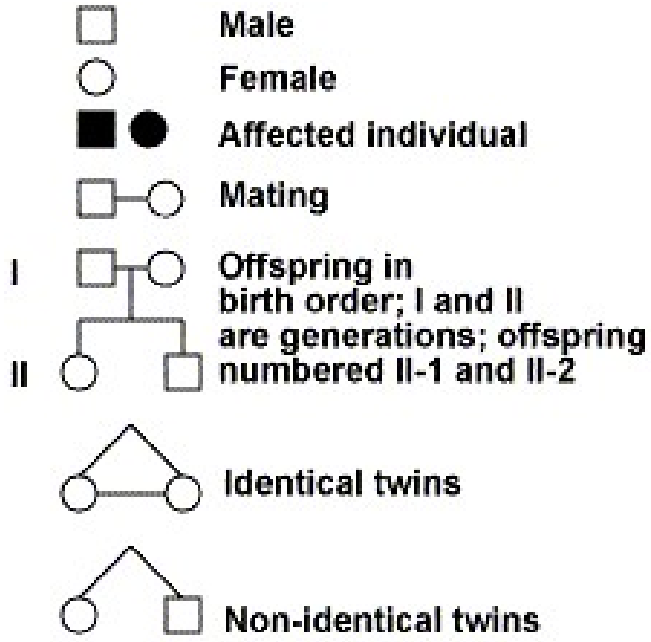
Haplotype vs genotype

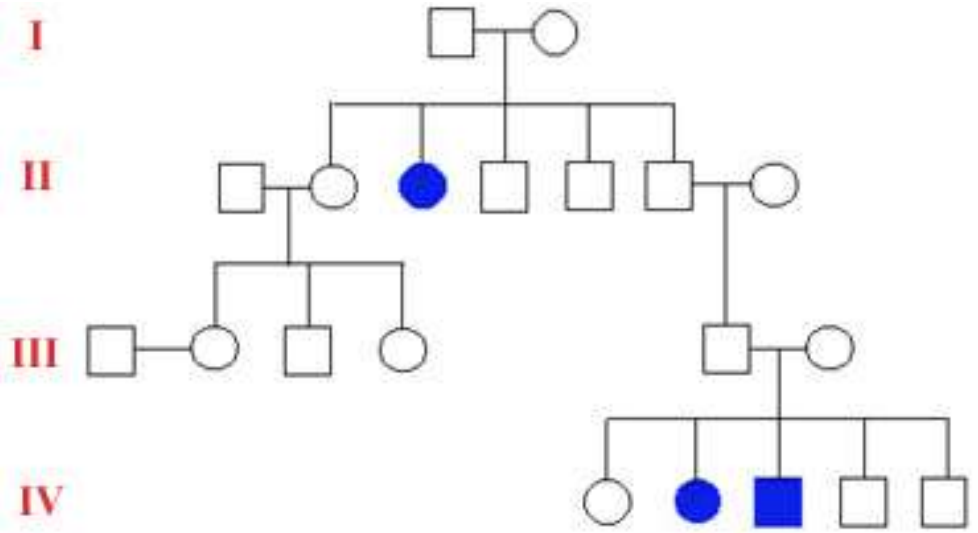
- When we know the allelic composition of multiple alleles in an individual, can we partially reconstruct the chromosomes ?

Zhou and Wang *BMC Bioinformatics* 2007 **8**:484

- How many possible haplotype pairs based on genotype of set of loci, n of which heterezygous?

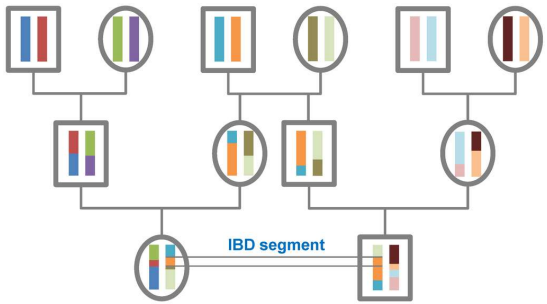
Pedigree : modelling relations between individuals





IBD and IBS in alleles

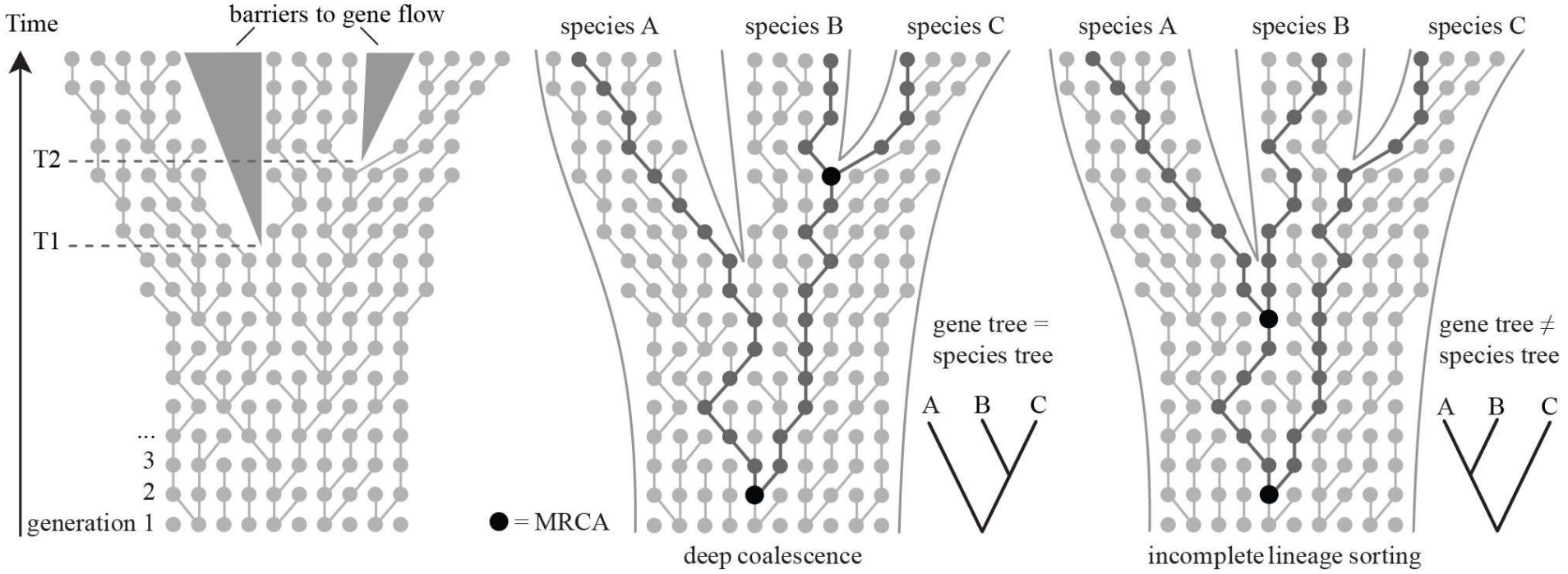
- Identity by state (IBS) : if two alleles are identical based on their sequence
- Identity by descent (IBD) : if two alleles are identical based on common ancestor, with no recombination or mutation affecting descent
- IBD implies IBS, but not the other way around



Wikipedia

Coalescent

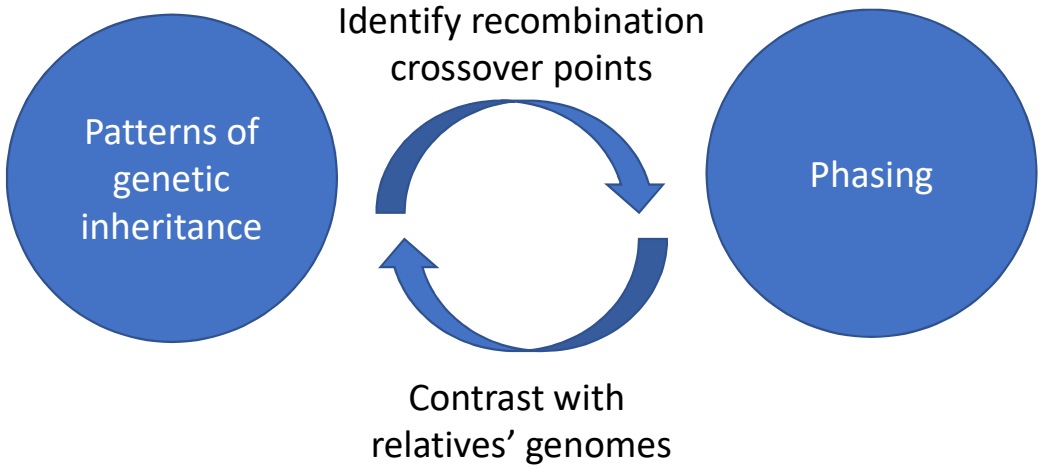
- Shows evolutionary history of a population of alleles



https://frederikleliaert.wordpress.com/

Phasing

- The process of identifying the haplotypes
  - Long read sequencing
  - Observed patterns in short read sequencing
  - Sequencing related individuals



Epistasis

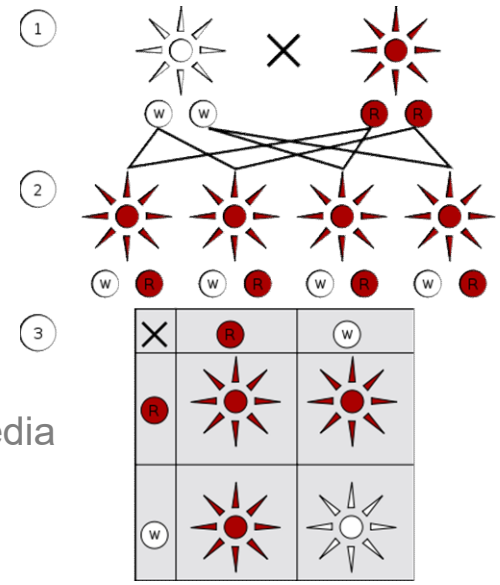
- How does the genotype affect the phenotype ?
- Mendelian genetics : allele → phenotype mapping is relatively simple
- Notion of dominance
  - Really a special case of a G2P function
  - Other variations : incomplete dominance, co-dominance, multi-locus (complex) trait

G2P models : epistasis

- Phenotype corresponding to a particular locus is modified by alleles / genotype at other loci
  - may or may not be additive : epistasis models
  - which is the primary locus ?

G2P models: pure dominance

- Two allele, two phenotype model
- Phenotype of one will mask the phenotype of the other : Mendelian model : dominant & recessive



Wikipedia

Mendelian model : parental, first filial (F1) and F2 generation

A → dominant

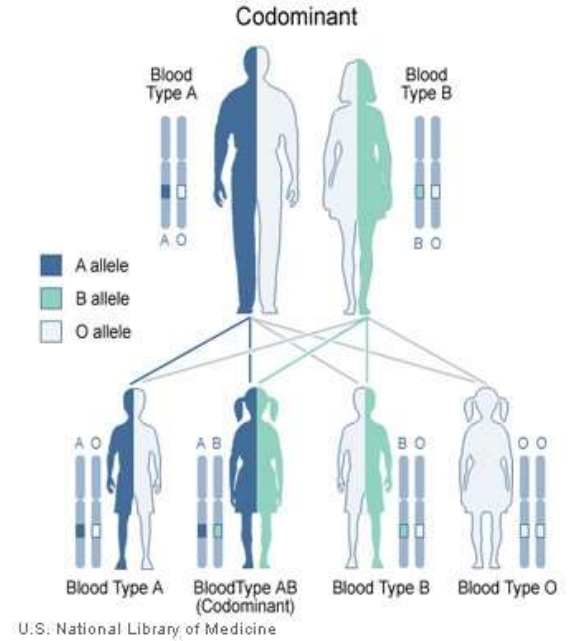
a → recessive

G2P models: pure dominance

- Before genotyping / haplotyping / genome sequencing
  - allele frequency of pure dominant / recessive bi allelic traits were frequently determined by analyzing the phenotype frequencies

G2P models : co-dominance

- Both alleles equally affect phenotype



Real valued traits

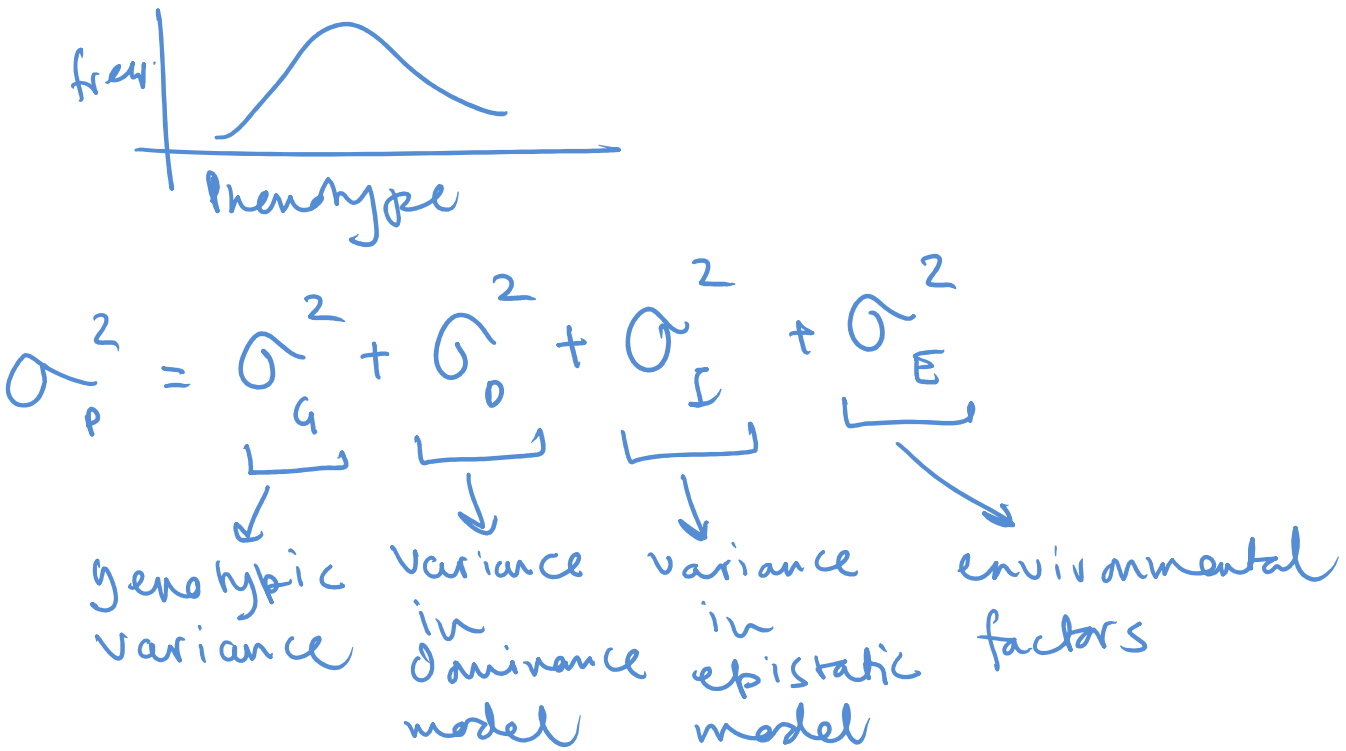
- So far, we have talked about binary or meristic (countable) traits

- How about continuous phenotypes
  - continuous valued phenotype (height, weight , …)
  - continuous liability (eg. disease susceptibility)

Dominance models for real valued traits

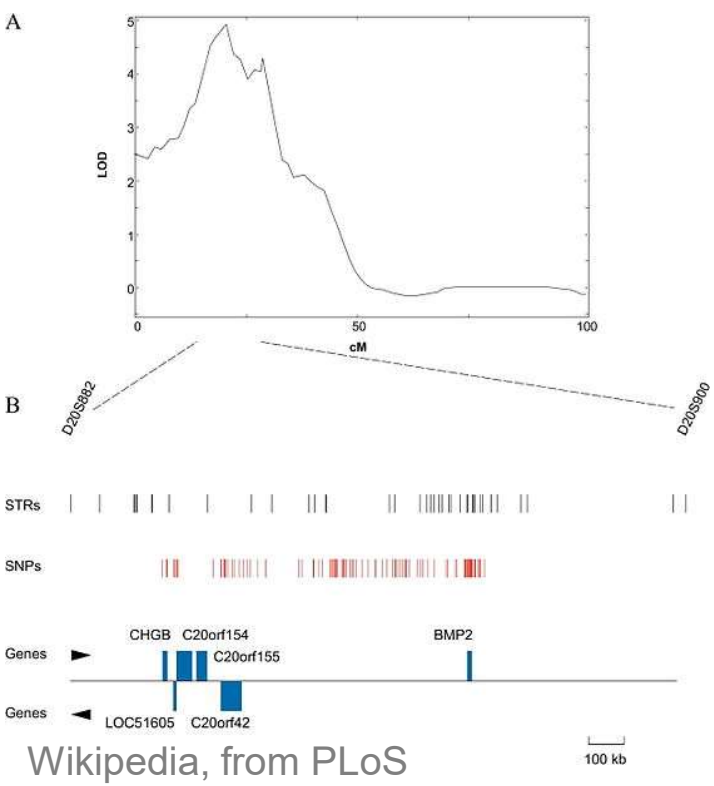
- Typically, neither co-dominant or pure dominant / recessive, somewhere in the middle
  - Weakly dominant (dominance models) : weighted additive effect models for simplicity

Explaining phenotypic variance



Predicting the phenotype: QTL

- Complex, (usually) real valued traits :
  Quantitative trait locus
  - additive model to explain trait based on the loci
- Multiple loci implicated



Wikipedia, from PLoS

Implicating loci : association studies

- How to find the implicated loci in 1st place ?
  - find loci which are jointly discriminative wrt phenotype (independently discriminative is a sp case)
- Notion is to pick loci that reduce entropy in P(phenotype | genotype)
- Discrimination is a necessary condition for implication, is it sufficient ?

Forces shaping allele frequency

Variation increasing changes

- Mutation: de novo random change across generation

Variation reducing changes

- Selection: directed change across generation
- Drift: random change across generation not explained by mutation & selection (finite sampling)

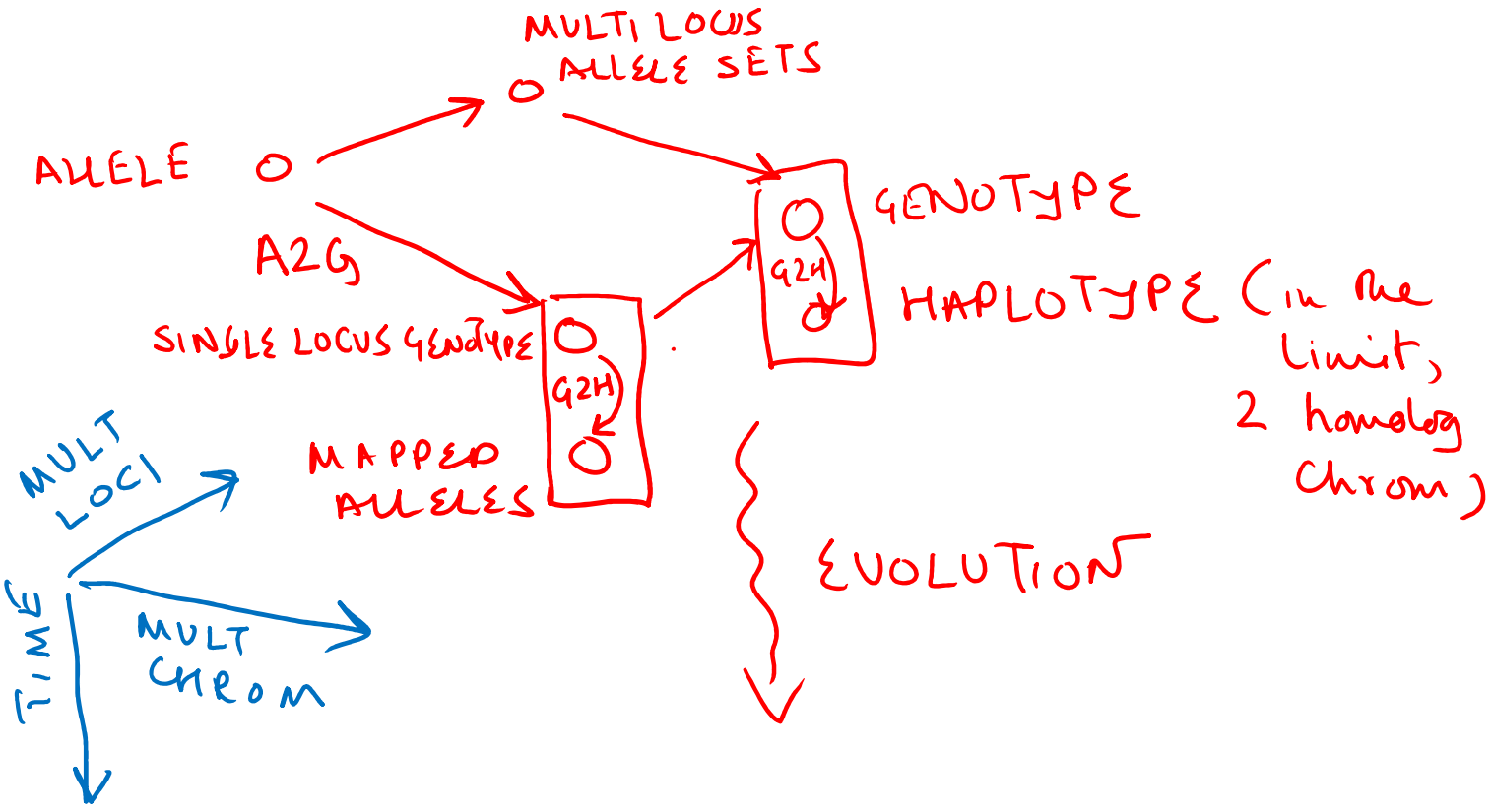
Variation increasing or reducing

- Genetic flow / migration: one time change of frequencies and population size
- Population: effect of finite (small/large) population size

HETEROZYGOTIC ADVANTAGE SELECTION MODELS INCREASE VARIATION

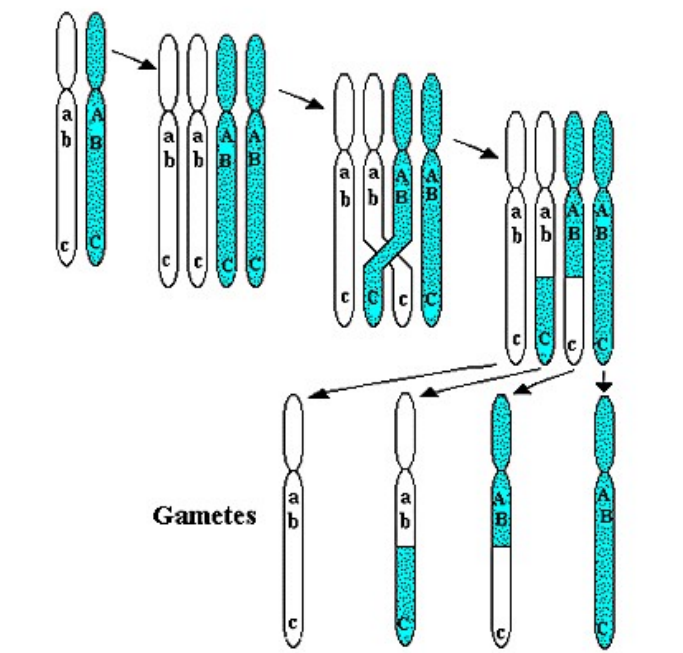
CAN INCR. VARIATION BETWEEN POPULATIONS, NOT WITHIN POPULATIONS

Modelling a population



Recombination

- When modelling multiple loci
  - recombination increases variation



Crossing-over and recombination during meiosis

A few other relevant concepts

- Linkage disequilibrium (LD)
  - How to calculate LD ?
  - What causes LD ?
  - Why is it useful ?

- Recombination rate
  - Estimating it from population genetic data