A *very* basic introduction

# THE WRIGHT-FISHER MODEL

# The paper and the idea

# The Population Genetics of dN/dS

**Sergey Kryazhimskiy[1], Joshua B. Plotkin[1,2]***

1 Biology Department, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, 2 Program in Applied Mathematics and Computational Science, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America

## Abstract

Evolutionary pressures on proteins are often quantified by the ratio of substitution rates at non-synonymous and synonymous sites. The dN/dS ratio was originally developed for application to distantly diverged sequences, the differences among which represent substitutions that have fixed along independent lineages. Nevertheless, the dN/dS measure is often applied to sequences sampled from a single population, the differences among which represent segregating polymorphisms. Here, we study the expected dN/dS ratio for samples drawn from a single population under selection, and we find that in this context, dN/dS is relatively insensitive to the selection coefficient. Moreover, the hallmark signature of positive selection over divergent lineages, dN/dS>1, is violated within a population. For population samples, the relationship between selection and dN/dS does not follow a monotonic function, and so it may be impossible to infer selection pressures from dN/dS. These results have significant implications for the interpretation of dN/dS measurements among population-genetic samples.
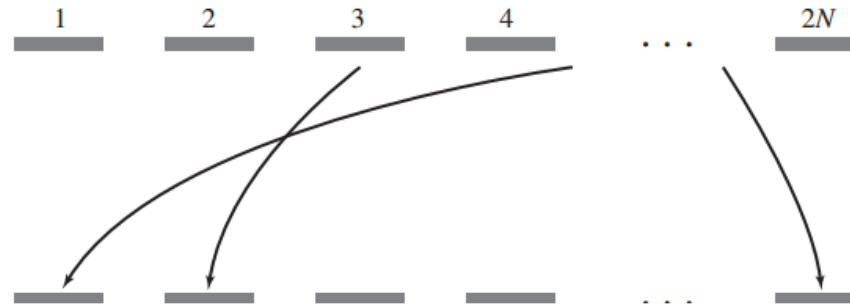
# Meet the model…

- describing the genealogical relationship among genes
- dynamic description of the evolution of an idealised population and the transmission of genes from one generation to the next.
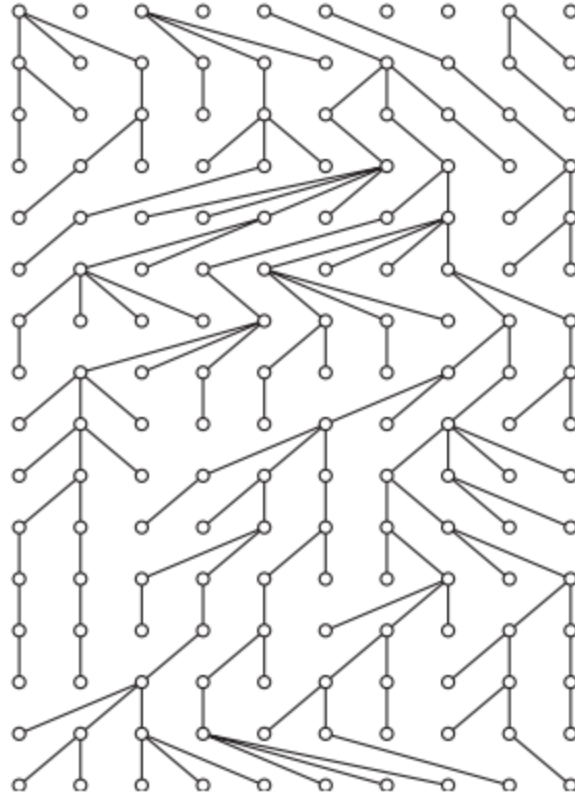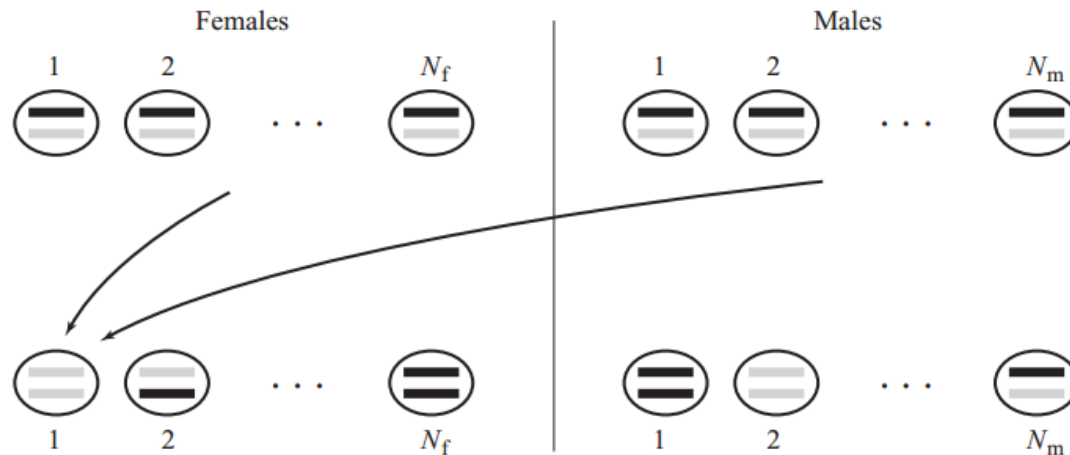
# Schematic overview

**Haploid**



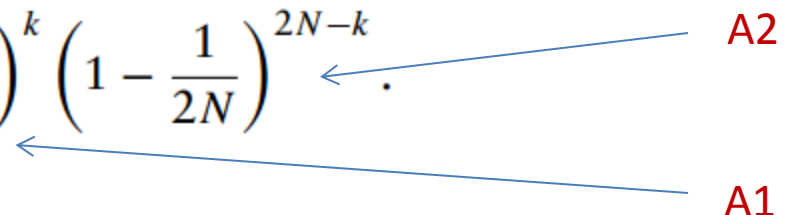**Diploid**

# Assumptions of the model

- Discrete and non-overlapping generations

- Haploid individuals or two subpopulations (males and females)

- The population size is constant

- All individuals are equally fit

- The population has no geographical or social structure

- The genes (or sequences) in the population are not recombining

→ **Modelling Genetic Drift! (but selection can be included)**

# How the model works

- We consider a diploid population of size N
- At a given locus, there could be either one of the two alleles A1 and A2
- The population reproduces in discrete time steps
- the alleles in generation n+1 (children) are chosen by <u>random sampling with replacement</u> from the ones in generation n (parents)

Let vi be the number of descendants of gene i in generation t, then

$$P(v_i = k) = \binom{2N}{k}\left(\frac{1}{2N}\right)^k\left(1 - \frac{1}{2N}\right)^{2N-k}.$$

A2

A1

This is an example of the binomial distribution, Bi(m,p), with parameters:

m=2N and p=1/(2N)

The number of genes descending from a given gene is binomially distributed