



La sélection naturelle en évolution moléculaire

histoire, statut et perspectives

Guillaume Achaz

Atelier de Bioinformatique, UMR 7205, Museum National d'Histoire Naturelle, Paris
SMILE, UMR 7241, Collège de France, Paris

Molecular Evolution

Molecular component of evolution
(mostly genetics)
to
assess the evolution **of** and **from** molecules

Micro-evolution (populations)

Polymorphisms (transient states)

Population genetics
Phylogeography

Macro-evolution (species)

Divergence (fixed differences)

Speciation process
Phylogeny

Molecular Evolution (*sensu lato*)

Historical Glimpse

Origin of Population Genetics (20's-30's)

The Modern Synthesis (40's-60's)

Neutral Theory (70's-80's)

Contemporary View

Tools

Mathematical models

deterministic (*i.e.* selection)

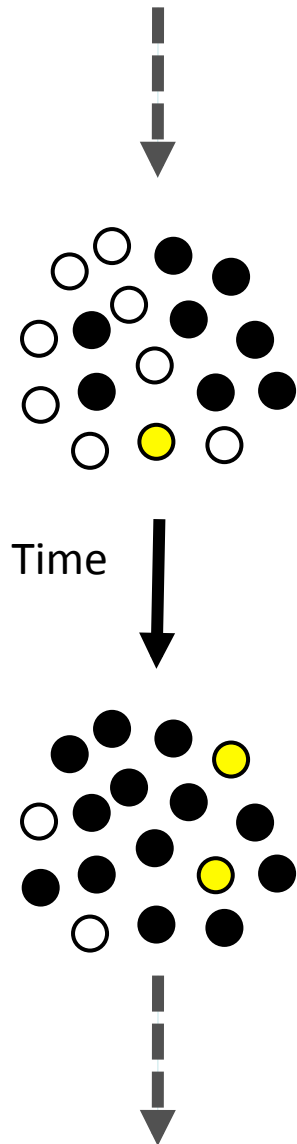
stochastic (*i.e.* genetic drift)

Data

Intra-specific homologous loci

polymorphism (>1 allele)

Population Genetics



Selection

s : selection coefficient
 ρ : rate of selection events

Mutation

μ : mutation rate

Genetic drift

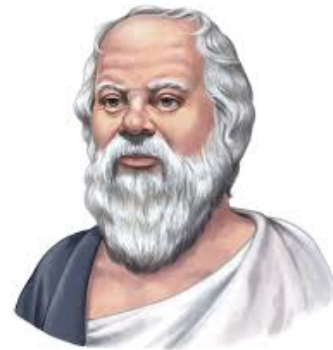
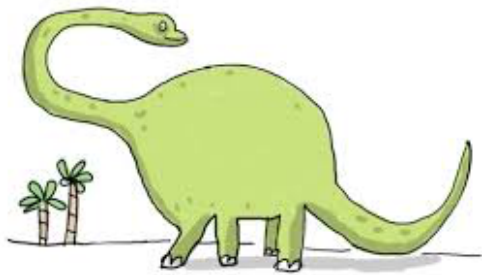
N : census population size
 N_e : effective population size

Misc

Population structure
Demography

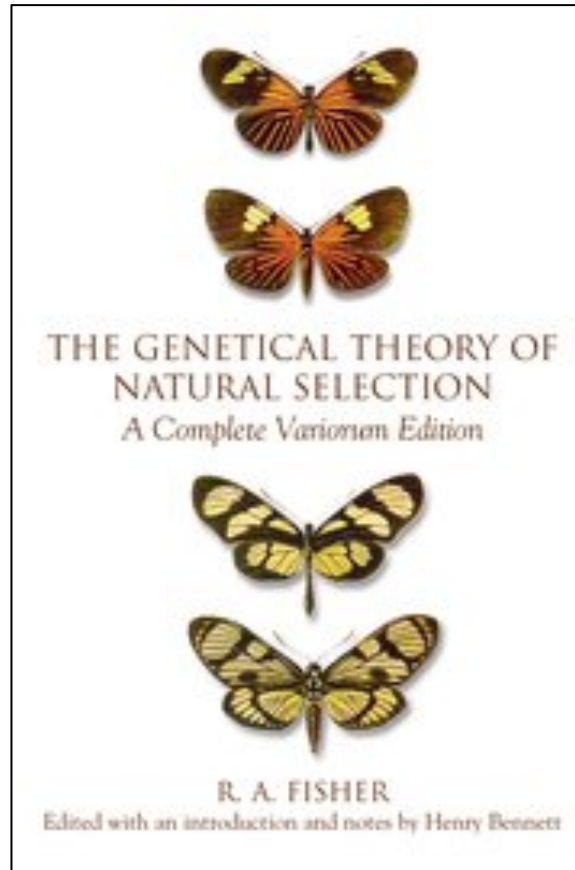
...

A quick historical perspective



Sir Ronald A. Fisher

(1890-1962)



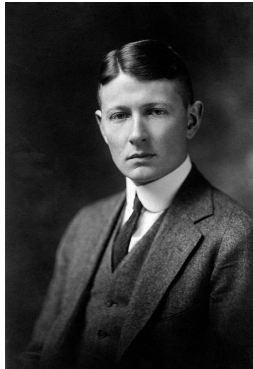
CONTENTS

List of Illustrations	xiii
I. The Nature of Inheritance	1
The consequences of the blending theory, as drawn by Darwin. Difficulties felt by Darwin. Particulate inheritance. Conservation of the variance. Theories of evolution worked by mutations. Is all inheritance particulate? Nature and frequency of observed mutations.	
(II) The Fundamental Theorem of Natural Selection	22
The life table and the table of reproduction. The Malthusian parameter of population increase. Reproductive value. The genetic element in variance. Natural Selection. The nature of adaptation. Deterioration of the environment. Changes in population. Summary.	
III. The Evolution of Dominance	48
The dominance of wild genes. Modification of the effects of Mendelian factors. Modifications of the heterozygote. Special applications of the theory. The process of modification. Inferences from the theory of the evolution of dominance. Summary.	
(IV) Variation as determined by Mutation and Selection	70
The measurement of gene frequency. The chance of survival of an individual gene; relation to Poisson series. Low mutation rates of beneficial mutations. Single origins not improbable. Distribution of gene ratios in factors contributing to the variance. Slight effects of random survival. The number of the factors contributing to the variance.	
(V) Variation &c. (continued)	97
The observed connexion between variability and abundance. Stable gene ratios. Equilibrium involving two factors. Simple metrical characters. Meristic characters. Biometrical effects of recent selection. Summary.	
VI. Sexual Reproduction and Sexual Selection	121
The contrast between sexual and asexual reproduction. The nature of species. Fission of species. Sexual preference. Sexual selection. Sex limitation of modifications. Natural Selection and the sex ratio. Summary.	
VII. Mimicry	146
The relation of mimicry theory to the parent theory of Natural Selection. Theories of Bates and Müller. Supposed statistical limitation of Müllerian theory. Observational basis of mimicry theory. The evolution of distastefulness. The theory of saltations. Stability of the gene-ratio. Summary.	

1930

Sewall Wright

(1889-1988)



EVOLUTION IN MENDELIAN POPULATIONS

SEWALL WRIGHT
University of Chicago, Chicago, Illinois

Received January 20, 1930

TABLE OF CONTENTS

	PAGE
Theories of evolution	97
Variation of gene frequency	100
Simple Mendelian equilibrium	100
Mutation pressure	100
Migration pressure	100
Selection pressure	101
Equilibrium under selection	102
Multiple allelomorphs	104
Random variation of gene frequency	106
Rate of decrease in heterozygosis	107
The population number	110
The distribution of gene frequencies and its immediate consequences	111
No mutation, migration or selection	111
Nonrecurrent mutation	116
Reversible recurrent mutation	121
Migration	126
Irreversible recurrent mutation	128
Selection	129
General formula	133
The distribution curves	134
Dominance ratio	137
The mean and variability of characters	139
The evolution of Mendelian systems	142
Classification of the factors of evolution	142
Lability as the condition for evolution	147
Control of evolution	151
Agreement with data of evolution	153
"Creative" and "emergent" evolution	154
SUMMARY	155
LITERATURE CITED	158

THEORIES OF EVOLUTION

One of the major incentives in the pioneer studies of heredity and variation which led to modern genetics was the hope of obtaining a deeper insight into the evolutionary process. Following the rediscovery of the Mendelian mechanism, there came a feeling that the solution of problems of evolution and of the control of the process, in animal and plant breeding

GENETICS 16: 97 Mr 1931

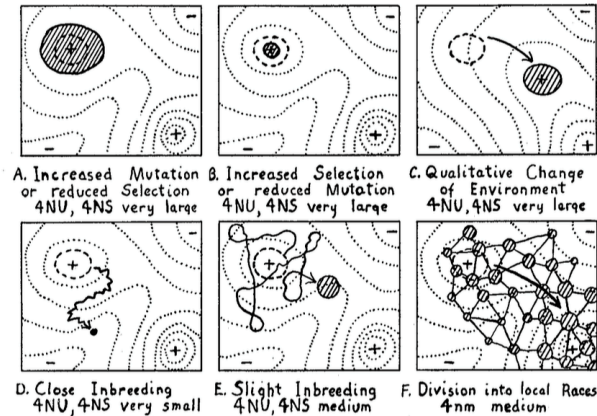
1930

356

PROCEEDINGS OF THE SIXTH

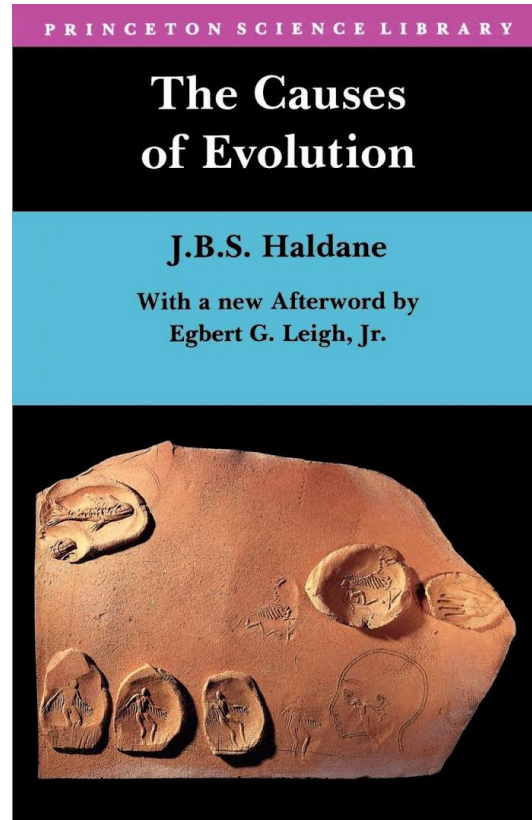
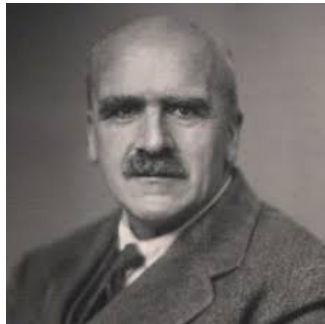
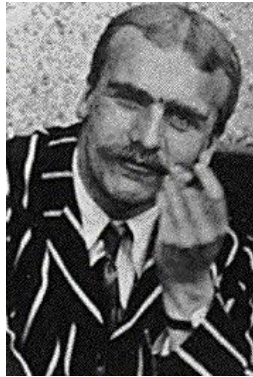
THE ROLES OF MUTATION, INBREEDING, CROSSBREEDING AND SELECTION IN EVOLUTION

Sewall Wright, University of Chicago, Chicago, Illinois



1932

John B.S. Haldane (1892-1964)

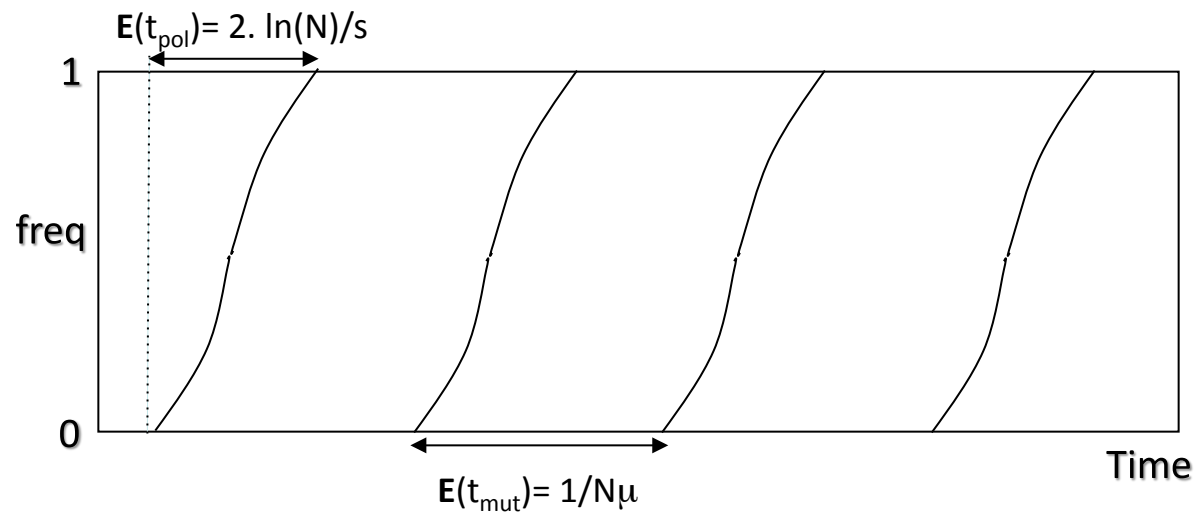


Contents

<i>Preface, by J. B. S. Haldane</i>	vii
<i>Preface to the 1990 Edition, by Egbert G. Leigh, Jr.</i>	viii
<i>Introduction by Egbert G. Leigh, Jr.</i>	ix
CHAPTER ONE Introduction	1
CHAPTER TWO Variation Within a Species	19
CHAPTER THREE The Genetical Analysis of Interspecific Differences	35
CHAPTER FOUR Natural Selection	46
CHAPTER FIVE What is Fitness?	61
CHAPTER SIX Conclusion	78
<i>Appendix</i>	92
<i>Bibliographical References</i>	126
<i>Afterword, by Egbert G. Leigh, Jr.</i>	130
<i>Index to the Text</i>	213

Adaptation

Macromolecules are constantly adapting to their environment



Polymorphisms result from a *selection-mutation* equilibrium

Various pattern of selection are described (positive, purifying, balancing, sexual, ...)

The modern synthesis (40's-60's)

Promoted by

J. Huxley

Evolution: The modern synthesis, 1942

E. Mayr

G.L. Stebbins

T. Dobzhansky

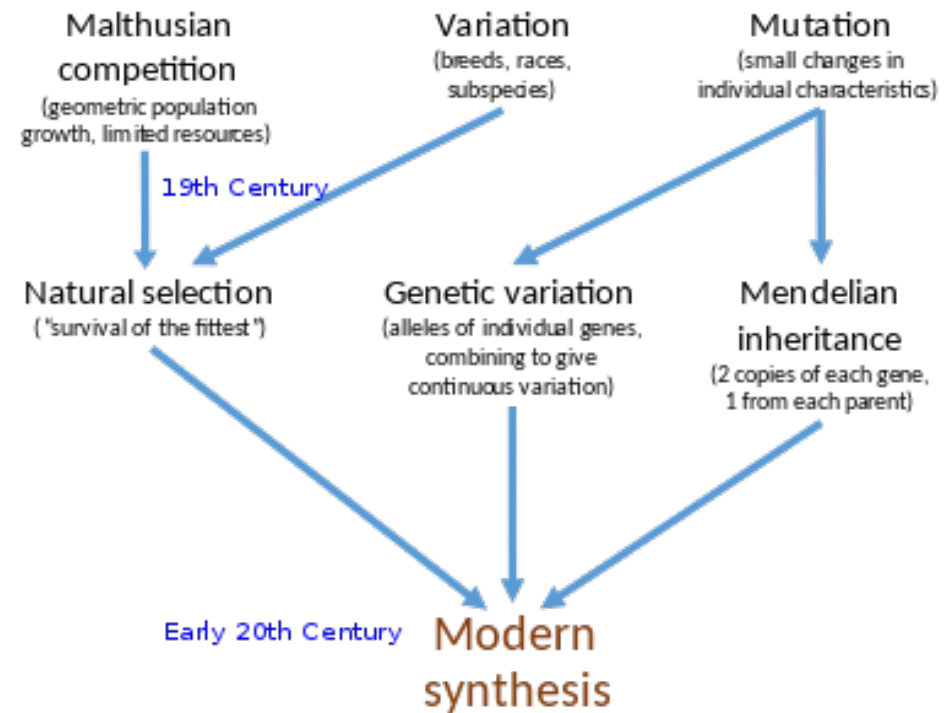
E.B. Ford

B. Rensch

I. Schmalhausen

G.G. Simpson

(among others)



From wikipedia « Modern Synthesis »

A paradigm shift

The “reference” model

1859 - ~1970: Evolution is driven by adaptation

1970 – today: Molecular evolution is driven by genetic drift

Kimura (1968) *Evolutionary Rate at the Molecular level*

Jukes and Kings (1969) *Non Darwinian Evolution*

An influential figure

M Kimura

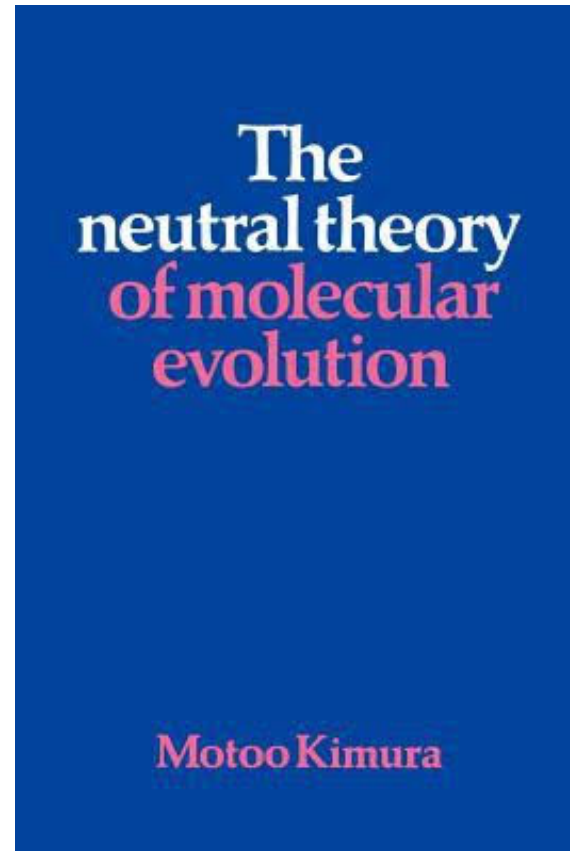
1950-1970 : major mathematical outbreaks

1970-1994 : causes of molecular evolution

The Neutral Theory of Molecular Evolution (1983)

Motoo Kimura

(1924-1994)

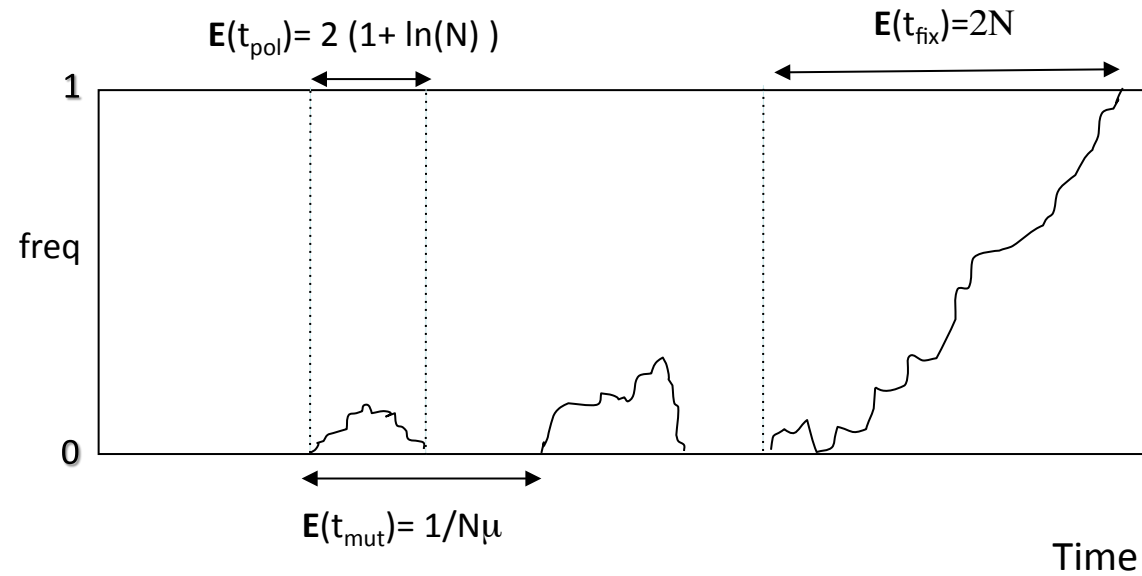


<i>Preface</i>	ix
<i>Introduction</i>	xi
1 From Lamarck to population genetics	1
2 Overdevelopment of the synthetic theory and the proposal of the neutral theory	15
2.1 Formation of the synthetic theory as the orthodox view	15
2.2 Proposal of the neutral mutation-random drift hypothesis	25
3 The neutral mutation-random drift hypothesis as an evolutionary paradigm	34
3.1 Chance acting on selectively equivalent alleles	34
3.2 Random genetic drift due to finite population size	36
3.3 Effective population size	40
3.4 Neutral and nearly neutral mutations	43
3.5 Population dynamics of mutant substitution	46
3.6 On some misunderstandings and criticisms	50
4 Molecular evolutionary rates contrasted with phenotypic evolutionary rates	55
4.1 Some features of phenotypic evolution	55
4.2 Rate of evolution at the molecular level	65
5 Some features of molecular evolution	98
6 Definition, types and action of natural selection	117
6.1 Meaning of natural selection	117
6.2 Phenotypic selection on quantitative characters	119
6.3 Genotypic selection	121
6.4 Other terminologies and concepts for natural selection	125
6.5 Genetic load	127
6.6 Fisher's model of adaptive process	135
6.7 Relation between phenotypic selection and genotypic selection	137

1983

The Neutral Theory

All mutations are neutral



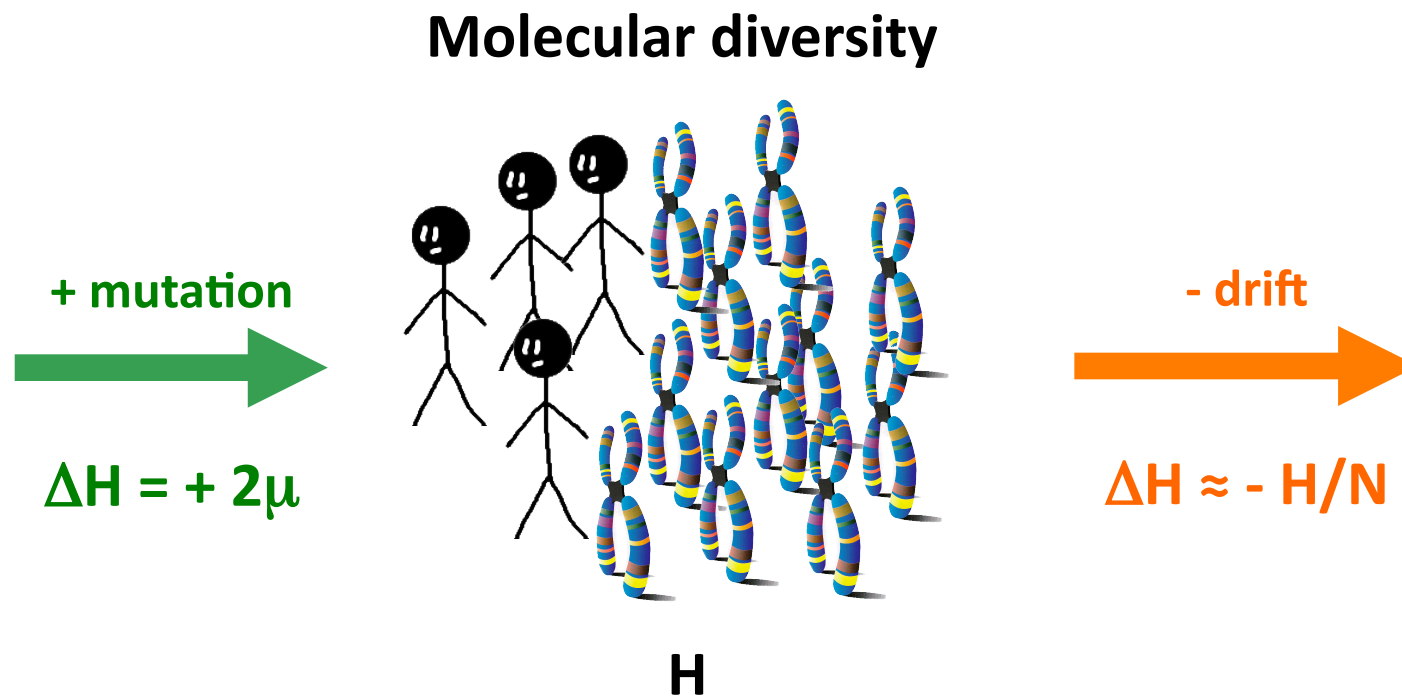
Polymorphisms result from a *mutation-drift* equilibrium

More polymorphisms are expected under neutrality, for a given mutation rate

Deeper into the so-called **Standard Neutral Model**

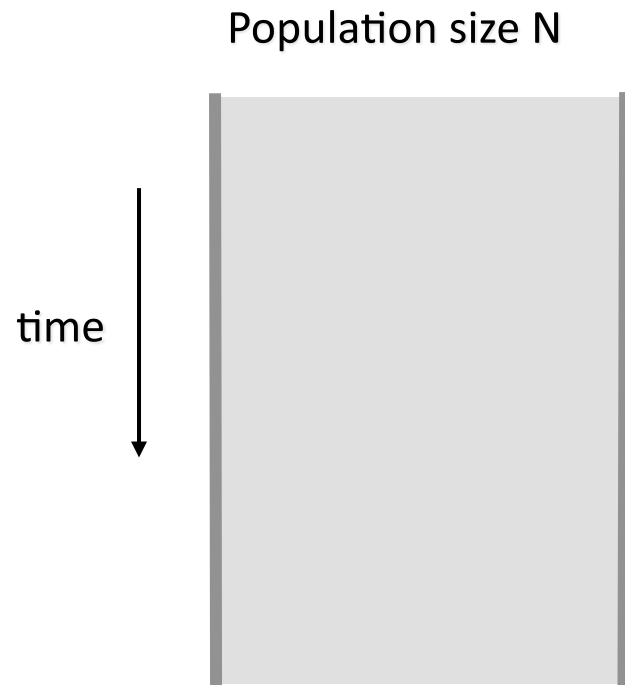


The mutation-drift paradigm (H0)



At equilibrium, $\Delta H = 0 \Rightarrow H^* = 2N\mu$

Standard Neutral Models



Assumptions

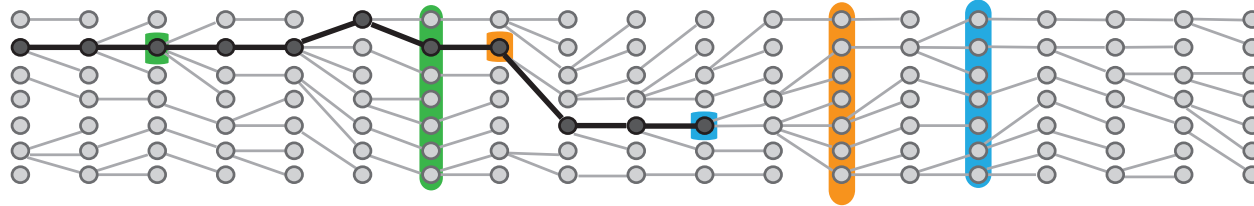
No selection
Constant Population Size
Strict Panmixia

Consequences

X: the number of descendants
is distributed with

$E[X] = 1$, for all N individuals

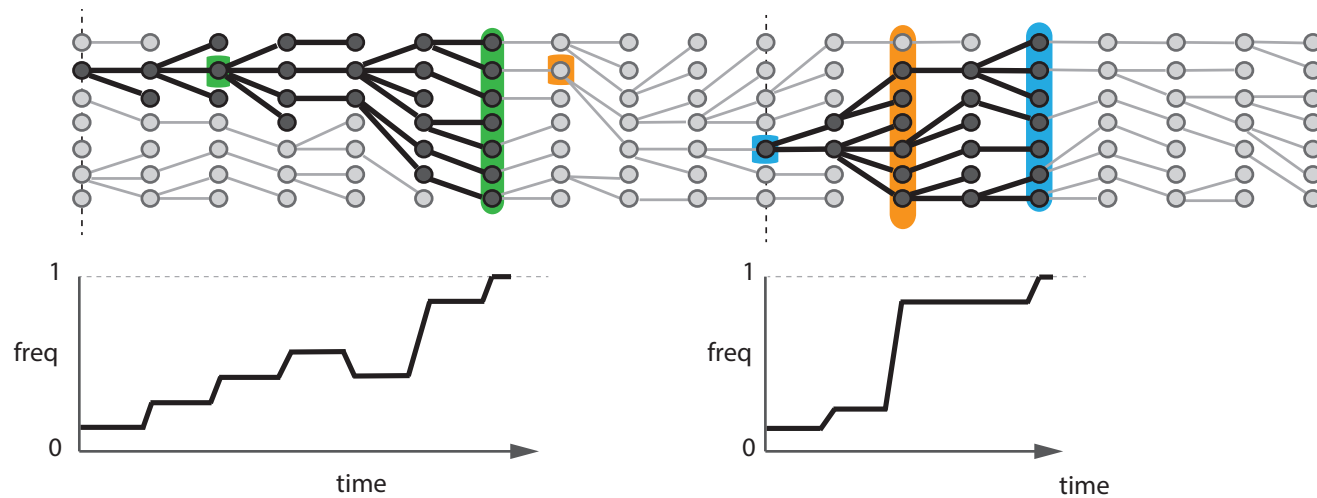
The Wright-Fisher model



1 generation = all individuals die and are replaced by a random sample

The fixation process

Forward time



From a random time

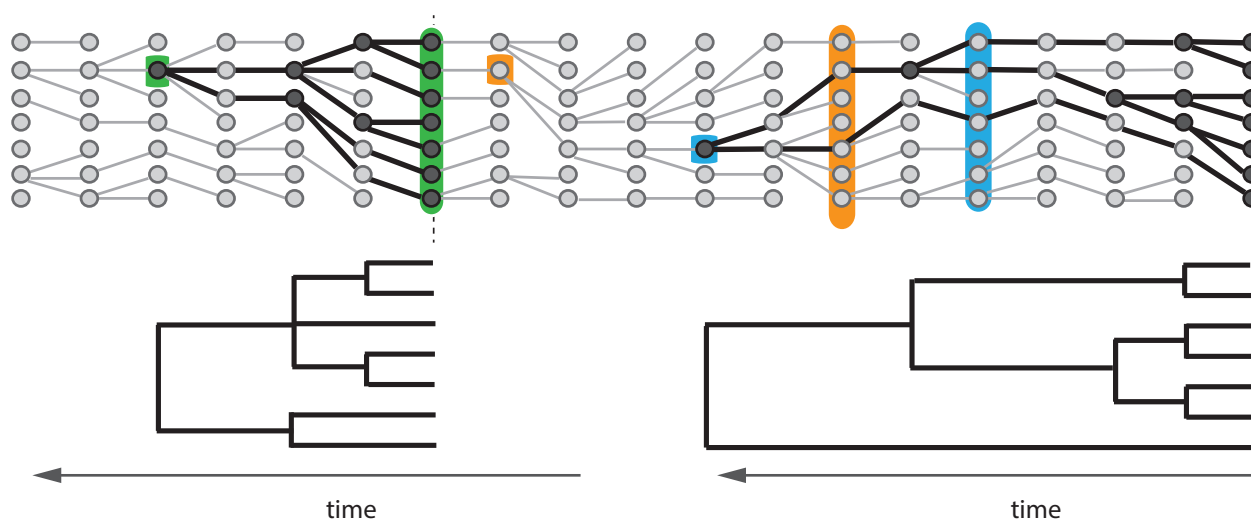
$$E[t_{\text{fix}}] = 2N$$

From the MRCA

$$E[t_{\text{fix}}] = N$$

The coalescent process

Backward time



From the fixation time

$$E[t_{\text{MRCA}}] = N$$

From a random time

$$E[t_{\text{MRCA}}] = 2N$$

The current paradigm

H0

=

standard neutral model

=

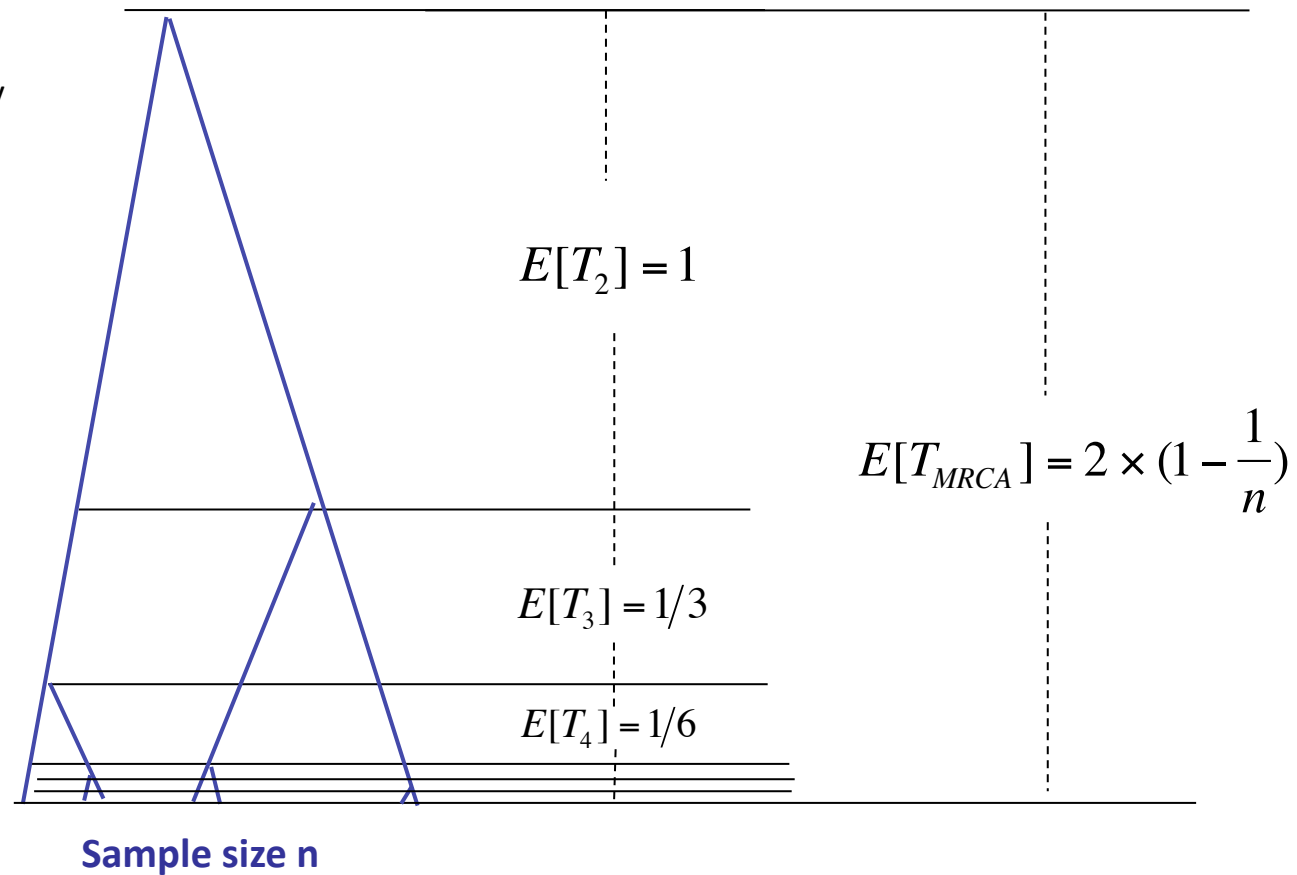
Kingman coalescent

Kingman coalescent trees

T_i = time while there are exactly i lineages.

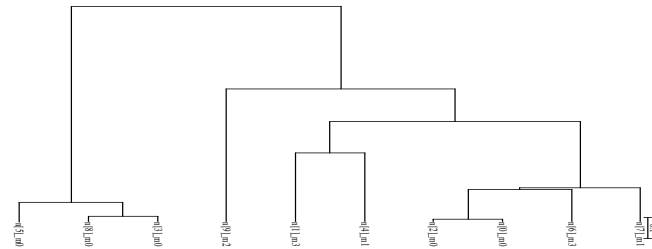
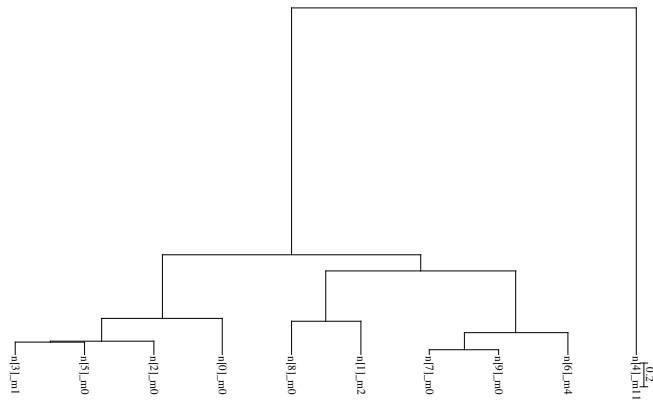
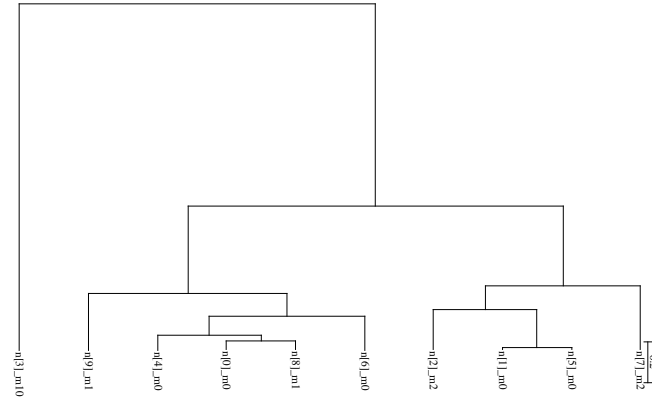
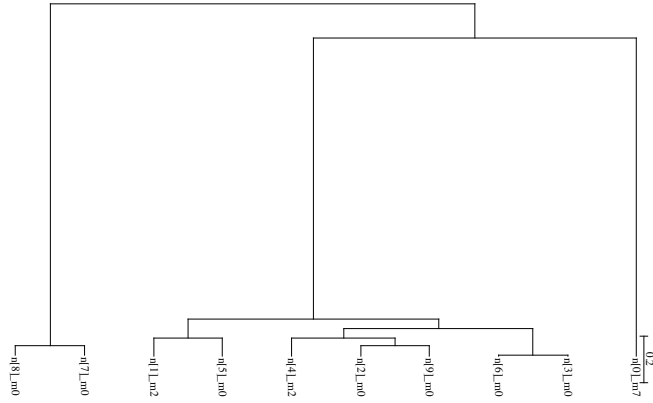
$$f(T_i) = \lambda_i e^{-\lambda_i T}$$

$$\lambda_i = \frac{i(i-1)}{2}$$

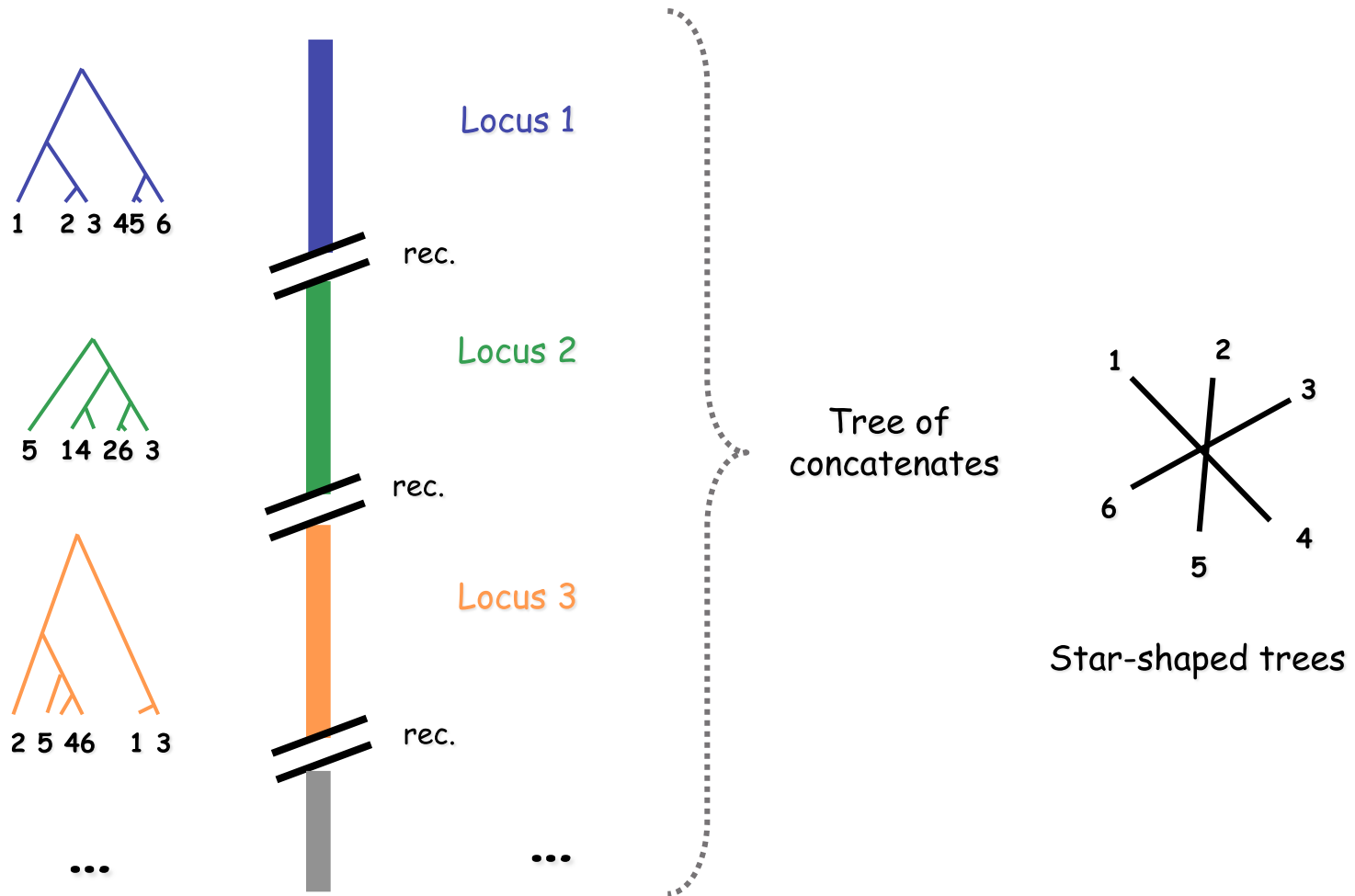


Time is counted in N generations ; $N \rightarrow \infty$

Kingman trees



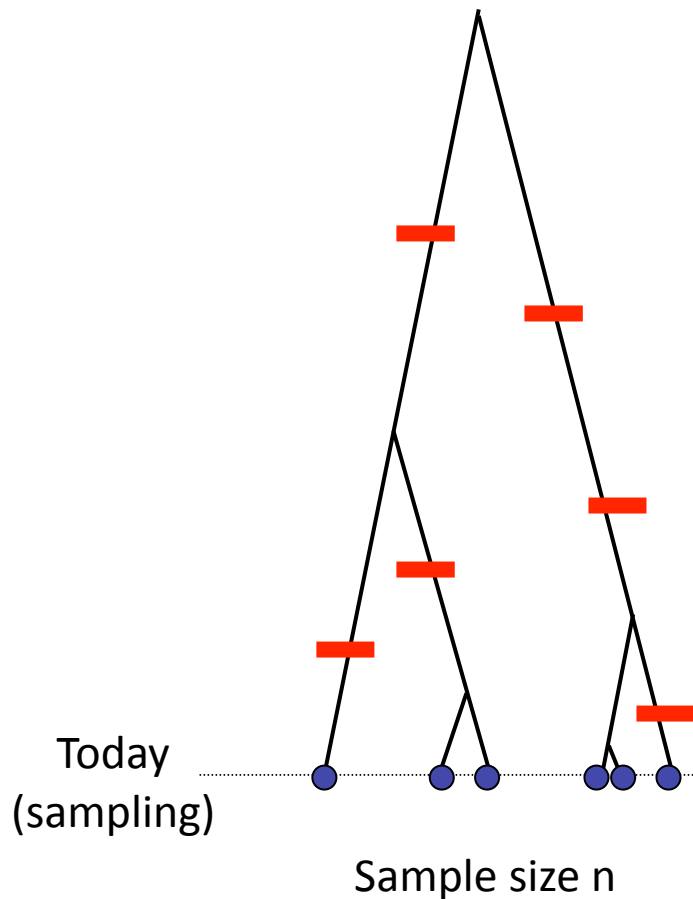
With recombination...



No single tree can be inferred but...

Genome-wide = expected diversity

Mutations are Poisson distributed on the branch lengths ; Tree space is averaged



Mutations

S : total # of mutations:

$$E[S] = 2 N \mu \times (\sum 1/i) \quad (\text{Waterson, 1975})$$

Other measures of diversity:

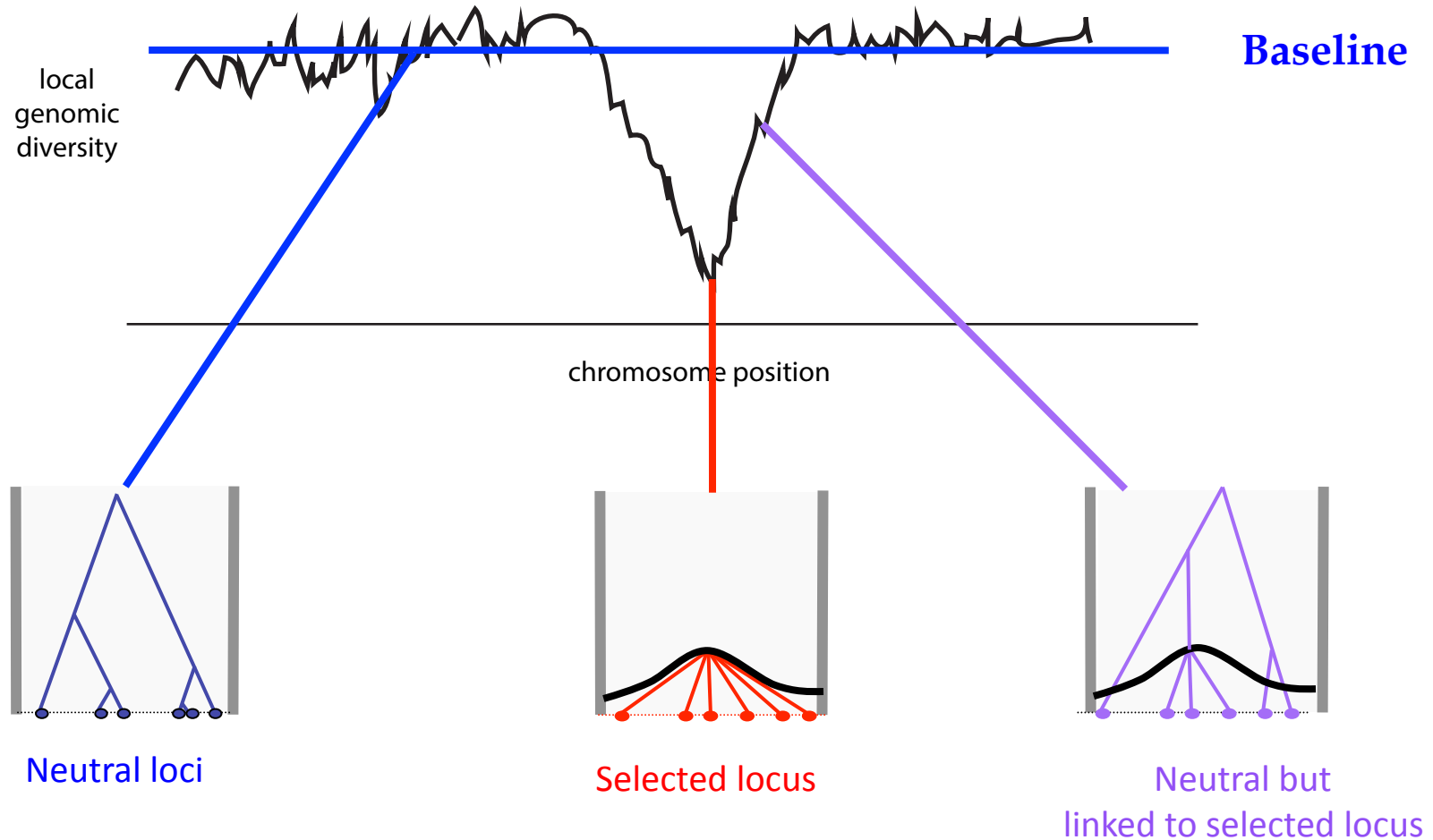
$$E[\pi] = 2 N \mu \quad (\text{Tajima, 1983})$$

$$E[\xi_1] = 2 N \mu \quad (\text{Fu and Li, 1993})$$

$$E[\xi_i] = 2 N \mu / i \quad (\text{Fu, 1995})$$

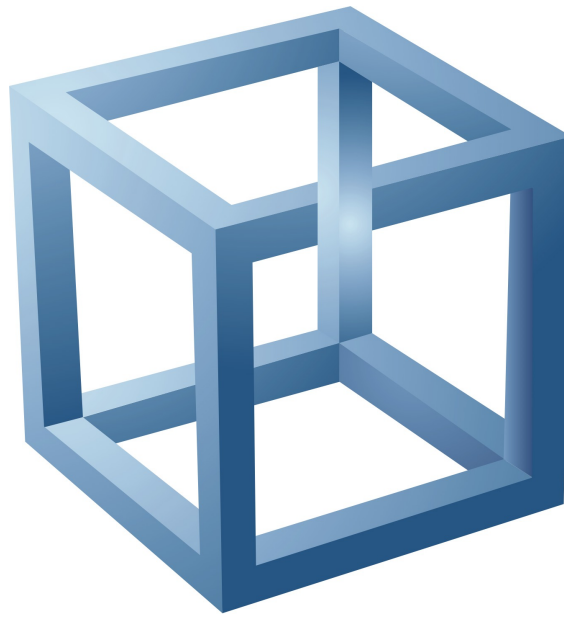
...

Mostly neutral with exceptions



Genome-wide hunts for selective sweep (Lactase, Immune system, etc.)

The case of
The Ne (des)illusion



Let's Pause and Ponder



Can we evaluate the Neutral model?

Within species nucleotide diversity

Effective population size

Mutation frequencies

Distribution of mutation frequencies

At the locus scale

Large variance

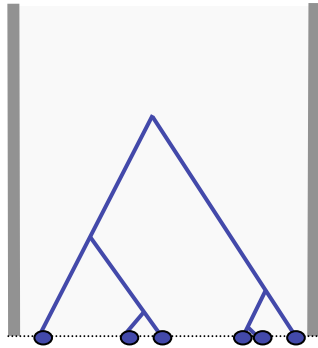
Hard to reject H_0

At the genome scale

Recombination = average many loci
current approach

From model to real populations

Population size N



Model Population

No selection
Constant Population Size
Strict Panmixia

$$E[t_2] = N$$

$$E[\pi] = 2 N \mu$$

(in a Wright-Fisher model)

“Effective” population size N_e



Realistic Population

Selection
Variable Population Size
Demography

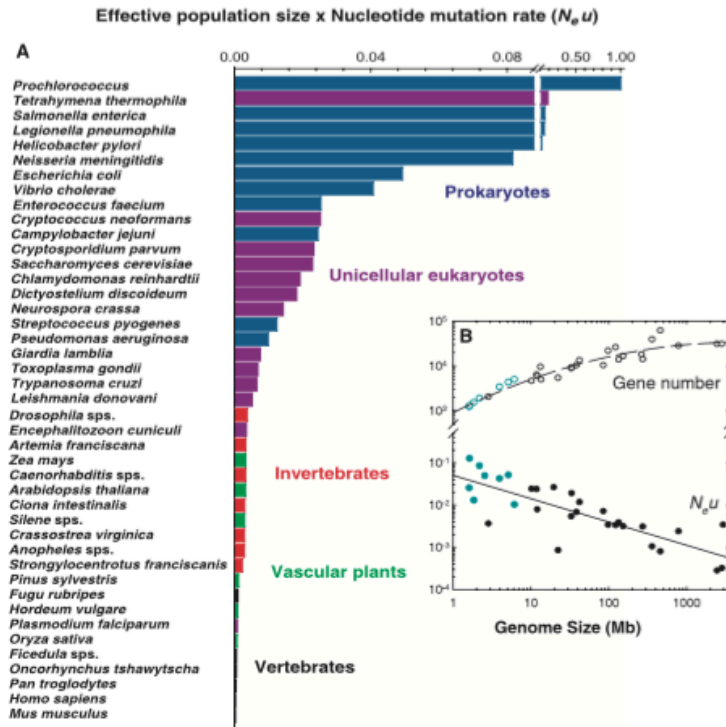
$$\bar{t}_2 = N_e$$

$$\hat{N}_e = \pi / 2 \mu$$

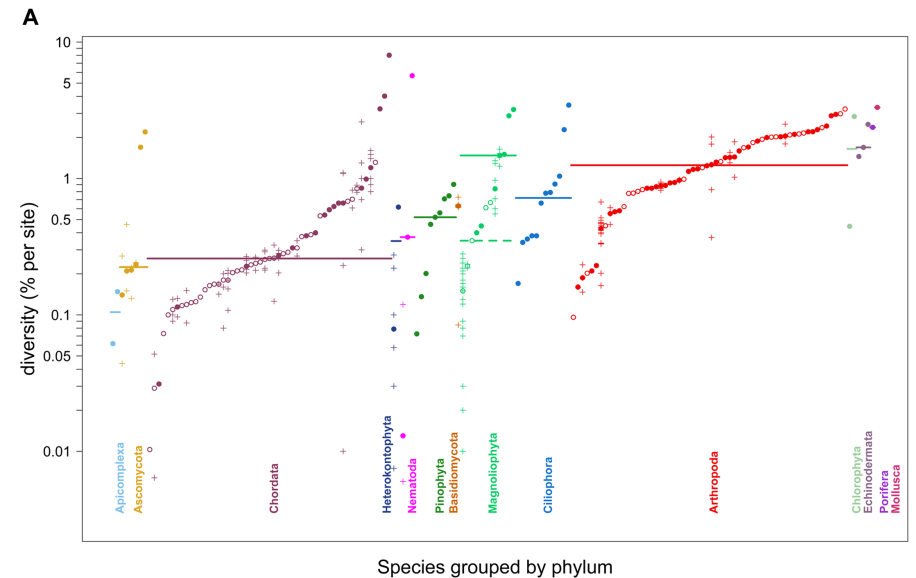
(π : pairwise differences ; μ mutation rate)

Assessing species diversity

(Lewontin & Hubby, 1966)



Lynch and Connery, 2003



Lefler et al., 2012

Why diversity does not scale linearly with N? (Lewontin's variation paradox, 1974)

Examples of N_e vs N

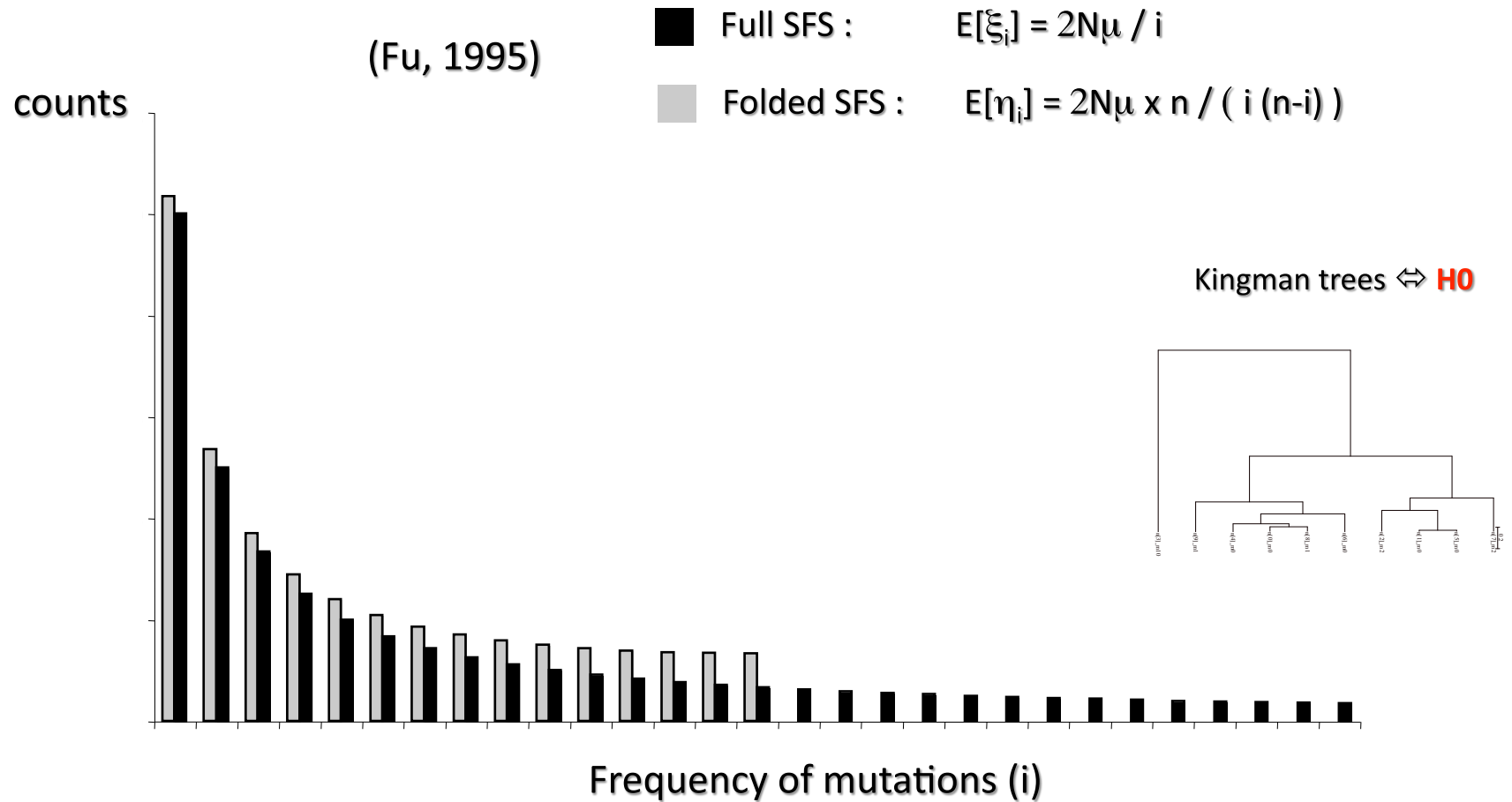
Species	N (census size)	N_e
<i>H. sapiens</i>	$7 \cdot 10^9$	10^4
<i>G. gorilla</i>	10^5	10^3
<i>D. melanogaster</i>	?	10^6
<i>C. elegans</i>	?	10^5
<i>A. thaliana</i>	?	10^5
<i>P. kergelensis</i>	?	10
<i>F. Psychrophilum</i>	10^9 /ml of cult.	10^6
<i>E. coli</i>	10^9 /ml of cult.	10^8
HIV (within patient)	10^{10}	10^3

Why is N_e unrelated to current census size? Demography?

The case of the Yoruba demography




Expected Site Frequency Spectrum (H0)



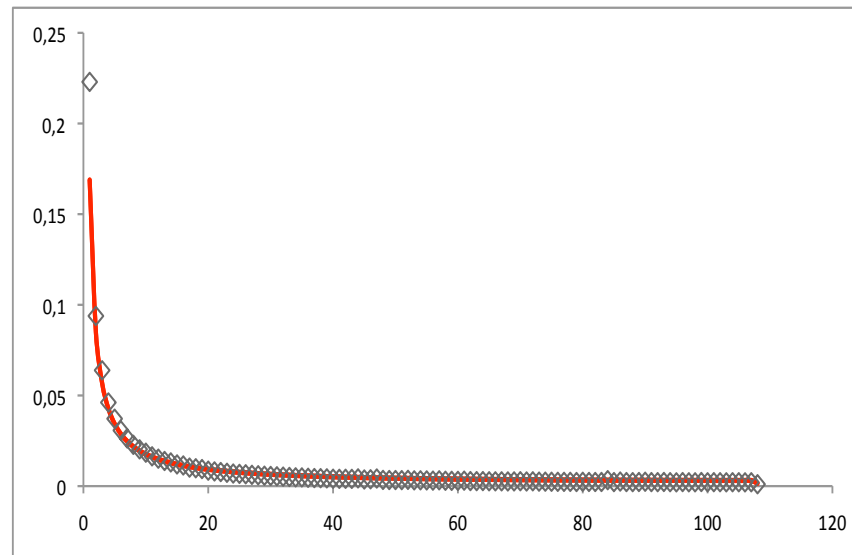
Under H0, full SFS is proportional to 1/i

Visual test for H0

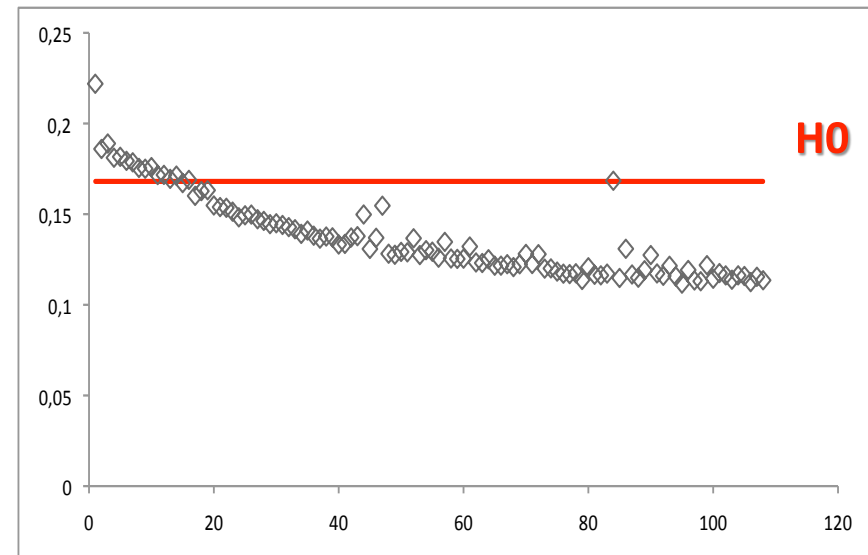
(Nawa & Tajima 2008, Achaz 2009, Lapierre et al. 2017)

(108 Yoruba individuals  from the 1,000 genomes)
(~81 10⁶ SNPs)

Folded SFS, normalized

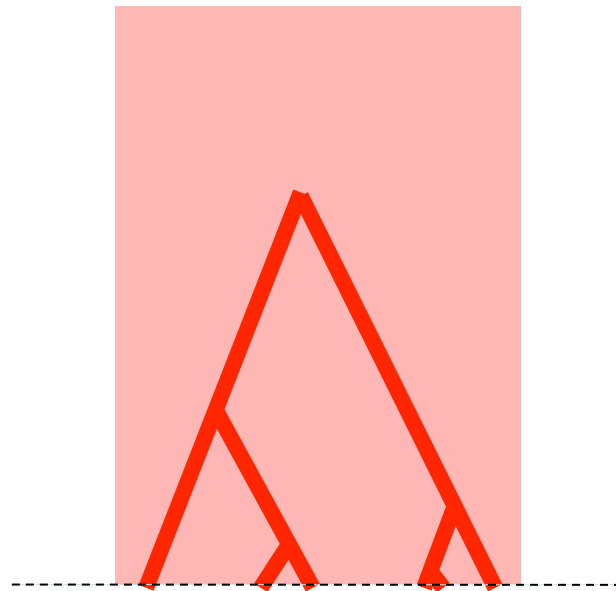


... and transformed

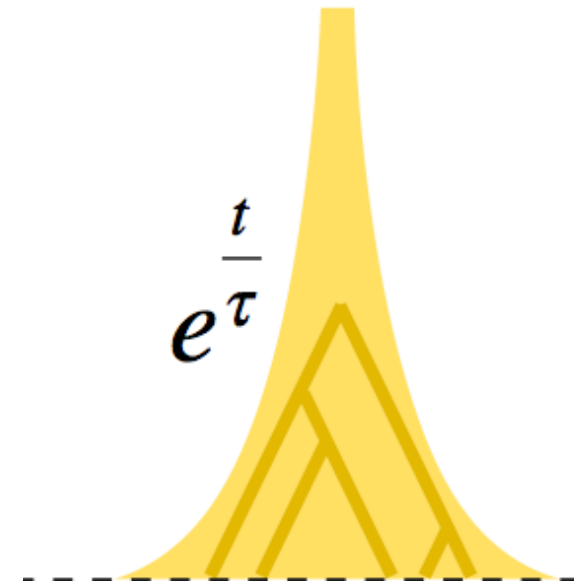


Departure from H0: an excess of *low frequency alleles*

SFS with demography



Constant Size

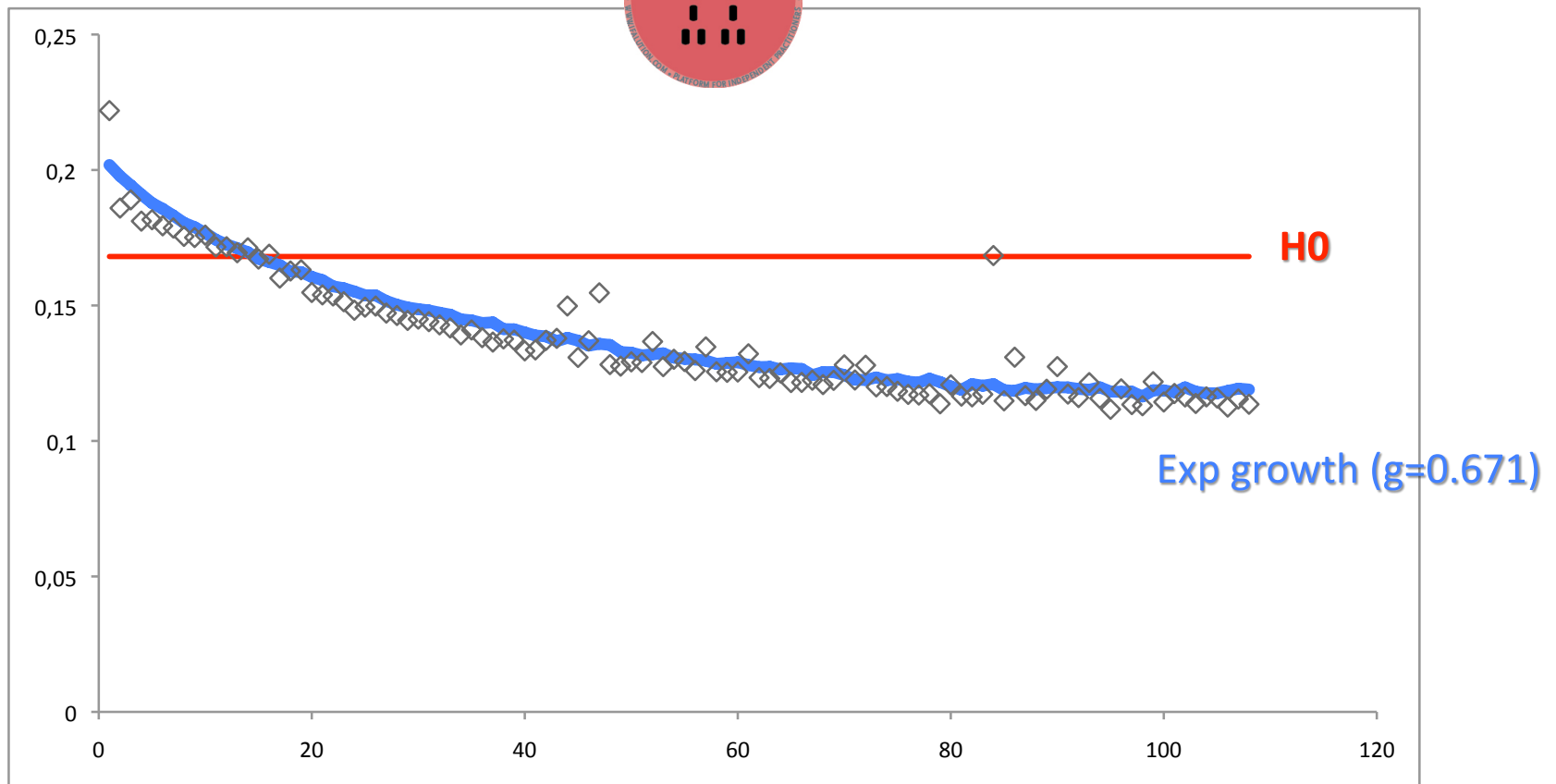


Exponential

SFS with demography (e.g. exp growth or any scenario)

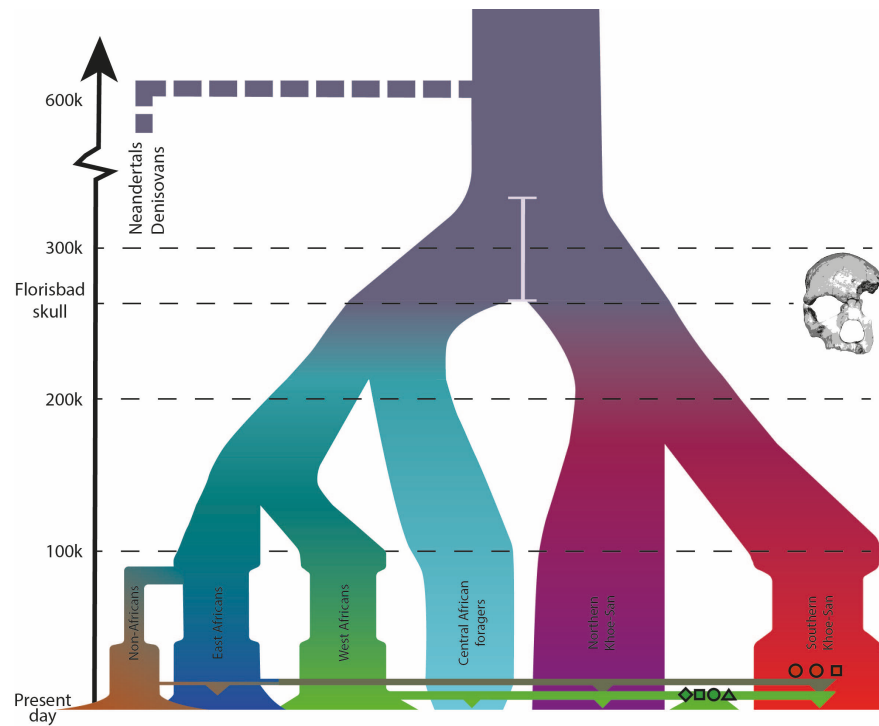
Demography with SFS

(e.g. Nielsen 2000, Gutenkunst et al. 2009, Liu and Fu 2015, Lapierre et al. 2017)

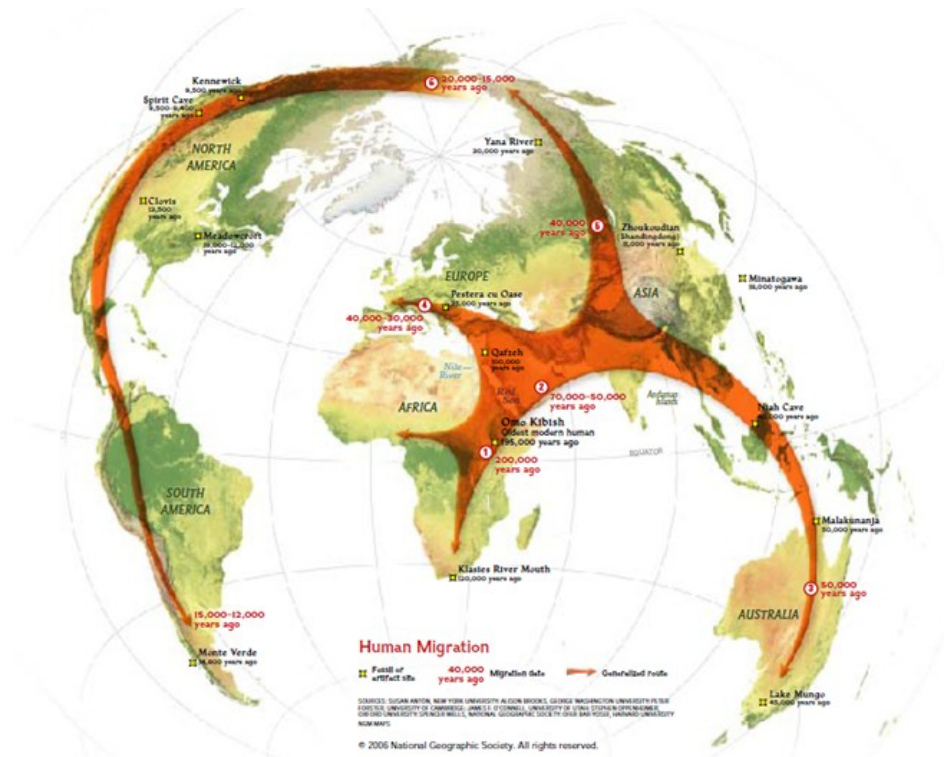


Adding demography greatly improves the fit

Human demography and migrations



Schlebusch et al., 2017



National Geographic, 2006

One of the favorite “game” of human population geneticists...

Do demography explains diversity?

Nice fit to data

Demographical inferences work approximately well provided the « correct N_e » is used.

Several scenarios are indistinguishable (Lapierre et al., 2017)

Structure is completely neglected (Mazet et al. 2016)

What is N_e , when accounting for demography?

$N_e(0)$, the "*current effective population size*" would be...

33,000 ???

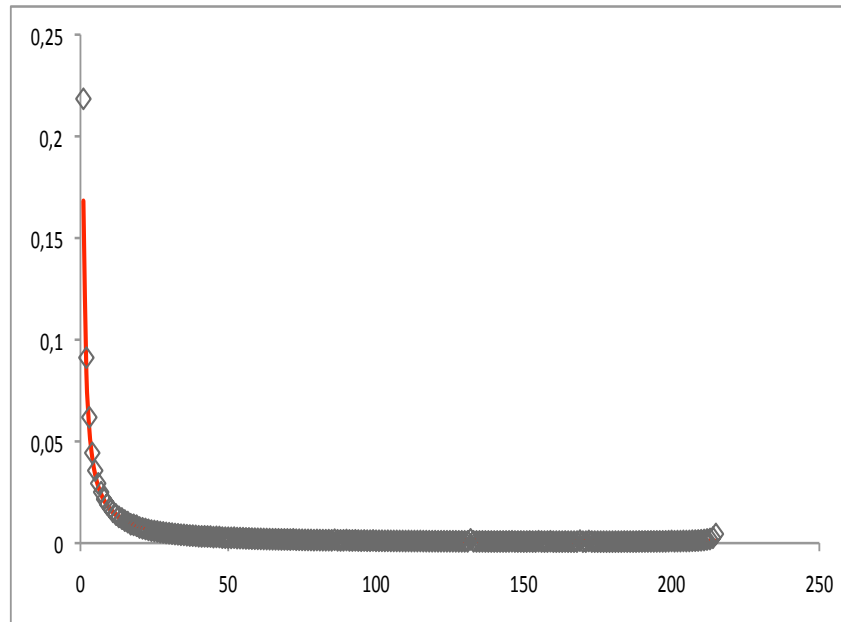
The case of the **Global species diversities**

(ongoing work with F Freund, S Matuszewski, J Jensen, A Lambert)

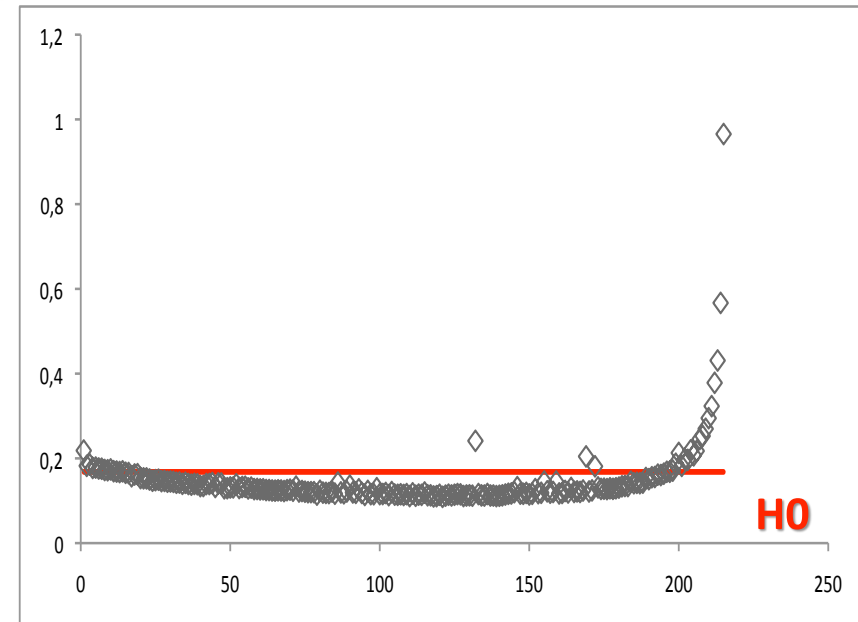


Then came the U spectrum

Unfolded SFS normalized

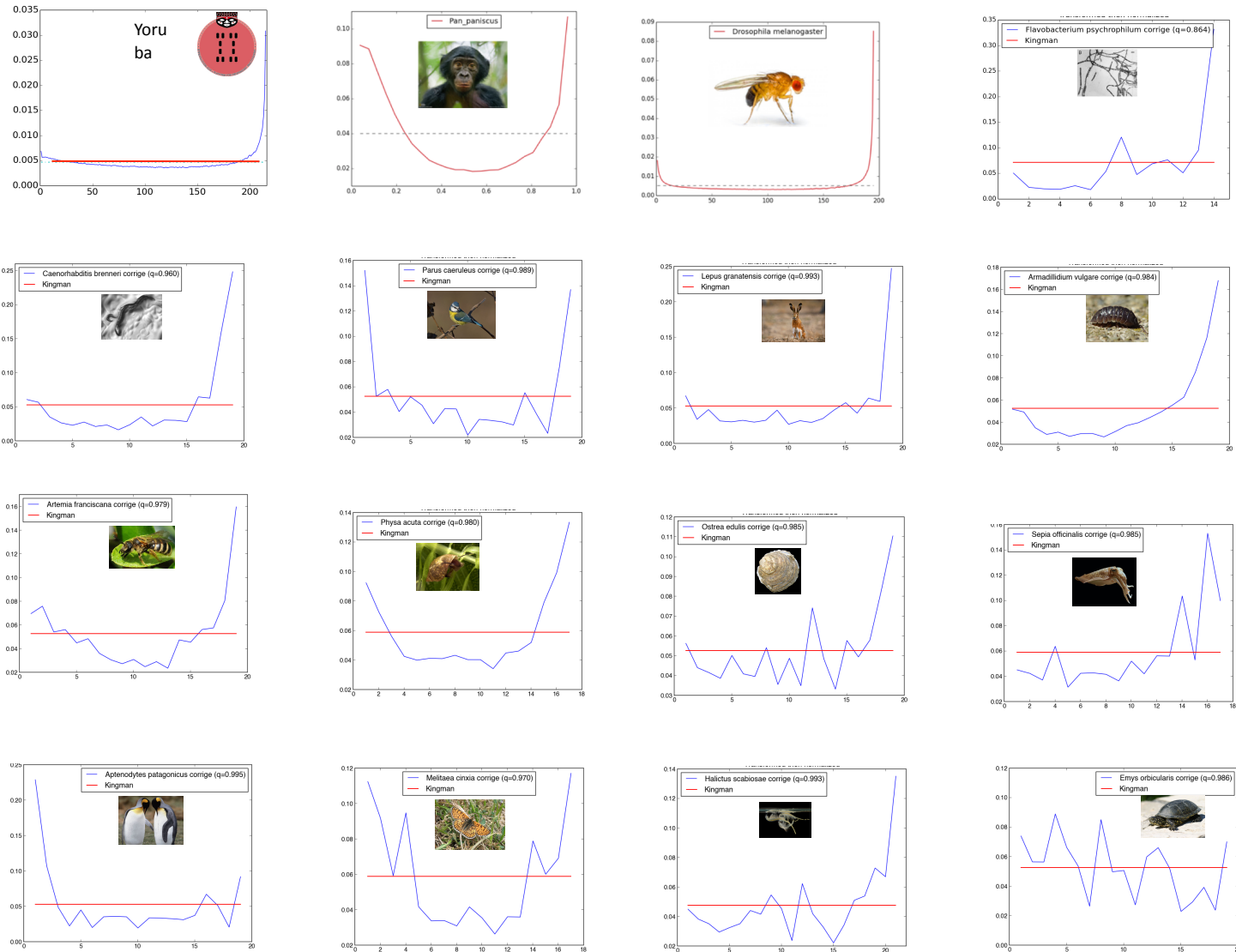


... and transformed



Departure from H0: an excess of *low* & *high* frequency alleles

... in all species

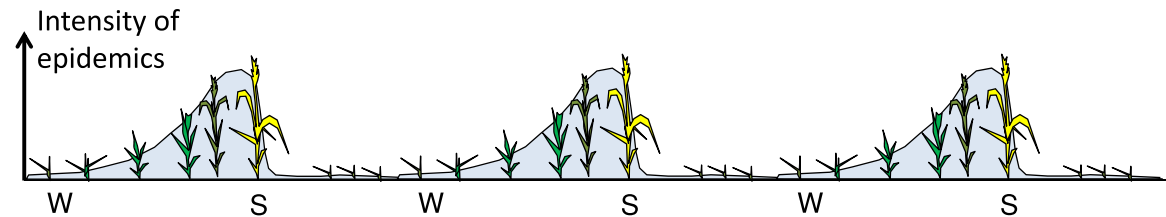


No simple demography can account for the U-shape!

Multiple Merger => U-shaped spectra

few individuals have many offsprings

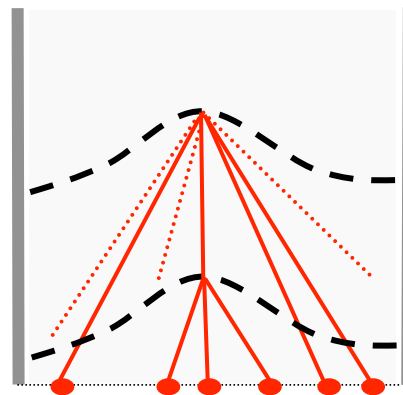
Neutral variance
in offspring numbers



(from Tellier and Lemaire, 2014)

Very large population size
(genetic draft)

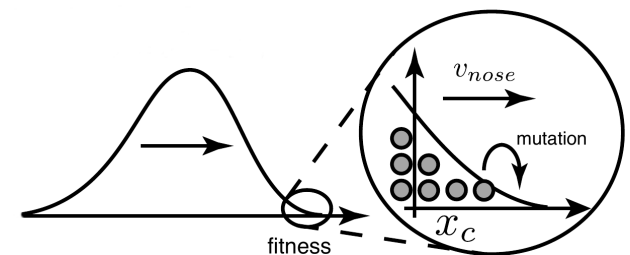
Selection + Recombination



Recurrent partial sweep

(Gillespie 2000a, 2000b, ...)

Never-ending adaptation

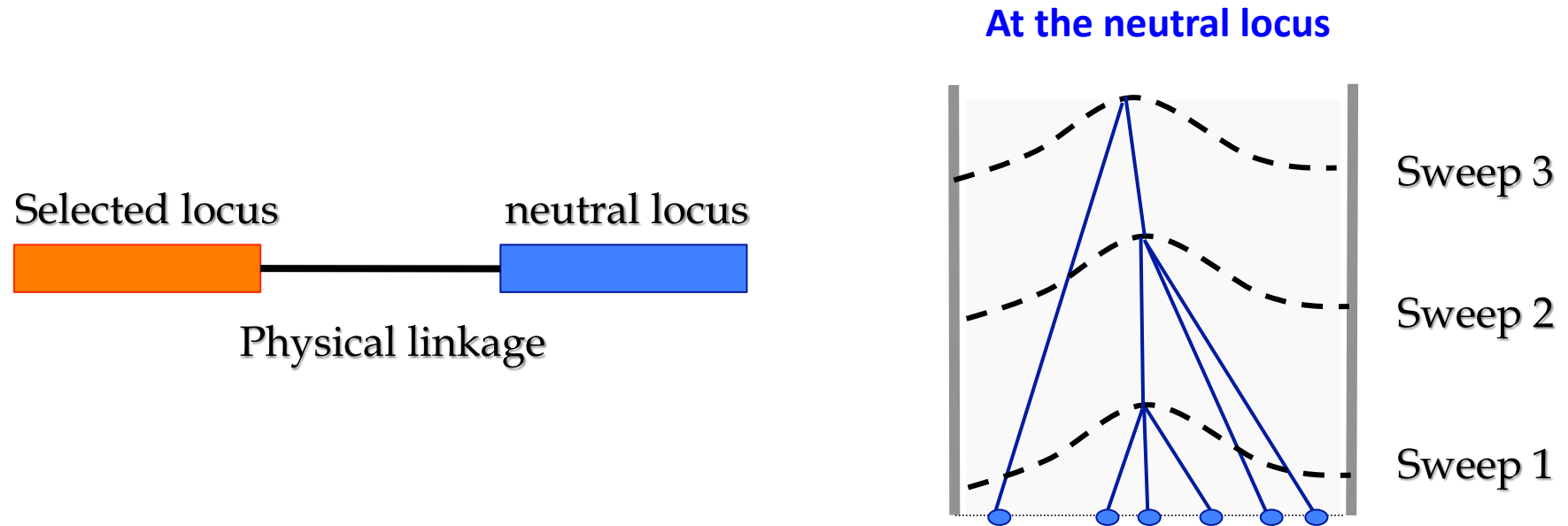


(reviewed in Neher 2013)

The genetic draft

(Recurrent selective sweeps in very large populations)

(Maynard Smith and Haigh 1974, ..., Gillespie 2000a, 2000b, ...)



Selection would be the cause of (low) genetic diversity

From data to models, and vice-versa

Observations

Sequences do change
Homologous loci show diversity

The (unknown) Cause of Molecular Evolution

Neutral theory (standard neutral models, H_0)
Adaptation theory (multiple mergers coalescent)
Demography
Population structure

Ultimately, assess the Biological Relevance of models

Kingman (H_0)	-- small stable populations--
Multiple mergers,	-- large populations—
...	

The relative role of selection and drift needs careful rethink

