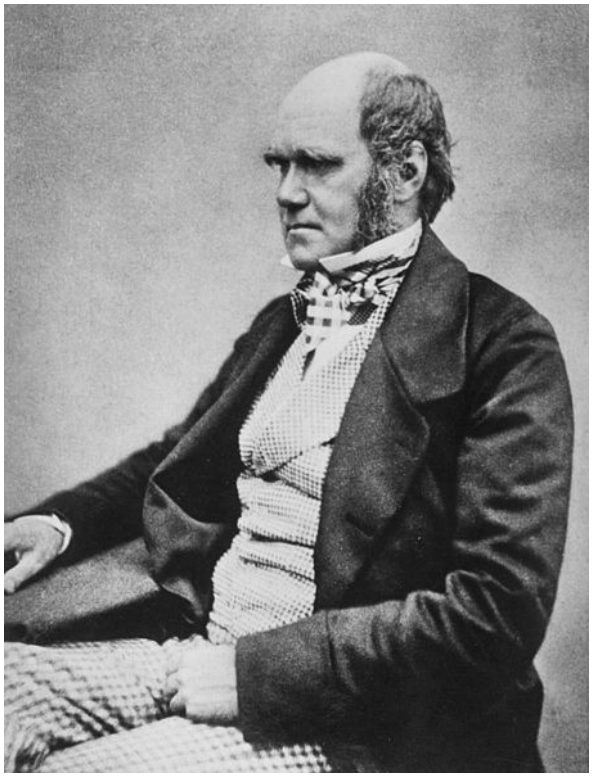
The Mathematics of Evolution

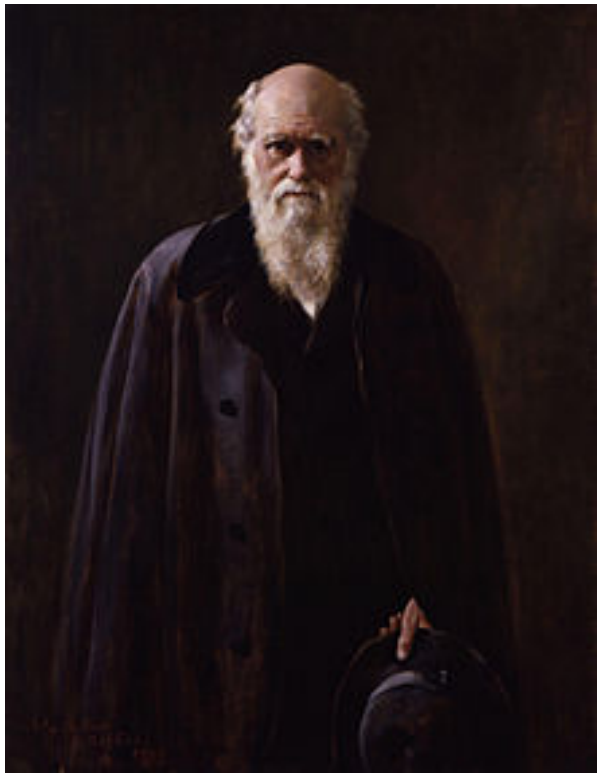
Robert Griffiths

University of Oxford

CRM Clay Senior Scholar

Charles Darwin 1809-1882





Origin of the Species

Darwin's Finches

Darwin: Natural selection

Seeing this gradation and diversity of structure in one small, intimately related group of birds, one might really fancy that from an original paucity of birds in this archipelago, one species had been taken and modified for different ends.

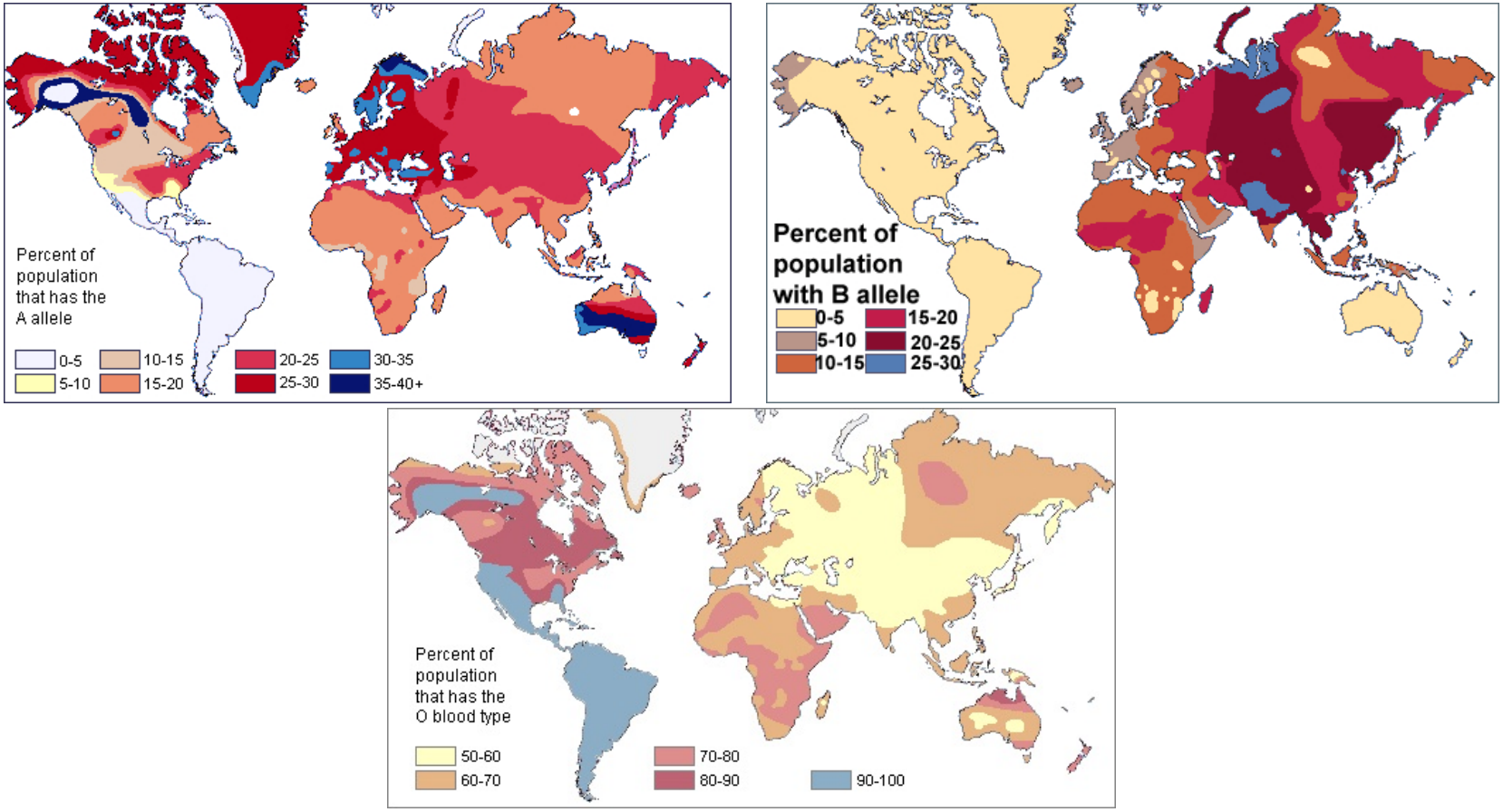
G. H. Hardy Mendelian proportions in a mixed-population 1908 Science 49-50.

- Gregor Mendel Pea experiments 1886.
- Two types of genes a and A in a population.
- $p =$ proportion of a genes.
- $q =$ proportion of A genes.

  Genes occur in pairs in individuals.
- Hardy-Weinburg principle.

Pair	aa	aA	AA
Frequency	p^2	$2pq$	q^2

Blood Group frequencies



Attribution: Muntuwandi

Sewall Wright 1889-1988.
Pioneer in Population Genetics.

90th Birthday. Photo by Hildegard Adler.

Sewall Wright 1889-1988.

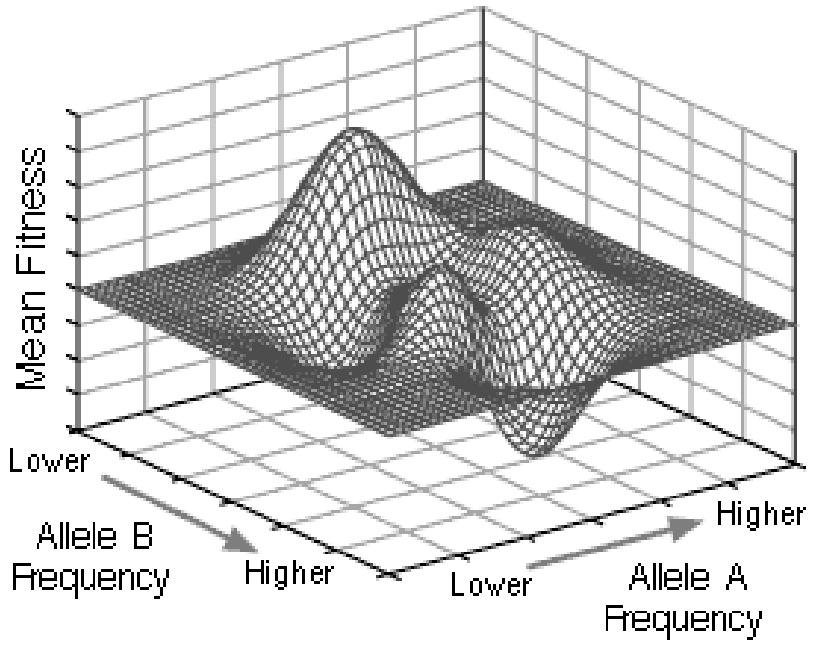
- Died at 99 - Slipped on ice and broke his pelvis.
- At 7 years old he wrote The wonders of nature.
- Department of Agriculture 1915-1925.
- University of Chicago 1926. Studied Guinea Pigs.
- University of Wisconsin. Age 65+ for 30 years.

Sewall Wright Awards.

- National Academy of Sciences
- American Philosophical Society
- Weldon medal, Oxford University
- Darwin medal from the Royal Society
- National Medal of Science

Sewall Wright's mathematics

- Random drift of gene frequencies
- Shifting Balance theory



Attribution: Adam M. Goldstein shiftingbalance.org

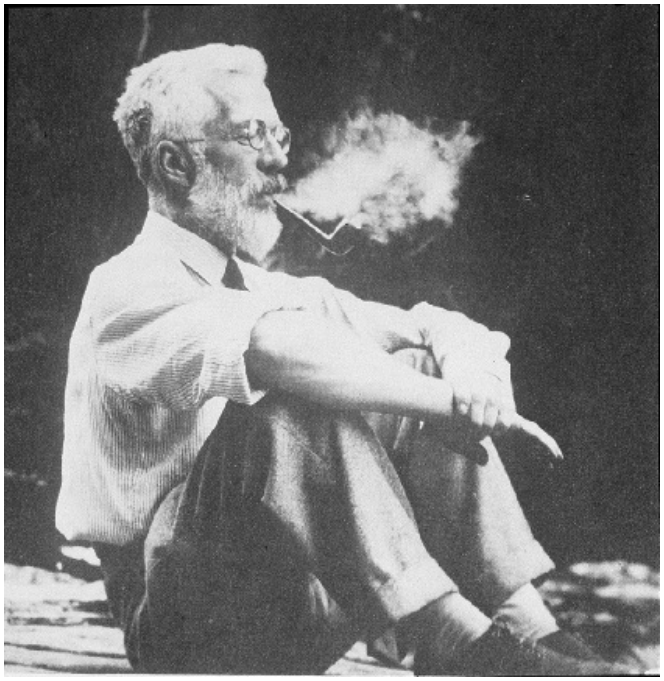
Sir Ronald Fisher 1890-1962.

English Statistician, Evolutionary Biologist, Geneticist, Eugenicist.

Sir Ronald Fisher History.

- Worked on a farm in Canada 1913-1919
- Statistician, Investment firm 1913-1919
- Rothamsted Agricultural Station 1919-1933
- Published Genetical theory of natural selection 1930
- Galton Professor of Eugenics,
  London University College
- Cambridge University 1945-1957
- Adelaide University 1957-1962
- 1917 Married Ruth Guinness 17 years old without her mother's knowledge. 2 sons, 7 daughters.

Fisher on smoking

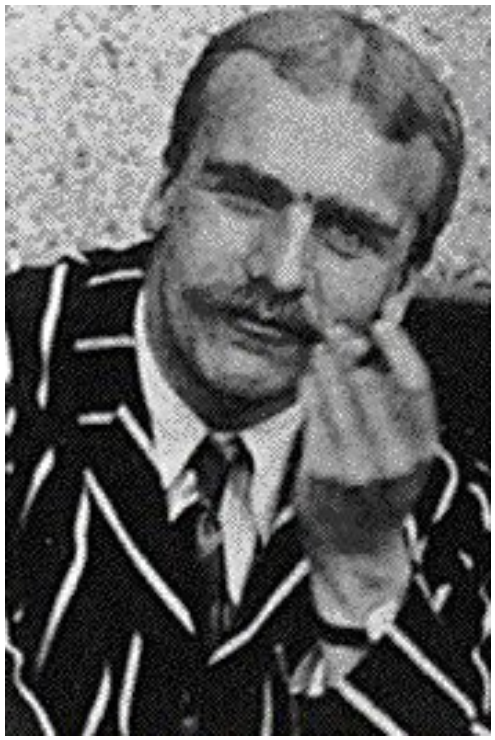


Fisher was opposed to the conclusions of Richard Doll and Austin Hill that smoking causes lung cancer. He argued that correlation did not prove causation.

Sir Ronald Fisher 1890-1962.
Important mathematics

- Analysis of variance
- Method of maximum likelihood
- Fiducial inference
- Population Genetics
- Wright-Fisher model of evolution
- Fisher 1918 studied Mendel's results
- Fundamental theorem of natural selection

J. B. S. Haldane 1892-1964



British born Geneticist from Oxford University, Evolutionary Biologist.

J. B. S. Haldane

- Studied at Oxford University 1919-1922
- Biochemist
- Cambridge University 1923-1932
- First Weldon Professor of Biology,
  University College London 1933.
- Marxist, critical of Britian's role in Suez Crisis,
  Moved to India and became an Indian citizen 1956,
  Worked at the Indian Statistical Institute.

J. B. S. Haldane's mathematics

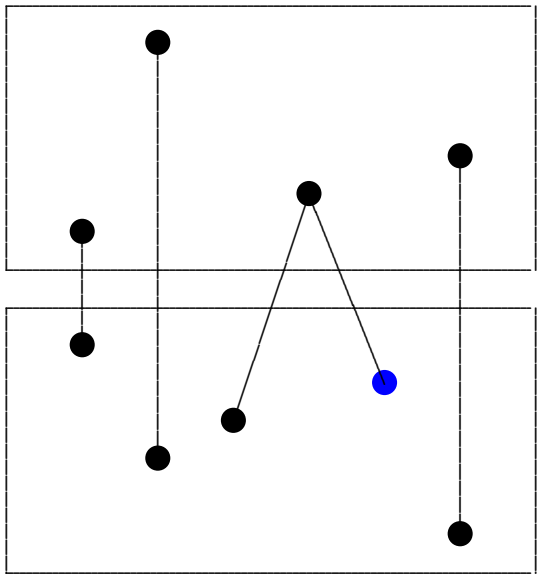
- Series of 10 papers

  A Mathematical theory
  of Natural and Artificial Selection

- Haldane's rule for speciation

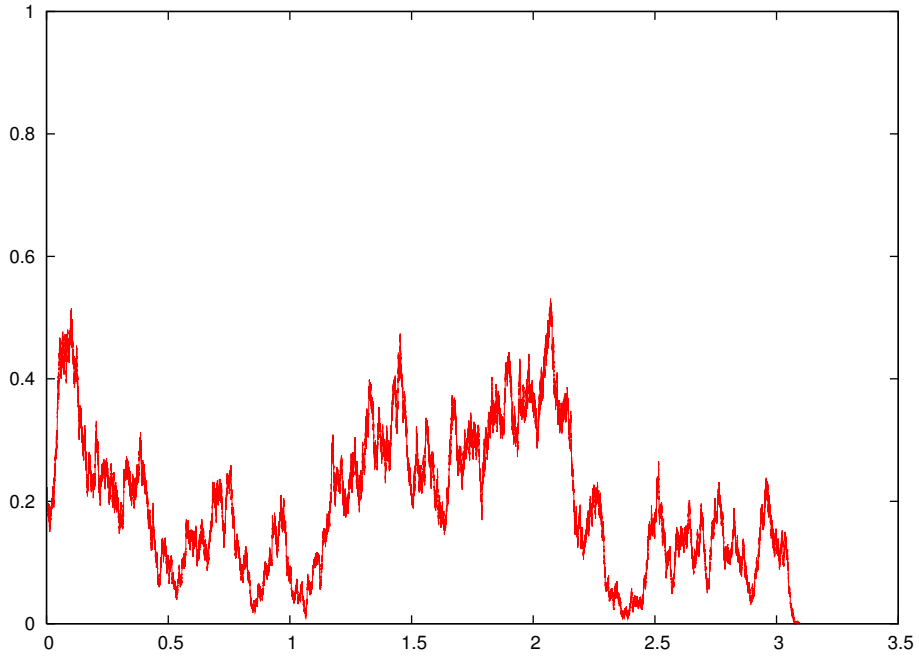
Wright-Fisher model of evolution

Current generation



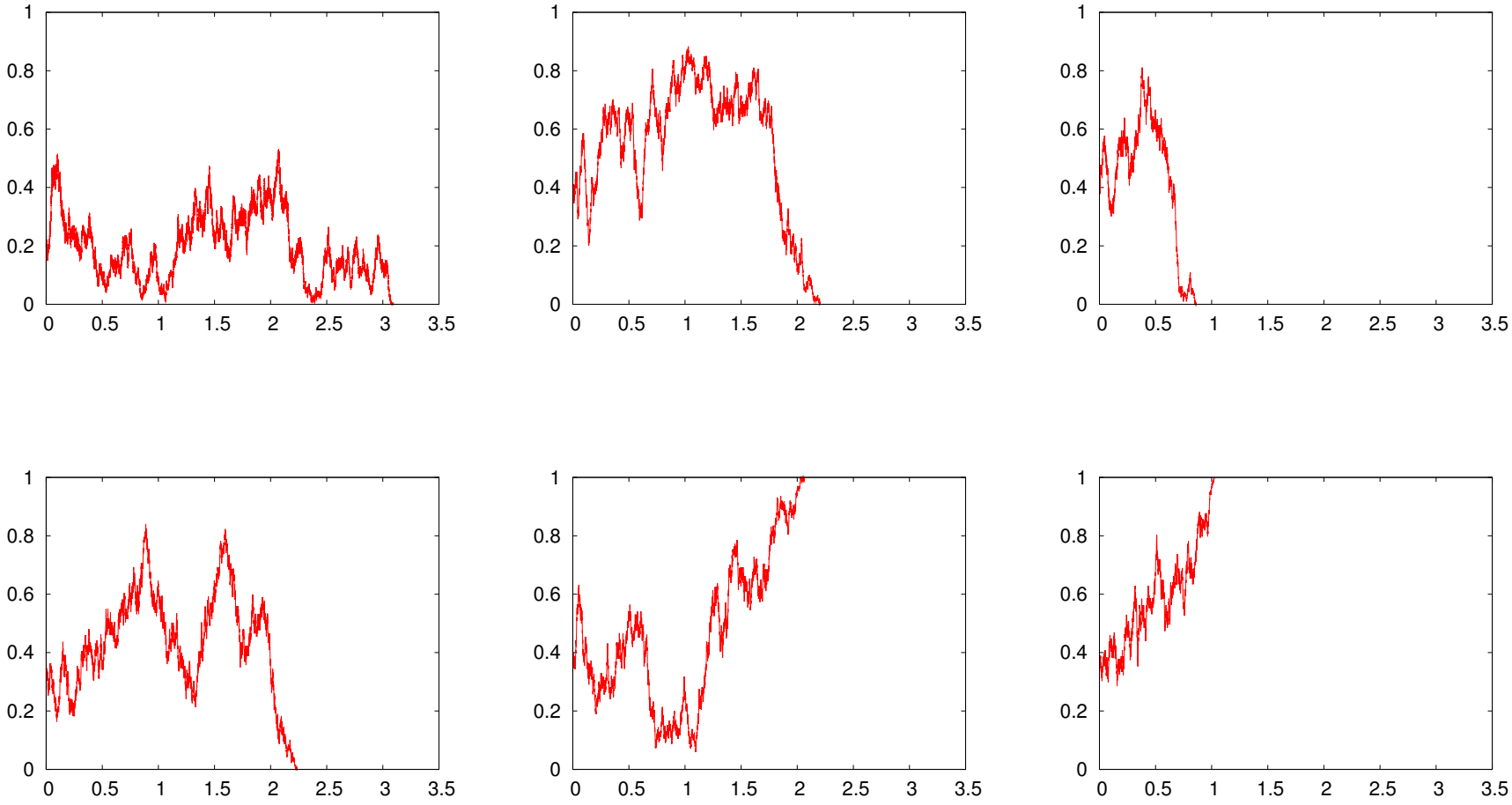
Next generation

A population of N genes. Generations $0, 1, \ldots$.
The next generation is formed by choosing N offspring at random from the previous generation.

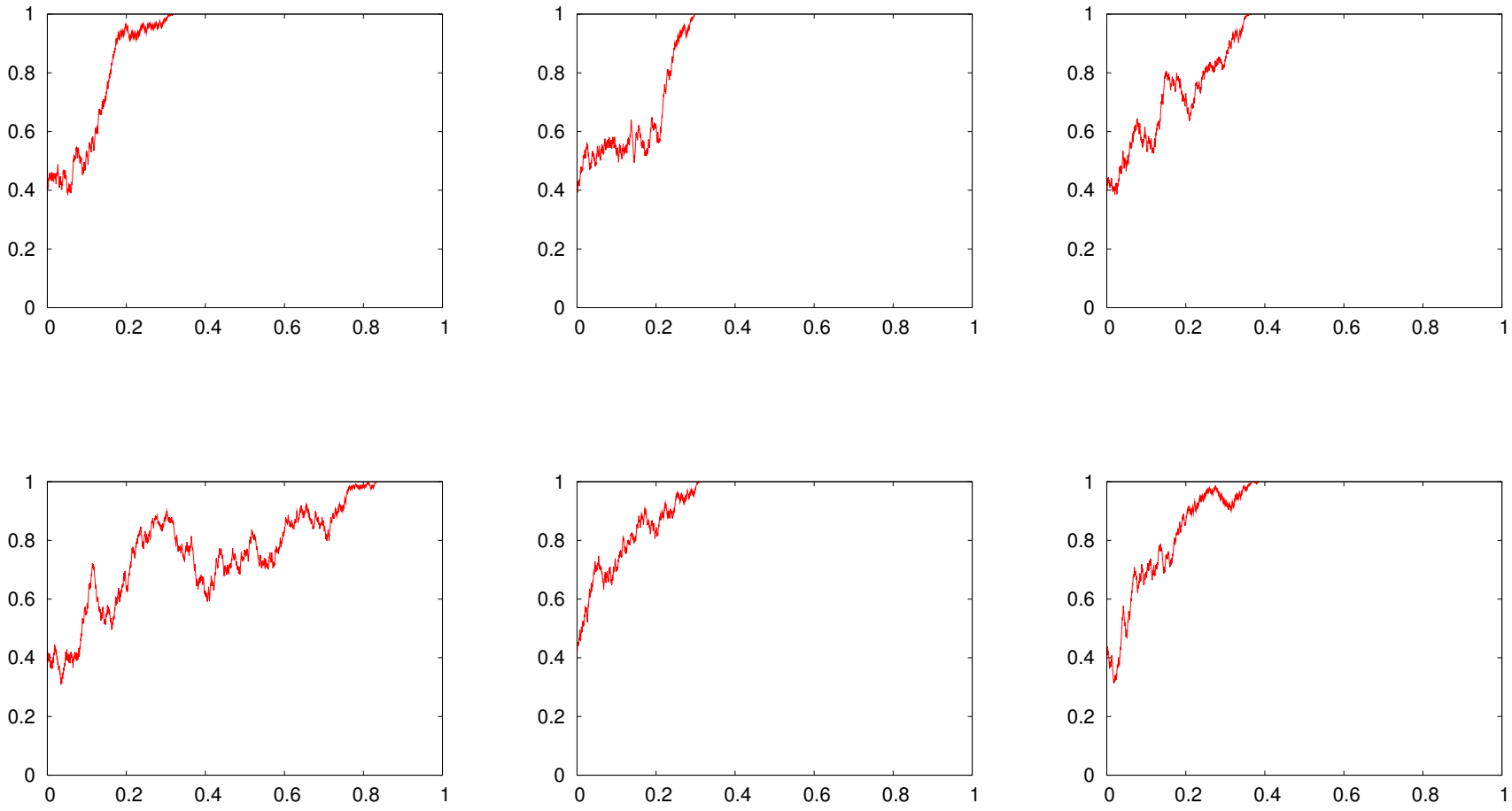


$X(t)$ is the gene frequency of type a genes at time t, measured in units of N generations. N is large.

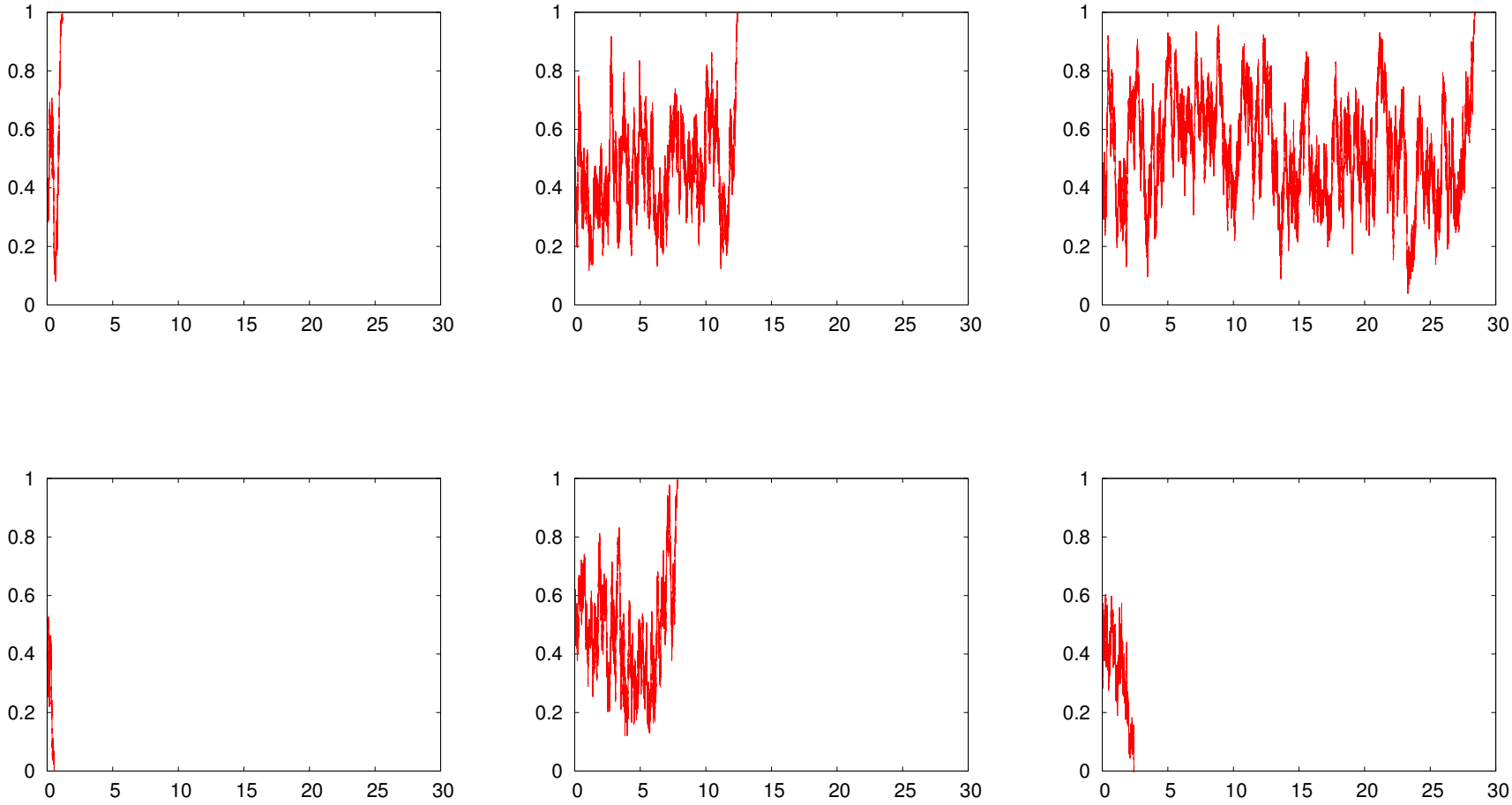
$X(t)$ is a Wright-Fisher Diffusion process which is very important in probability theory.



Plots of $X(t)$ from different populations in a neutral model.

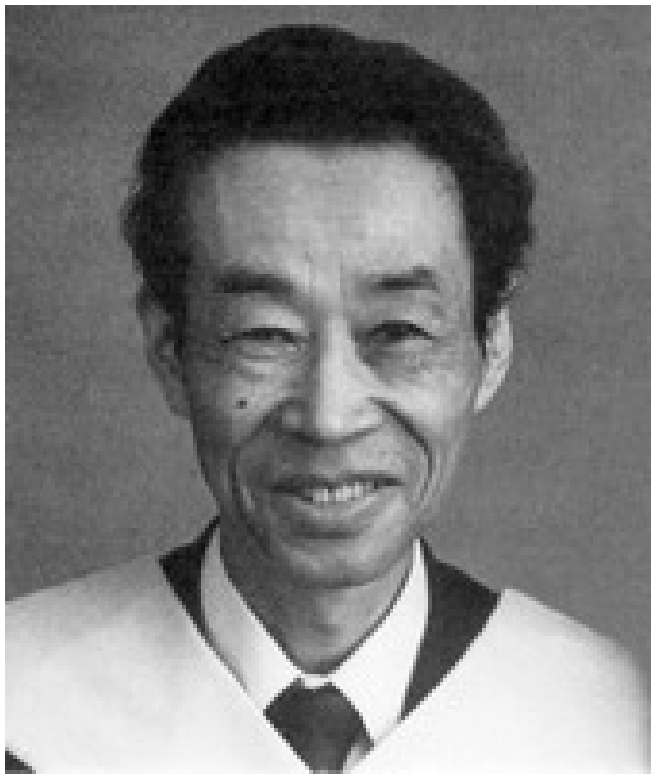


Plots of $X(t)$ when a is favoured.

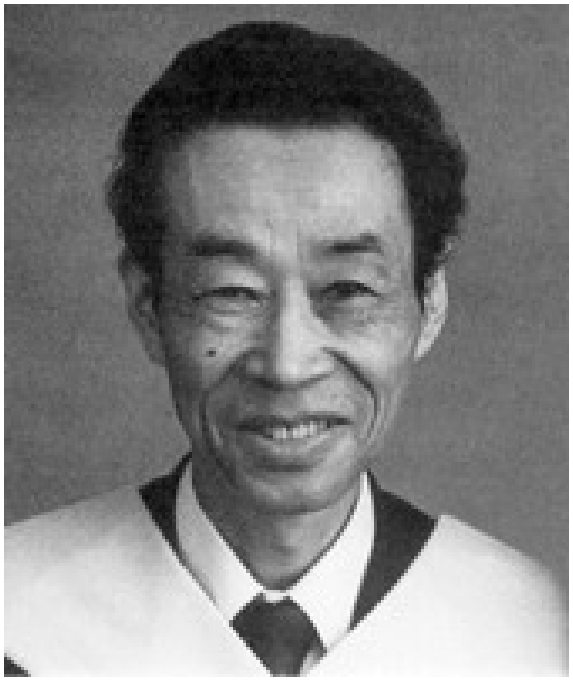


Plots of $X(t)$ with balancing selection

Motoo Kimura 1924-1994
Neutral Evolution Theory.



Tomoko Ohta and Motoo Kimura



James Crow, Ronald Fisher, Motoo Kimura 1961

Motoo Kimura

- Theoretical population genetics.
- National Institute of Genetics, Mishima, Japan 1949
- University of Wisconsin 1954

Honours

- 1959 Genetics Society of Japan
- Weldon Prize, Oxford University
- 1973 Foreign member of National Academy of Sciences, USA
- 1992 Darwin medal
- 1993 Foreign member of the Royal Society

Motoo Kimura's mathematics

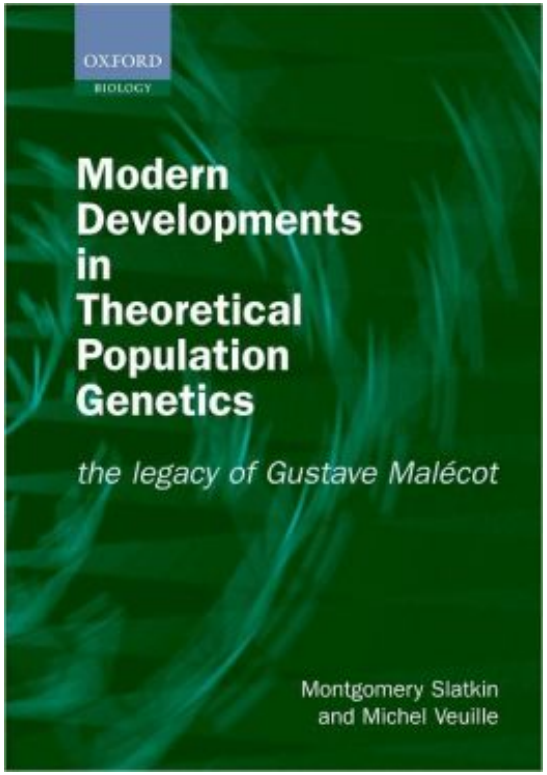
- The neutral theory of evolution
- Diffusion processes in population genetics
- The theory of molecular evolution
- The molecular clock
- 660 papers and 6 books

Discovery of DNA

James Watson and Francis Crick (Rosalind Franklin)

- DNA double helix Nature 1953
- Nobel prize 1962

Gustave Malécot 1911-1998, French Mathematician



Gustave Malécot 1911-1998

- Université de Lyon 1946-1998
- The mathematics of heredity
- Spatial populations of gene frequencies
- Identity by descent of genes, looking back in time.

Warren Ewens 1937-

- Theoretical Population Genetics
- Ewens' Sampling formula 1972
- Mathematical Population Genetics: Book 2004

Warren Ewens

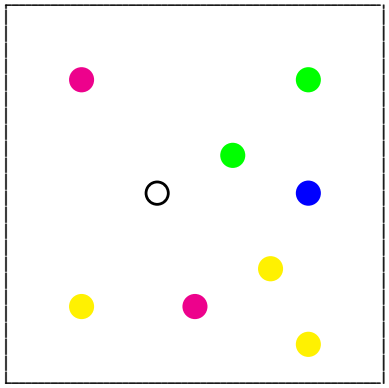
- La Trobe University 1967-1972
- University of Pennsylvania 1972-1977
- Monash University 1978-1996
- University of Pennsylvania 1997
- Fellow of the Royal Society
- Fellow of the Australian Academy of Science.
- Australian Statistical Society's E.J. Pitman Medal
- Weldon Memorial Prize, Oxford University.

Ewens' sampling formula

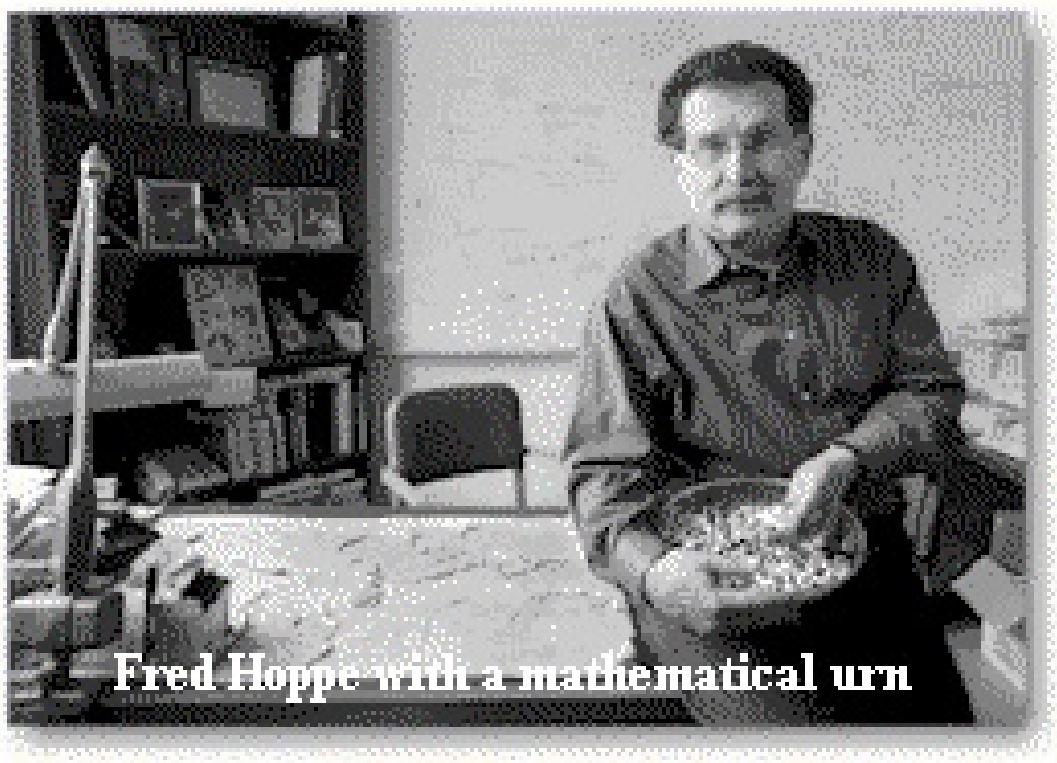
- Mutation maintains different types of genes in a population.
- In the infinitely-many-alleles model every mutation produces a new type.
- The probability distribution of the number of genes of different types in a sample of n is the Ewens' sampling formula.
- AAA, BB, $CCCCCCC$, $DDDDD$

$$\frac{17 \times 16 \times \cdots \times 1}{3 \times 2 \times 7 \times 5} \times \frac{\theta^4}{\theta \times (\theta+1) \times \cdots \times (\theta+17-1)}$$

Hoppe's (1987) urn model



- Start with 1 white ball ○ of mass θ in the urn.
- Select a ball from the urn. If it is white return it with a ball of a new colour, if not add a ball of mass 1 of the same colour as the ball drawn.
- Stop when n non-white balls.
- The probability distribution of the coloured balls is the same as Ewens' sampling formula.



Fred Hoppe with a mathematical urn

THEORETICAL POPULATION BIOLOGY **49**, 154–188 (1996)
ARTICLE NO. 0010

Motoo Kimura's Use of Diffusion Theory in Population Genetics

G. A. WATTERSON

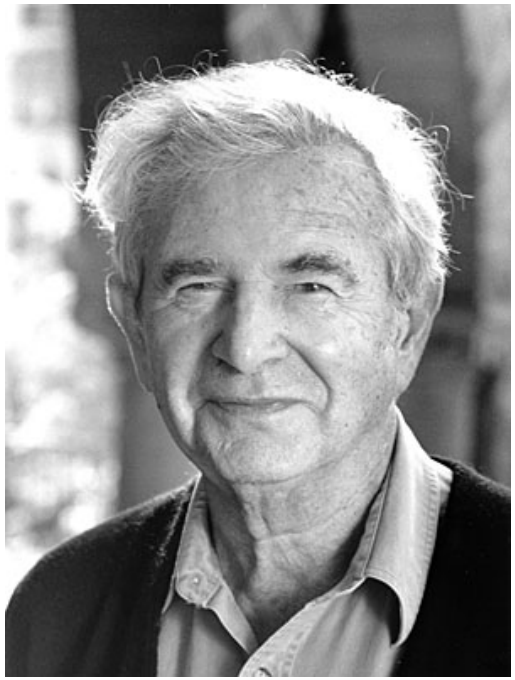
11 Fairview Avenue, Wheelers Hill, Victoria, Australia 3150

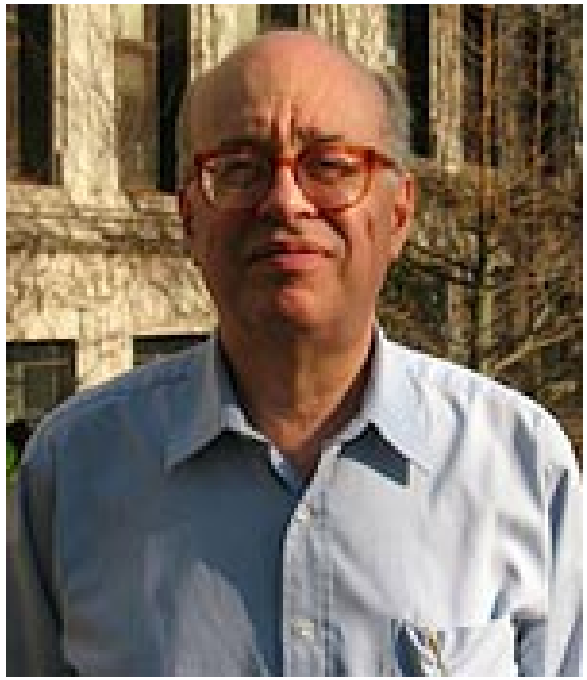
Received May 2, 1995

1. INTRODUCTION

Only late in our careers did I have the privilege of meeting Motoo Kimura. Two occasions stand out in my memory. One was a celebration of his 60th birthday at Mishima in 1984, a very happy occasion. The other was at the symposium in Tokyo following his being awarded the International Prize for Biology, November, 1988. In his response to being awarded the prize, Kimura said

> During the course of my academic endeavors, I encountered the molecular revolution in biology. This led me to attempt the construction of a new theory of population genetics, incorporating new knowledge from molecular biology. I feel very lucky that this revolution occurred just at the time when my theoretical work was ready for it. I was therefore able to publish the first version of my neutral theory of molecular evolution just twenty years ago.... After the expression, "Survival of the Fittest," which epitomizes the Darwinian theory of natural selection, I have proposed "Survival of the Luckiest" as a phrase that best characterizes my Neutral Theory. (Kimura, 1990a)





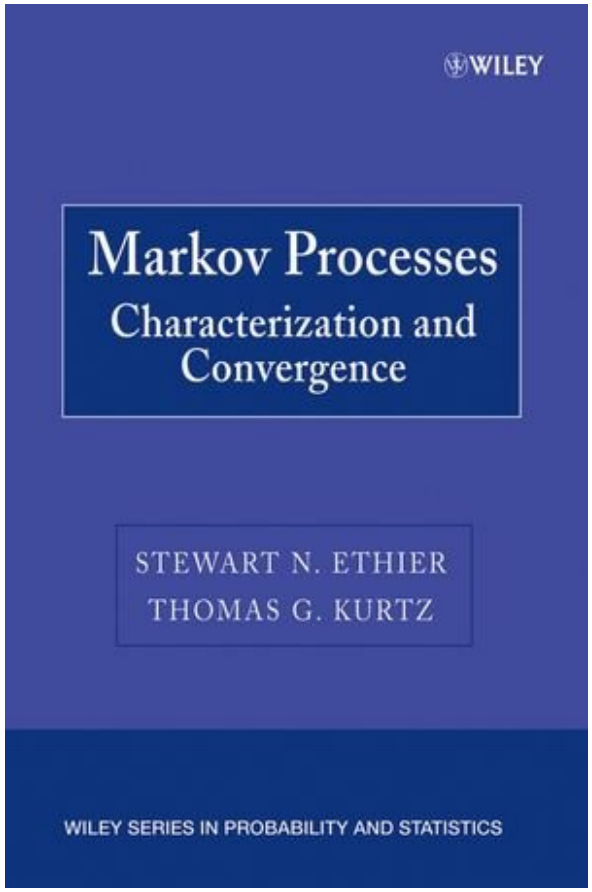
Sam Karlin, Pat Moran, Tom Nagylaki

Measure Valued Diffusions

- $(X(t), Y(t), Z(t), \ldots)$ are gene-frequencies at time t with $X(t) + Y(t) + Z(t) + \cdots = 1$.
- A measure-valued diffusion process associates random positions with gene frequencies $\Big\{(\alpha, X(t)), (\beta, Y(t)), (\gamma, Z(t)), \ldots\Big\}$.

  A random probability distribution

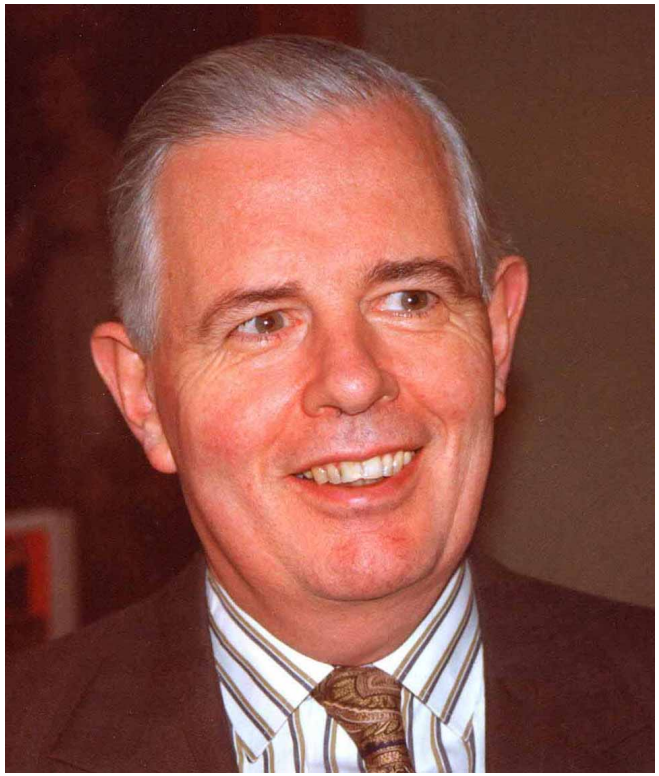
- The probability of taking a position α is $X(t)$; β is $Y(t)$; γ is $Z(t)$; ....
- Measure-valued diffusion processes are important in modern probability theory.



Stuart Ethier and Tom Kurtz worked on measure-valued diffusion processes.

Sir John Kingman 1939-

- British Mathematician
- The Coalescent

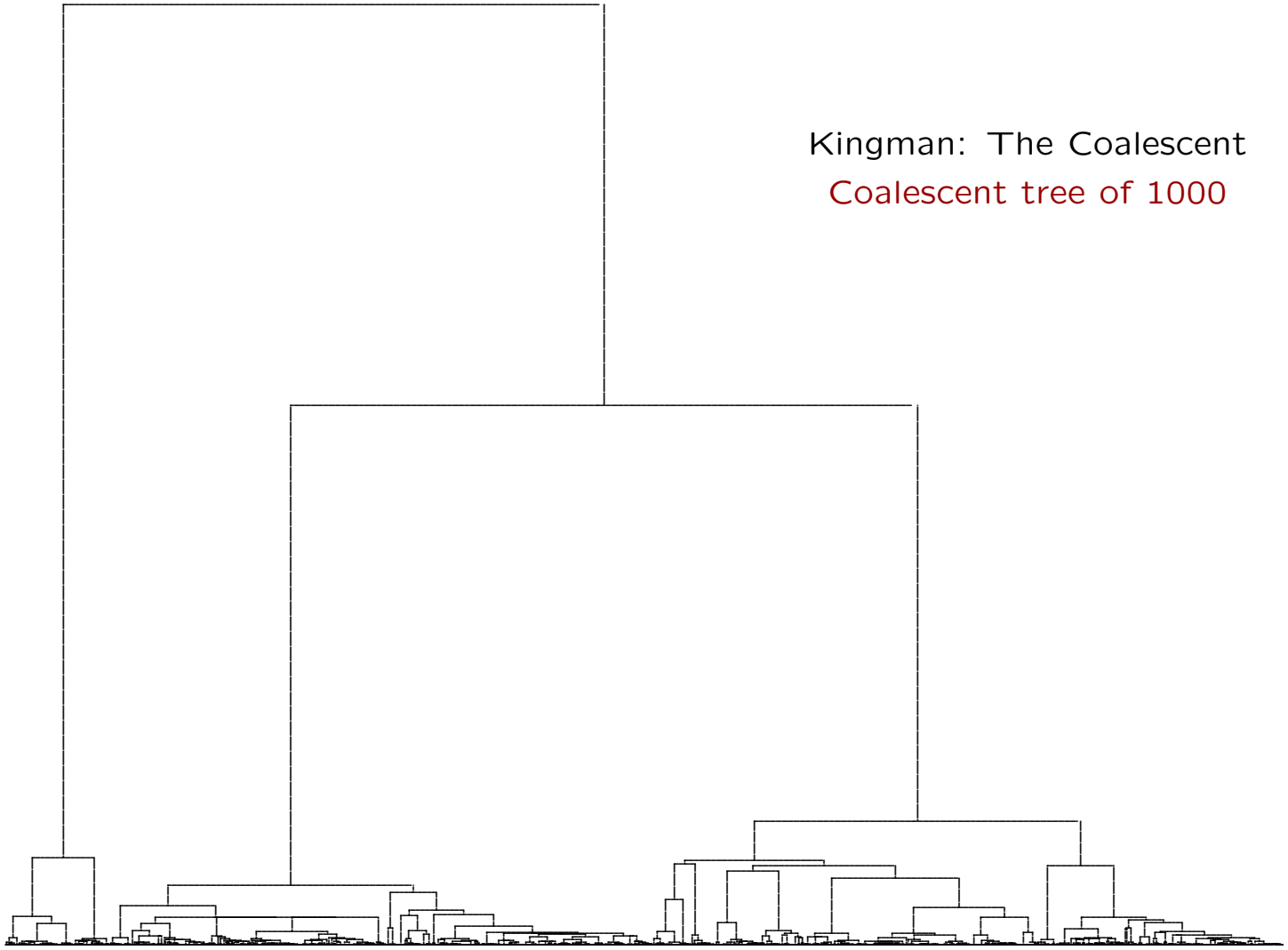


Sir John Kingman

- University of Sussex 1965-69
- University of Oxford 1969-85
- SERC 1981-85
- Knighted in 1985
- Vice-Chancellor, University of Bristol 1985-2001
- Director of the Isacc Newton Institute
  University of Cambridge 2001-06

Kingman: The Coalescent

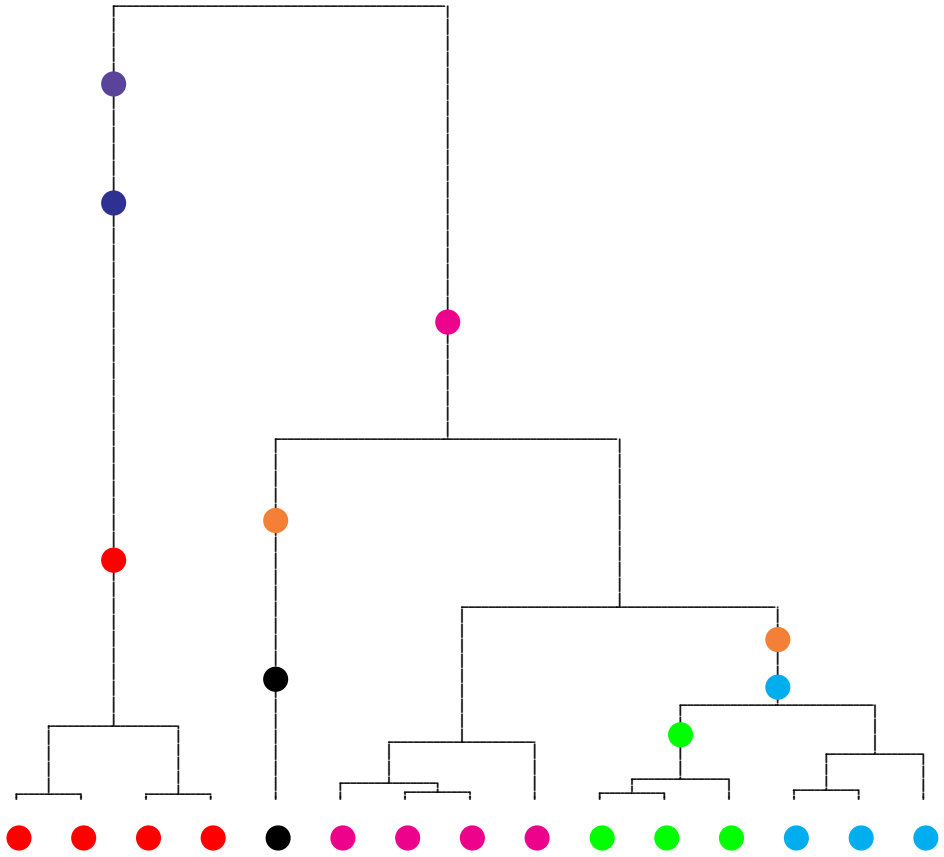
Coalescent tree of 1000



The Coalescent 1982

- The genealogical tree of n genes in a population can be traced back in time to a common ancestor.
- Ancestral lines in the tree coalesce when there is a common ancestor of two lines.
  The rate of coalescence is 1 for each pair of lines at any given time.
- An exact mathematical description gave a way to look back in time and make deductions from data.

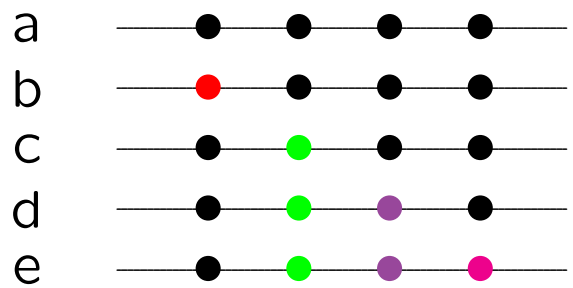
Infinitely-many-alleles-model: unique mutations



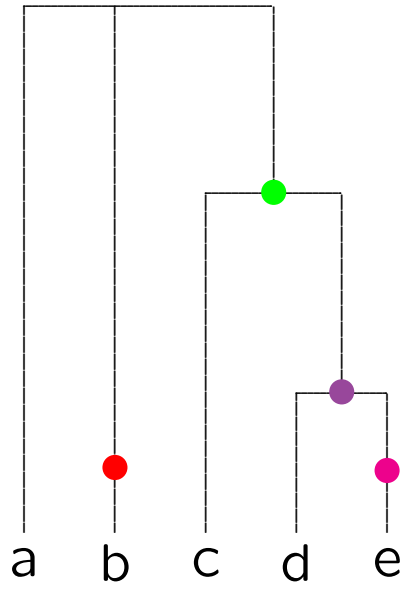
Sample of genes 4 A, 1 B, 4 C, 3 D, 3 E.

Infinitely-many-sites-model

DNA sequences



Hammer's Y tree

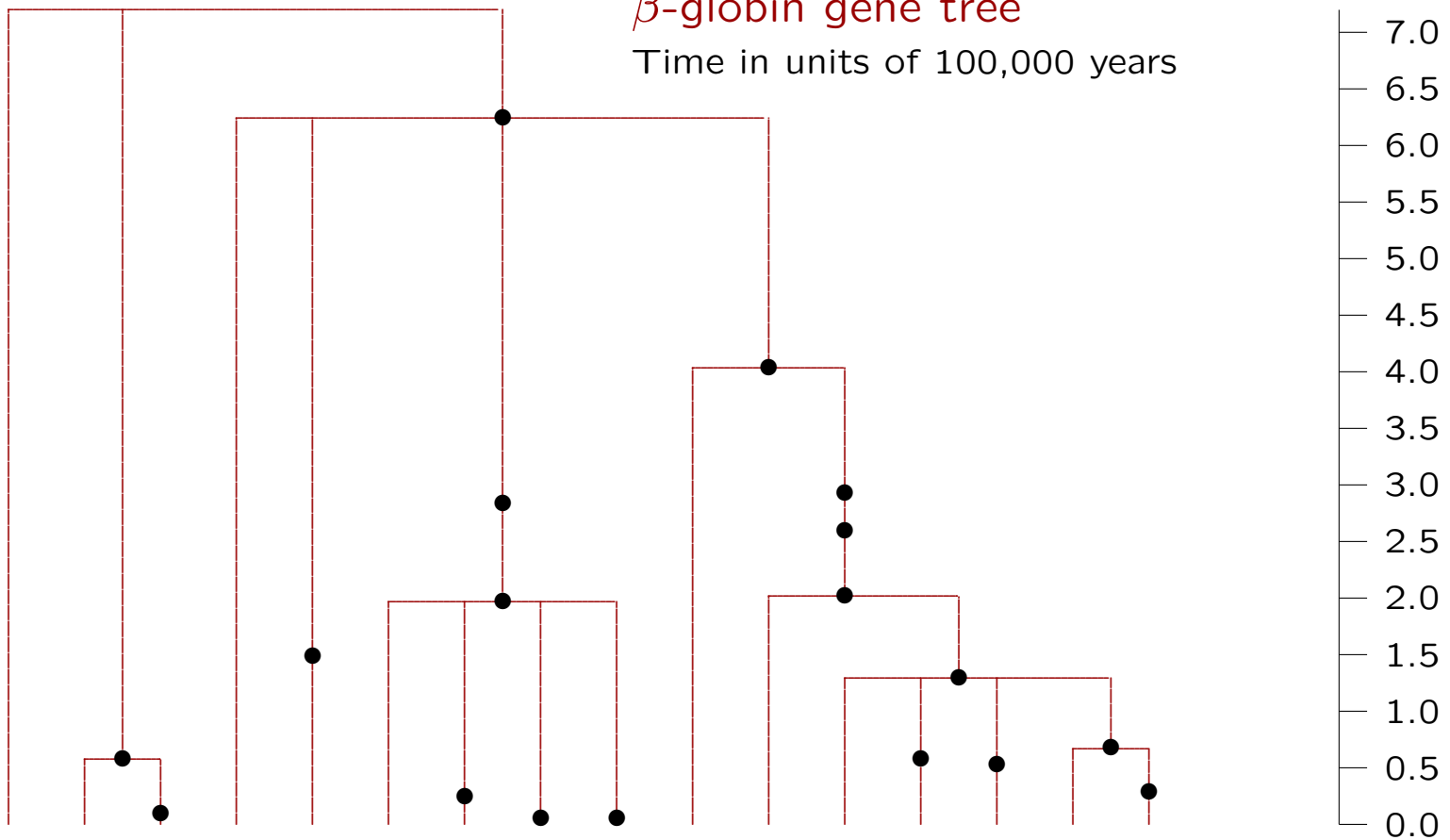


A recent common ancestry for human Y chromosomes

Michael F. Hammer, Nature 1995.

β-globin data sequences

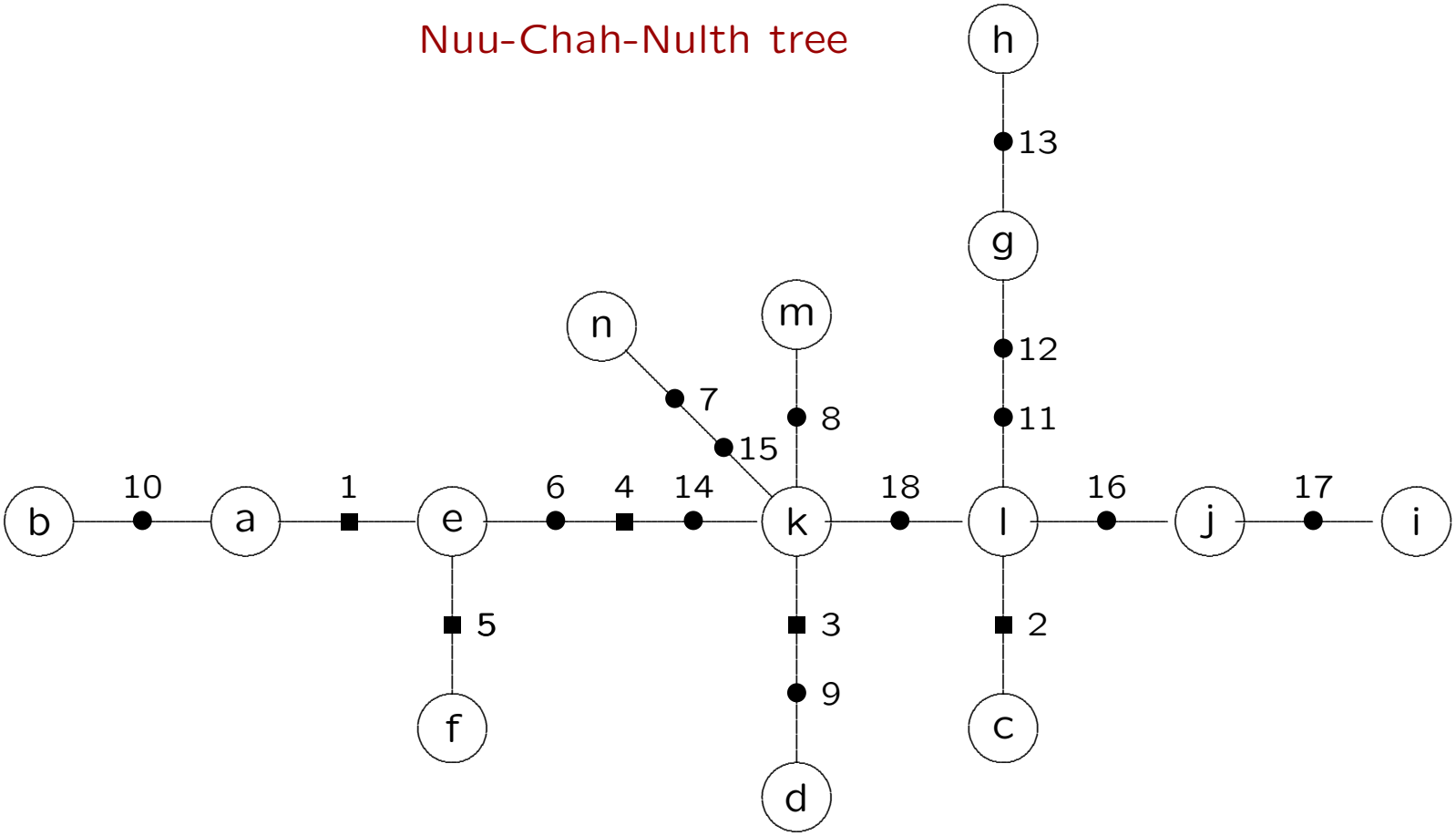
Root	C	T	T	T	A	C	C	T	T	C	T	G	G	G	C	A	G	T	T	Freq
A1	A	T	T	T	A	C	C	T	G	C	T	G	G	G	C	A	G	T	G	104
A2	A	T	T	C	A	C	C	T	G	C	T	G	G	G	C	A	G	T	G	1
A3	A	T	T	T	A	C	C	T	G	C	T	A	G	G	C	A	G	T	G	8
A4	A	T	T	T	G	C	C	T	G	C	T	G	G	G	C	A	G	T	G	1
B1	A	T	T	T	A	C	C	T	T	C	T	G	G	G	C	T	G	T	T	79
B2	C	T	T	T	A	C	C	T	T	C	T	G	G	G	C	A	G	T	T	18
B3	A	T	T	T	A	C	C	T	T	C	T	G	G	G	C	A	G	T	T	9
B4	C	T	T	T	A	C	C	T	T	C	T	G	G	G	C	A	G	C	T	3
B9	A	T	T	T	A	C	C	T	T	C	T	G	G	G	A	A	G	T	T	2
B11	C	T	T	T	A	C	C	T	T	C	T	G	G	G	C	A	A	C	T	1
C1	A	C	C	T	A	T	G	T	T	C	C	G	G	G	A	A	G	T	T	48
C2	A	T	C	T	A	T	G	T	T	C	C	G	G	G	A	A	G	T	T	9
C3	A	T	C	T	A	T	C	T	T	C	C	G	G	G	A	A	G	T	T	10
C7	A	T	C	T	A	T	G	T	T	C	C	G	G	A	A	A	G	T	T	19
D1	A	T	C	T	A	T	G	T	T	T	C	G	C	G	A	A	G	T	T	13
D2	A	T	C	T	A	T	G	T	T	T	C	G	G	G	A	A	G	T	T	1



World	18	3	1	9	79	104	8	1	1	2	10	9	48	19	1	13
Pygmies	4	·	·	1	6	9	1	·	·	1	·	·	·	·	·	·
Gambia	6	3	·	2	5	8	2	·	·	·	·	·	1	·	1	·
Kenya	8	·	1	6	9	12	5	·	·	·	·	·	·	·	·	1
Mongolia	·	·	·	·	3	3	·	·	1	·	·	2	4	6	·	3
Amerind	·	·	·	·	2	15	·	·	·	·	·	6	22	·	·	1
PNG	·	·	·	·	12	1	·	·	·	·	7	·	·	4	·	·
Sumatra	·	·	·	·	10	8	·	·	·	·	·	1	14	6	·	·
UK	·	·	·	·	16	23	·	·	·	1	·	·	·	2	·	4
Vanuatu	·	·	·	·	16	25	·	1	·	·	3	·	7	1	·	4

Mitochondrial DNA sample

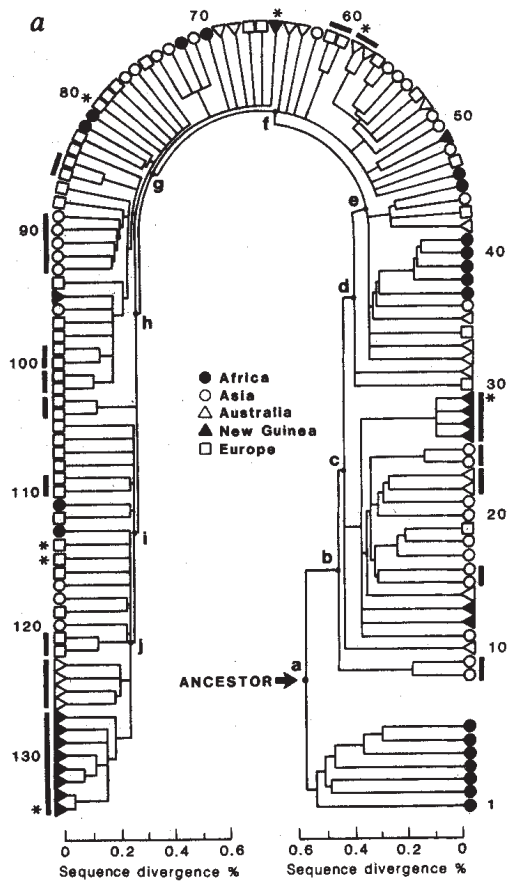
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Freq
a	A	G	G	A	A	T	C	C	T	C	T	T	C	T	C	T	T	C	2
b	A	G	G	A	A	T	C	C	T	T	T	T	C	T	C	T	T	C	2
c	G	A	G	G	A	C	C	C	T	C	T	T	C	C	C	T	T	T	1
d	G	G	A	G	A	C	C	C	C	C	T	T	C	C	C	T	T	C	3
e	G	G	G	A	A	T	C	C	T	C	T	T	C	T	C	T	T	C	19
f	G	G	G	A	G	T	C	C	T	C	T	T	C	T	C	T	T	C	1
g	G	G	G	G	A	C	C	C	T	C	C	C	C	C	C	T	T	T	1
h	G	G	G	G	A	C	C	C	T	C	C	C	T	C	C	T	T	T	1
i	G	G	G	G	A	C	C	C	T	C	T	T	C	C	C	C	C	T	4
j	G	G	G	G	A	C	C	C	T	C	T	T	C	C	C	C	T	T	8
k	G	G	G	G	A	C	C	C	T	C	T	T	C	C	C	T	T	C	5
l	G	G	G	G	A	C	C	C	T	C	T	T	C	C	C	T	T	T	4
m	G	G	G	G	A	C	C	T	T	C	T	T	C	C	C	T	T	C	3
n	G	G	G	G	A	C	T	C	T	C	T	T	C	C	T	T	T	C	1


Nuu-Chah-Nulth tree
h
13
g
12
11
n
m
7
8
15
10
1
6
4
14
18
16
17
b
a
e
k
l
j
i
5
3
2
9
f
c
d
• C, T sites; ■ A, G sites

Mitochondrial Eve lived 200,000 years ago in Africa

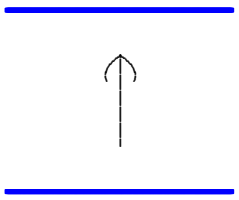
Mitochondrial DNA and Human Evolution, Nature, 1987

Rebecca Cann, Mark Stoneking and Alan C. Wilson.

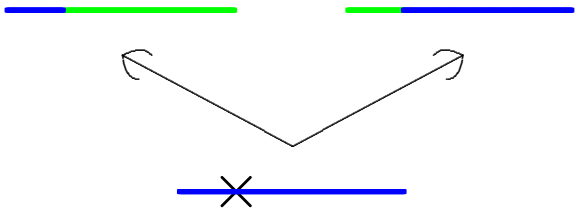


Recombination

In a Wright-Fisher model with N genes an individual has a single parent with probability $1 - r$ and two parents with probability r.



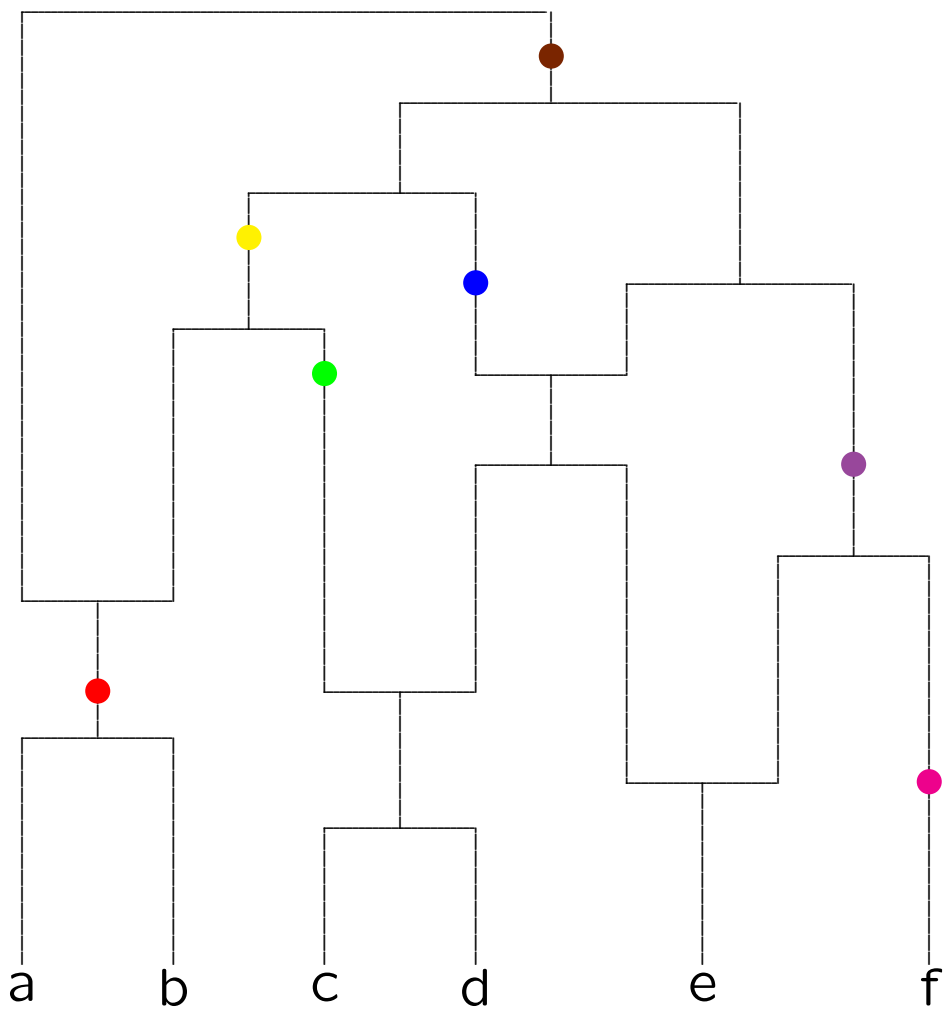
Probability $1 - r$



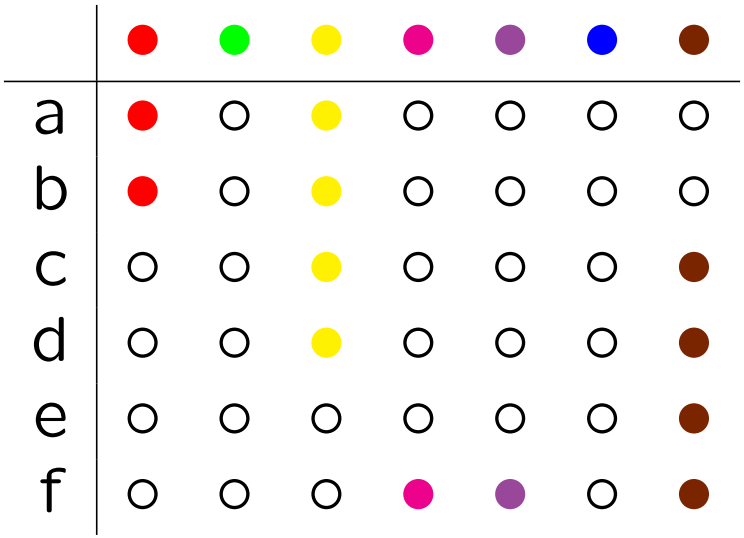
Probability r

A recombination break is chosen along the sequences according to a probability distribution.
The blue parts of the sequences are ancestral.
The green parts of the sequences are non-ancestral.

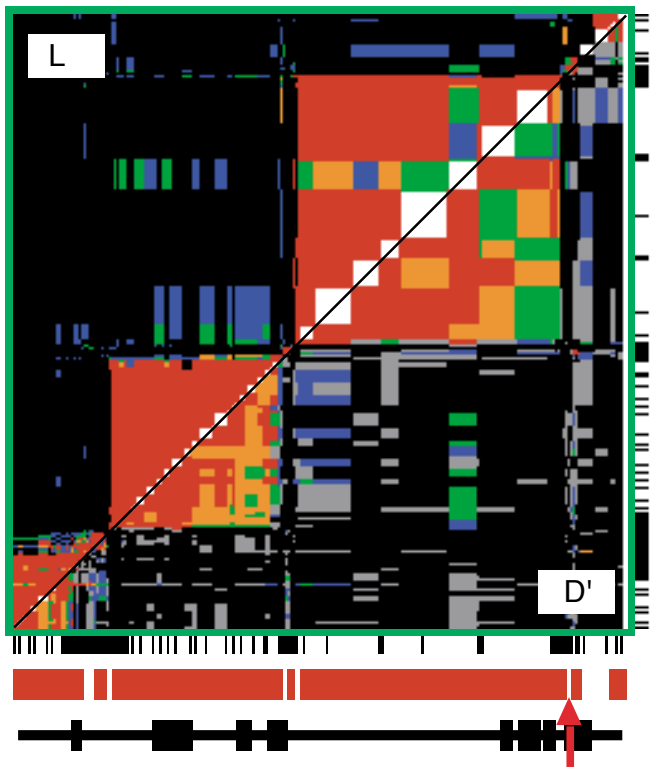
Ancestral Recombination Graph



Pattern of mutations on DNA sequences



Disequilibrium plot



Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex.
Alec J. Jeffreys, Lisa Kauppi and Rita Neumann. Science 2001

Experimental study: 8 British Semen Donors, 216 Kb region.

The International HapMap Project

Second generation human haplotype map, Nature 2007

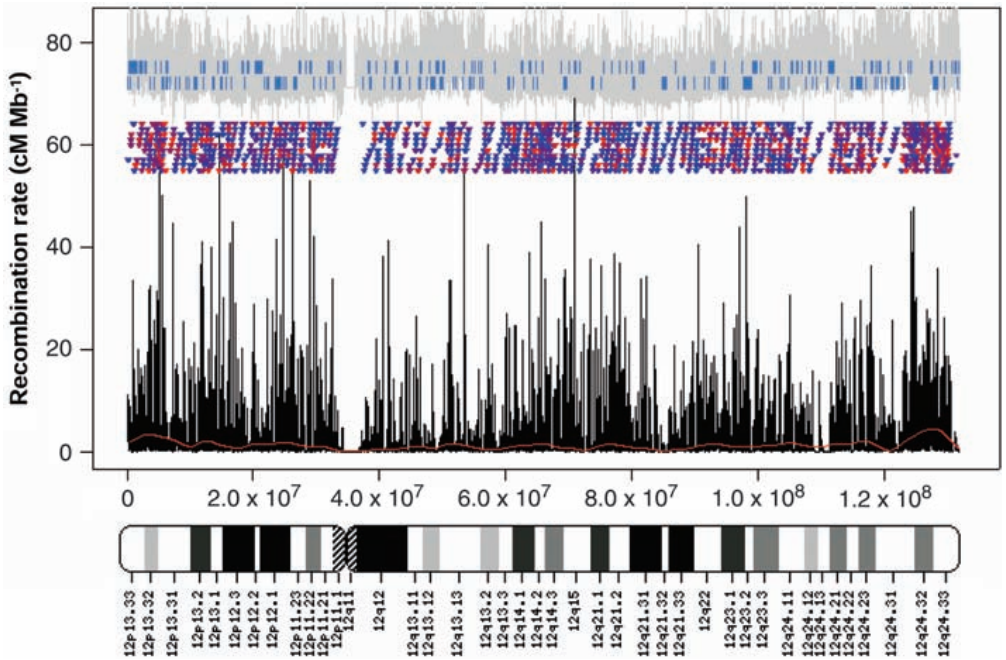
The International HapMap Consortium

- The goal of the International HapMap Project was to determine the common patterns of DNA sequence variation in the human genome and to make this information freely available in the public domain.
- 3.1 million human single nucleotide polymorphisms (SNPs) genotyped in 270 individuals from four geographically diverse populations.
- The objective was to genotype at least one common SNP every 5 kilobases (kb).

Geographic populations

- 30 mother-father-adult child trios from the Yoruba in Ibadan, Nigeria;
- 30 trios of northern and western European ancestry living in Utah from the Centre d'Etude du Polymorphisme Humain collection
- 45 unrelated Han Chinese individuals in Beijing, China
- 45 unrelated Japanese individuals in Tokyo, Japan

Recombination rate variation along chromosome 12



A Fine-Scale Map of Recombination Rates and Hotspots Across the Human Genome.

Simon Myers, Leonardo Bottolo, Colin Freeman, Gil McVean, Peter Donnelly. Science 2005

A Fine-Scale Map of Recombination Rates and Hotspots Across the Human Genome.

Simon Myers, Leonardo Bottolo, Colin Freeman, Gil McVean, Peter Donnelly. Science 2005

Hotspots are associated with a motif

CCTCCCT

Chimpanzees do not have hotspots with this motif.

The landscape of recombination in African Americans.
David Reich and Simon Myers, many others
Science 2011.

Hotspots in African Americans are associated with a motif

CCCCAGTGA

Hotspots are associated with PRDM9 genes.

Simon Myers, Gil McVean
Peter Donnelly, David Reich