Statistical Properties of Genealogical Trees

Introduction

Distribution of repetitions

Comparison between two trees

Galton Watson process and the renormalization group

Genealogy of a chromosome

PRL 99
Physica A 2000
J. Th. Biol. 2000

JSP 99

S. Manrubia
D. Zanette

B. Jung Muller

References

B. Derrida, S. Manrubia, D. Zanette

Phys. Rev. Lett. 82, 1987 (1999)

J. Th. Bio 203, 303 (2000)

Physica A 281, 1 (2000)

B. Derrida, B. Jung Muller

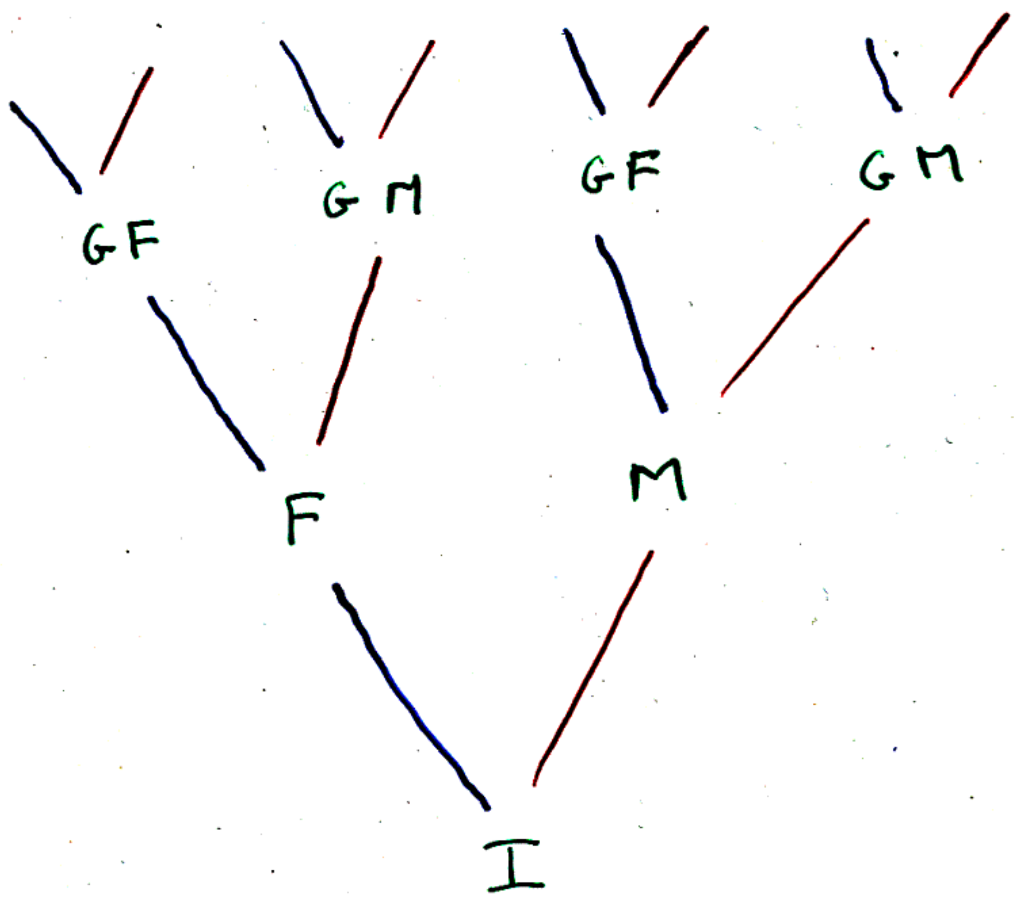
J. Stat. Phys. 94, 277 (1999)

J. T. Chang

Ann. Appl. Prob 31, 1002 (1999)

C. Wiuf and J. Hein

On the number of ancestors to a DNA sequence Genetics 147, 1459 (1997)


GF
GM
GF
GM
F
M
I

Repetitions in a genealogical tree

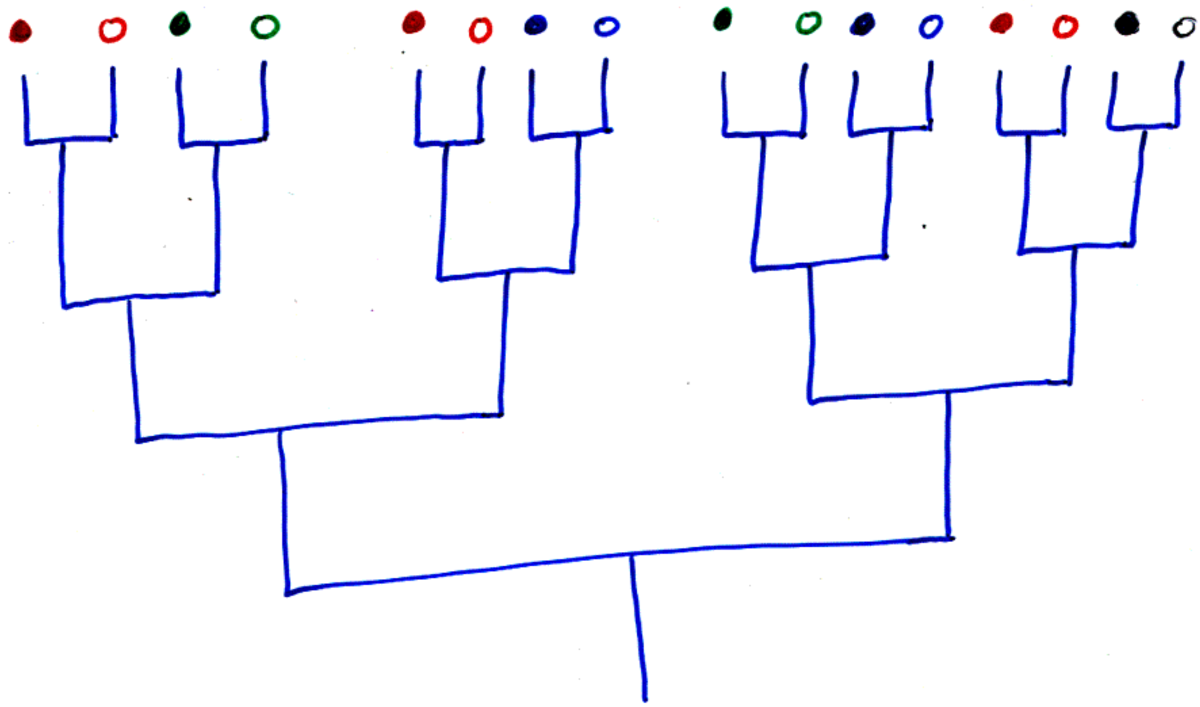
mass of earth $\sim 10^{26}$ h.b.

10^{26} ancestors $\sim$ 90 generations

90 generations $\sim$ 2000 years

$\Downarrow$

Lots of repetitions

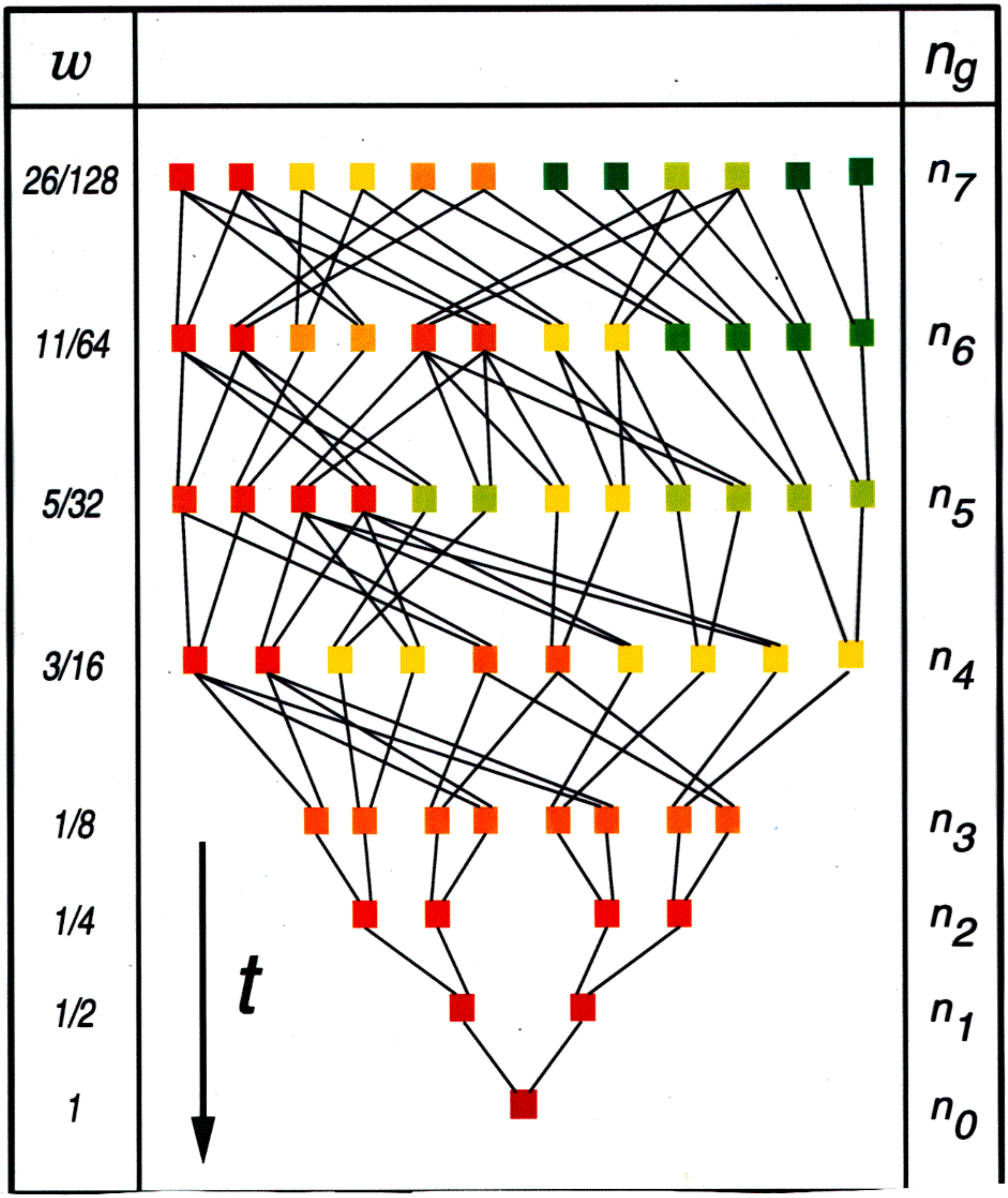


$$r(\bullet) = r(\circ) = 3$$

$$r(\bullet) = r(\circ) = 2$$

$$r(\bullet) = r(\circ) = 2$$

$$r(\bullet) = r(\circ) = 1$$



w
n_g
26/128
11/64
5/32
3/16
1/8
1/4
1/2
1
n_7
n_6
n_5
n_4
n_3
n_2
n_1
n_0
t

Model

- N individuals at each generation g
- Each individual at generation g has 2 parents chosen at random in the previous generation $g+1$

$H_g(r)$ = probability that an ancestor at generation g in the past is repeated r times

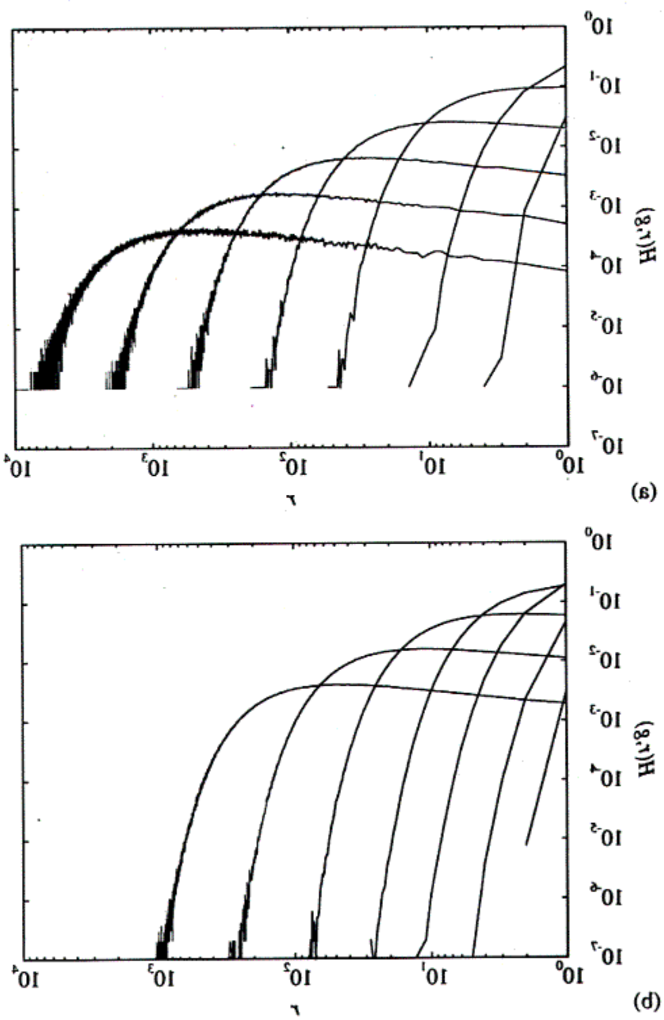


Fig. 1. Probability distribution $H(r,g)$ of r repetitions after g generations ($H(0,g)$ is not s
$g = 5, 9, 12, 14, 16, 18$, and 20 for a population of constant size. In Fig. 1a, $N = 1000$ and in
$N = 10\,000$. Both figures show averages over 1000 samples.

$H_g(r) \to$ stationary shape

r small $\quad \boxed{H(r) \sim r^{\beta}}$

$\beta = .2991138\ldots$ (exact)

Distribution of repetitions

$H_g(r)$ = probability that an ancestor is repeated r times at generation g in the past

$\boxed{\text{Large } g: \text{ all the } H_g(r) \text{ have the same shape}}$

$$r_i(g+1) = \sum_{\substack{j \text{ children} \\ \text{of } i}} r_j(g)$$

Rescaling:

$$w_i(g) = \frac{N}{2^g} r_i(g)$$

$$\boxed{w_i(g+1) = \frac{1}{2} \sum_{\substack{j \text{ children} \\ \text{of } i}} w_j(g)}$$

$$\boxed{w_i(g+1) = \frac{1}{2} \sum_{\substack{j \text{ children} \\ \text{of } i}} w_j(g)}$$

Large N:

- the $w_j(g)$ become uncorrelated
- the probability p_n of having n children is

$$\boxed{p_n = \frac{2^n}{n!} e^{-2}}$$

Generating function

$$F_g(\lambda) = \langle e^{\lambda w_i(g)} \rangle$$

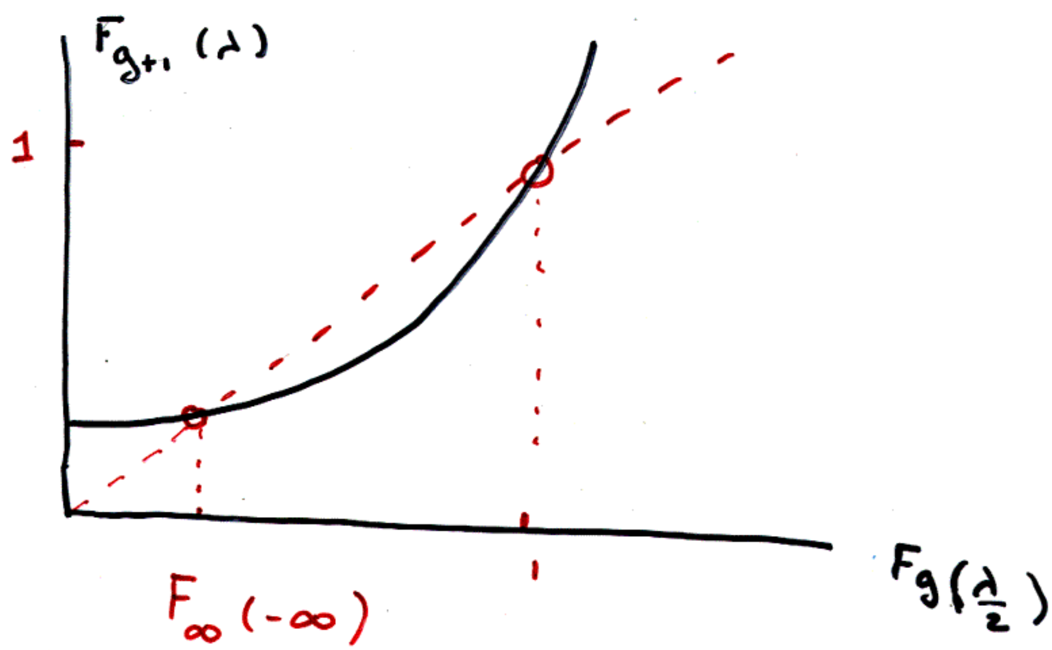
Then

$$F_{g+1}(\lambda) = \sum_n p_n \left[F_g\left(\frac{\lambda}{2}\right) \right]^n$$

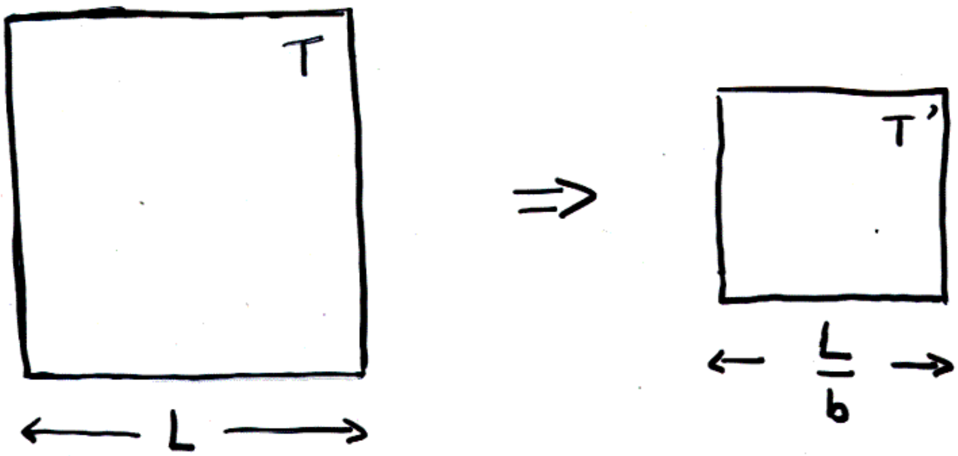
Generating function

$$F_g(\lambda) = \langle e^{\lambda w_i(g)} \rangle$$

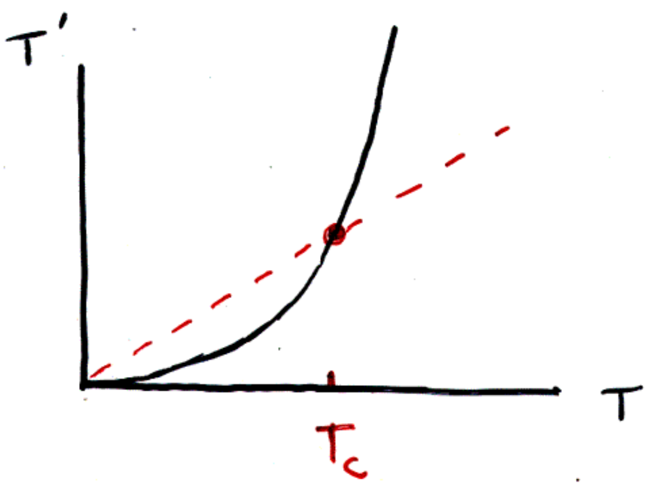
$$\boxed{F_{g+1}(\lambda) = \exp\left[2 F_g\left(\frac{\lambda}{2}\right) - 2 \right]}$$



Renormalisation



$$\boxed{T' = R_b(T)}$$



Unstable fixed point $\equiv T_c$

Exponent $\boxed{b^{1/\nu} = \left.\frac{dT'}{dT}\right|_{T_c}}$

Generalisation

$2 \longrightarrow p$ parents

$$F_{g+1}(\lambda) = e^{p F_g\left(\frac{\lambda}{p}\right) - p}$$

$p = 1 + \epsilon \Rightarrow \epsilon$ expansion . . .

$\Longleftrightarrow$ • 2 parents

• Population increasing exponentially

$$N_t = \frac{p}{2} N_{t-1}$$

Comparing two trees

$w_i^{(\alpha)}(g)$: weight of ancestor i at generation g in the pas in the tree of α

$$q_{\alpha\beta}(g) = \frac{\sum_i w_i^{(\alpha)}(g)\, w_i^{(\beta)}(g)}{\left(\sum_i w_i^{(\alpha)}(g)\right)^{1/2}\left(\sum_i w_i^{(\beta)}(g)\right)^{1/2}}$$

$$\boxed{g_c = \frac{\log N}{\log 2} - 1}$$

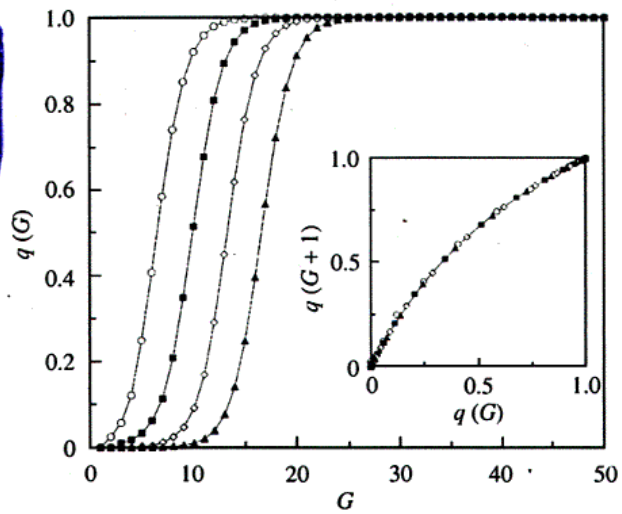


FIG. 4. The averaged overlap $q(G)$ as a function of the number of generations G. The results of simulations for different sizes of the population $N = 100$ (○), 1000 (■), 10000 (◇), 100 000 (▲) agree with this prediction, up to small finite-size corrections only visible for $N = 100$. The inset shows the results of simulations and the prediction (8).

$$q(g) = \frac{1}{1 + 2^{g_c - g}}$$

Galton Watson process 1874

T.E. Harris:
"The Theory of Branching Processes" 1963

Each individual has n descendants with probability p_n

	children	
	0	p_0
Parent	1	p_1
	2	p_2
	$\vdots$	

What is the number N_g of descendants after g generations?

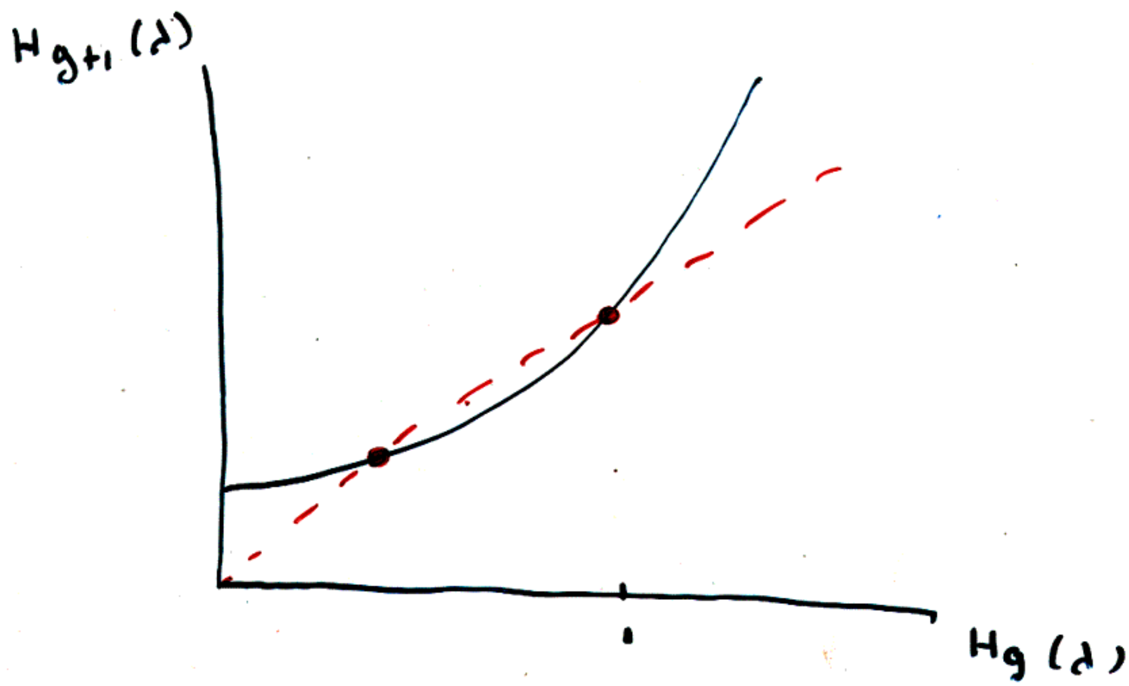
Generating function:

$$H_g(\lambda) = \langle \lambda^{N_g} \rangle$$

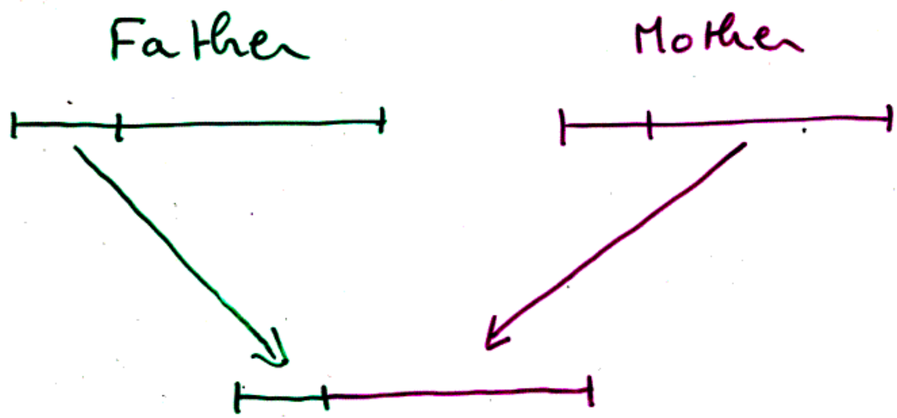
If: $f(\lambda) = \sum_n p_n \lambda^n$

Then

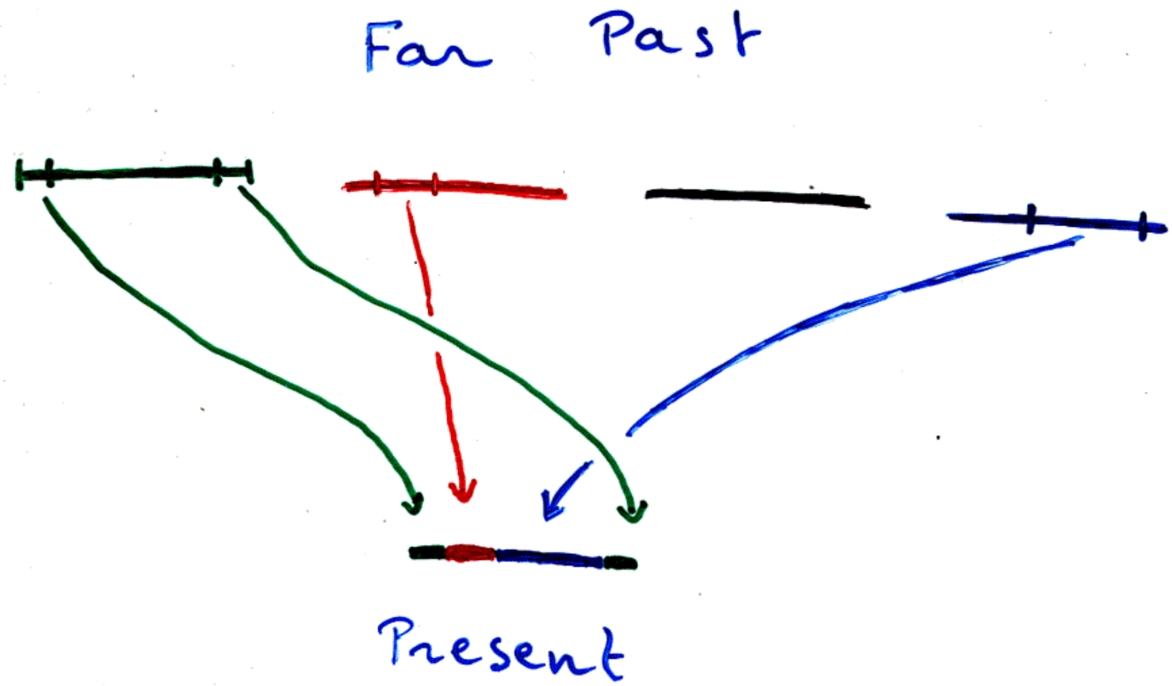
$$\boxed{H_{g+1}(\lambda) = f(H_g(\lambda))}$$



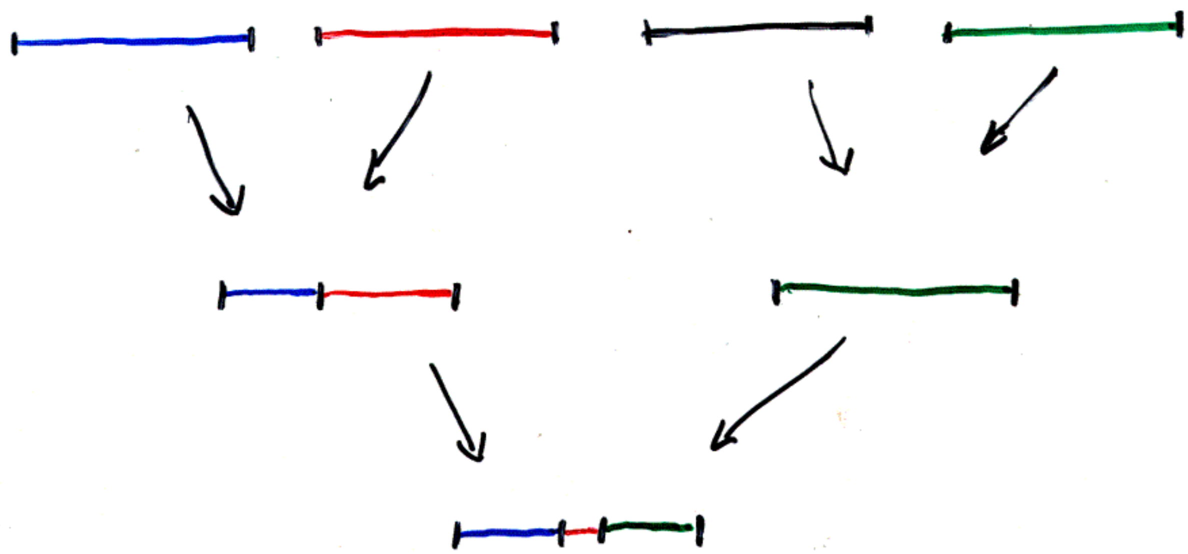
Genealogy of a chromosome



B. Jung Muller

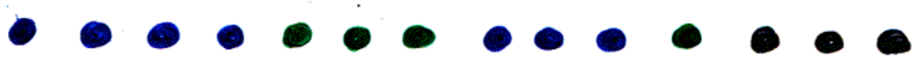


Example



Wiuf - Hein 97

Dynamics during time dt ($dt \ll 1$)



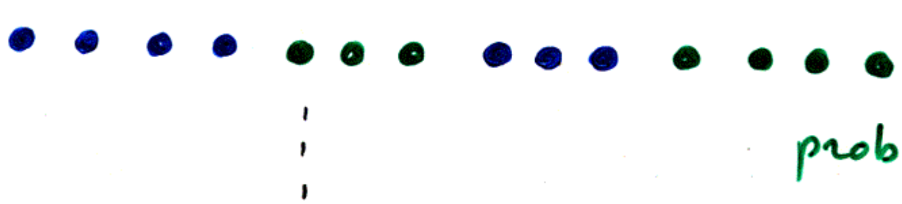
recombination : prob $17\, r\, dt$

($17 = 9 + 6 + 2$)

prob. $4\, r\, dt$

coalescence : prob $3\, dt$

(3 = nb of pairs of colors)



prob. dt

unchanged : prob $1 - (17 r + 3)\, dt$

L sites , Ω_L configurations

L	1	2	3	4	...	10	...
Ω_L	1	2	5	15	...	115975	...

- Solve exactly for L small

- Calculate the small r expansion for all L

- Calculate exactly simple correlation functions

Correlation functions

$$P_{ij} = \frac{1}{1 + r\,|j-i|}$$

$$P_{ijk} = F^{n}(r, |j-i|, |k-j|)$$

Quantities of interest

Q = number of ancestors
(≡ nb of colors)

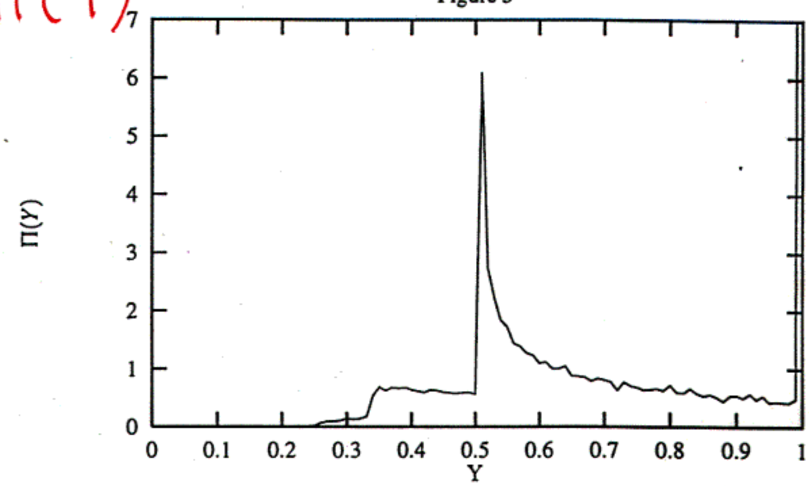
$$Y = \sum_{\alpha} (W_\alpha)^2$$

W_α : fraction of the length having color α

rL = 1

Π(Y)

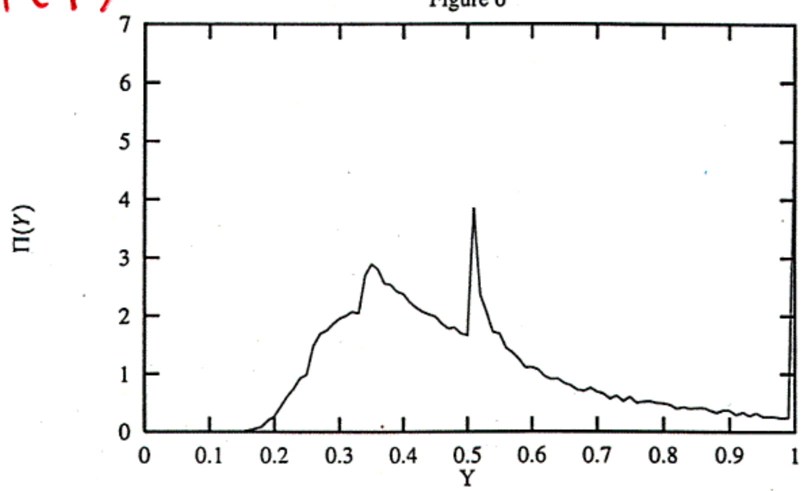
Figure 5



Y

Π(Y)

Figure 6



Y

rL = 4