PUBLIC HEALTH
TEXAS A&M HEALTH SCIENCE CENTER

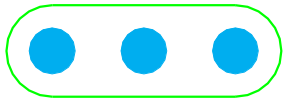
Shedding new light on random chromosome segregation

Qi Zheng
Department of Epidemiology and Biostatistics

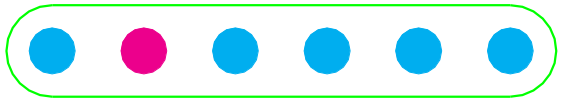
Motivation

- Bacteria were thought to be mono-ploid (haploid), but not any more
- Ploidy value c is the number of chromosomes per cell
- c could be $>$ 100
- If one copy of a gene is a mutant in a cell of $c = 20$, what will be the ploidy "landscape" 20 generations later?

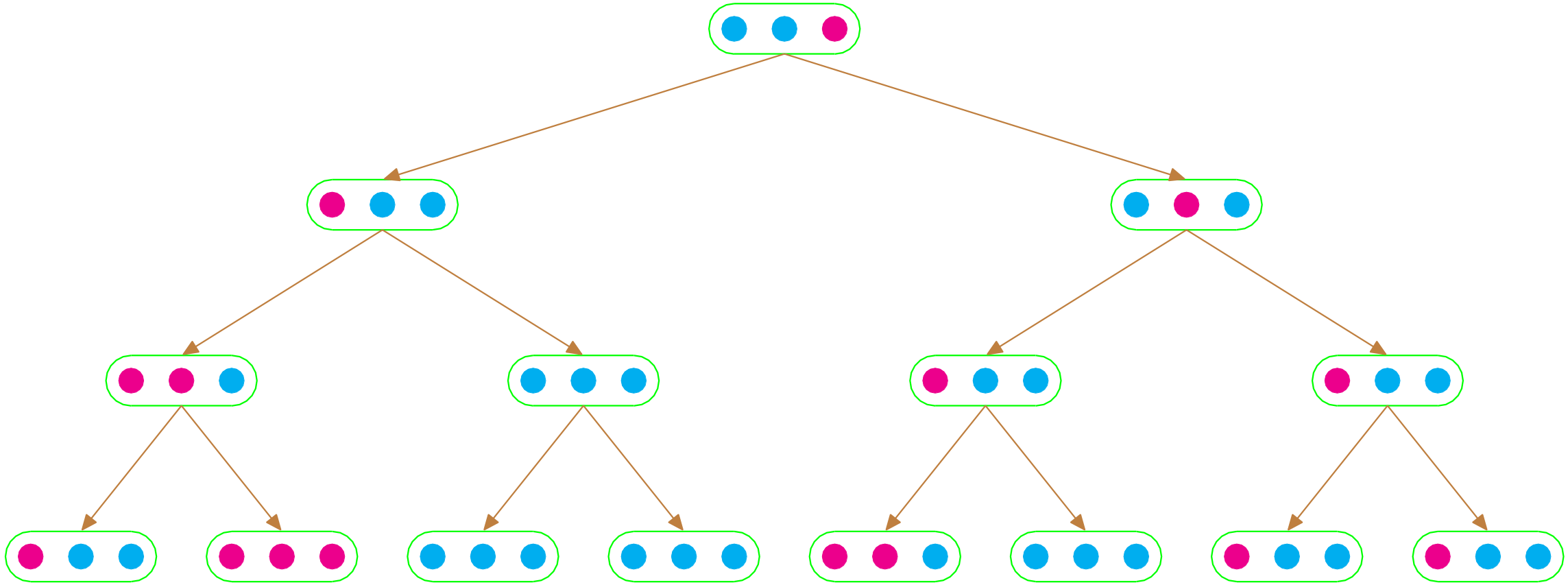
$c = 3$



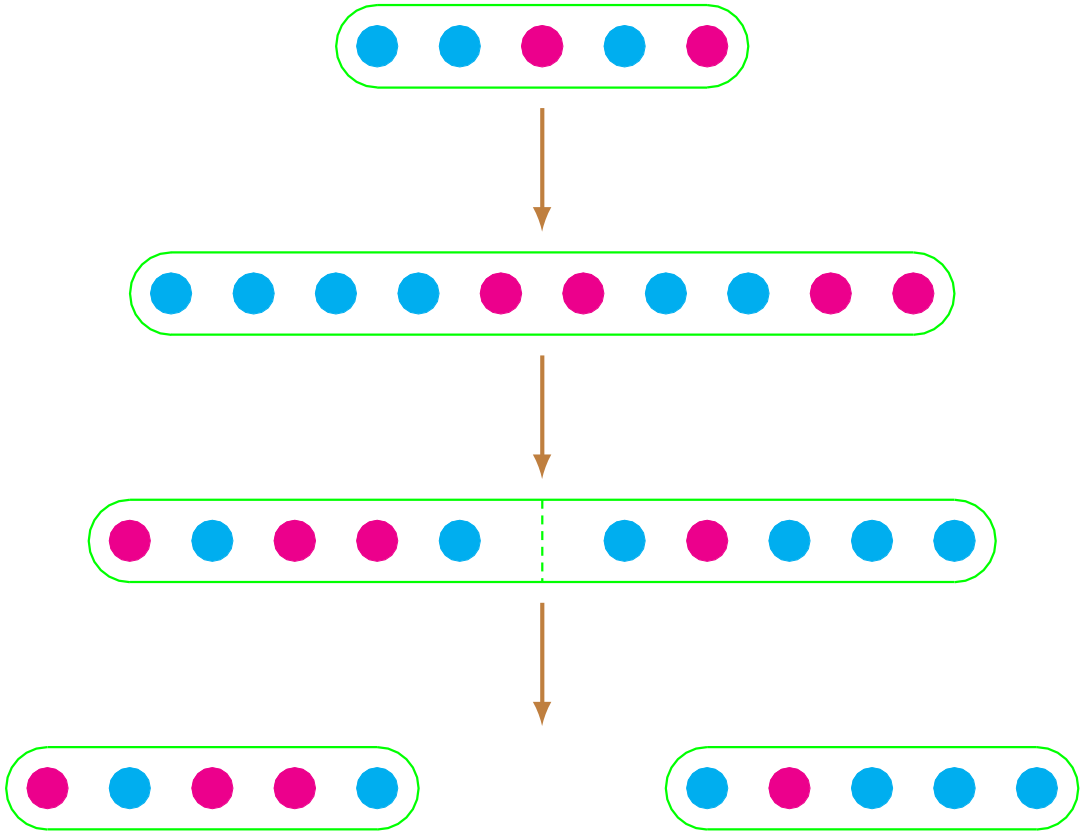
$c = 6$ with a mutant



It's conceptually simple, but ...



Does the random segregation hypothesis make sense?



- It agrees with our intuition.
- But it seemingly contradicts an observation.

Waiting for homozygosity

- a cell is said to be homozygous if all the genes are mutant.
- Let L_c be the average waiting time for homozygosity

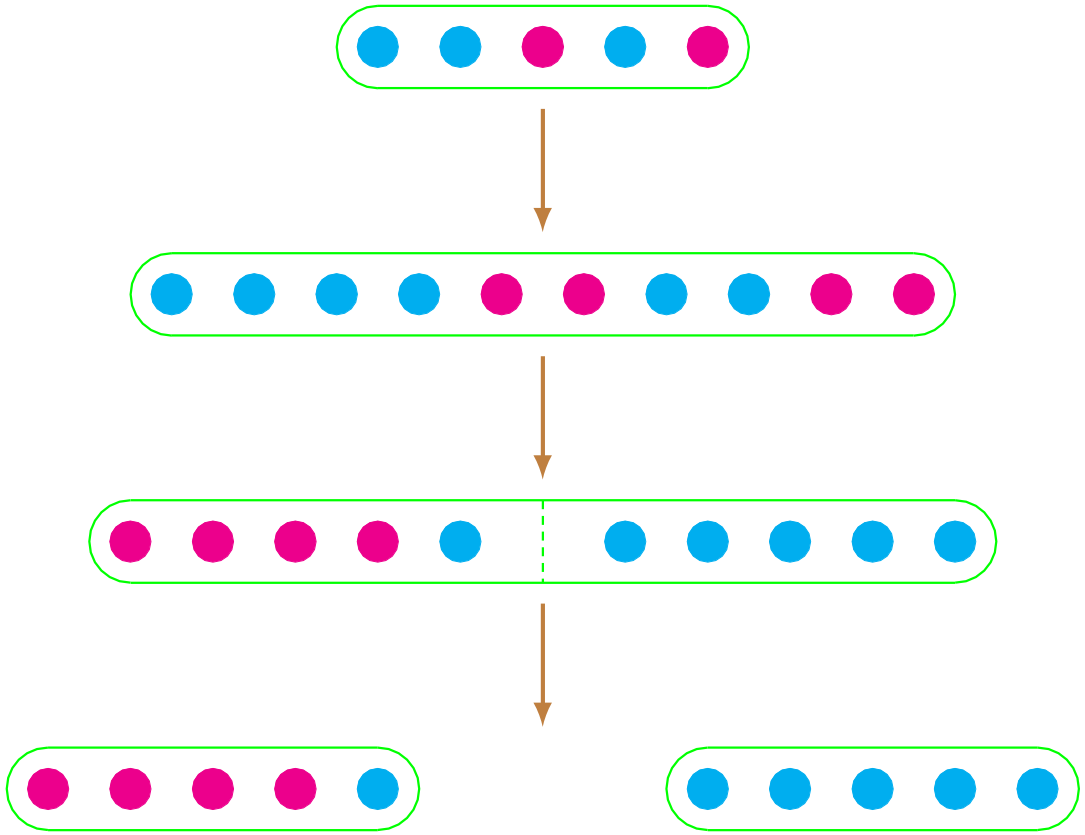
Ploidy (c)	Min	Median	Max	Mean (L_c)
2	1	2	5	2.02
3	2	3	6	3.47
4	2	5	7	4.80
5	3	6	8	5.77
6	3	7	10	6.72
7	4	8	10	7.60
8	4	8	11	8.37
9	5	9	12	9.08
10	5	10	13	9.73

A big dilemma

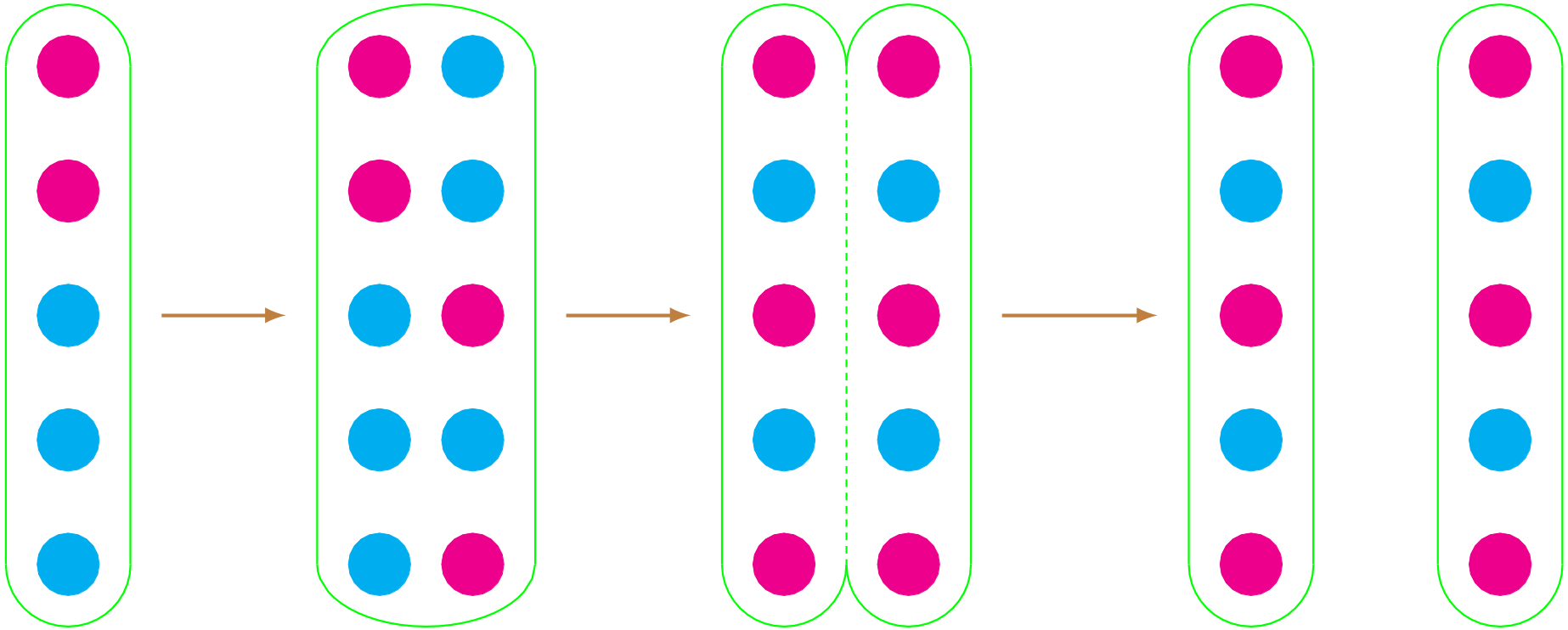
- Let $C = 100$ and assume $L_c = c$
- if a 100-ploidy cell contains just one mutant gene, will we see a homozygous cell?
- $2^{100} \approx 1.267 \times 10^{30}$
- according to Whitman et el. (1998, *PNAS*), the number of prokaryotes on earth is $4 \sim 6 \times 10^{30}$ cells.

What is nonrandom chromosome segregation?

- according to Lee & Haughn (1980, ***Genetics***), waiting time is reduced to $L_C = \log_2 C$ ($\log_2 100 \approx 7$).
- Is this biological plausible?



Gene conversion is another mechanism



- if conversion is unbiased, little can be accomplished
- if conversion is biased, what directs the conversion?

Why simulation?

- mathematically, it is a $(c + 1)$-type branching process
- even for $c = 3$, little is known mathematically about a 4-type branching process
- with $c = 100$, one faces intractable mathematics
- In simulation, each cell can be represented as an "agent" so the precise dynamics can be examined

Preliminary, important insights

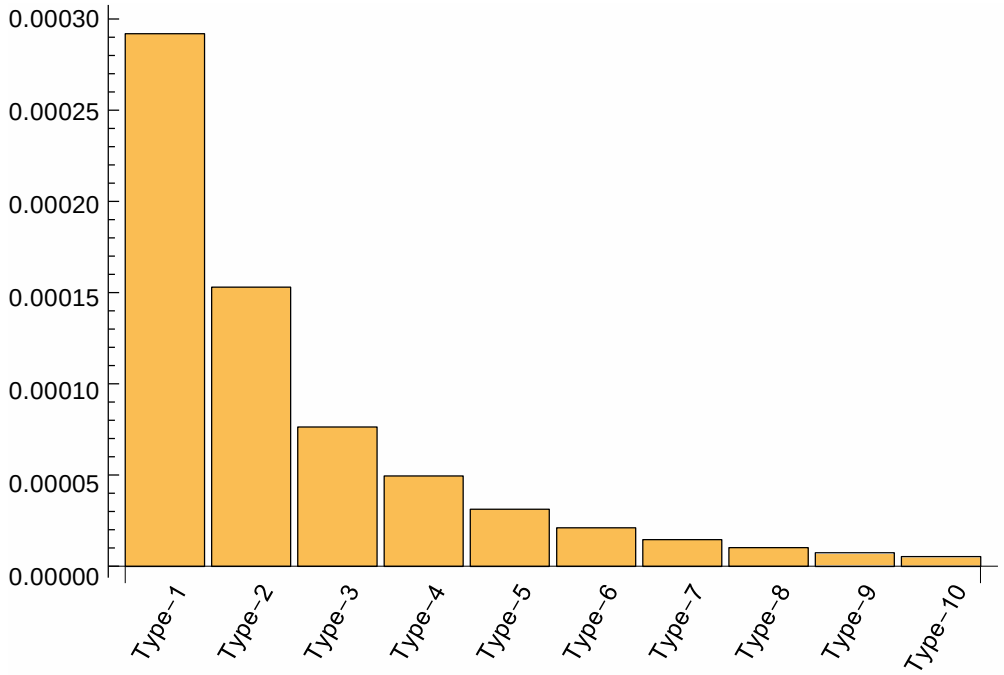
- definition: a type-k cell contains k mutant genes
- let a type-0, 100-ploidy cell multiply for 20 generations, with a mutation rate $p = 1.0 \times 10^{-6}$
- in the 20th generation, the population size is $2^{20} =$ 1048576
- there are 17 type-5 cells in one simulated scenario

1048139, 240, 98, 38, 30, 17, 4, 3, 2, 2, 3, 0, 0, 0, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0

Two contrasting observations

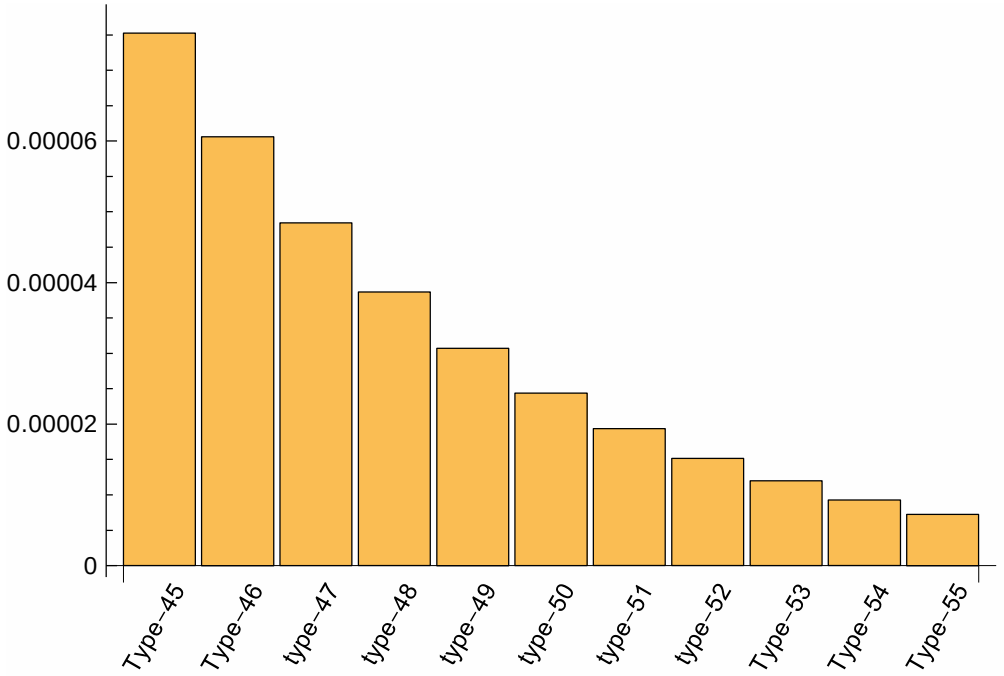
- 5-type cells are in the minority in a population of about 1 million cells (about 0.003%)
- but they are almost certain to occur (with a probability of 0.9996, based on 5000 simulations)



Now start with a type-5 cell

- 4998 out of 5000 simulations include at least one 50-type cell
- there are 56 type-50 cells in one simulated scenario

473176, 41929, 51289, 47724, 45247, 41648, 38118, 35007, 31348,
28190, 25378, 22968, 20245, 18079, 16185, 14454, 12687, 11320, 9730,
8553, 7532, 6707, 5768, 4910, 4284, 3731, 3237, 2817, 2478, 2158,
1858, 1516, 1345, 1137, 995, 758, 718, 558, 499, 410, 360, 280, 246,
184, 170, 124, 108, 92, 64, 58, 56, 32, 25, 12, 17, 16, 11, 9, 4, 6,
5, 2, 0, 1, 0, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0

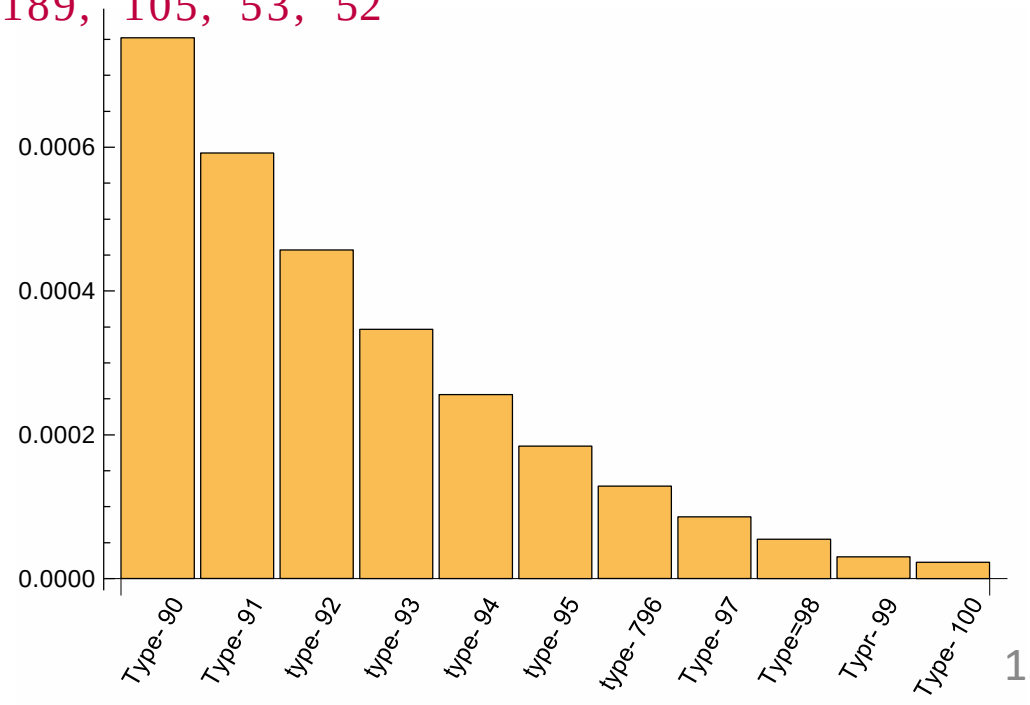


Now a type-50 cell does the magic work

- there are 52 type-100 cells in one simulated scenario

3, 4, 21, 36, 66, 92, 173, 256, 335, 465, 649, 799, 1054, 1359,
1647, 2052, 2497, 2873, 3605, 4142, 4812, 5458, 6424, 7116, 8029,
8843, 9751, 10735, 11839, 12705, 13872, 14793, 15656, 16552, 17658,
18467, 19472, 19900, 20877, 21256, 21946, 22602, 22996, 23507, 23704,
24079, 24426, 24517, 24596, 24350, 24167, 24131, 23785, 23150, 23059,
22713, 22251, 21916, 21143, 20772, 19997, 19438, 18572, 18292, 17333,
16822, 15826, 15031, 14156, 13559, 12672, 11992, 11446, 10407, 9621,

8966, 7914, 7435, 6604, 5970, 5439, 4780, 4246, 3702, 3280, 2806,
2326, 1929, 1647, 1451, 1100, 956, 743, 559, 424, 335, 248, 189, 105, 53, 52



Contribution to theoretical biology

- random segregation is still a viable model, although it has fallen into disfavor
- selection plays an inseparable role, which we should not disavow

Computing details

- simulation was coded in the R language
- each scenario required about 80 hours (wall time)
- hundreds of scenario were simulated to gain useful insight
- memory was the bottleneck – simulation stopped after 20 cell generations due to memory constraint