

The background of the slide is an aerial photograph of a river delta, likely the Nile Delta. It shows a dense, intricate network of dark blue water channels winding through a vast expanse of brown, textured land. The channels vary in width and form, creating a complex, organic pattern that fills the entire frame.

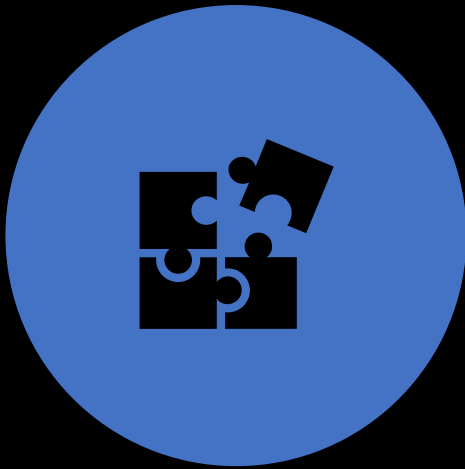
Genetic algorithms

Miloš Pivaš 12.2019.

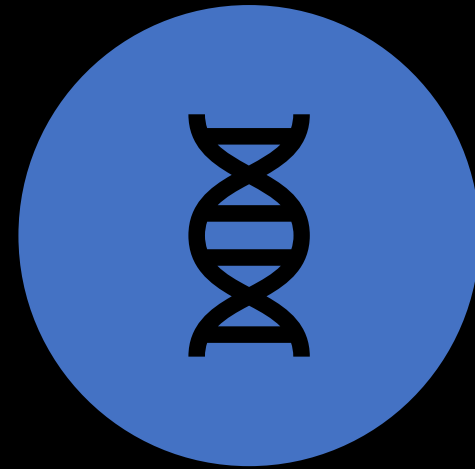
Problem

SEARCH/OPTIMIZATION IN HIGH
DIMENSIONAL SPACES

Idea



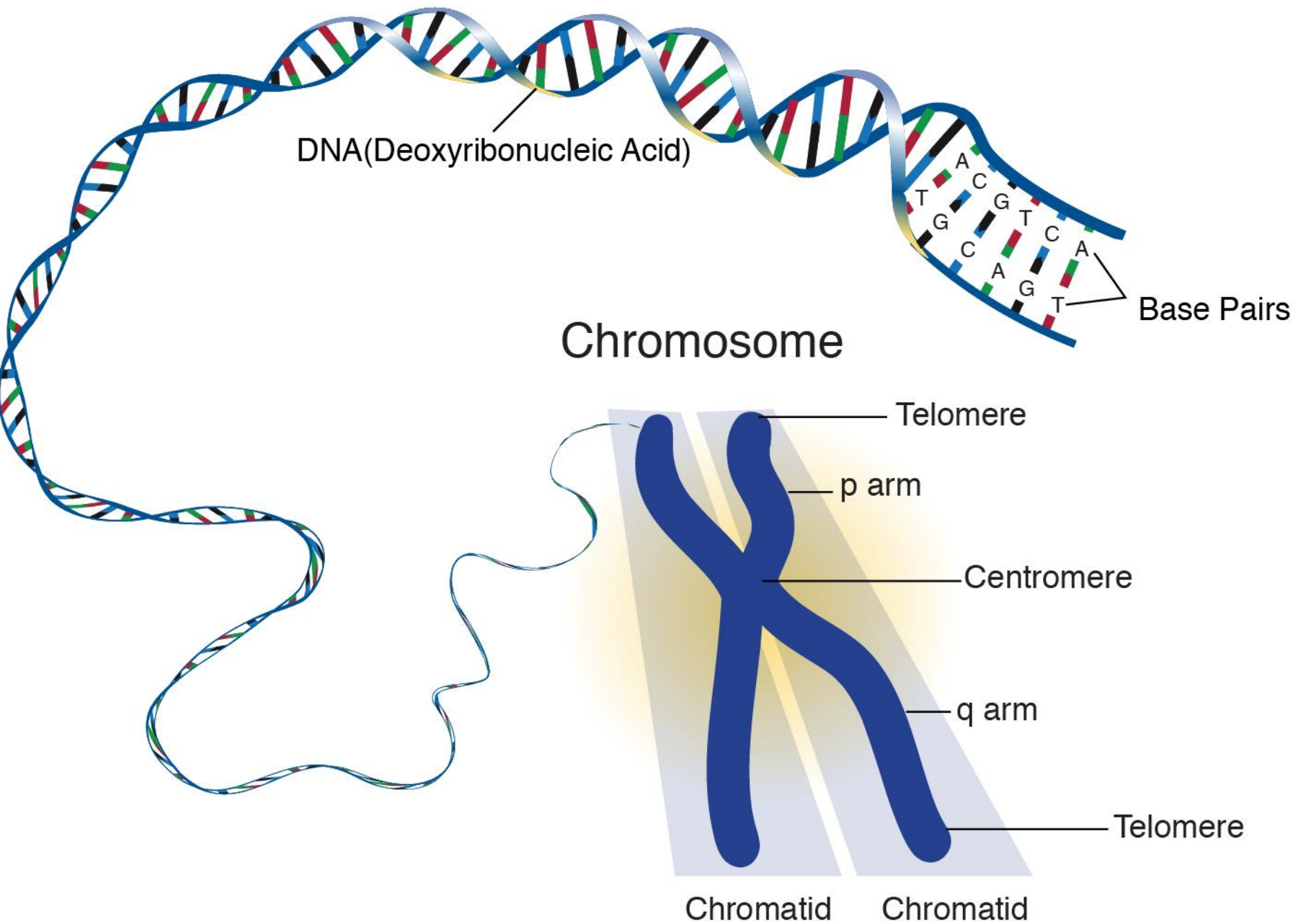
USE HEURISTICS BORROWED
FROM BIOLOGY



NAÏVE EVOLUTION, GENETICS,
SEXUAL REPRODUCTION

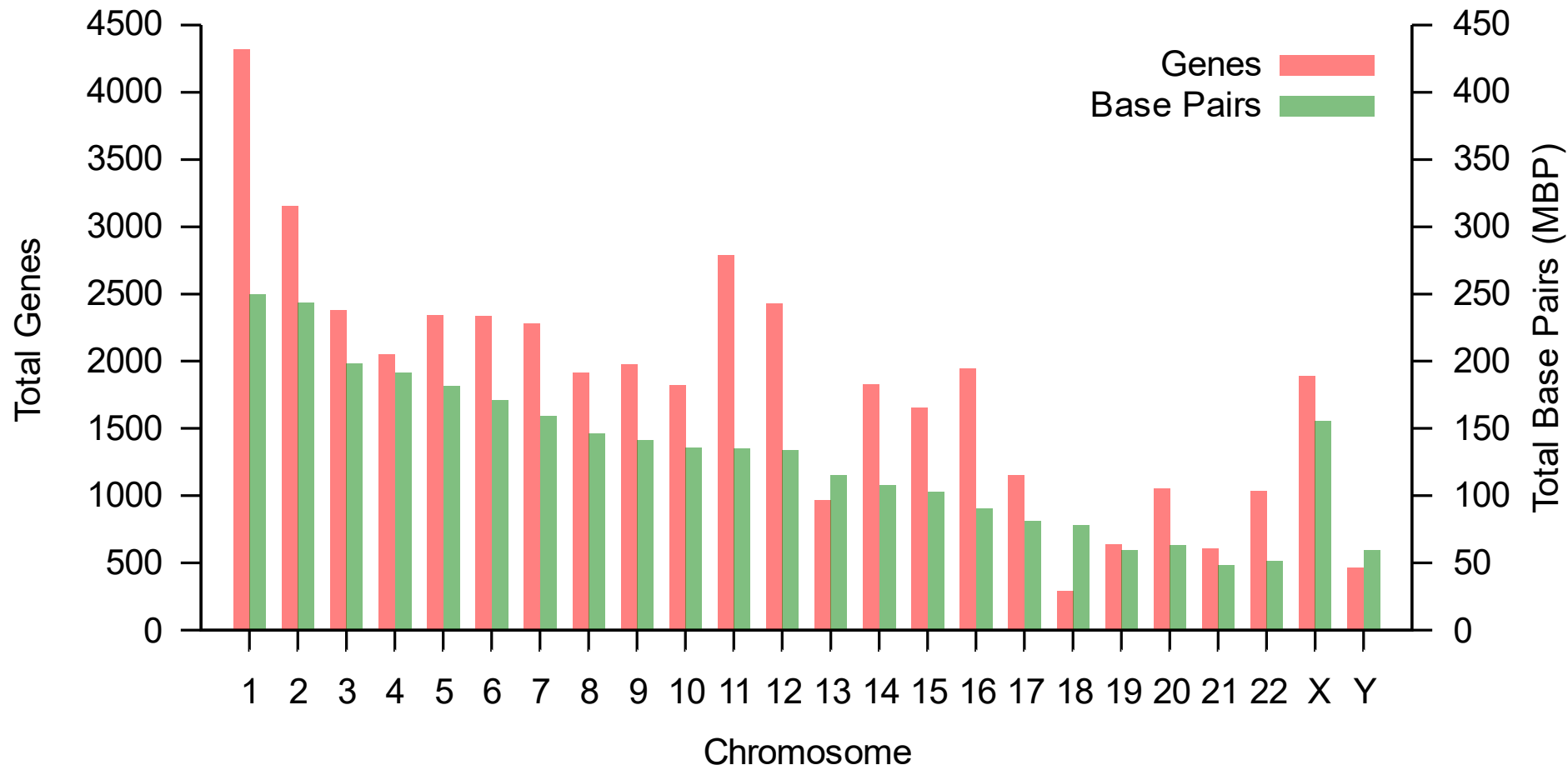


Some biology...



Human genome is huge

- Number of base pairs in one genome: $2 \cdot 3 \cdot 10^9 \approx 6\text{G}$
- Number of bits for coding one base pair: 2b
- Memory requirements for one genome: 12Gb = 1.5GB
- Dimensions of a chromosome are on the order of μm
- Current dimensions of a transistor (a bit): 5nm
- *We can put 1.5kb, in the same space nature puts 1.5GB*
- => A chromosome is a very good data storage device

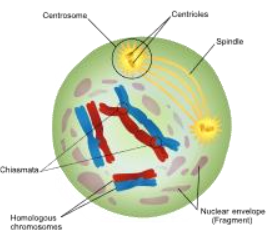




Mutations are a great source of random variation

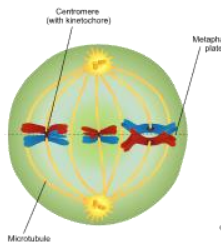
Sexual reproduction evolves better. So let's look at meiosis:

Prophase I



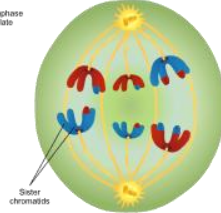
The chromosomes condense, and the nuclear envelope breaks down. Crossing-over occurs.

Metaphase I



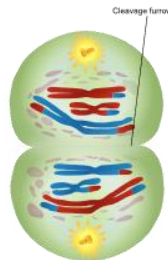
Pairs of homologous chromosomes move to the equator of the cell.

Anaphase I



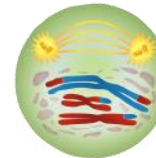
Homologous chromosomes move to the opposite poles of the cell.

Telophase I & cytokinesis

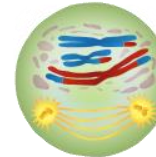


Chromosomes gather at the poles of the cells. The cytoplasm divides.

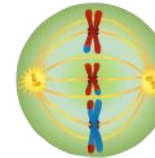
Prophase II



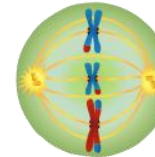
A new spindle forms around the chromosomes.



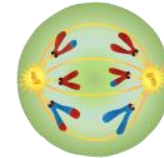
Metaphase II



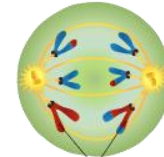
Metaphase II chromosomes line up at the equator.



Anaphase II

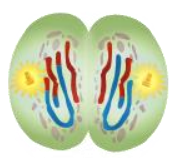


Centromeres divide. Chromatids move to the opposite poles of the cells.

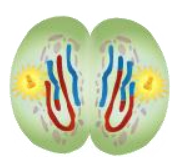


Sister chromatids separate

Telophase II & cytokinesis



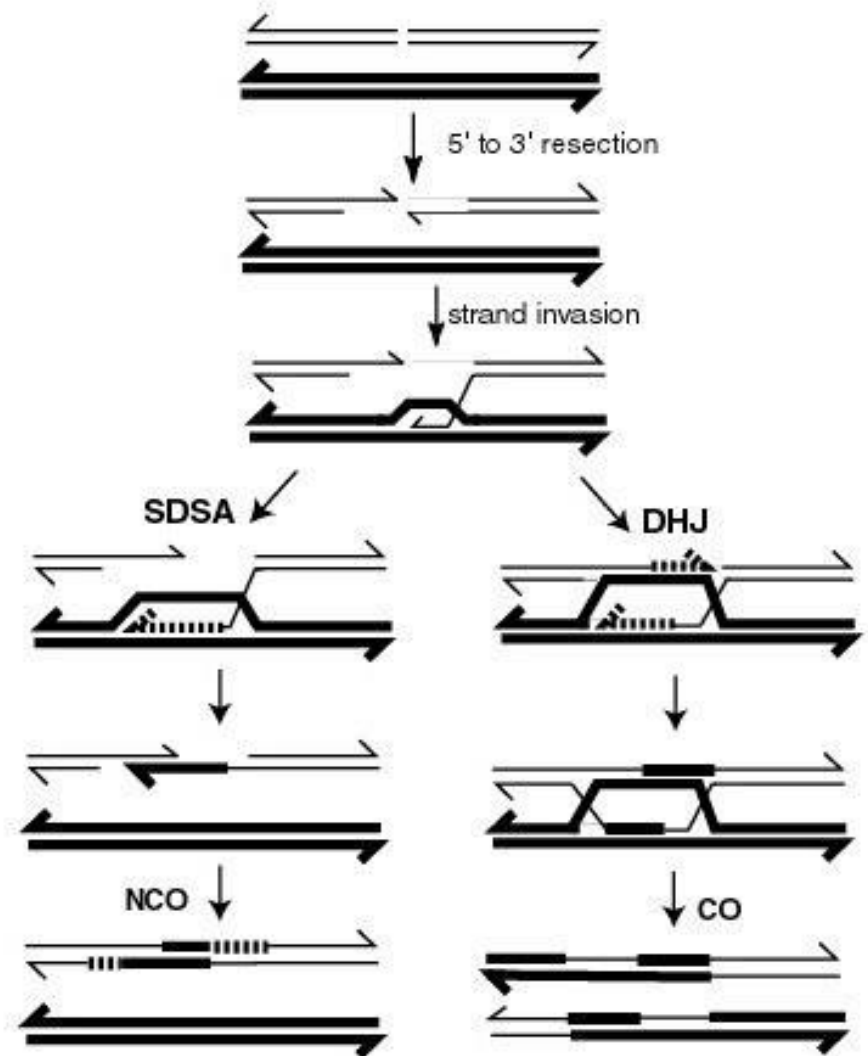
A nuclear envelope forms around each set of chromosomes. The cytoplasm divides.



Crossover is a key to the transfer of quality.

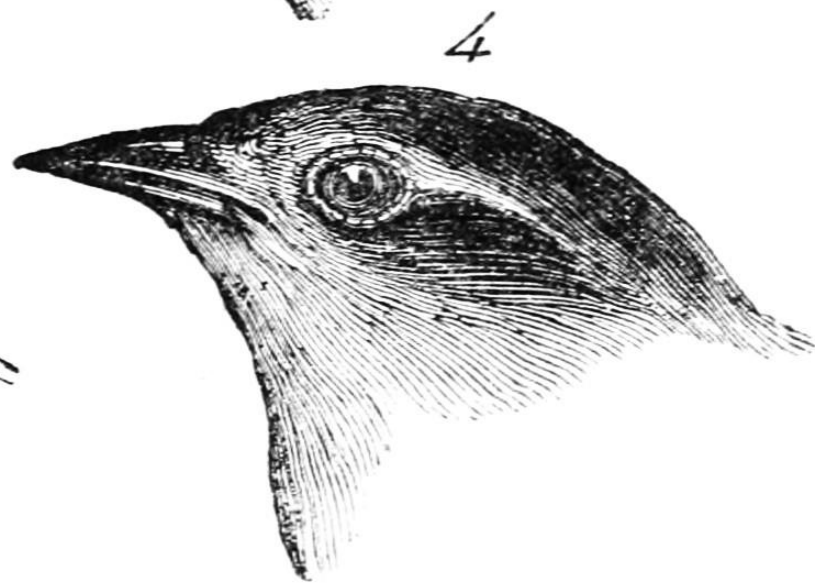
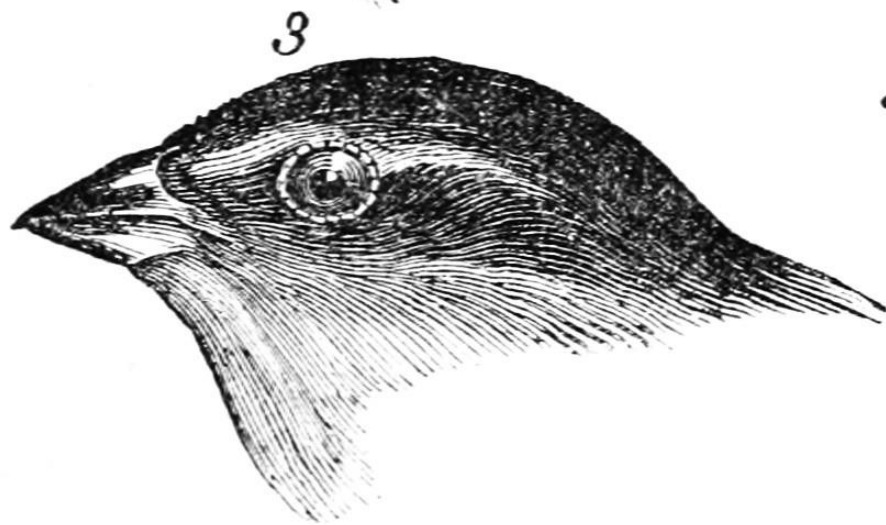
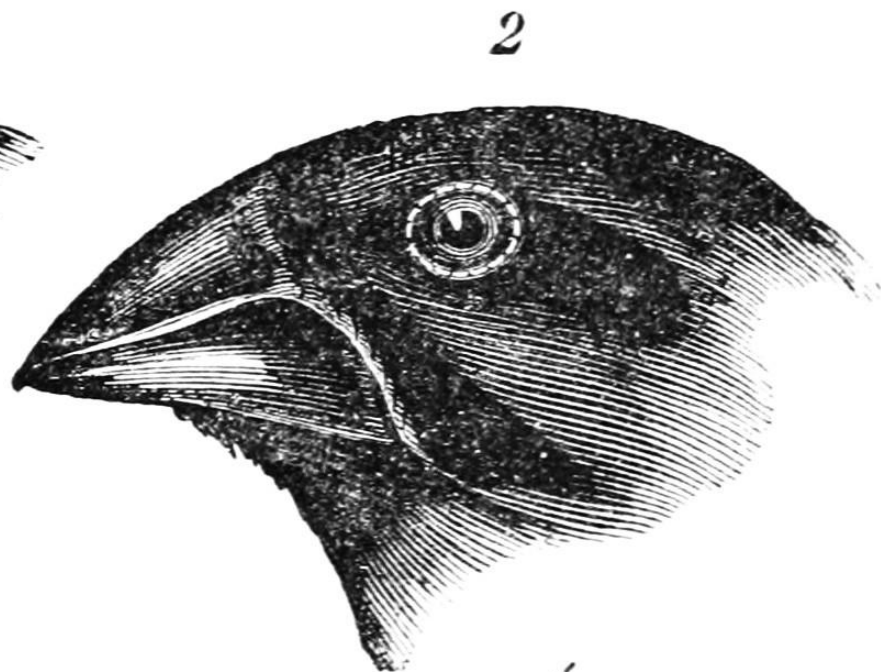
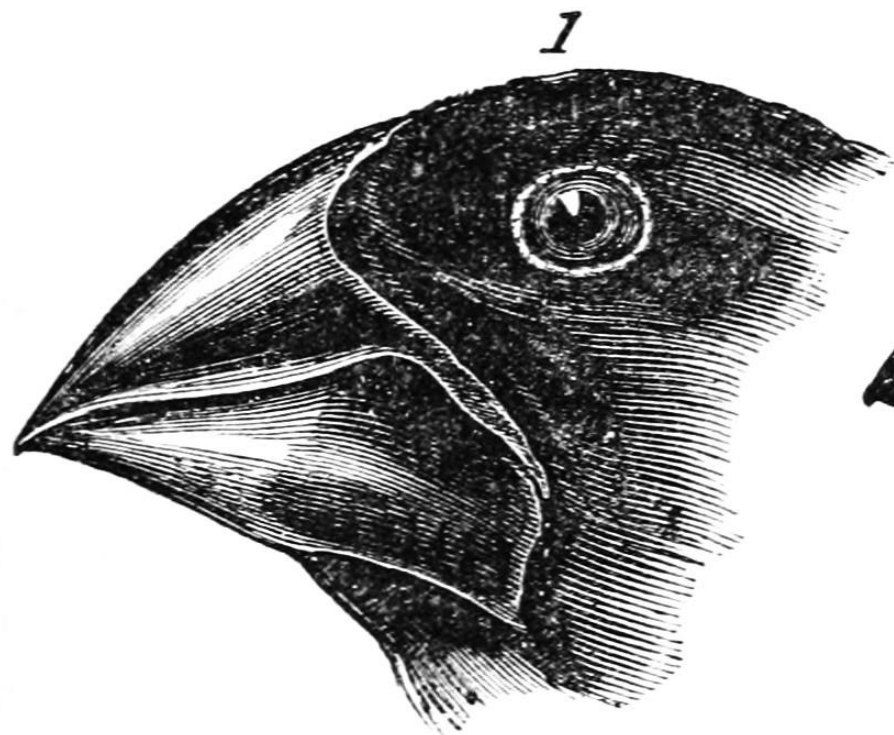


FIG. 64. Scheme to illustrate a method of crossing over of the chromosomes.



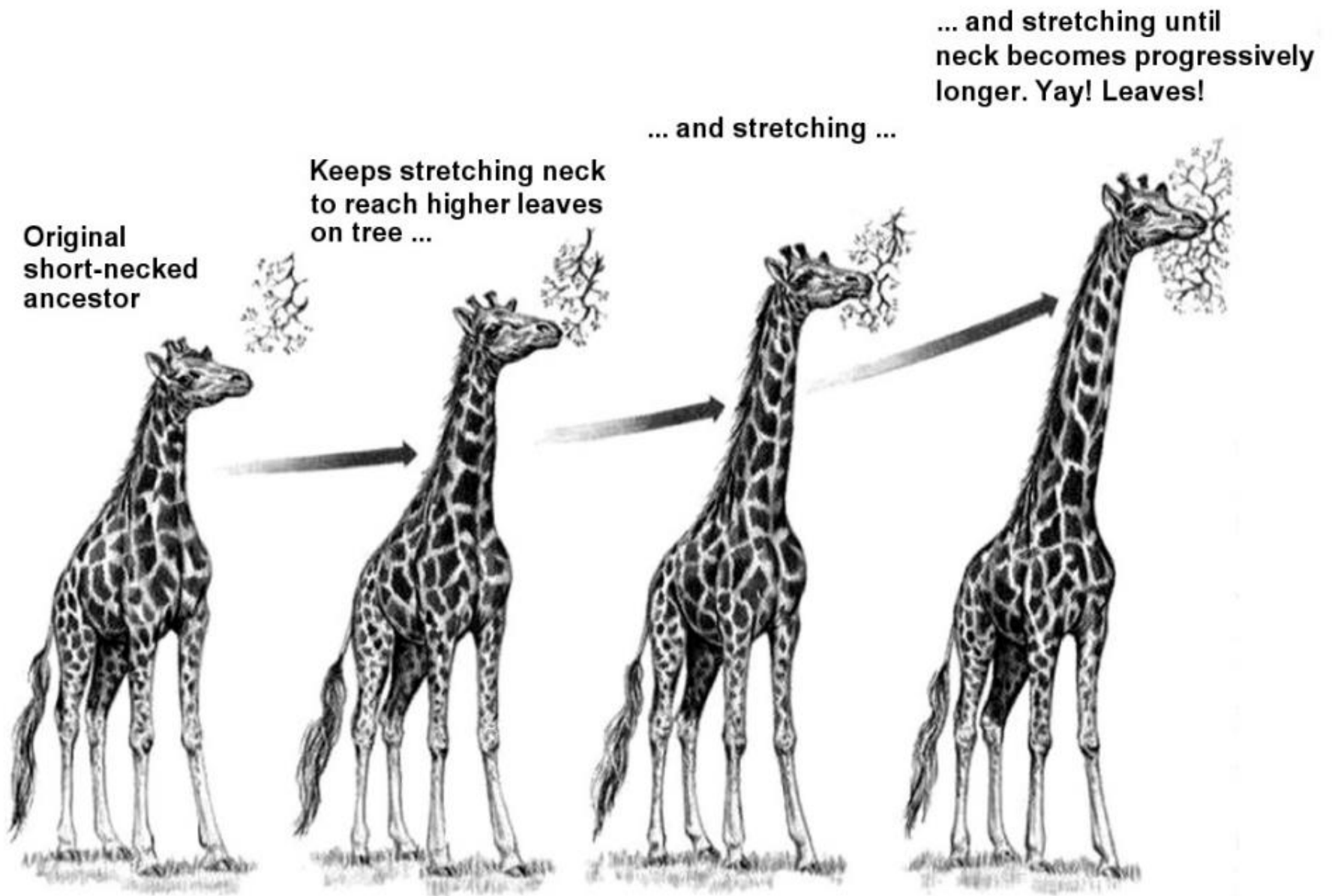


| Evolution is very complicated.



1. *Geospiza magnirostris*.
3. *Geospiza parvula*.

2. *Geospiza fortis*.
4. *Certhidea olivacea*.



Lamarck's Giraffe

How do we use
these heuristics?



Genome representation

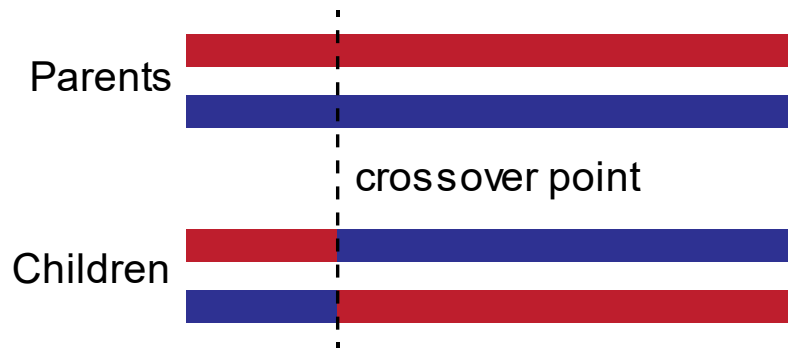
- Genome
 - Set of pairs of chromosomes -> one chromosome
- ATCG -> 01
- A chromosome is an array of variables that we wish to optimize

Mutations

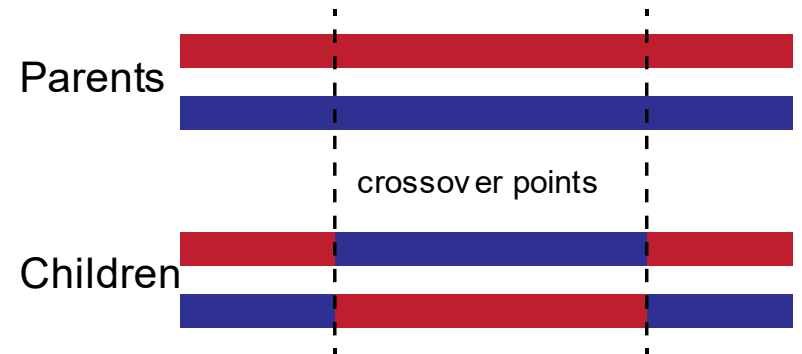
- Bit string mutations
- Flip bit mutations
- Boundary
- Uniform
- Gaussian
- Non-uniform

Crossover

Single-point



2-point, k-point....



Fitness function

- Depends on application
- Almost always non-linear, non-differentiable
- Usually it isn't even an explicit mathematical function, but a metric of some simulation or measurement...

How to turn fitness into probability of survival?

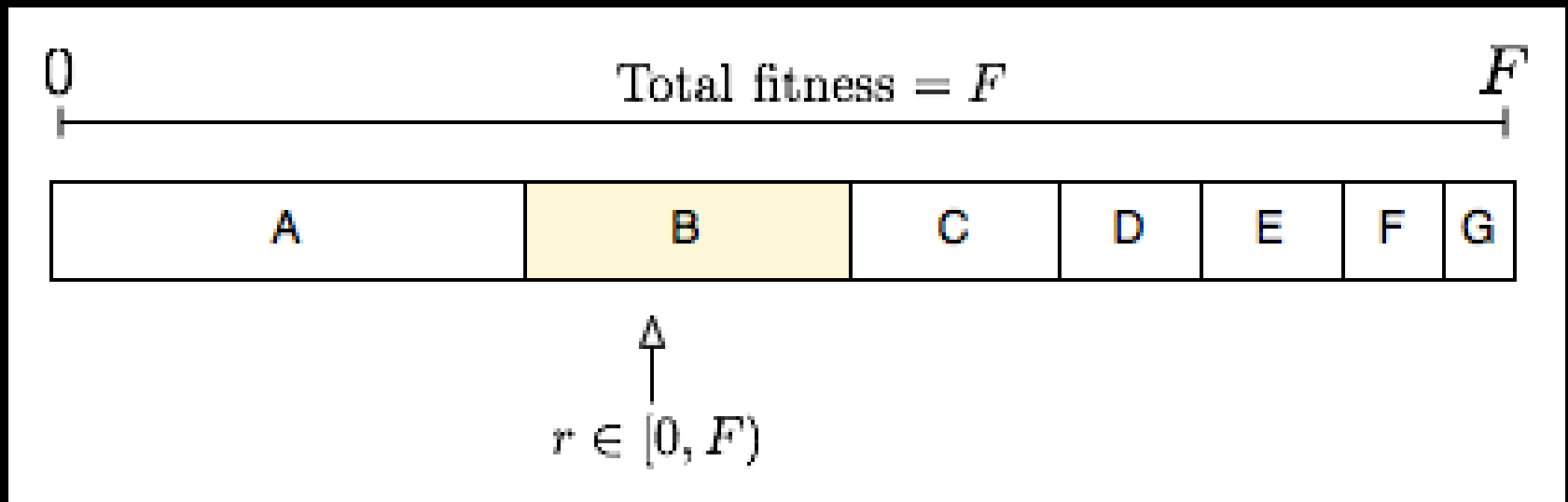
- Truncation selection
- Fitness proportionate selection (Roulette wheel selection)
- Stochastic universal sampling (*SUS*)
- Tournament selection
- **Rank-space selection*

Truncation selection

- Simple truncation of the worst individuals
- Only the best remain
- + Simple
- - Less sophisticated than other methods
- - Easiest to fall into local minima

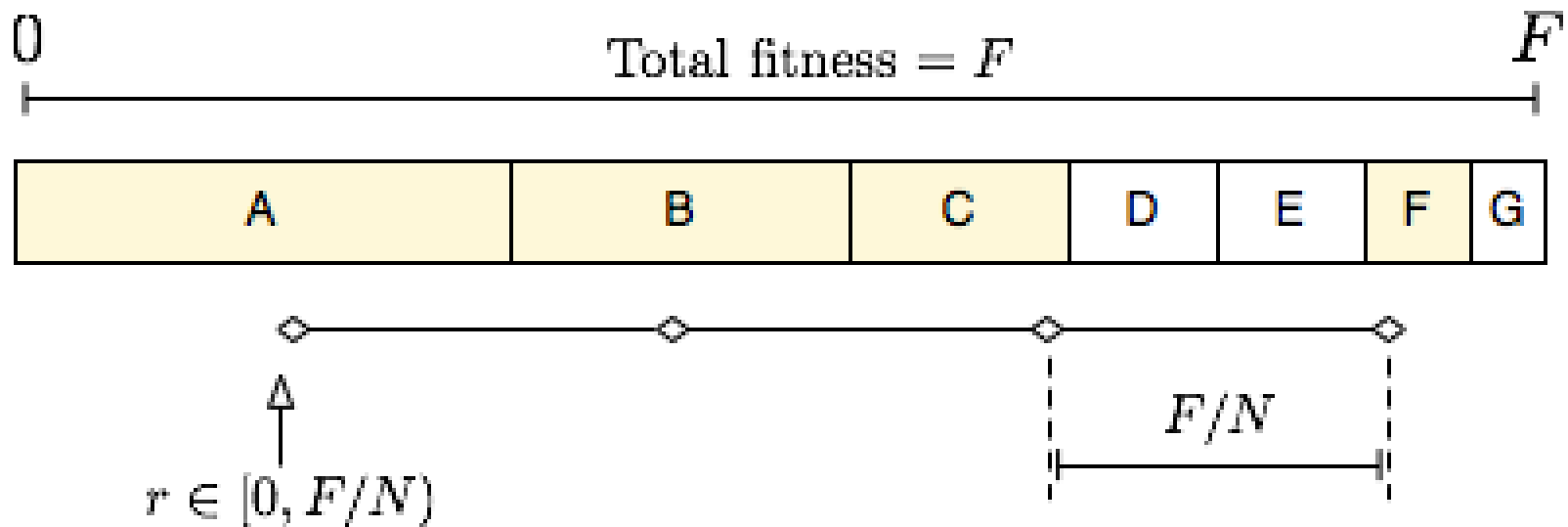
Fitness proportionate
selection (Roulette
wheel selection)

- Probability of survival is proportional to fitness
- $p_i = \frac{f_i}{\sum f_j}$
- Notice any flaws?



Stochastic universal sampling (SUS)

- Probability of survival is proportional to fitness
- But now we select all N individuals at once
- Flaws?



Tournament selection

- For the choosing one individual, a “tournament” is played:
 - Randomly choose K individuals
 - Sort by fitness
 - Go from best to worst and choose one with probability p , until chosen
- Result
 - Best individual is chosen with probability p
 - If the best isn't chosen, next one is chosen with probability $(1 - p)p$
 - Next one with probability $(1 - p)^2p$,
 - etc.
- For N individuals, the tournament is played N times
- + Fitness magnitude doesn't directly influence probability of selection
- + K and p can be tuned so that weaker individuals also have chances for survival in order to preserve diversity

*Rank-space selection

- First the best individual is chosen
- For every next one:
 - For every chromosome in population not already selected:
 - Calculate Euclidean distance (l_2 norm) d_{ij} between it and every already chosen chromosome
 - Sum of inverse square d values is used as a measure of **diversity**:
 - $$D_i = \frac{1}{\sum d_{i,j}^2}$$

*Rank-space selection

- Now we need to unify fitness and diversity into one metric based on which to apply selection:
 - Distance from best possible point: $(\max(f), \max(D))$
 - Some other methods?
(Scaling f and D to $[0,1]$, then F_1 score?)
- + Directly influences the diversity of the population
- - Computationally more expensive

Termination conditions

- Limited number of generations
- Limited execution time
- Change of maximum fitness value in the population less than given

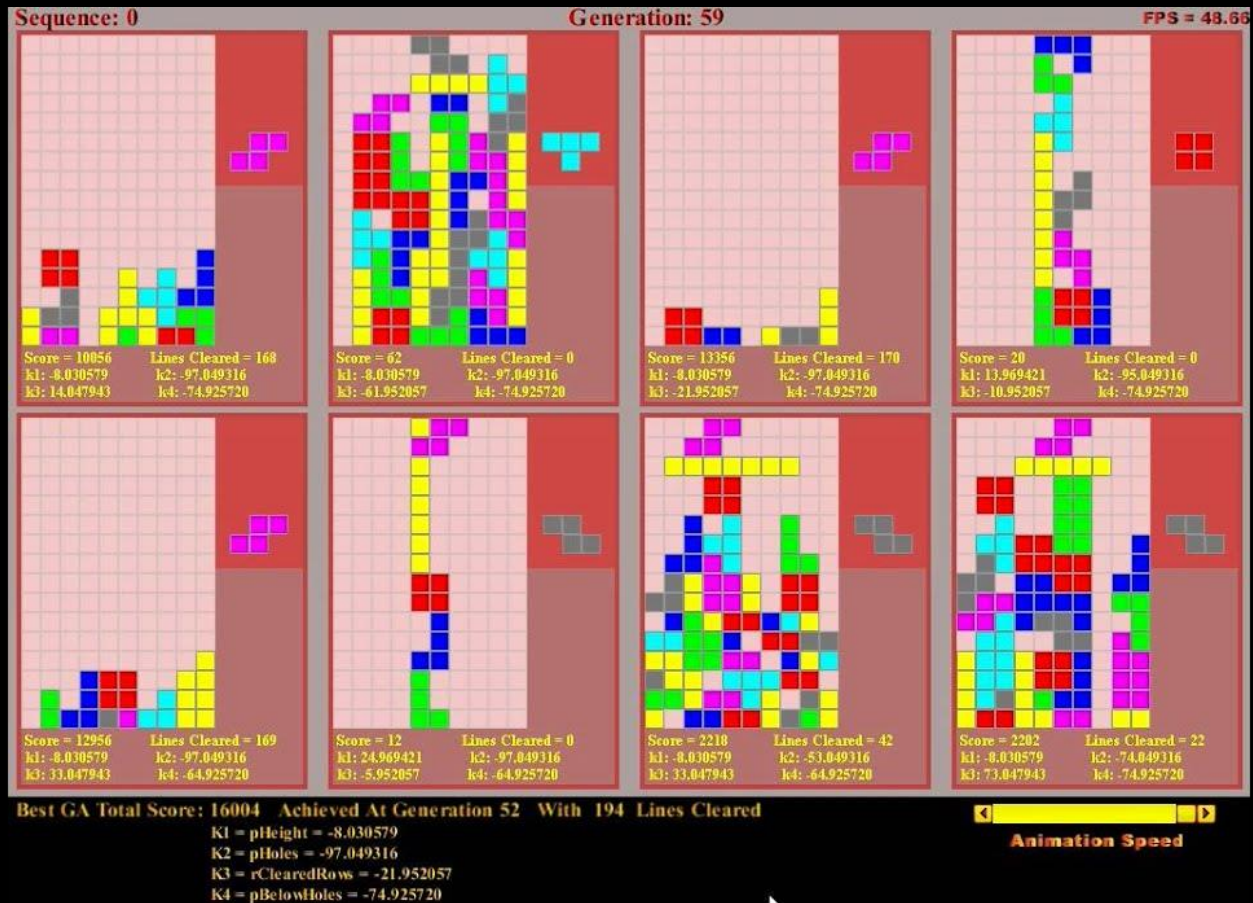
Some applications



Evolved
hardware



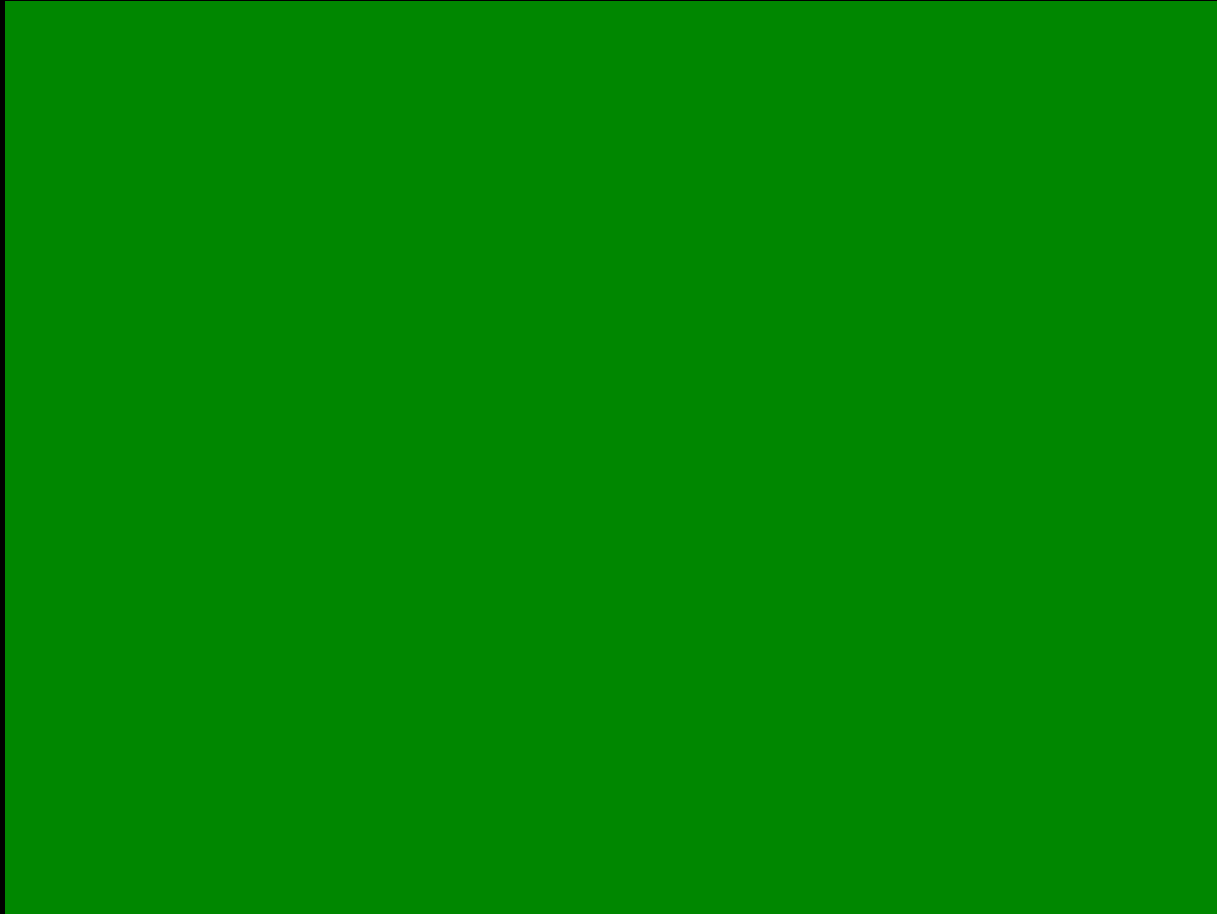
Evoluirana AI



Time (5:19)

Evolved Virtual Creatures

Karl Sims, 1994



Literature

- Predrag Janićić, Mladen Nikolić – Veštačka Inteligencija (2019. electronic edition)
- Patrick H. Winston - Artificial intelligence (1992. Addison-Wesley Pub. Co)
- <https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-034-artificial-intelligence-fall-2010/index.htm>