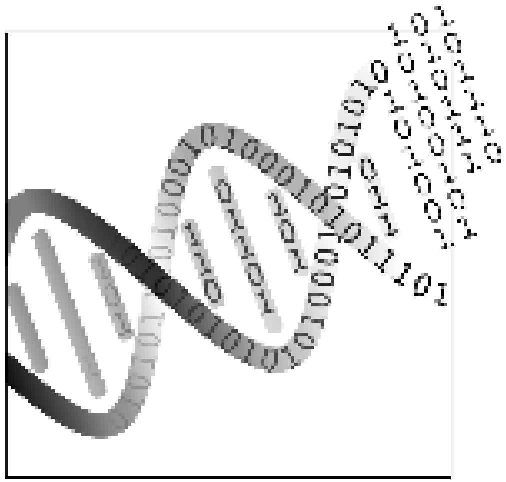
Evolutionary Computation
An overview



Companion slides for the book *Bio-Inspired Artificial Intelligence: Theories, Methods, and Technologies* by Dario Floreano and Claudio Mattiussi, MIT Press

What you will learn in this lecture

Evolutionary Computation in the Machine Learning landscape

Biological inspiration

Four pillars of natural and artificial evolution

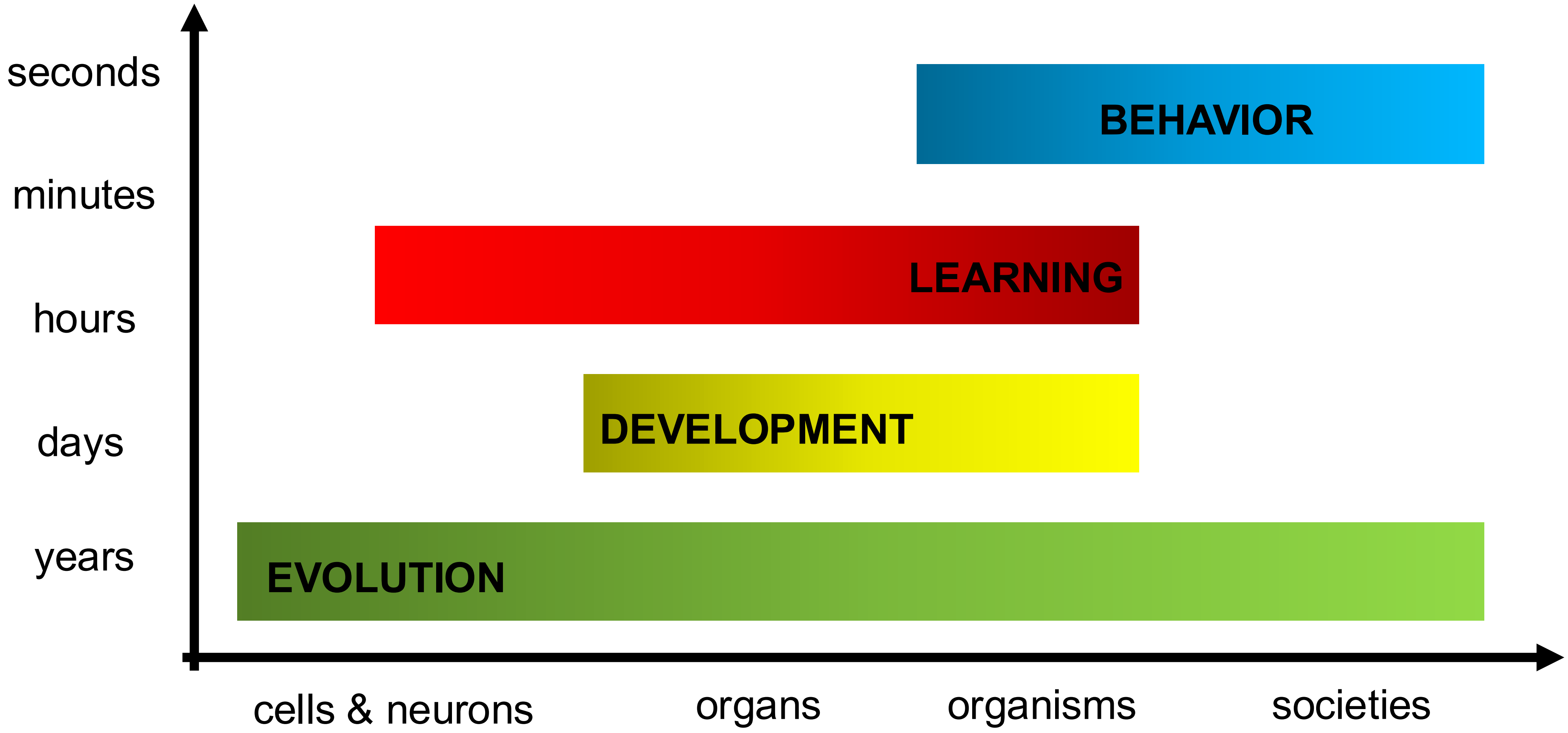
A simple Genetic Algorithm

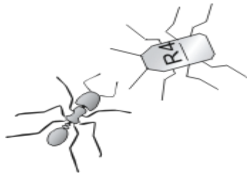
The fitness function

Function optimization with Genetic Algorithms

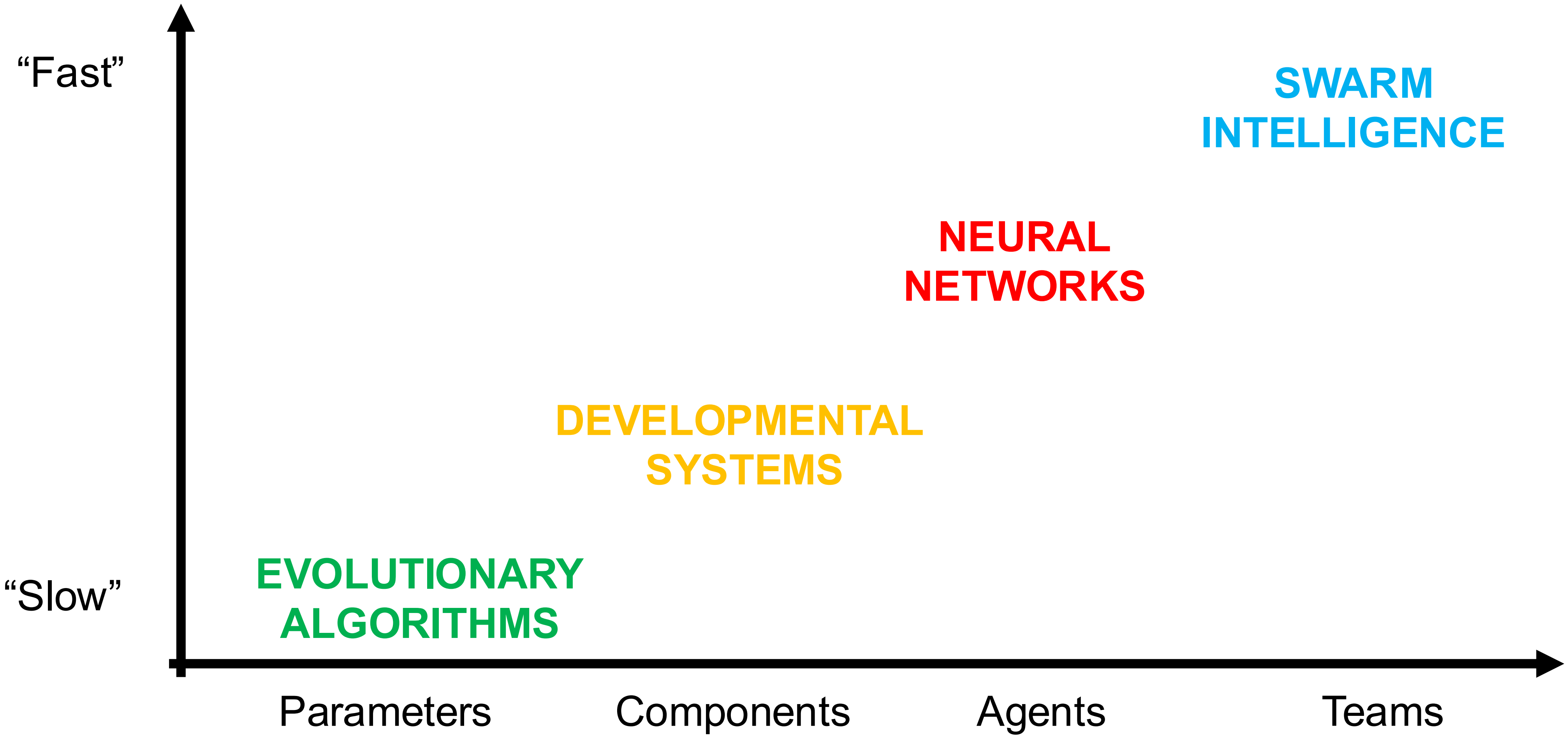
Assessing Evolutionary Runs

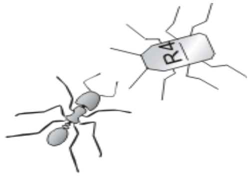
Spatio-Temporal Scales of BIOLOGICAL Adaptation





Spatio-Temporal Scales of Machine Learning

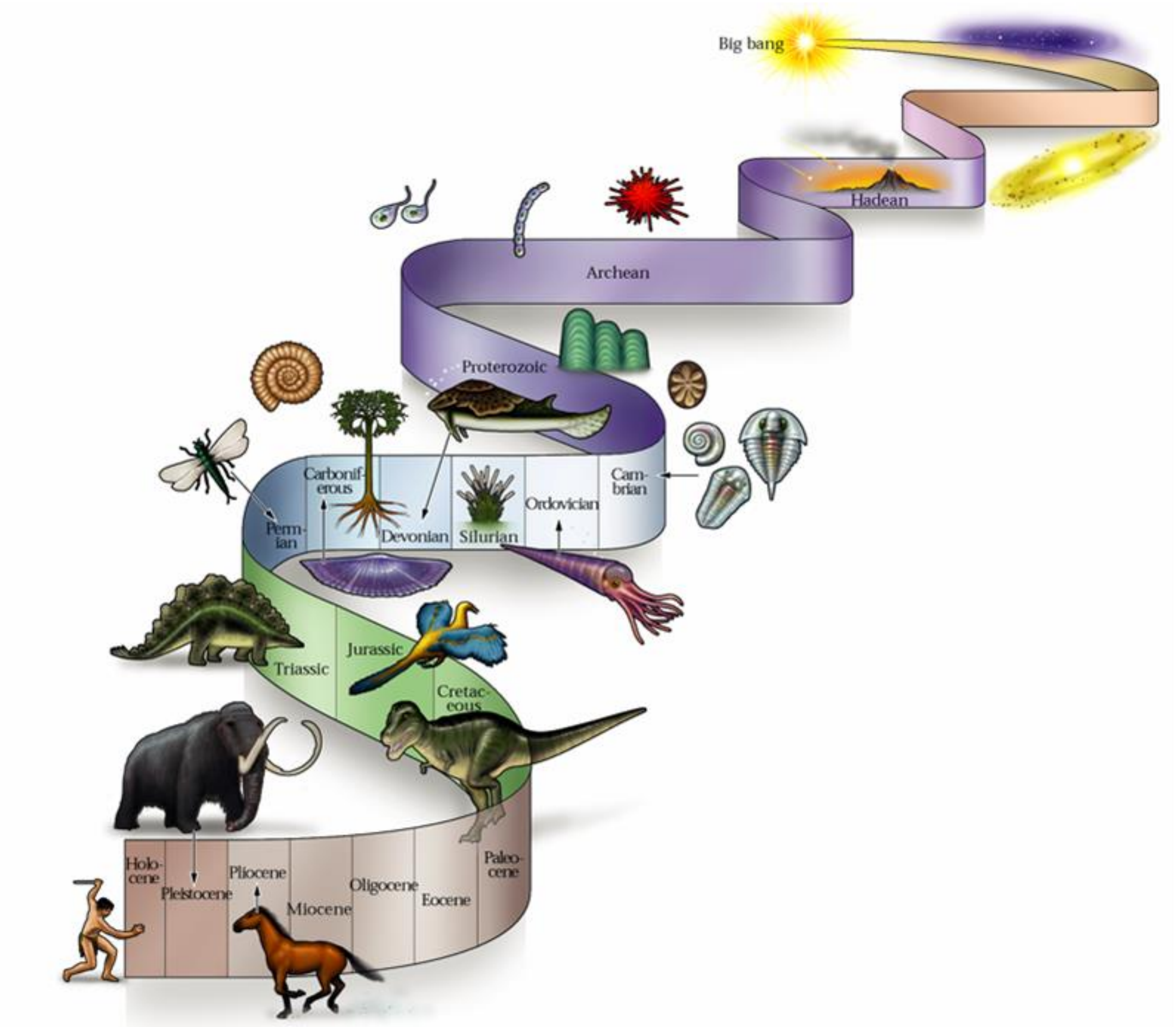


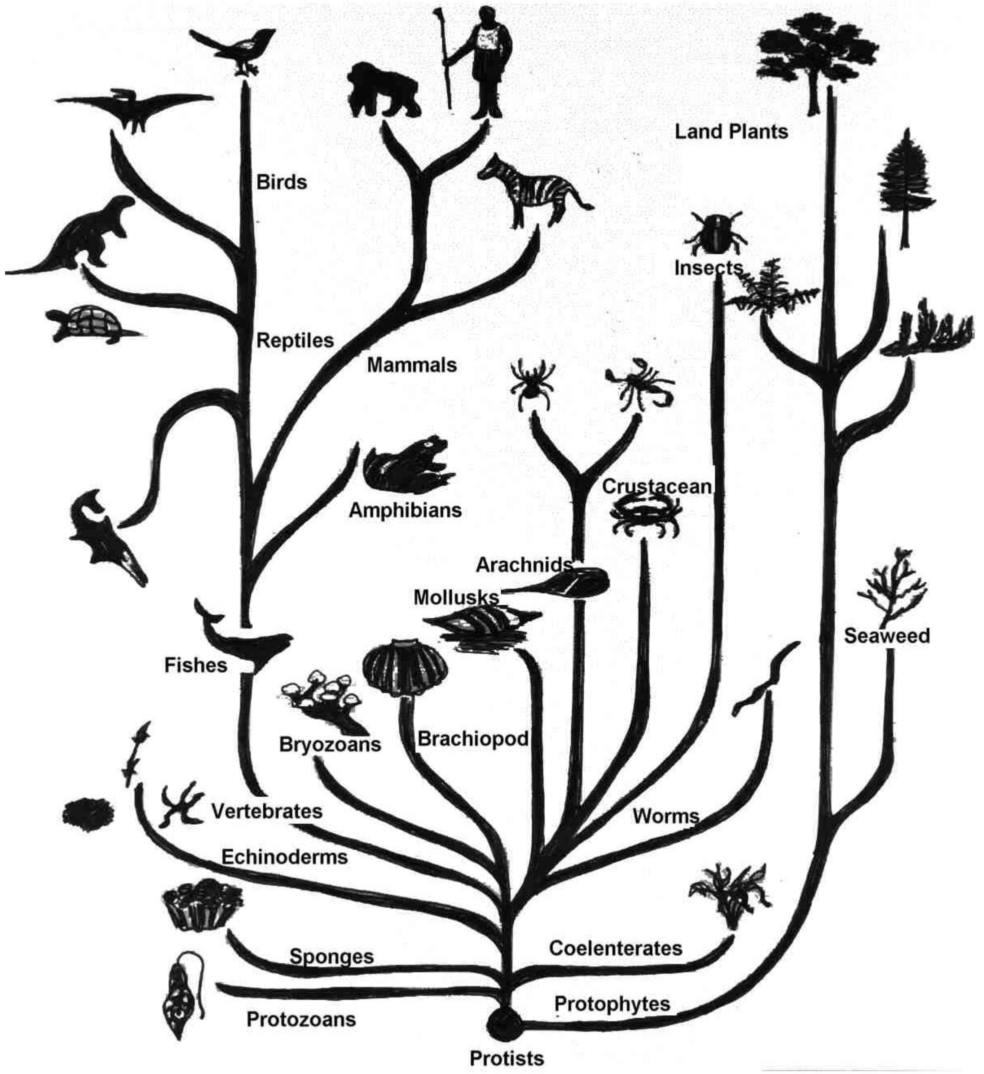


Popular Machine Learning Methods

- Convolutional Networks: Image recognition
  - Differentiable parameters, loss = teacher's output
- Diffusion Models: Image and video generation
  - Differentiable parameters, loss = teacher's output (noisy perturbations)
- Transformers: text recognition and generation
  - Differentiable parameters, loss = teacher's output (next token)
- Reinforcement Learning: problem solving
  - Differentiable parameters, loss = reward function
- Bayesian Optimization: problem solving
  - Black-box, builds probability distribution model of the function
- Evolutionary Algorithms: problem solving
  - Black-box, fitness function

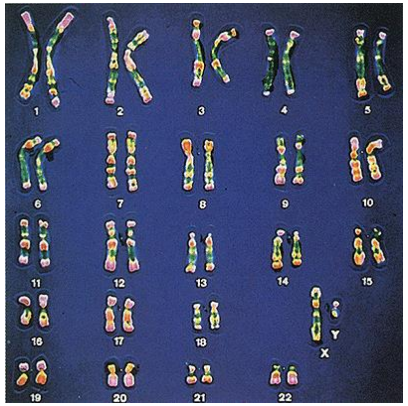
What is Evolution? 2 interpretations



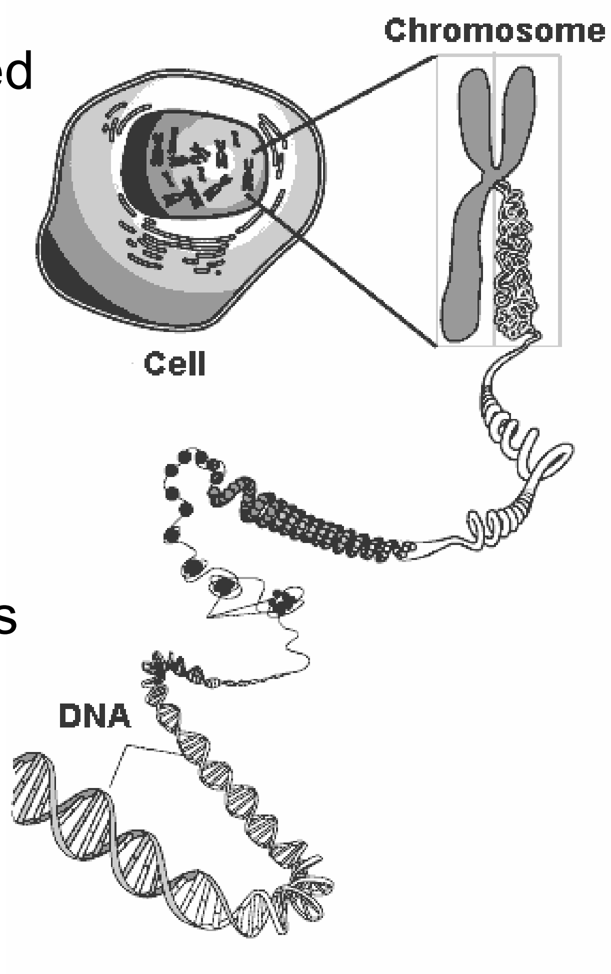


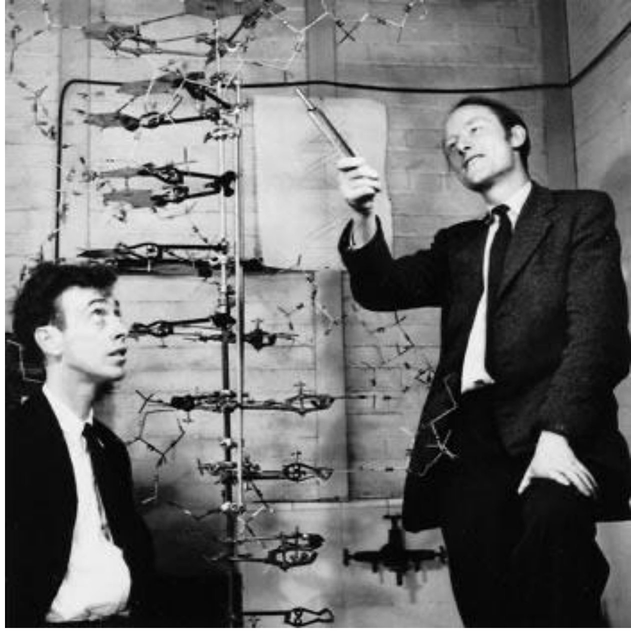
Genotype: DNA *(DeoxyriboNucleic Acid)*

Long molecule, twisted in spiral, and compressed



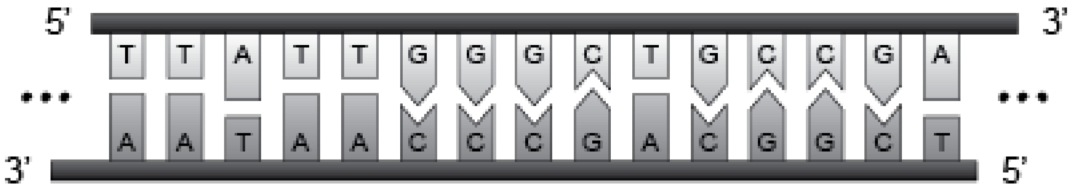
Humans have 23 pairs of DNA molecules (*chromosomes*)



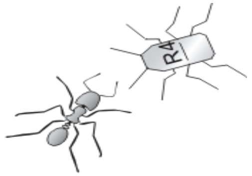


Crick 1953 Watson
Discovery of DNA structure

DNA is composed of 2 complementary sequences (*strands*) of 4 nucleotides (A, T, C, G), which bind together in pairs (A-T and C-G)



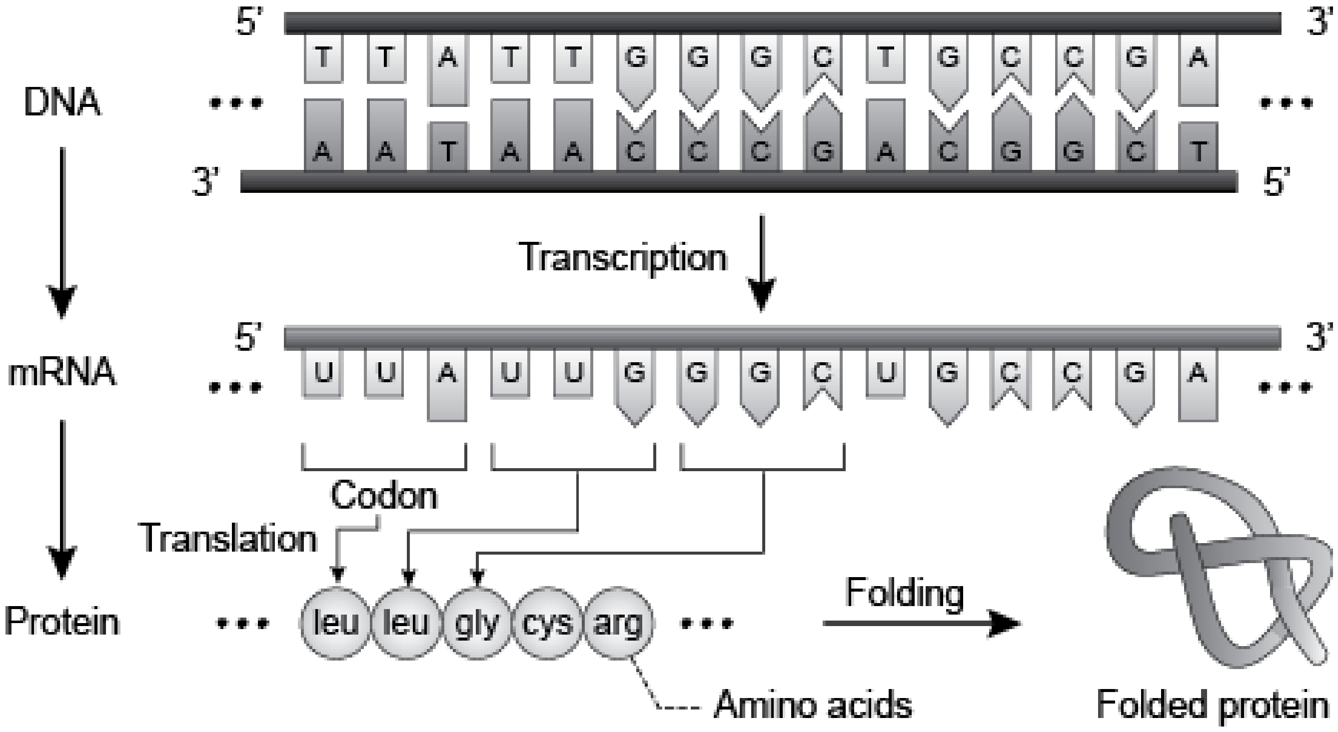
A **gene** is a sequence of several nucleotides that produce a protein



From Genotype to Phenotype: Gene Expression

Proteins are folded molecule chains whose shape define the type and function of cells (some proteins affect gene expression)

The sequence of nucleotides in one strand defines the type of protein. The expression of the gene into a protein is mediated by another molecule, known as messenger RNA.



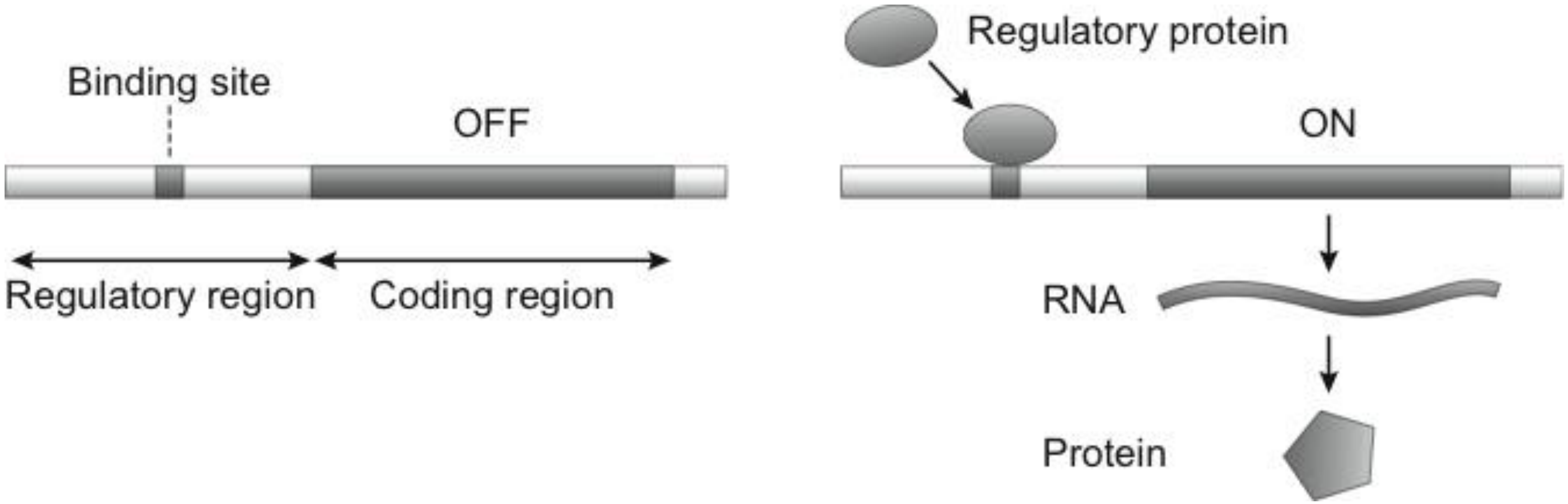
Google DeepMind

AlphaFold, 2021
Computer prediction of 3D protein folding

Regulation of Gene Expression

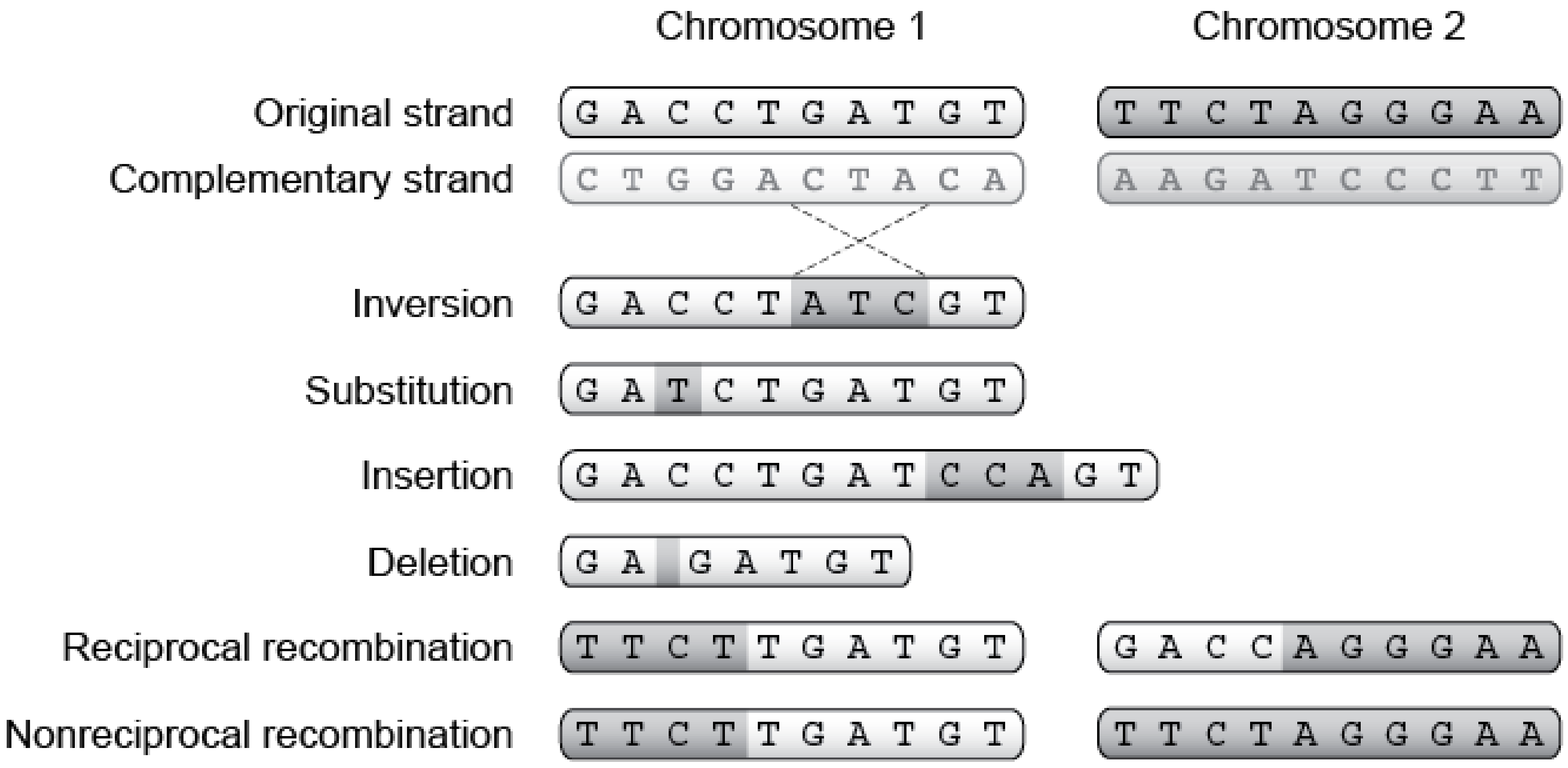
Genes are composed of a regulatory region and of a coding region.

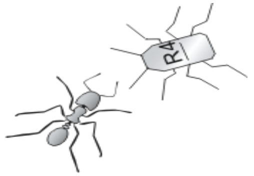
The coding region is translated into a protein if another protein binds onto the regulatory region.
Regulation can also be negative (i.e., inhibition of protein production).
Genome is a self-regulatory code



Genetic Mutations

- Genetic mutations occur during cell replication (4^{-10} per nucleotide per year)
- Those that occur in sex cells can affect evolution
- Sexual recombination is a mutation that affects two homologous chromosomes

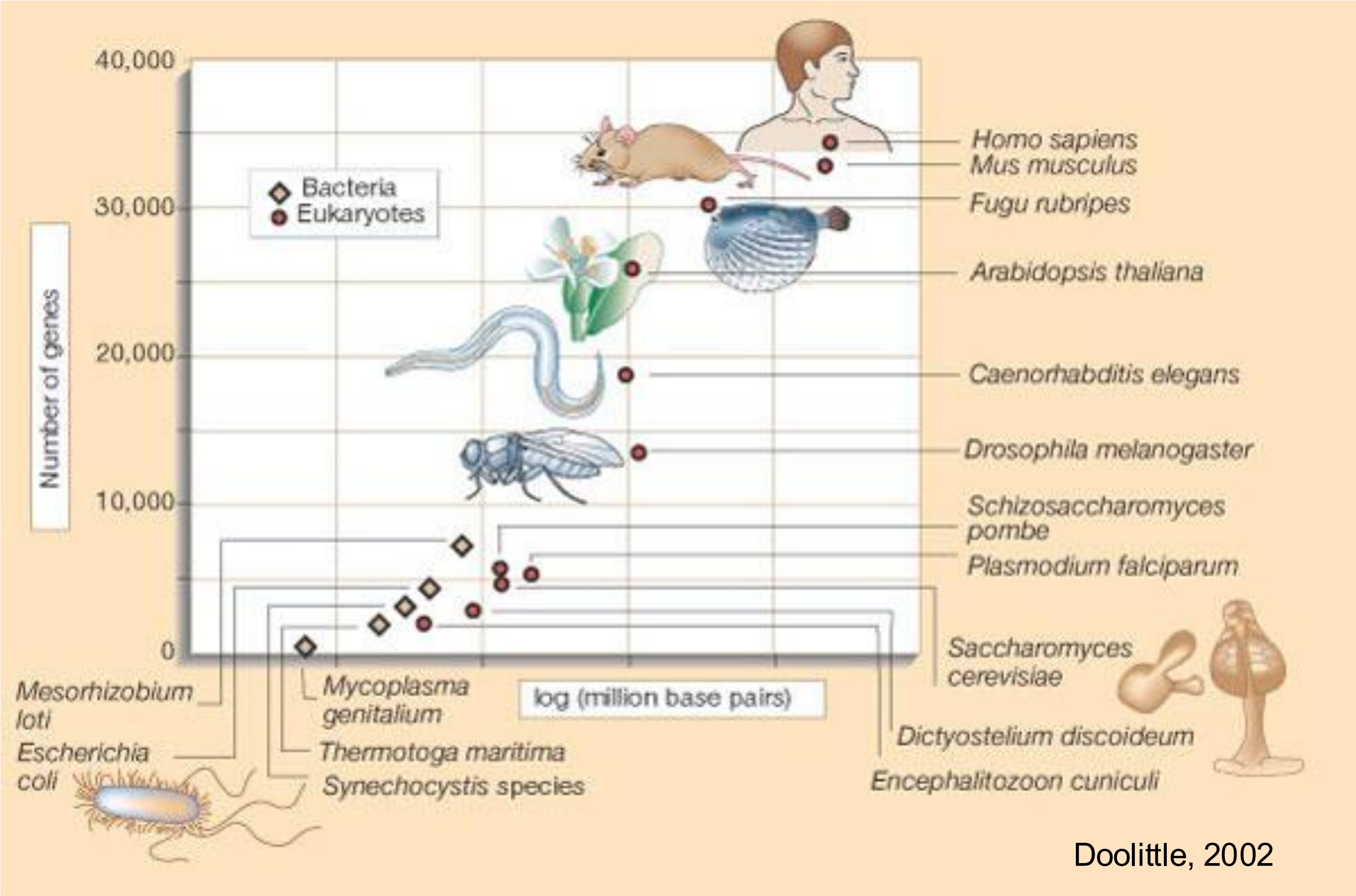




Genome Size

Genome size is equal for all individuals of a species, but it greatly varies across species

www.genomesize.com for comparisons



Doolittle, 2002

Genome includes:

- **Genic** DNA
- **Nongenic** DNA (pseudogenes)

Nongenic DNA **arises** from:

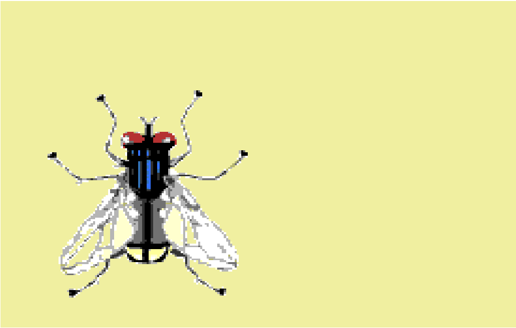
- insertion/deletion mutations
- gene duplication

Nongenic DNA can have **adaptive value**:

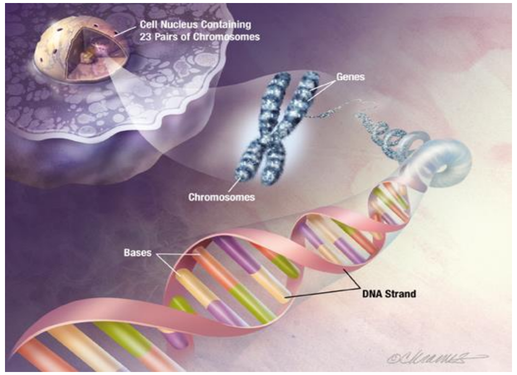
- pseudogenes may be re-activated
- pseudogenes may transform into new genes by several neutral mutations

Phenotype & Genotype

Phenotype
The organism (physical instantiation, behavior, etc.)
Selection operates on phenotype
It is affected by environment, development, and learning



Genotype
The genetic material of an organism.
Selection does not operate directly on genotype
It is affected only by mutations

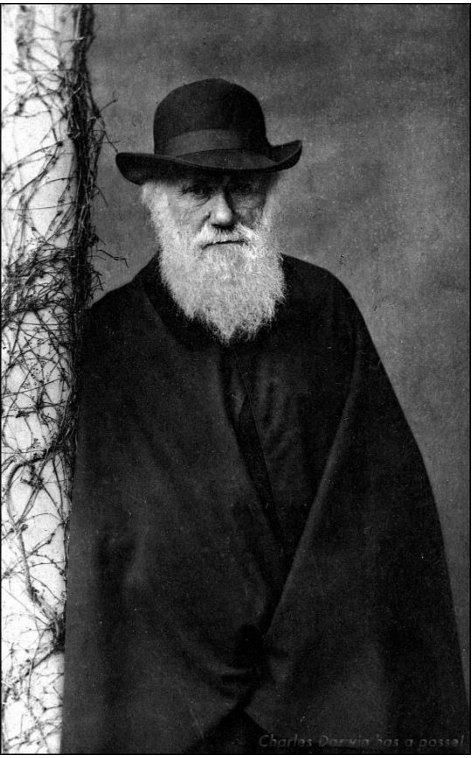


Gregor Mendel, 1858
Genetic basis of inheritance

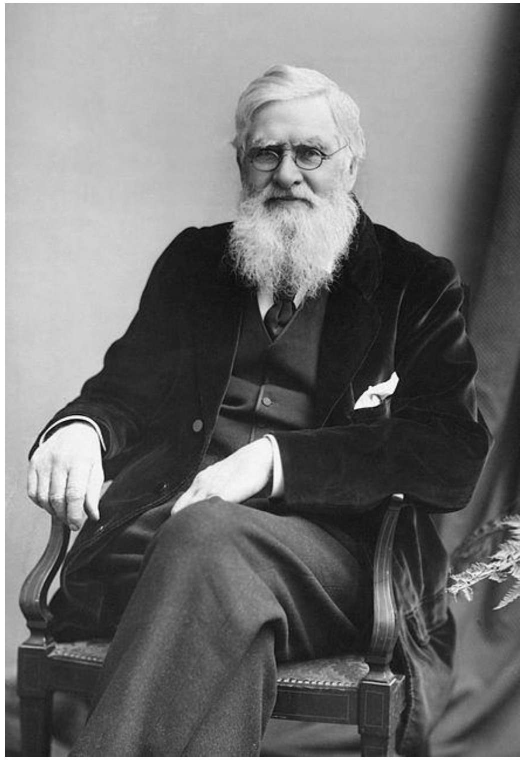
Huxley, 1940
Modern synthesis of evolution

Four Pillars of Evolution

Charles Darwin, 1859
On the Origins of Species



Alfred Russel Wallace, 1858



1. Population
Group of several individuals

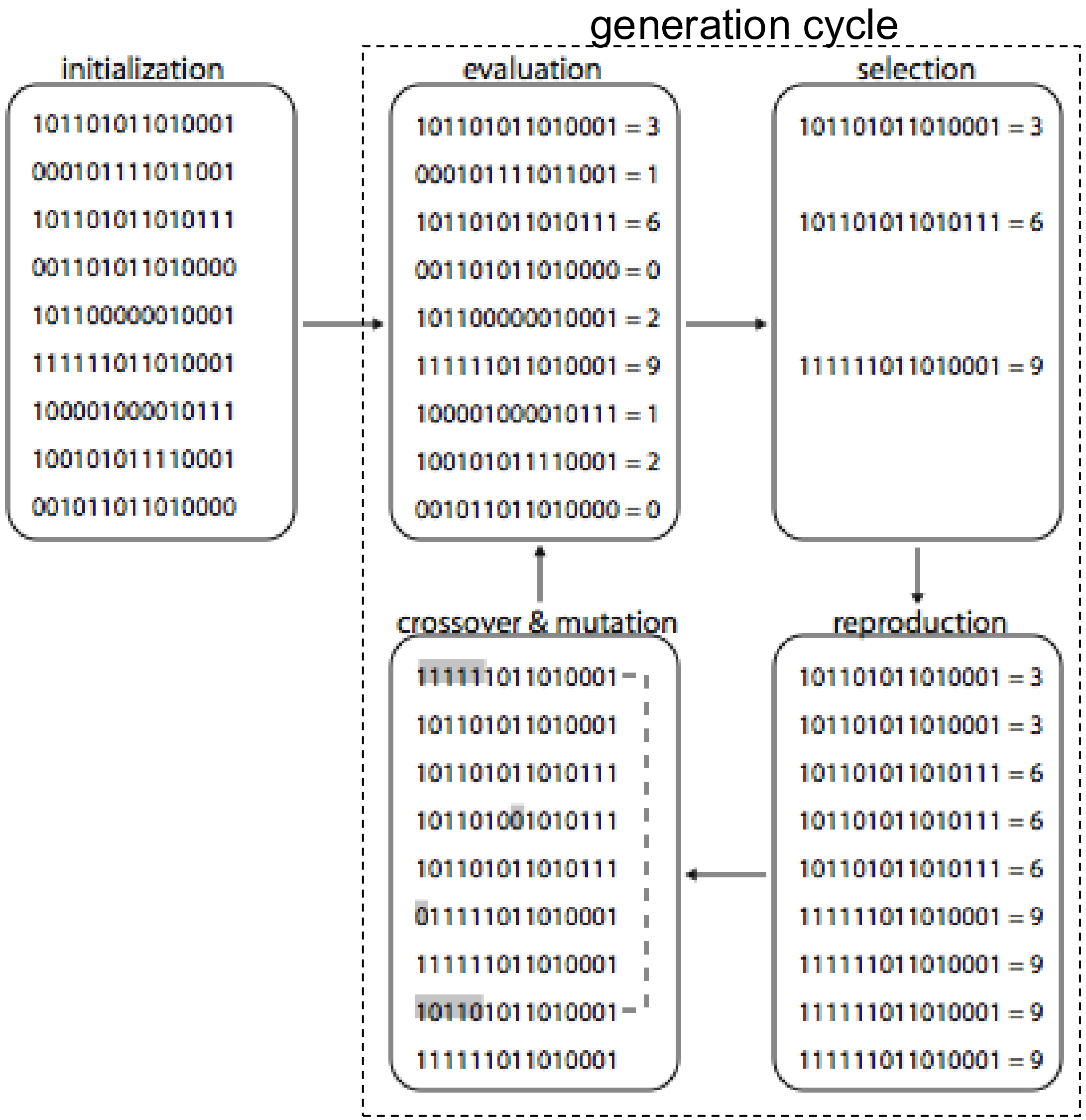
2. Diversity
Individuals have different characteristics

3. Heredity
Characteristics are transmitted over generations

4. Selection
- Individuals make more offspring than the environment can support
- Comparatively better ones have higher probability of reproducing

A simple Genetic Algorithm

1. Devise genetic representation: binary
2. Build a population: uniform sampling
3. Choose selection method: truncated, rank-based
4. Crossover: pairwise swap
5. Mutation: bit switch, p=0.01



<u>Repeat</u> generation cycle until:
- maximum fitness value is found
- solution found is good enough
- no fitness improvement for several generations

Fitness Function

It measures the performance of each individual phenotype

- It can be detailed like an RL reward function or a single number (success, failure)
- Requires extensive tests (rollouts) of an individual from different initial conditions
- Warning: You Get What You Evaluate!

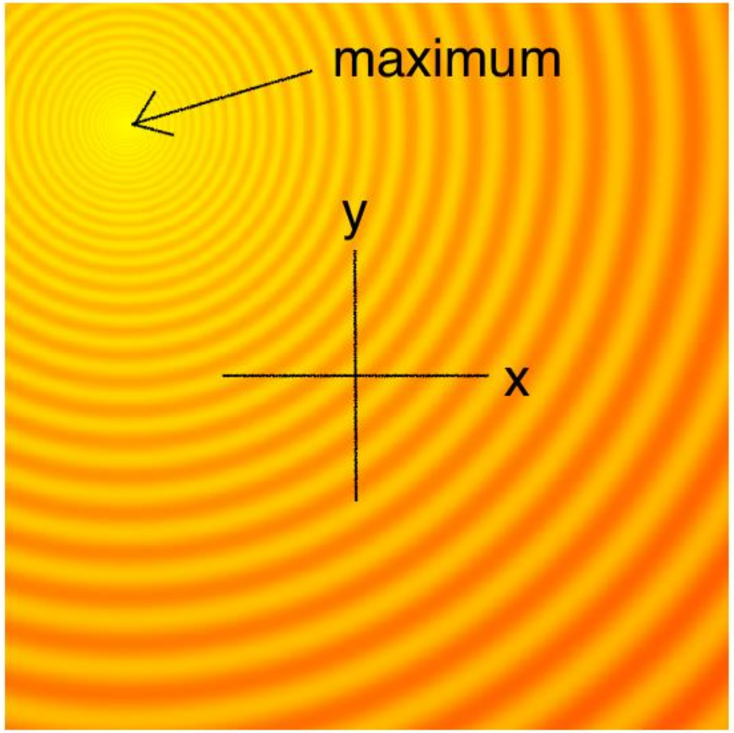
Subjective fitness: select phenotype by visual inspection

- Used when properties cannot be quantified objectively, such as aesthetics
- Can be combined with objective fitness function

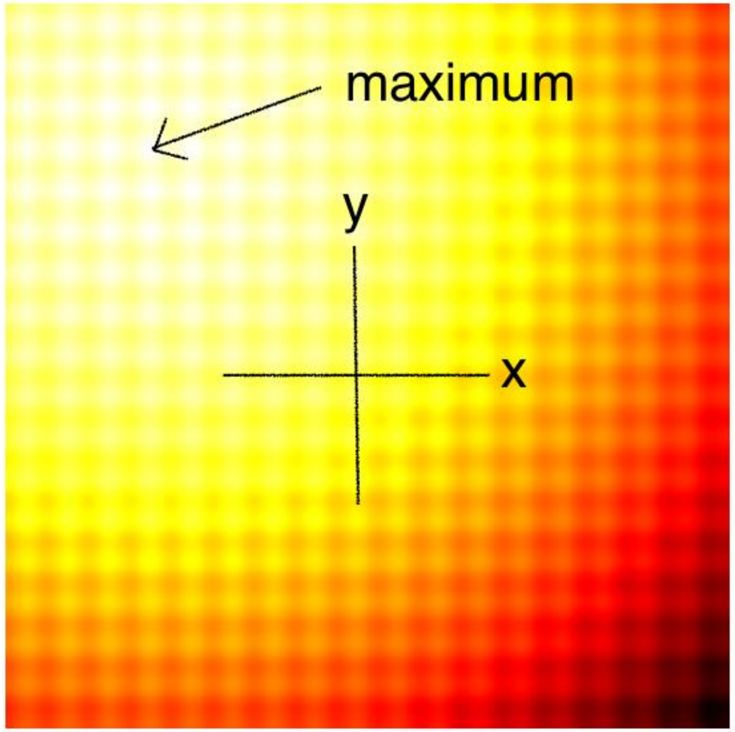
"A-Volve", Sommerer and Mignonneau,
NTT ICC Tokyo Opera House, www.ntticc.or.jp

Function Optimization

Shifted Schaffer-2D function



Shifted Rastrigin-2D function

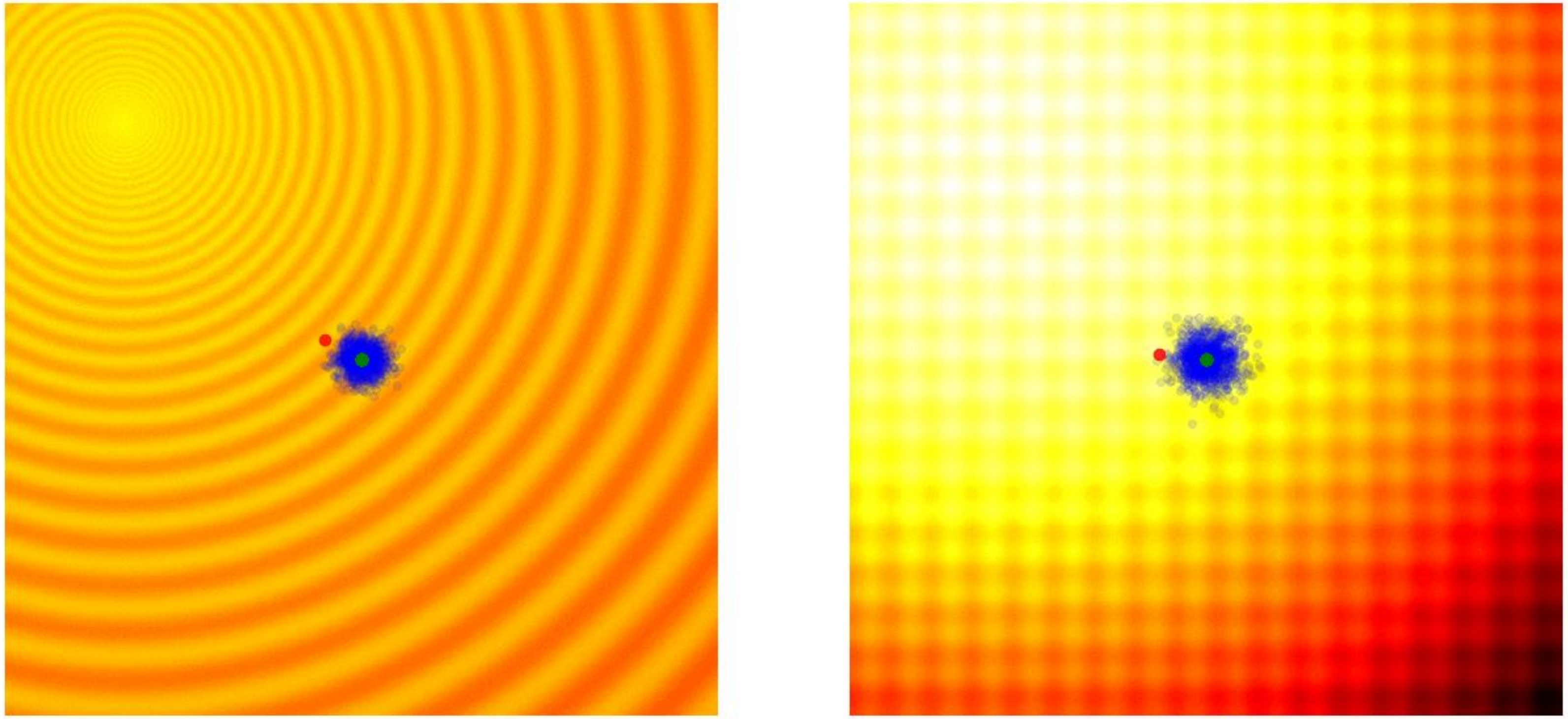


Goal: find a set of *parameters* (*x*,*y*),
such that *F*(*x*,*y*) is as close as possible to the global maximum

More test functions: https://en.wikipedia.org/wiki/Test_functions_for_optimization

Source: http://blog.otoro.net/2017/10/29/visual-evolution-strategies/

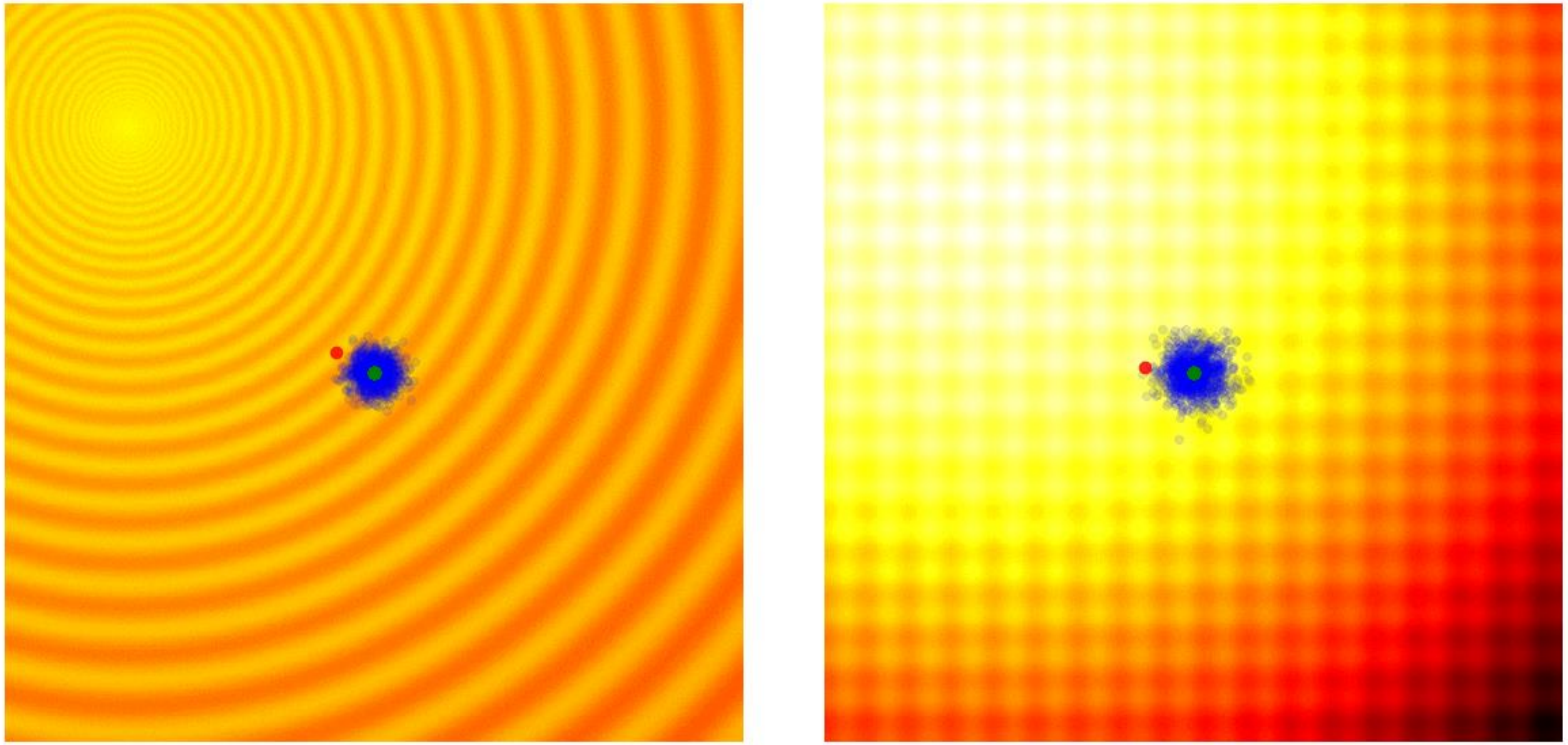
A Simple Genetic Algorithm



1. Population: Generate random binary string and create copies with low mutation probability
2. Select best 10% individuals and make copies (delete old individuals, but keep best one)
3. Crossover and mutate
4. Repeat steps 2-3 until satisfactory solution is found

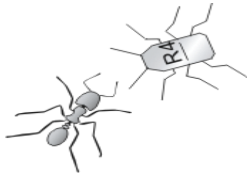
Source: http://blog.otoro.net/2017/10/29/visual-evolution-strategies/

20 generations



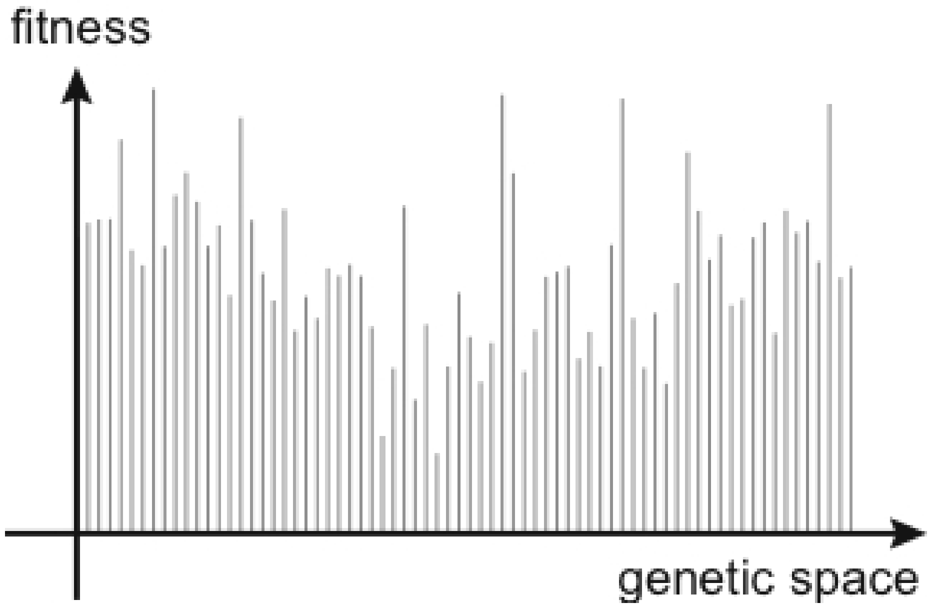
Blue dots show the individuals of the current generation
Red dot shows the best individual of the current generation
Green dots show the selected parents of the previous generation

Source: http://blog.otoro.net/2017/10/29/visual-evolution-strategies/



The Fitness Landscape

Fitness landscape is a theoretical plot of fitness values associated to all genotypes
Landscape ruggedness helps identify population size, selection pressure, mutation

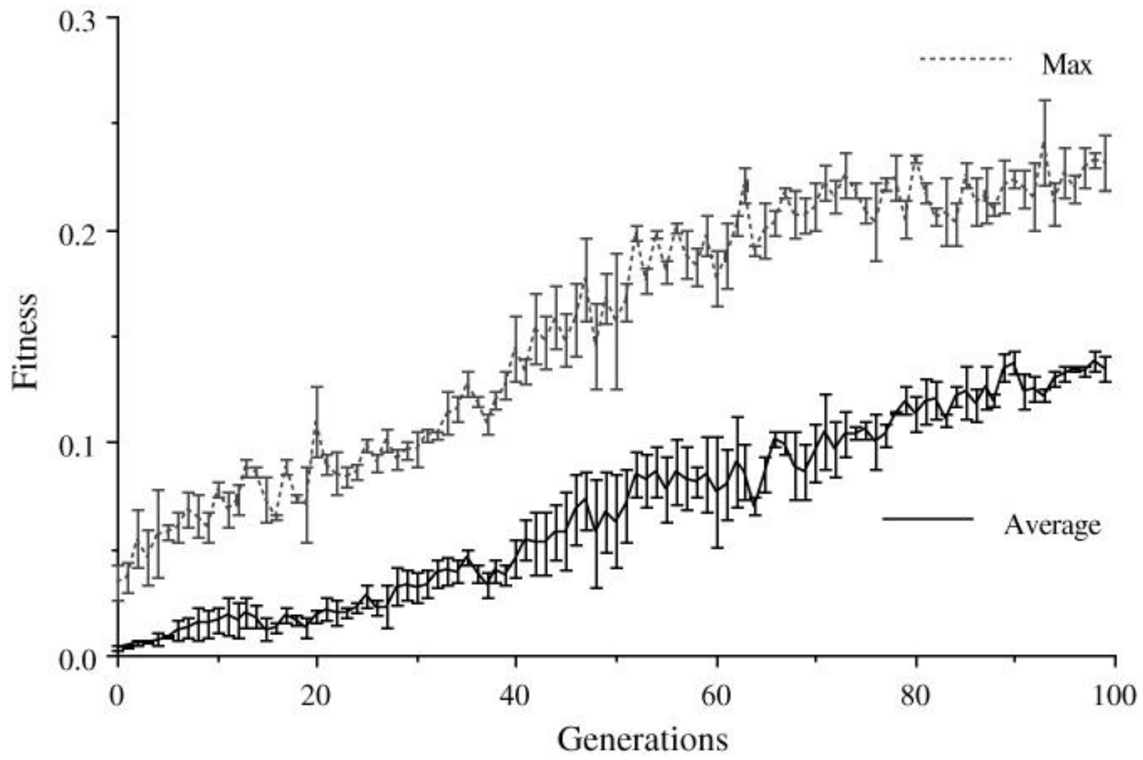


Estimating ruggedness of real landscape:

- Sample random genotypes: if flat, use large populations
- Explore surroundings of individual by applying genetic operators in sequence for fixed number of times: the larger the fitness improvement, the smaller the population size

Fitness Graph

Track average best individual fitness and mean population fitness over generations
Multiple runs are required: show average data and standard error

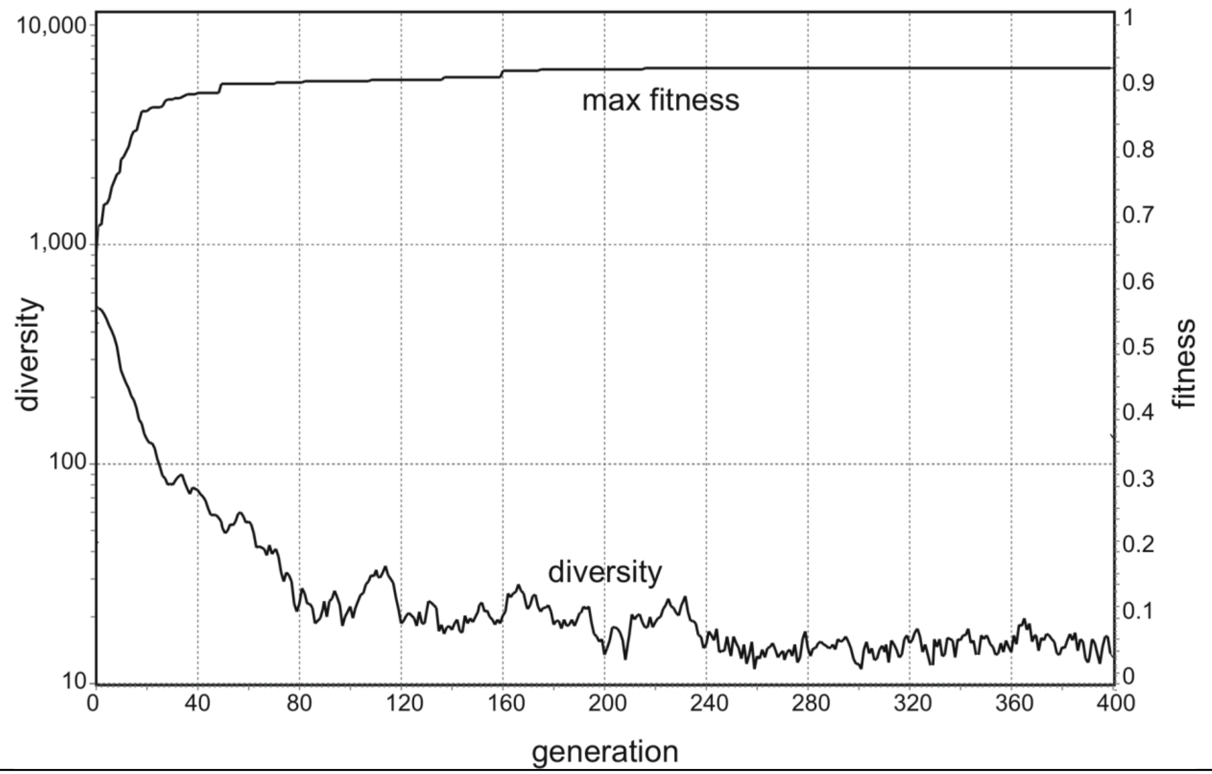


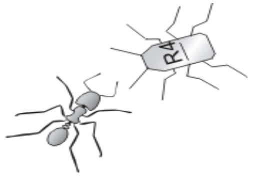
- Fitness graphs are meaningful only if the problem is stationary (no major modifications)
- Stagnation of fitness function may mean best solution is found or premature convergence

Measuring Diversity in Genotype Space

Diversity tells whether the population has potential for further evolution
Measures of diversity depend on genetic representation
E.g., for binary and real valued, use sum of Euclidean or Hamming distances

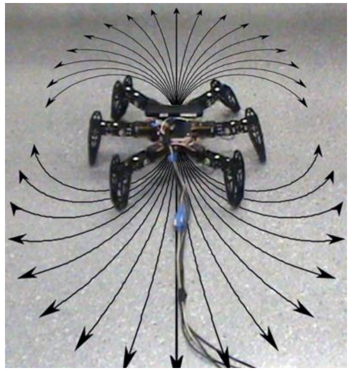
$$D_a(P) = \sum_{i,j \in P} d(g_i, g_j)$$

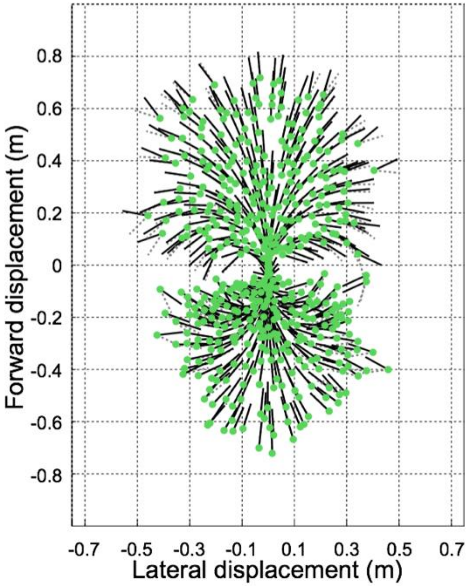


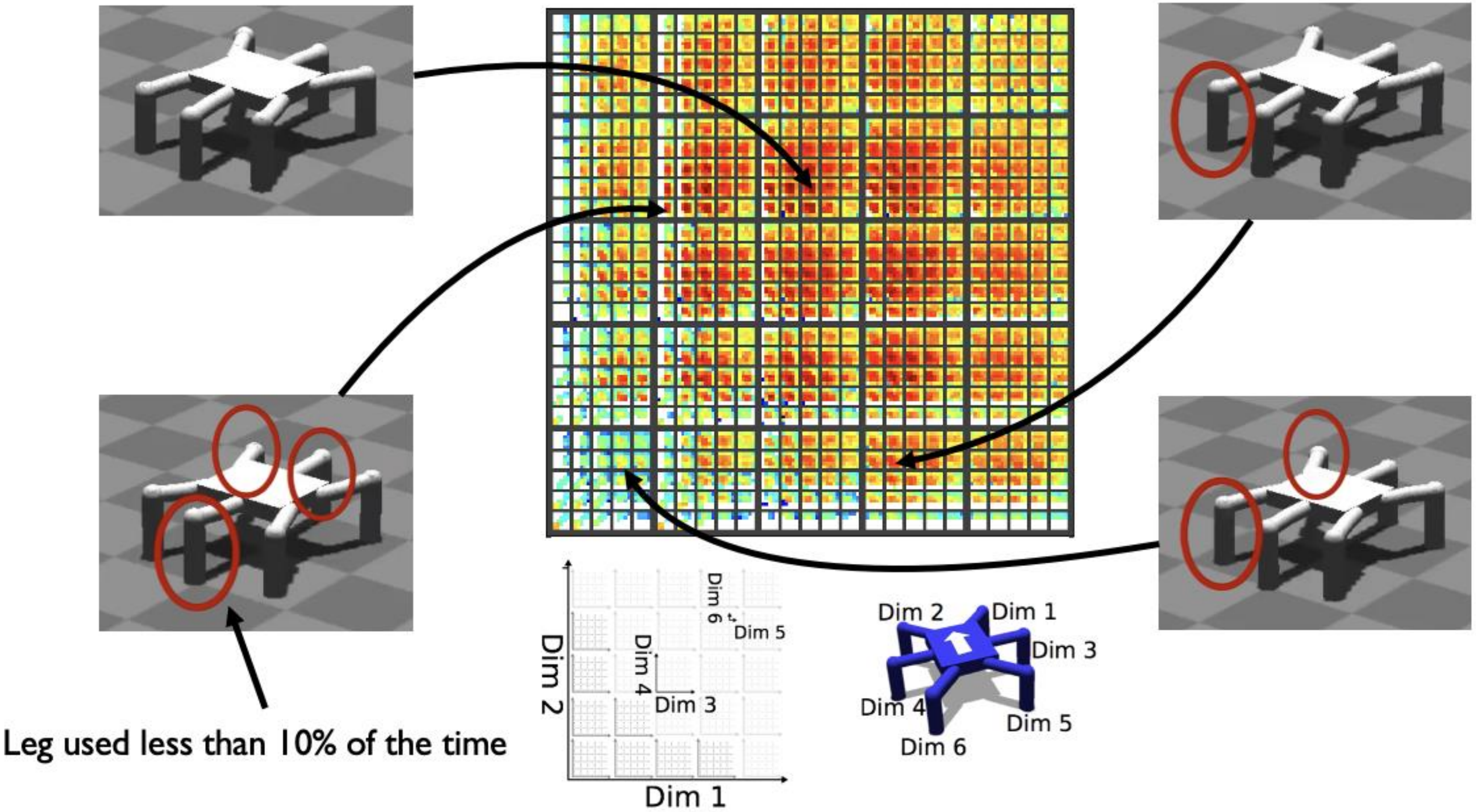


Measuring Diversity in Phenotype Space

Identify and track measures of morphological diversity (e.g., nr. of legs, mass, height, etc.) or measures of behavioral diversity (see two examples below). See also lecture on *Quality Diversity Optimization*







Cully, Mouret (2015). Evolving a Behavioral Repertoire for a Walking Robot. *Evolutionary Computation*

Cully, Clune, Tarapore, Mouret (2015). Robots that can adapt like animals. *Nature*

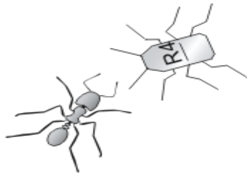
Artificial and Natural Evolution

<u>SIMILARITIES</u>

- Phenotype (computer program, object shape, electronic circuit, robot, etc.)
- Genotype (genetic representation of the phenotype)
- Population
- Diversity
- Selection
- Inheritance

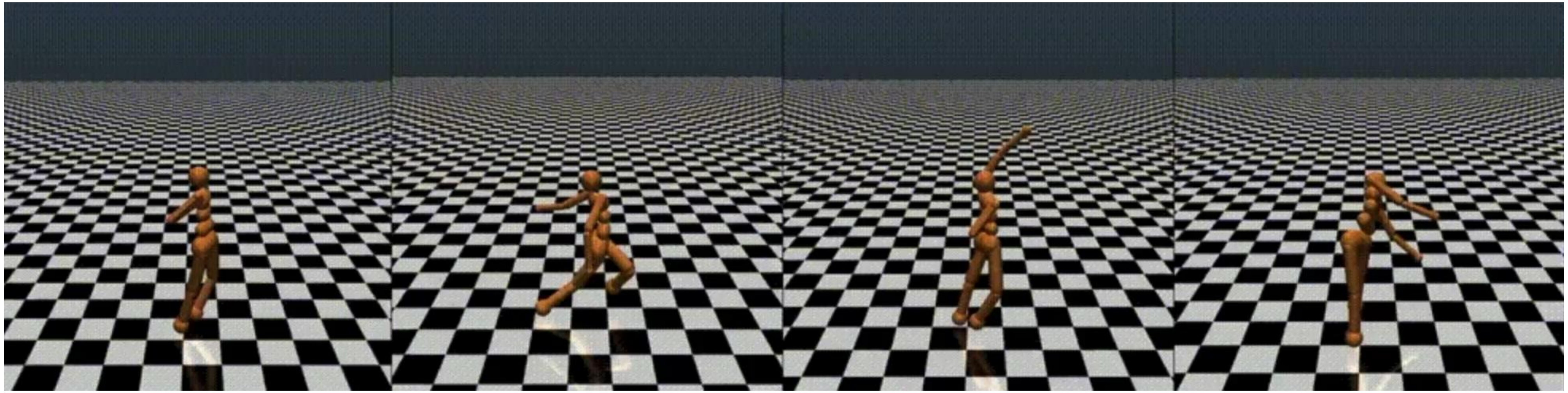
<u>DIFFERENCES</u>

- Fitness is measure of performance of the individual solution to the problem
- Selection of the best according to performance criterion (fitness function)
- Expected improvement between initial and final solution



What is Evolutionary Computation best for?

- Decision-making problems
- Sparse reward (only one number)
- Stochastic or no gradient
- Non-differentiable policies (e,g, binary neurons, spiking neurons)
- Suitable for large-scale hardware parallelization
- Want population of diverse solutions (instead of one solution)
- Co-adaptation of control policy and body design
- Multi-objective problems
- Combination with other ML methods



Checkpoints

- How does Evolutionary Computation relate to / differ from other Machine Learning methods?
- On what problems is Evolutionary Computation better than other ML methods?
- Describe the 4 pillars of natural and artificial evolution
- What is phenotype and genotype?
- Explain the information-carrying components of biological DNA
- Describe the types of biological mutations
- Describe the main steps of a generation cycle in evolutionary computation
- Similarities and differences between natural and artificial evolution
- What is a fitness function?
- What is the fitness landscape?
- What is the fitness graph?
- How is diversity monitored in the genotype and phenotype space?
- What are the differences between natural and artificial evolution?