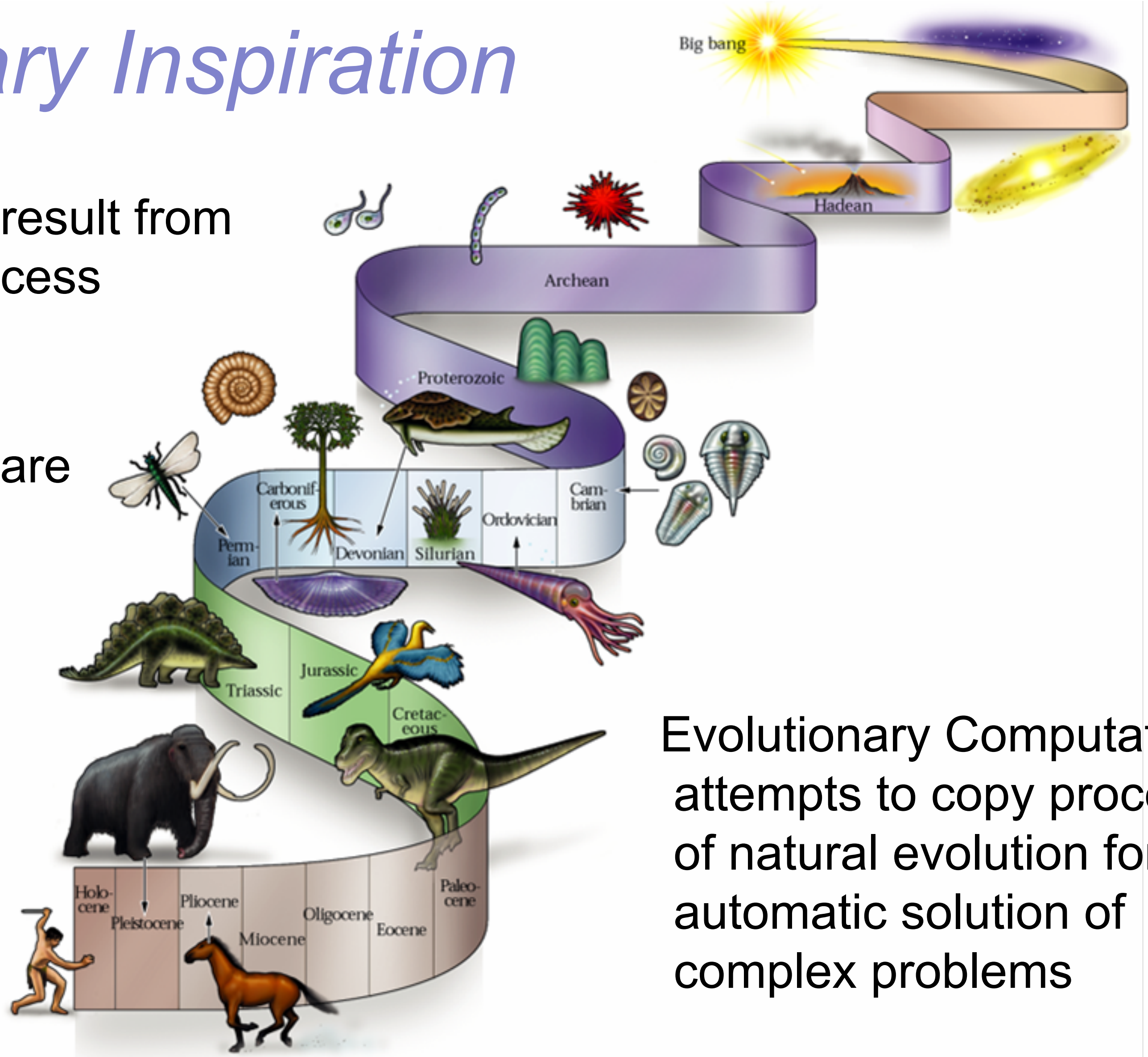
Evolutionary Systems

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

Evolutionary Inspiration

Biological systems result from an evolutionary process

Biological systems are
- robust
- complex
- adaptive



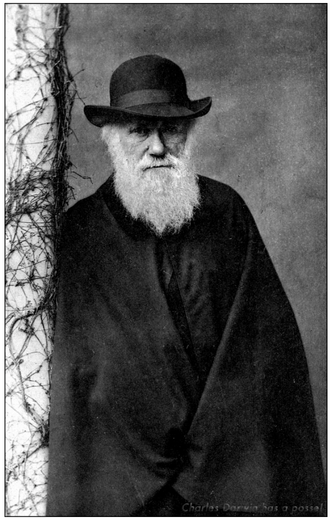
Evolutionary Computation attempts to copy process of natural evolution for automatic solution of complex problems

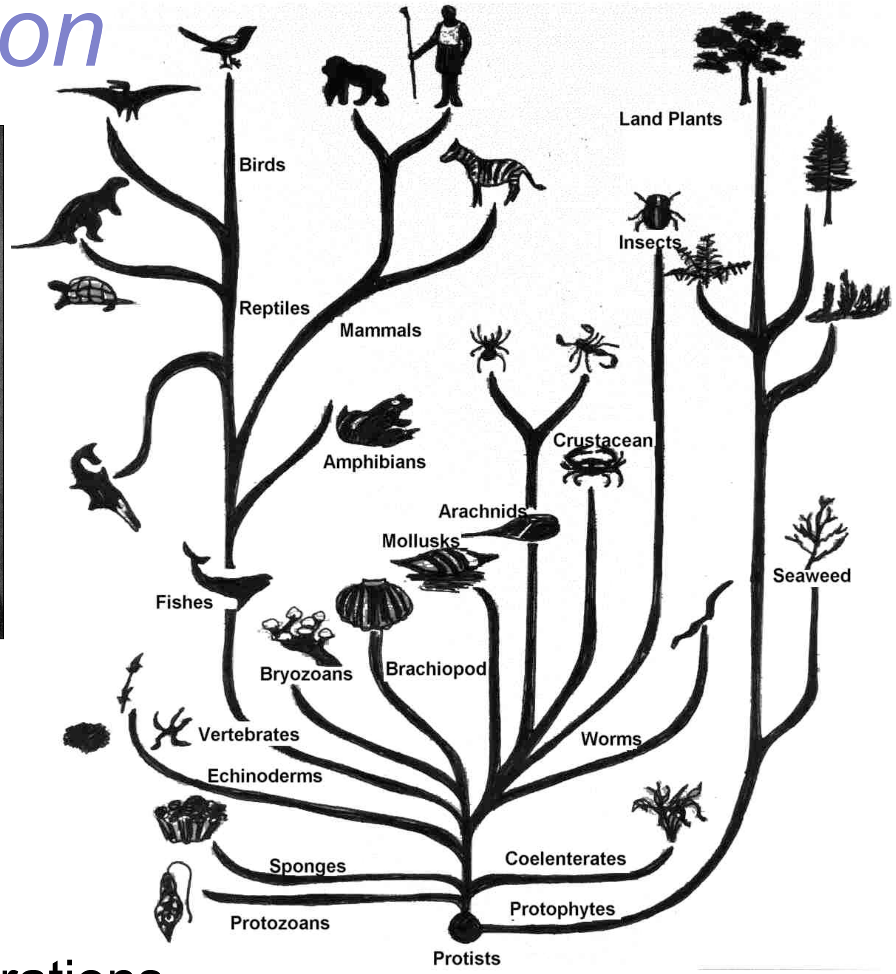
Does natural evolution generate increasingly complex systems?

The 4 Pillars of Evolution

All species derive from common ancestor

Charles Darwin, 1859
On the Origins of Species

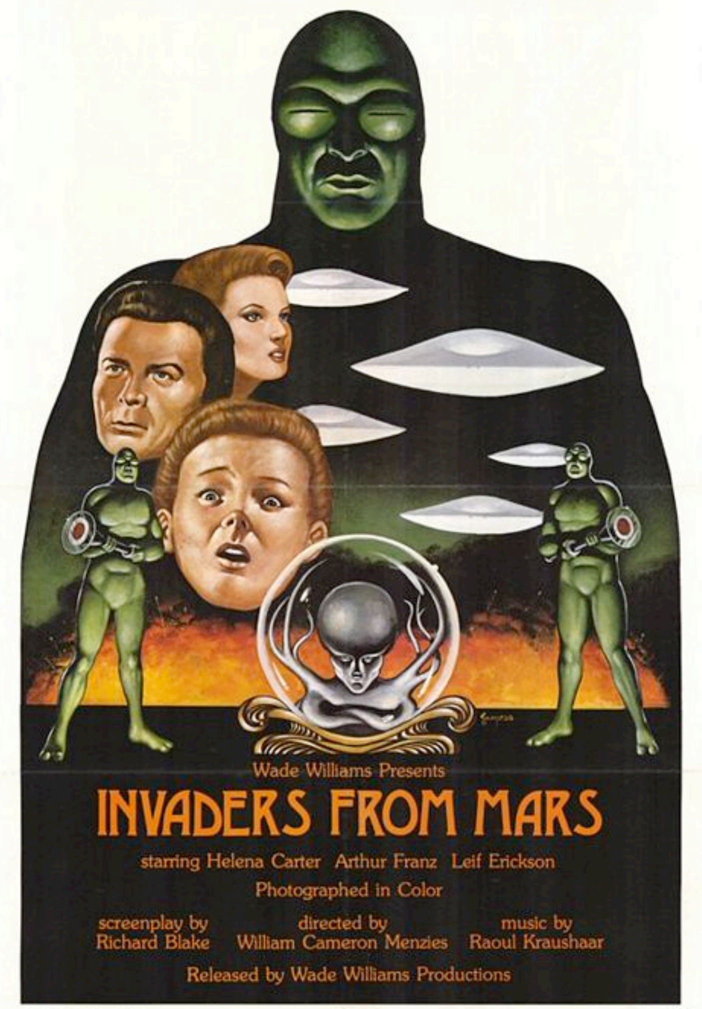




Population
Group of several individuals
Diversity
Individuals have different characteristics
Heredity
Characteristics are transmitted over generations
Selection
- Individuals make more offspring than the environment can support
- Better at food gathering = better at surviving = make more offspring

Evolution without Progress

… or "why we should not fear an invasion from Mars" (Gould, 1997)



Humans are not the top of the evolutionary ladder (misleading image of evolution with humans at top or end).

Evolution without Progress:

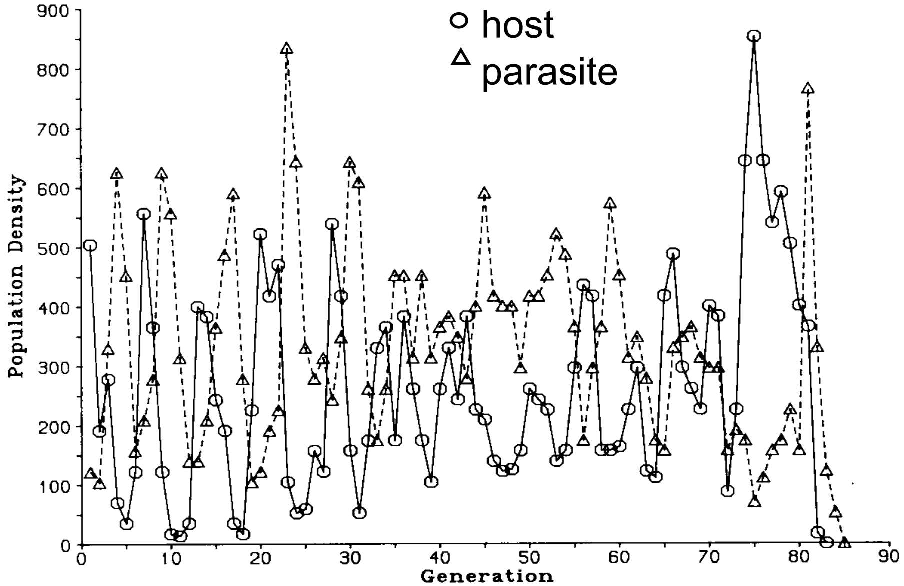
- If no competition, no selection of the fittest
- Individuals selected against current environment
- Accumulation of change with no cost or benefit (also known as *Neutral Evolution*)

Models of Evolution

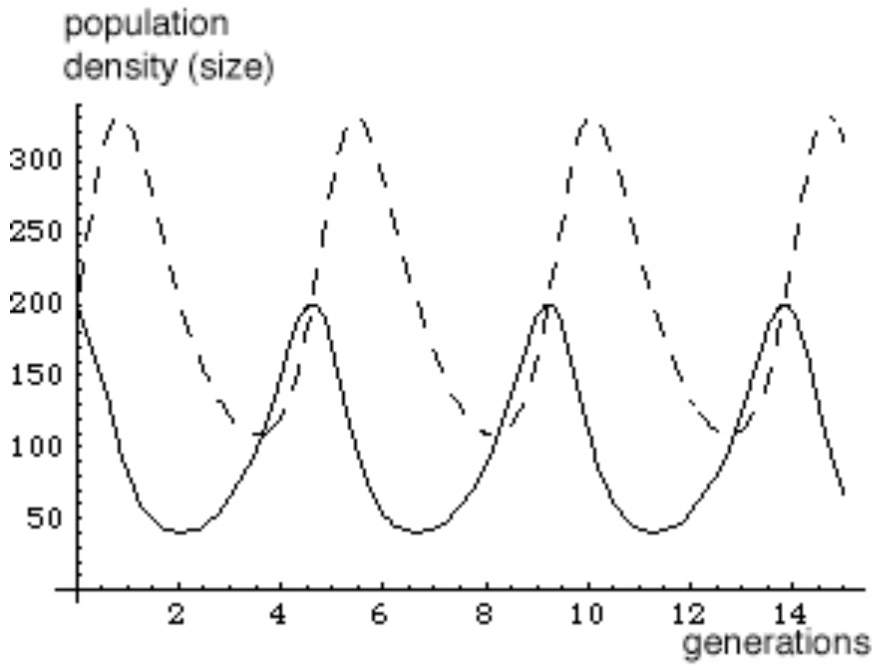
Biological models predict variations in population size or gene frequency, but not progress.

Ex: Lotka-Volterra model of competitive co-evolution

dN1/dt=N1 (r1-b1N2)
dN2/dt=N2 (-r2+b2N1)
where:
- N1, N2 are the two populations
- r1 is increment rate of prey without predators
- r2 is death rate of predators without prey
- b1 is death rate of prey caused by predators
- b2 is ability of predators to catch prey



(Utida, 1957)



Phenotype & Genotype

Phenotype = the manifestation of the organism (appearance, behavior, etc.).
Selection operates on the phenotype;
It is affected by environment, development, and learning

Genotype = the genetic material of that organism.
It is transmitted during reproduction;
It is affected by mutations;
Selection does not operate directly on it

Genetics = structure and operation of genes
Functional genomics = role of genes in the organism

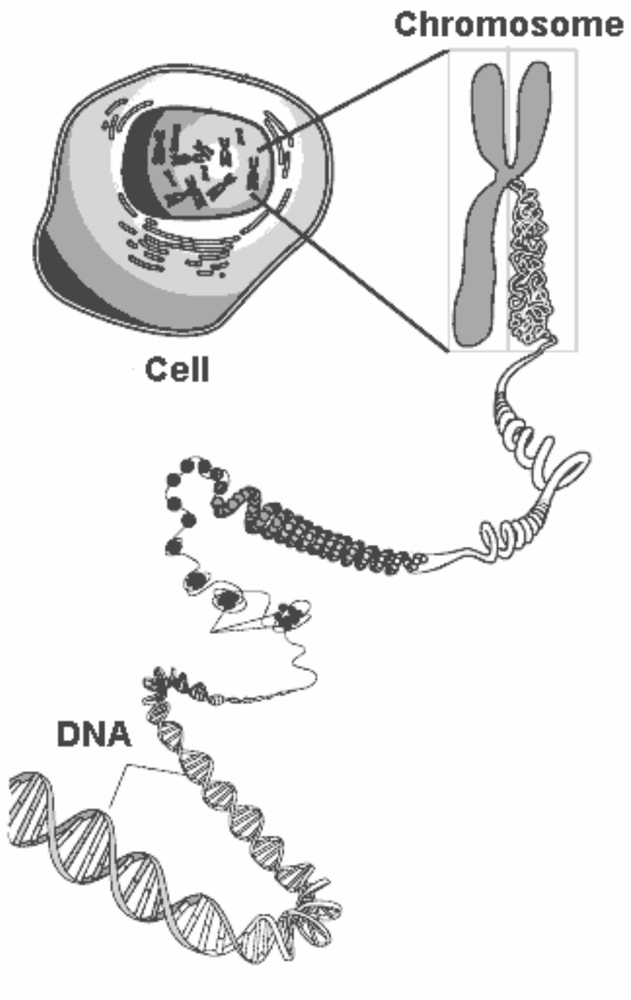
To what extent are we determined by genotype and
phenotype?

Jean-Felix & Auguste Piccard

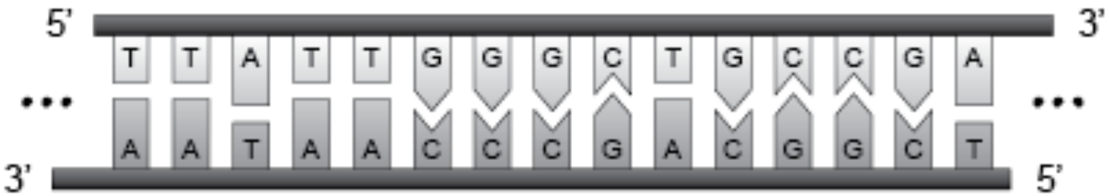
DNA *(DeoxyriboNucleic Acid)*

Long molecule, twisted in spiral, and compressed

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



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)

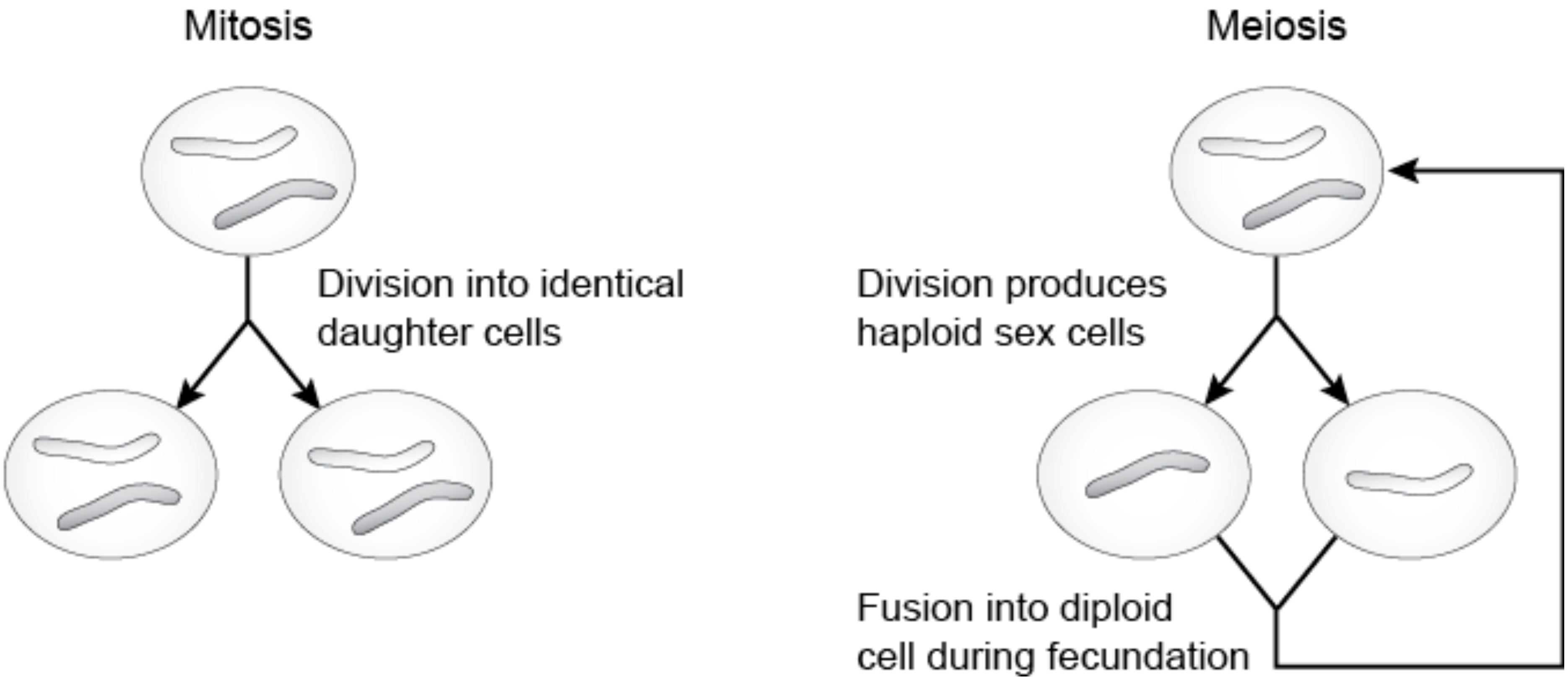


A **gene** is a sequence of several nucleotides that produce a protein

Cell Replication

Cells replicate in two ways:

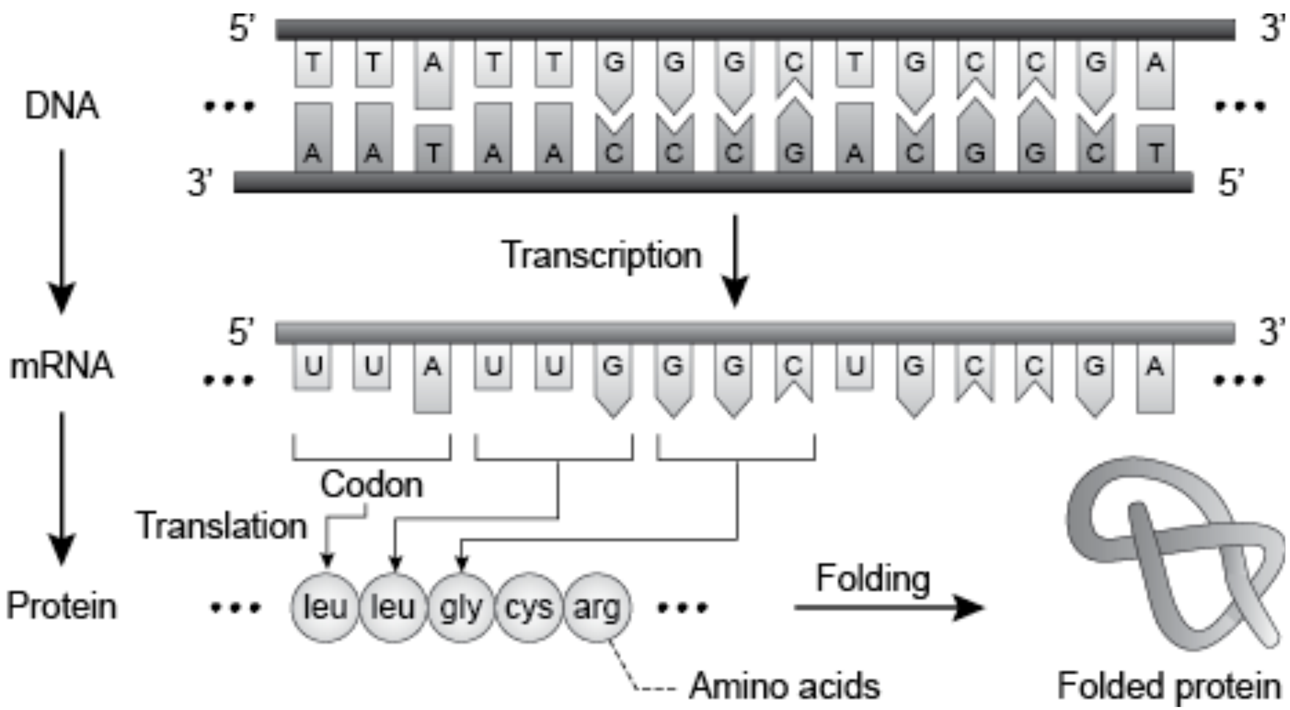
- **Mitosis**: during growth/maintenance of the organism
- **Meiosis**: during production of sex cells



From Genes to Proteins *(gene expression)*

Proteins are molecules that define the type and function of cells (hair and muscle cells are made of different proteins).

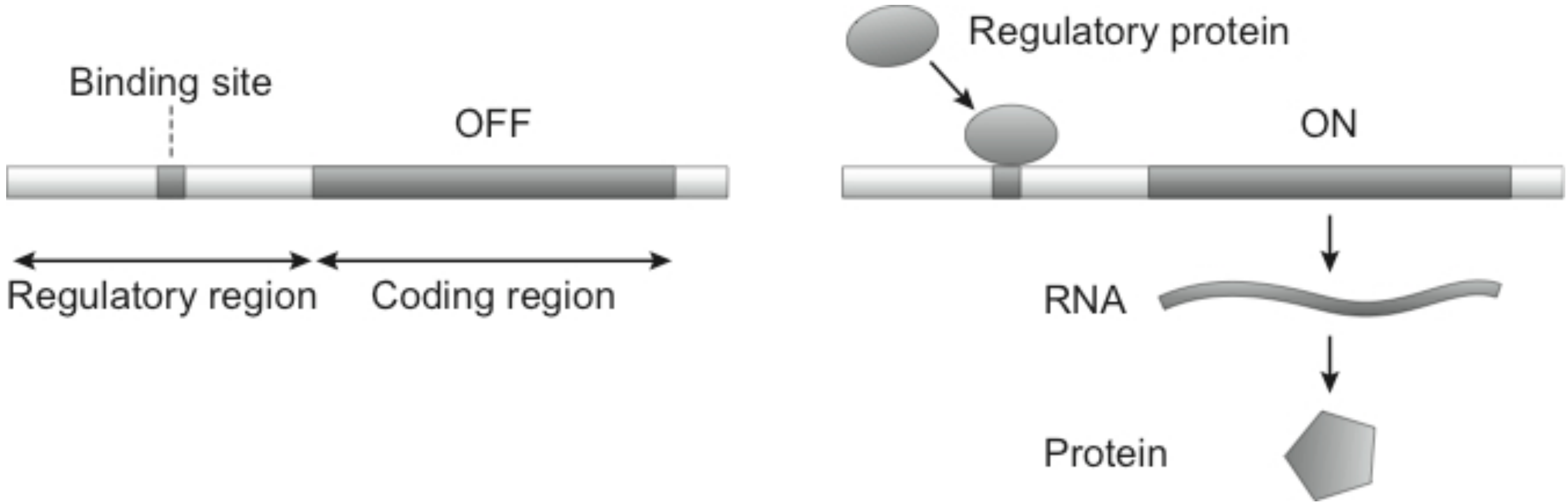
The sequence of nucleotides in one strand defines the type of protein. The expression of the gene into a protein is mediated by another molecule, known as messenger RNA.

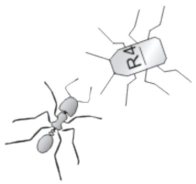


Gene structure

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

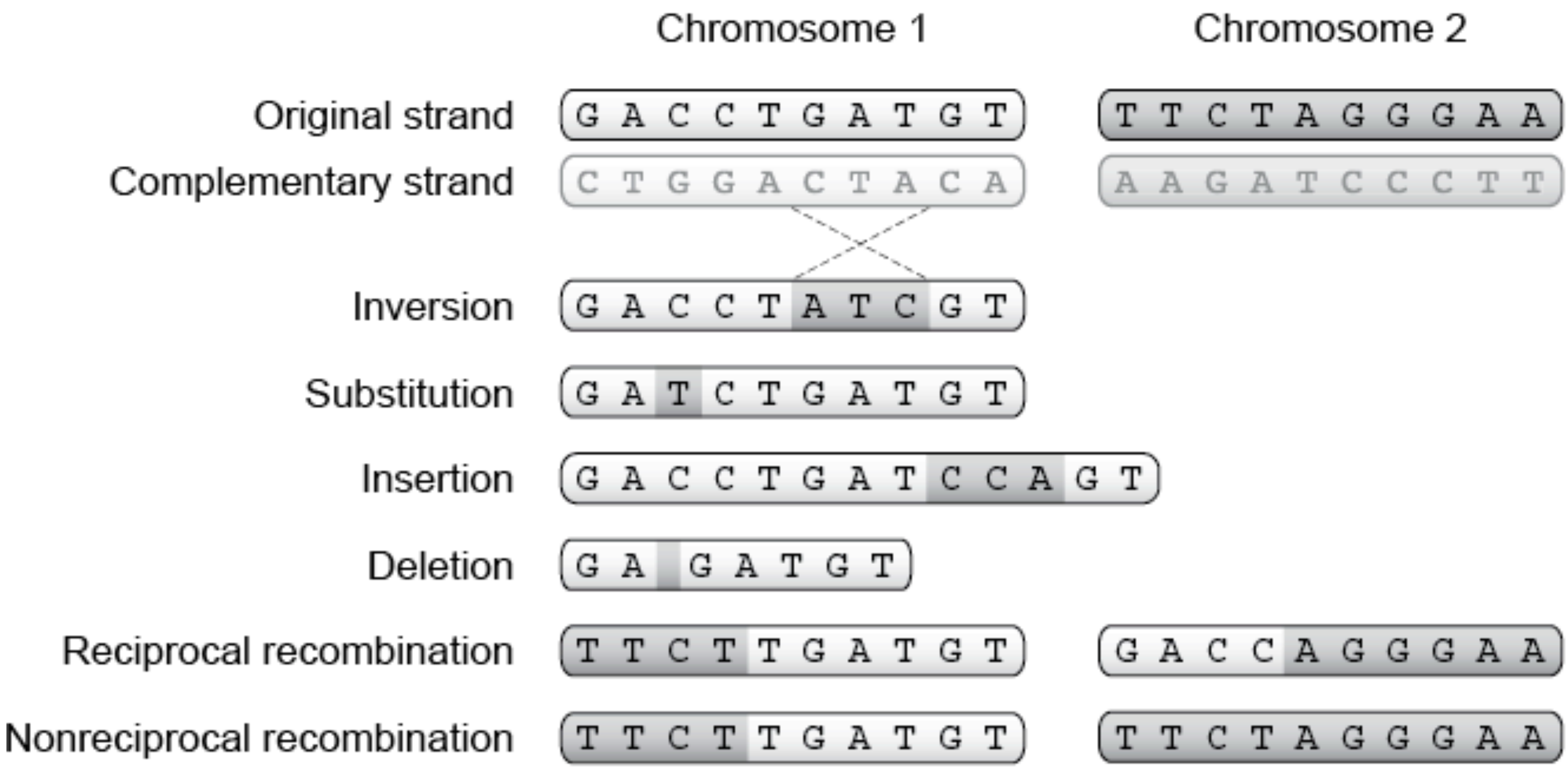
The coding region is translated into a protein if another protein binds onto the regulatory region. Regulation can also be negative (i.e., inhibition of protein production).





Genetic Mutations

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

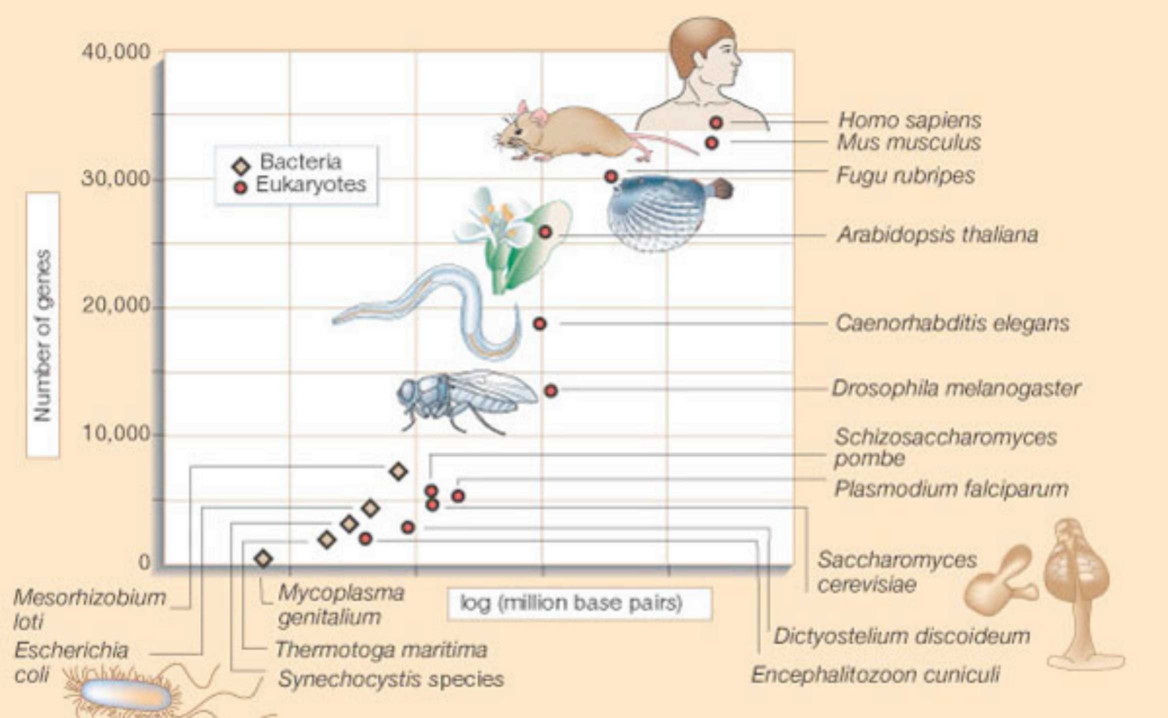


Companion slides for the book *Bio-Inspired Artificial Intelligence: Theories, Methods, and Technologies* by Dario Floreano and Claudio Mattiussi, MIT Press

Genome Size

Genome size within a species is constant (C-value, expressed in Mega bases), but it greatly varies across species www.genomesize.com for comparisons

Genome size is not related to complexity of phenotype

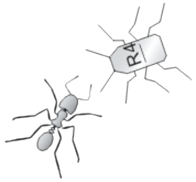


Doolittle, 2002

Genome contains:
- **Genic** DNA
- **Nongenic** DNA

Nongenic DNA **arises** from:
- insertion/deletion mutations
- gene duplication

Nongenic DNA may have an **adaptive value**:
- pseudogenes may be re-activated
- pseudogenes may transform into new genes by several neutral mutations



Artificial Evolution

Automatic generation of solutions to hard problems

<u>Similarities between natural and artificial evolution</u>:

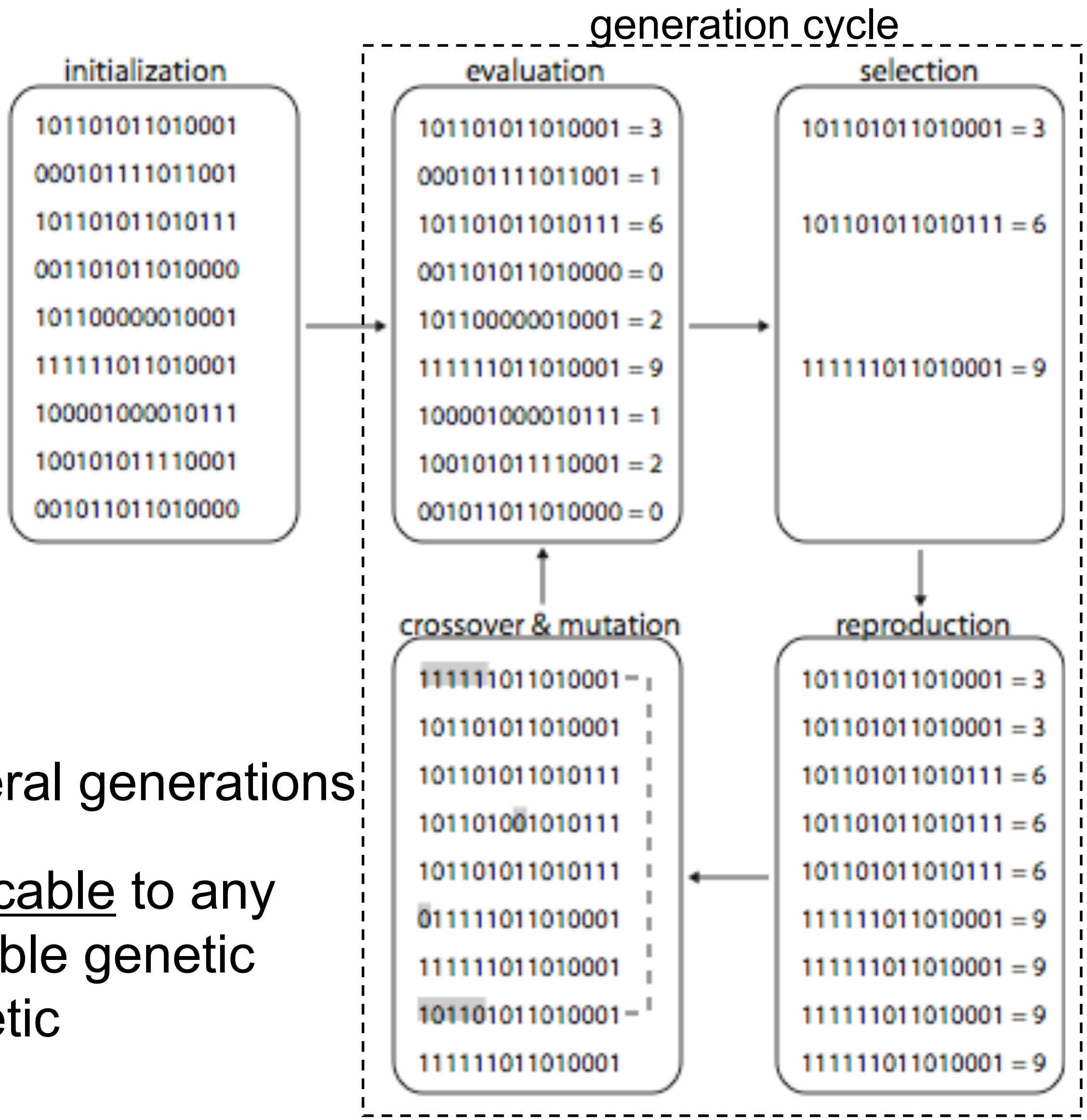
- Phenotype (computer program, object shape, electronic circuit, robot, etc.)
- Genotype (genetic representation of the phenotype)
- Population
- Diversity
- Selection
- Inheritance

<u>Differences between natural and artificial evolution</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

Evolutionary Algorithm

- Devise genetic representation
- Build a population
- Design a fitness function
- Choose selection method
- Choose crossover & mutation
- Choose data analysis method



<u>Repeat</u> generation cycle until:

- maximum fitness value is found
- solution found is good enough
- no fitness improvement for several generations

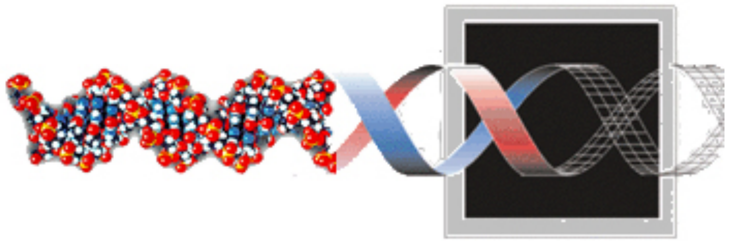
Evolutionary algorithms are <u>applicable</u> to any problem domain as long as suitable genetic representation, fitness, and genetic operators are chosen.

Genetic Representation

Describes elements of genotype and mapping to phenotype

- Must match genetic operators of recombination and mutation
- Set of possible genotypes should include optimal solution to the problem

Choice of representation benefits from domain knowledge:
- Encoding of relevant parameters
- Appropriate resolution of parameters



1001101010001

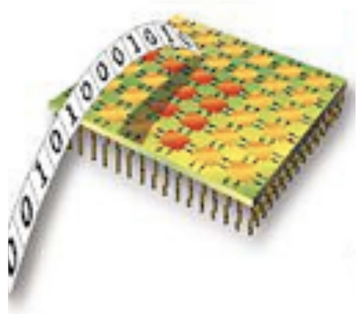
Great simplification of genetics:
- Single stranded sequence of characters (e.g., binary)
- Fixed length, only genic
- Often haploid structure and one chromosome
- Often one-to-one direct correspondence between gene and parameter
- Gene expression and genetic regulation used only in specific situations

Discrete Representations

A sequence of *l* discrete values drawn from alphabet with cardinality *k*

- E.g., binary string of 8 positions (l=8, k=2): 01010100
- Can be mapped into several phenotypes:

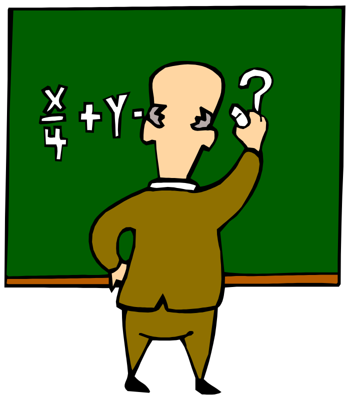
to configuration string of FPGA electronic circuits



01010100

to integer *i* using binary code

84

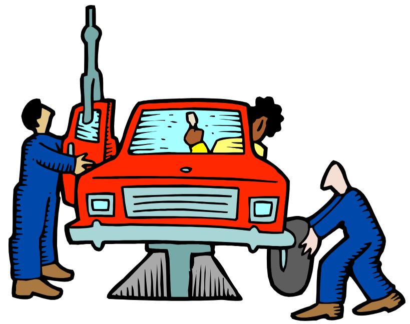


0.328125

to real value *r* in range [min, max]:

$r = min + (i/255)(max-min)$

Job	A.M.	P.M.
1	x	
2		x
3	x	
4		x
5	x	
6		x
7	x	
8	x	

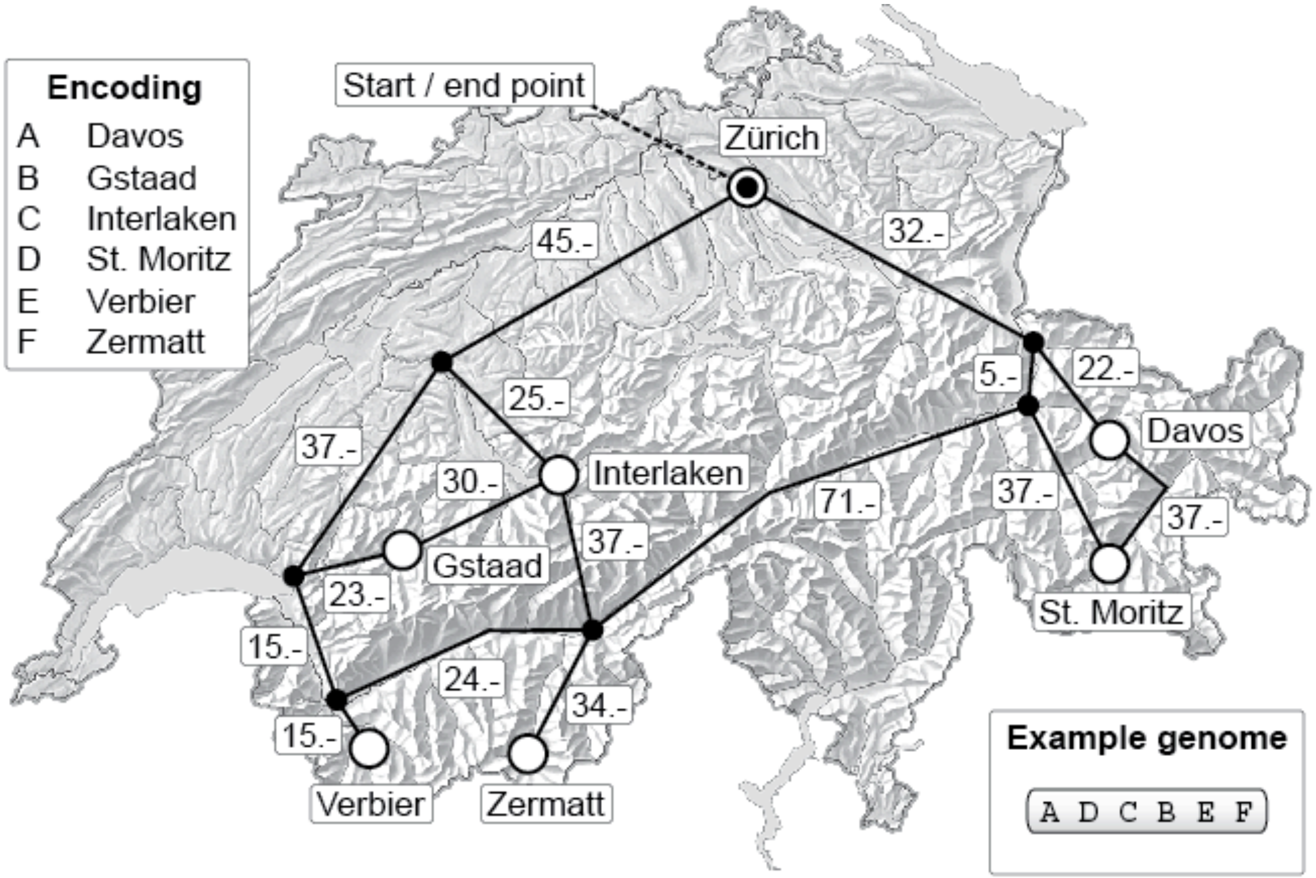


to job schedule:

- job=gene position
- time=gene value

Sequence Representation

It is a particular case of discrete representation used for class of Traveling Salesman Problems (plan a path to visit n cities under some constraints). E.g., planning ski holidays with lowest transportation costs

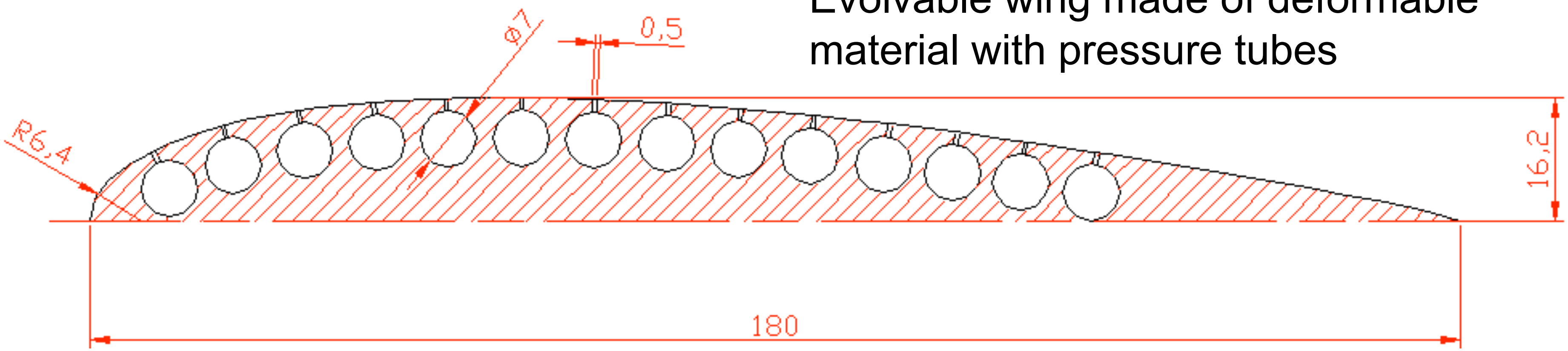


Real-Valued Representation

Genotype is sequence of real values that represent parameters

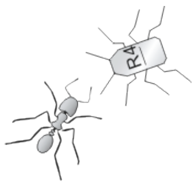
- Used when high-precision parameter optimization is required
- For example, genetic encoding of wing profile for shape optimization

Evolvable wing made of deformable material with pressure tubes



<u>Genotype</u>= pressure values of 14 tubes

Alternatively, encode values of variables of equations describing profile



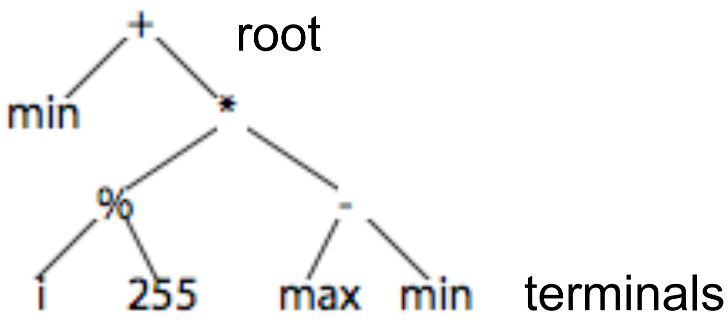
Tree-based Representation

Genotype describes a tree with branching points and terminals
Suitable for encoding hierarchical structures
E.g., used to encode computer programs

Computer program is made of:
• Operators (Function set: multiplication, If-Then, Log, etc.)
• Operands (Terminal set: constants, variables, sensor readings, etc.)

Expression r=min+(i/255)(max-min)

Nested list (+, min, (*, (/, i, 255), (-, max, min)))



• <u>Closure</u>: all functions must accept all terminals in Terminal set and outputs of all functions in Function set (e.g., protected division %)
• <u>Sufficiency</u>: elements in Function and Terminal sets must be sufficient to generate program that solves the problem

Initial Population

Sufficiently large to cover problem space (!), but sufficiently small for evaluation costs (typical size: between 10s and 1000s individuals)

Uniform sample of search space:

- Binary strings: 0 or 1 with probability 0.5
- Real-valued representations: uniform on a given interval if bounded phenotype (e.g., +2.0, -2.0); otherwise best guess or binary string with dynamic mapping resolution (Schraudolph and Belew, 1992; Dürr et al, 2007)
- Trees are built recursively starting from root: root is randomly chosen from function set; for every branch, randomly choose among all elements of function set and of terminal set; if terminal is chosen, it becomes leaf; set maximum depth of tree.

Mutated clones of previously evolved genotype or hand-designed genotype:

- Possible loss of genetic diversity
- Possible unrecoverable bias

Fitness Function

Evaluates **performance** of phenotype with a numerical score
- Choice of components; e.g., lift and drag of wing
- Combination of components; e.g. (lift + 1/drag) or (lift - drag)
- Extensive test of each phenotype
- Warning! You Get What You Evaluate (example in application, later)

Subjective fitness: select phenotype by visual inspection
- Used when aesthetic properties cannot be quantified objectively
- Can be combined with objective fitness function

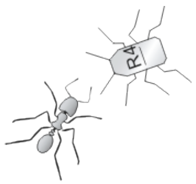
"A-Volve", Sommerer and Mignonneau,
NTT ICC Tokyo Opera House, www.ntticc.or.jp

Selection

A method to make sure that better individuals make comparatively more offspring

Used in artificial evolution and breeding

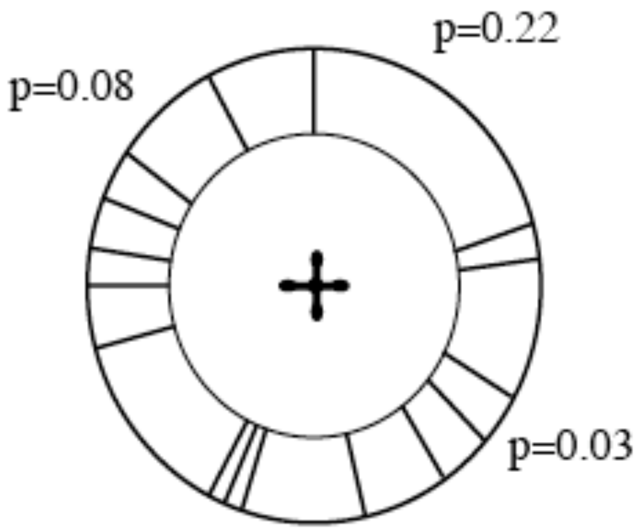
- Selection pressure is inversely proportional to nr. of selected individuals
- High selection pressure = rapid loss of diversity and premature convergence
- Make sure that also less performing individuals can reproduce to some extent

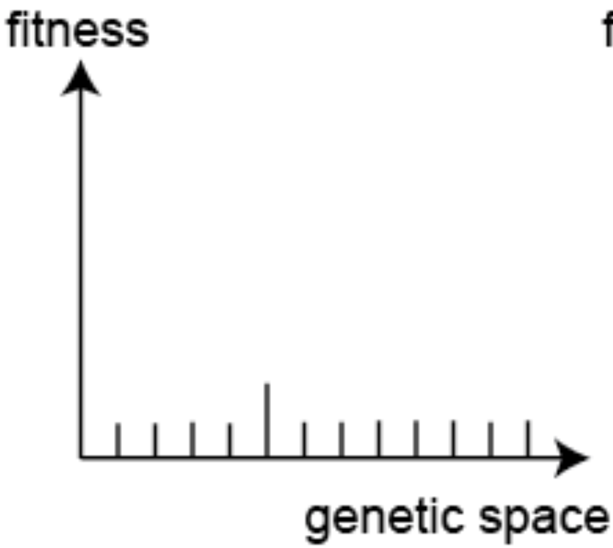


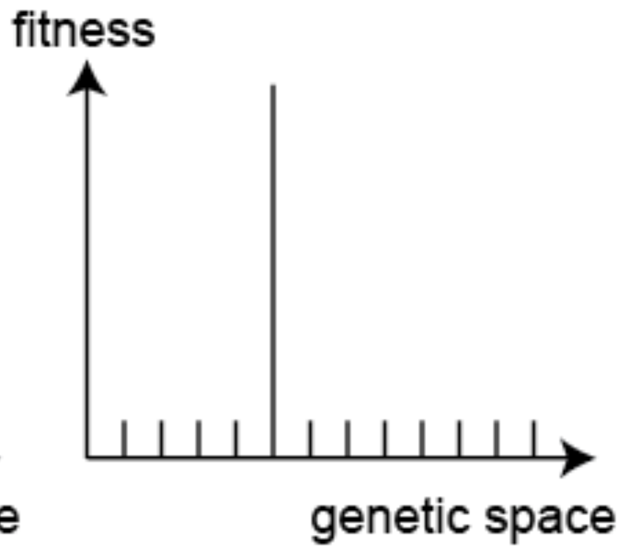
Proportionate Selection

The probability that an individual makes an offspring is proportional to how good its fitness is with respect to the population fitness: **$p(i) = f(i)/\Sigma f(i)$**

Also known as Roulette Wheel selection





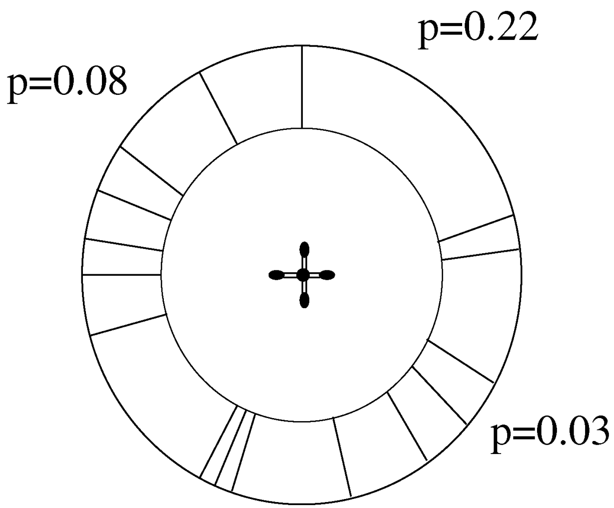


Problems:
Uniform fitness values = random search
Few high-fitness individuals = high selection pressure

Rank-based Selection

- Individuals are sorted on their fitness value from best to worse. The place in this sorted list is called the rank r.
- Instead of using the fitness value of an individual, the rank is used to select individuals: **p(i) = 1 - r(i)/Σr(i)**
- Use roulette wheel

individual	fitness	rank
A	5	5
B	7	3
C	8	2
D	2	8
E	3	7
F	9	1
G	7	4
H	4	6



Truncated Rank-based Selection

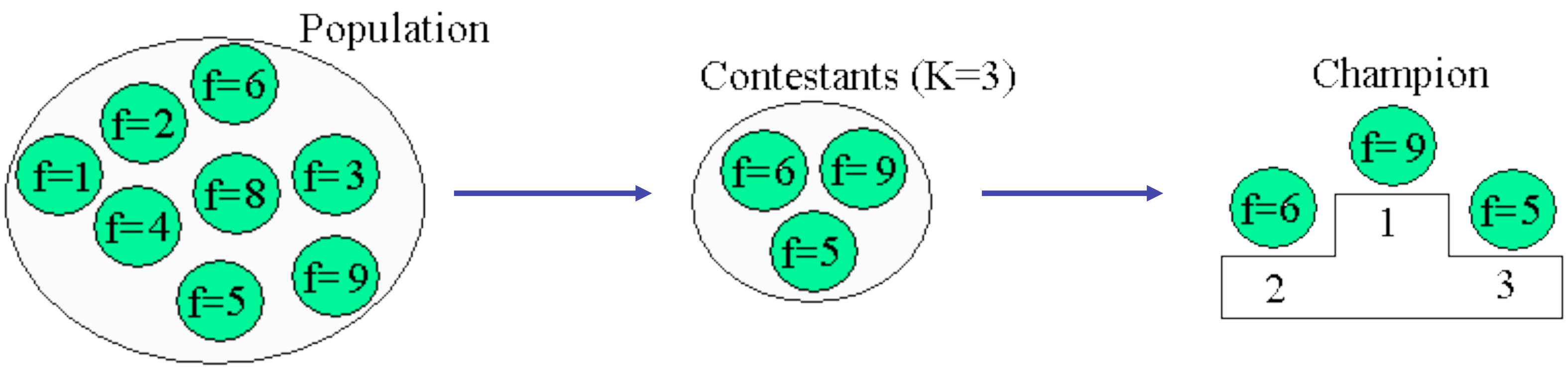
- Only the best **x** individuals are allowed to make offspring and each of them makes the same number of offspring: **N/x**, where **N** is the population size.
- E.g., in population of 100 individuals, make 5 copies of 20 best individuals

individual	fitness	rank	list
A	5	5	F
B	7	3	C
C	8	2	B
D	2	8	G
E	3	7	A
F	9	1	H
G	7	4	E
H	4	6	D

Tournament Selection

For every offspring to be generated:

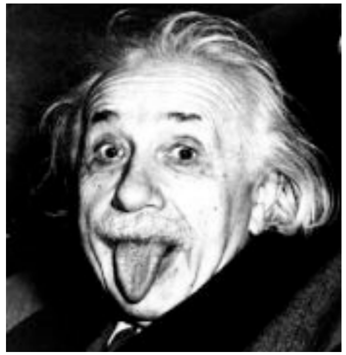
- Pick randomly ***k*** individuals from the population
- Choose the individual with the highest fitness and make a copy
- Put all individuals back in the population



k is the tournament size (larger size = larger selection pressure)

Replacement

Generational replacement: old population is entirely replaced by offspring (most frequent method)

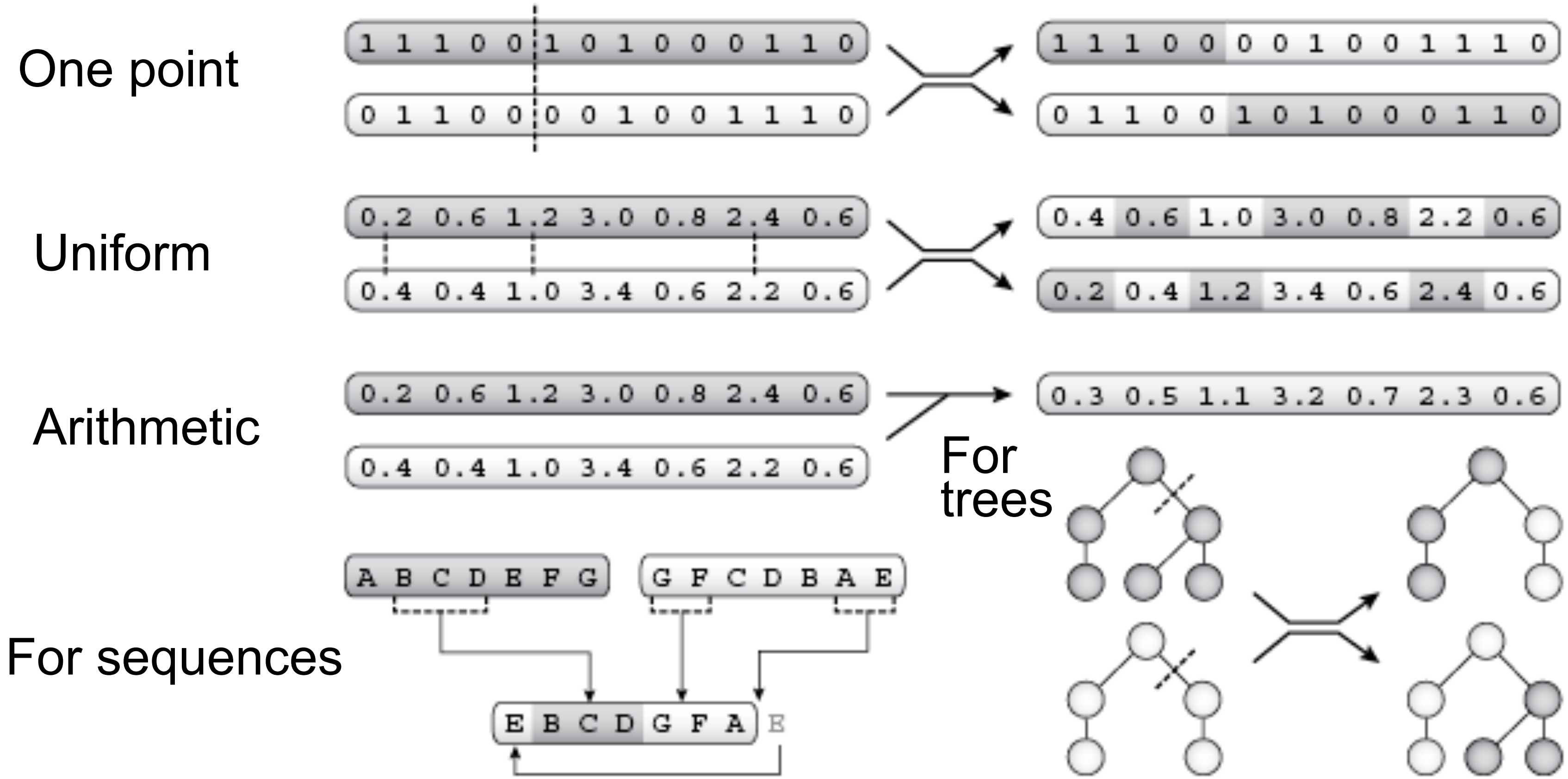


Elitism: maintain ***n*** best individuals from previous generation to prevent loss of best individuals by effects of mutations or sub-optimal fitness evaluation

Generational rollover: insert offspring at the place of worst individuals

Crossover

Emulates recombination of genetic material from two parents during meiosis
Exploitation of synergy of sub-solutions (building blocks) from parents
Applied to randomly paired offspring with probability p_c(pair)



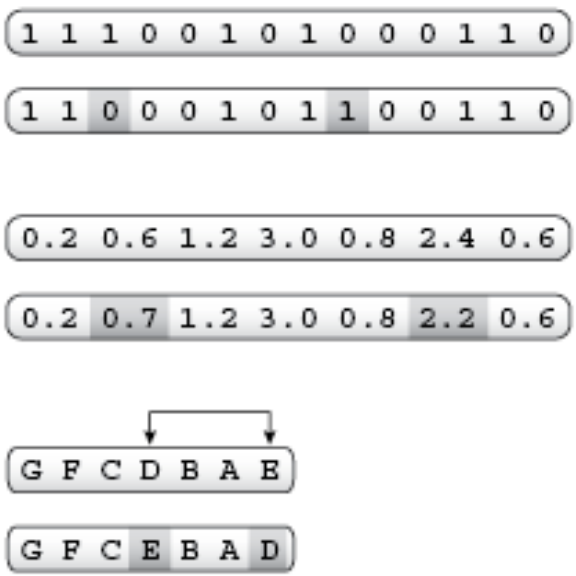
Mutation

Emulates genetic mutations
Exploration of variation of existing solutions
Applied to each character in the genotype with probability $p_m(char)$

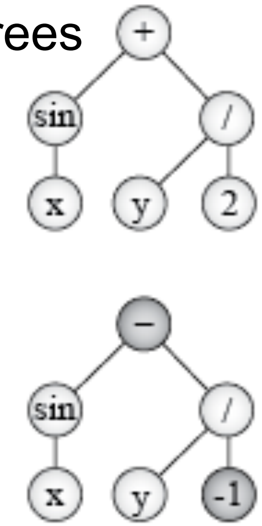
Binary genotypes

Real-valued genotypes

Sequence genotypes

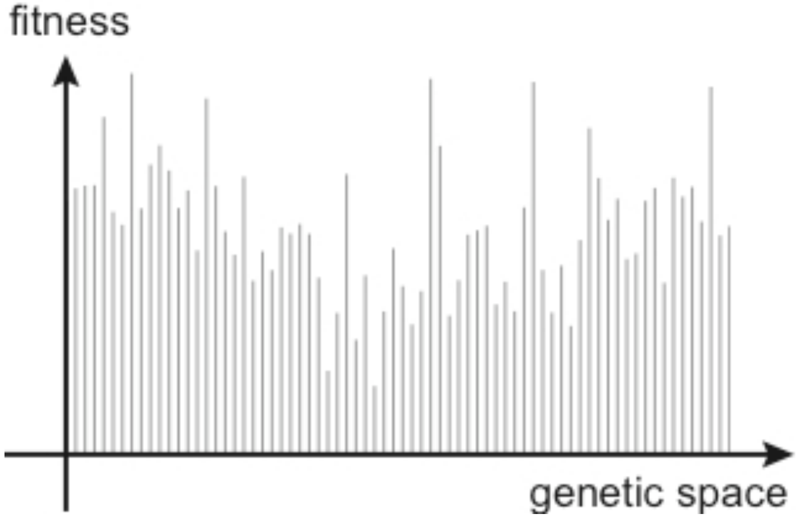


For trees



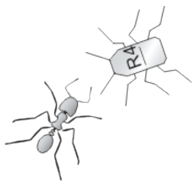
Assessing Fitness Landscape

Fitness landscape is a plot of fitness values associated to all genotypes
Real landscape is unknown; estimation helps to assess evolvability
Goal of evolution is to find genotype with best fitness
Navigation depends on genetic operator; landscape metaphor is misleading



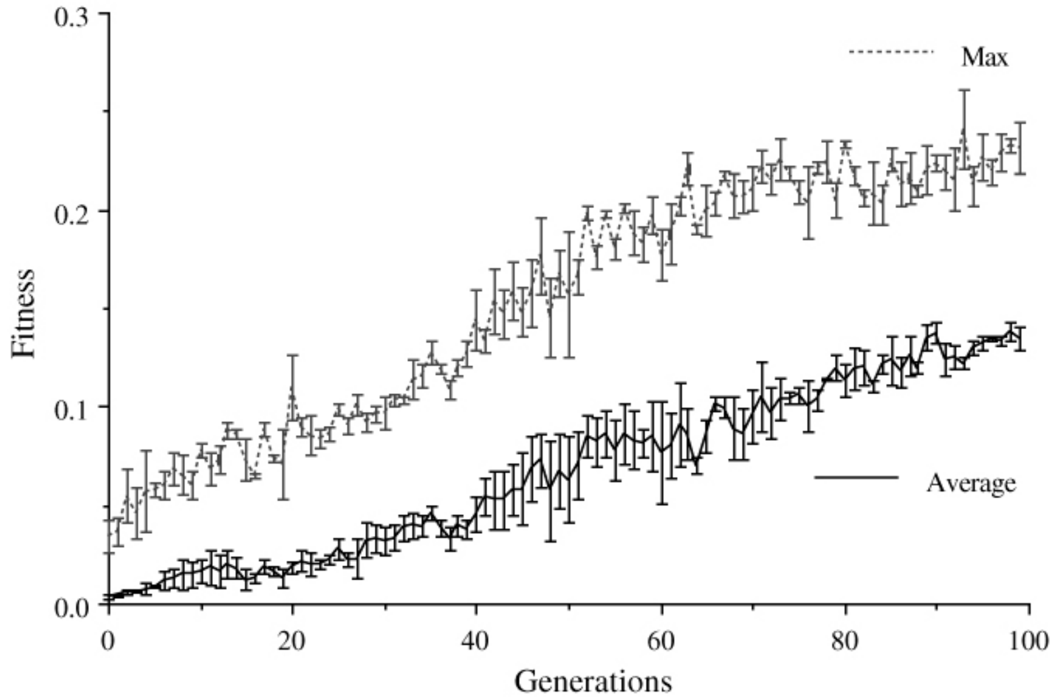
Estimating ruggedness of real landscape:

- 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 easier is to evolve

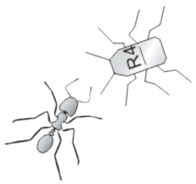


Monitoring Performance

Track best and population average fitness of each generation
Multiple runs are necessary: plot average data and standard error



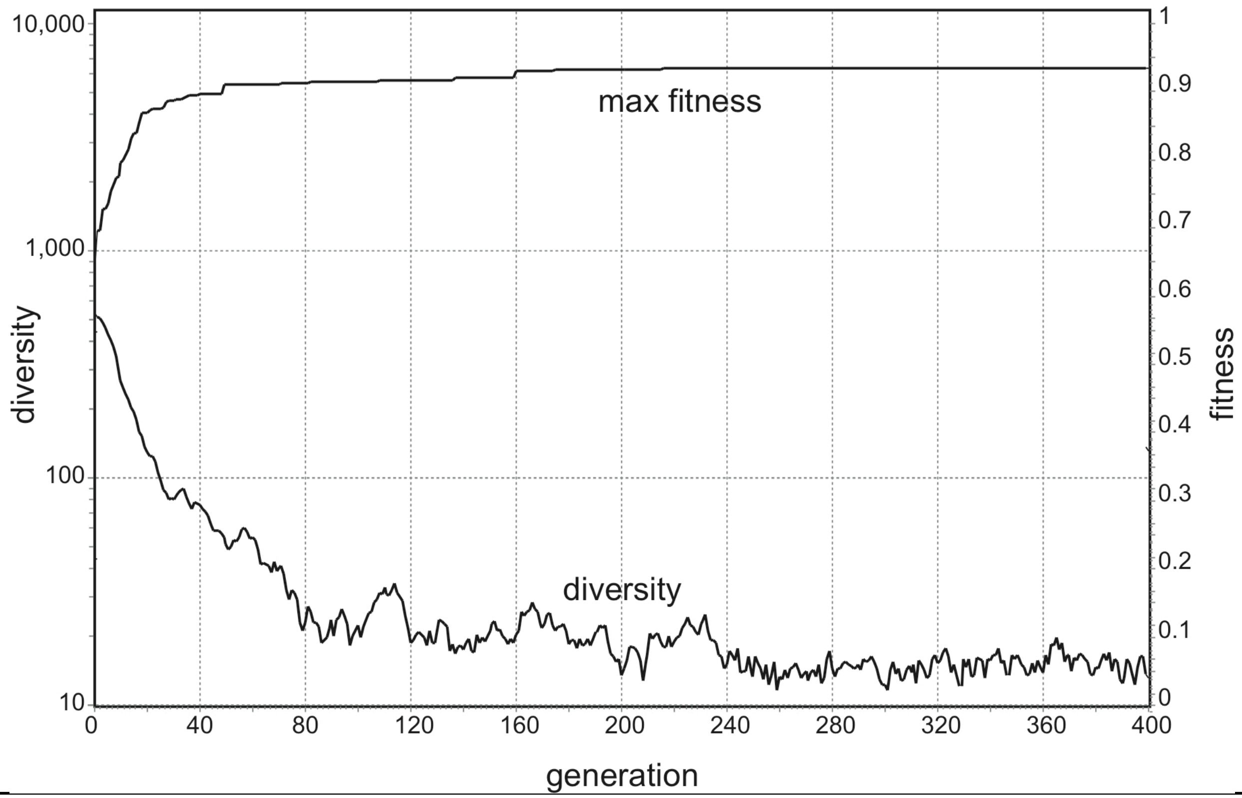
- Fitness graphs are meaningful only if the problem is stationary!
- Stagnation of fitness function may mean best solution found or premature convergence



Measuring Diversity

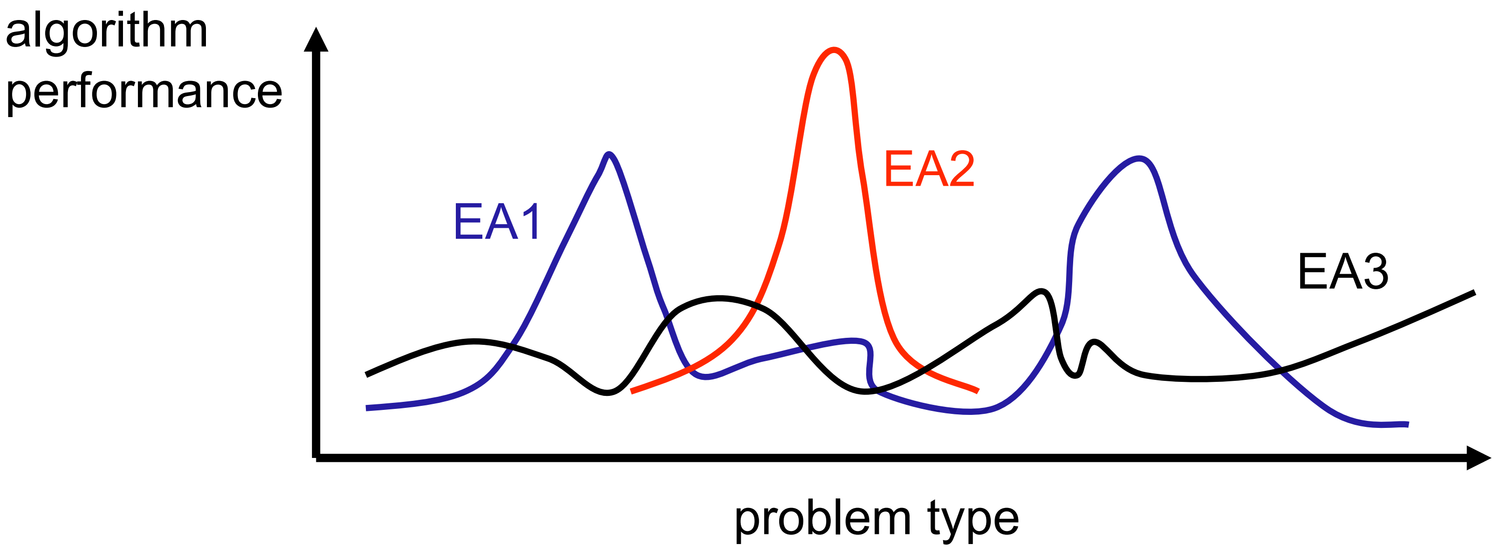
Diversity tells whether the population has potential for further evolution
Measures of diversity depend on genetic representation
E.g., for binary and real valued, use sum of Euclidean or Hamming distances

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



Applicability

- Evolutionary algorithms are used in a huge number of problems
- Biological inspiration is essential, but often distorted
- Different problems may require different algorithms



Knowledge of problem domain can help to choose or assemble best algorithm

Major Evolutionary Algorithms

• **Genetic Algorithms** (GA) - Holland, 1975
Binary genotypes, crossover and mutation

• **Genetic Programming** (GP) - Koza, 1992
Tree-based genotypes, crossover and mutations

• **Evolutionary Programming** (EP) - Fogel etal., 1966
Real-valued genotypes, mutations, tournaments, gradual pop. replacement

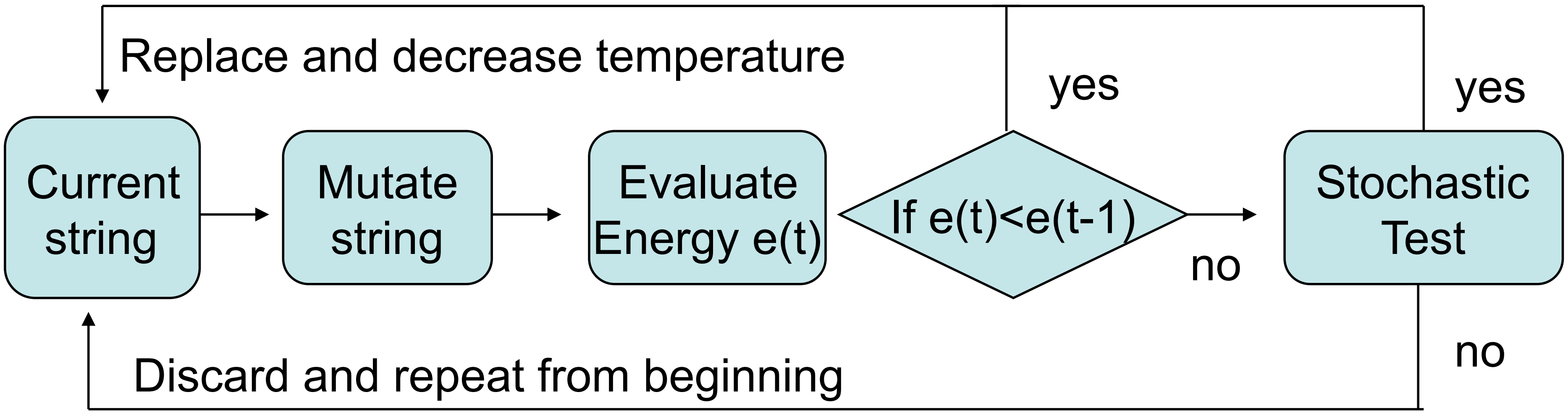
• **Evolutionary Strategies** (ES) - Rechenberg, 1973
As EP + mutation range encoded in genotype of individual

• **Island Models** – Whitley et al., 1998
Parallel evolving populations with rare migration of individuals

• **Steady-State Evolution** – Whitley et al., 1988
Gradual replacement: Best individuals replace replace worst individuals

Simulated Annealing

A heuristic process inspired on metallurgy
It works on mutations of a single string
It attempts to find string with minimum energy (1/fitness)



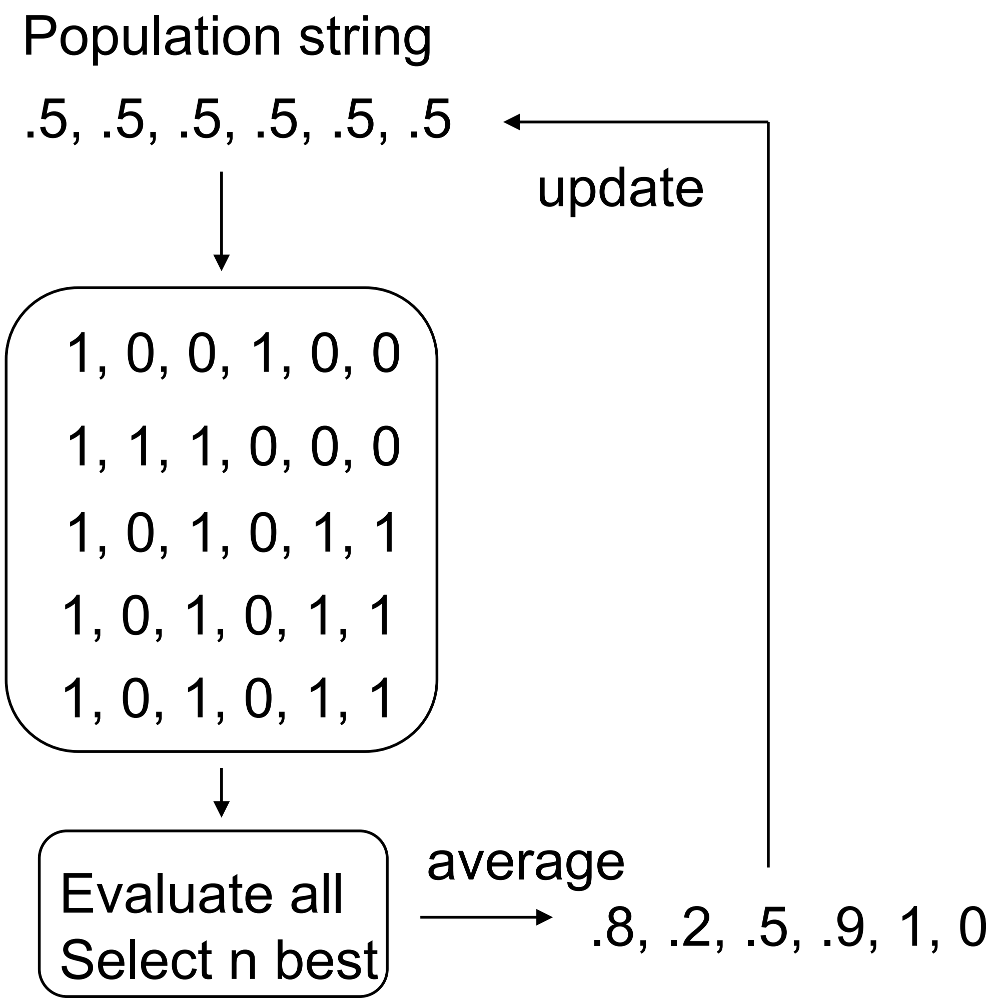
Stochastic test: p(Energy diff, Temp): high temperature, more likely to replace

Population-Based Incremental Learning

Binary representations

Population Vector

Sample, evaluate, update



Human-Competitive Evolution Prize

Annual Competition for best evolved solutions

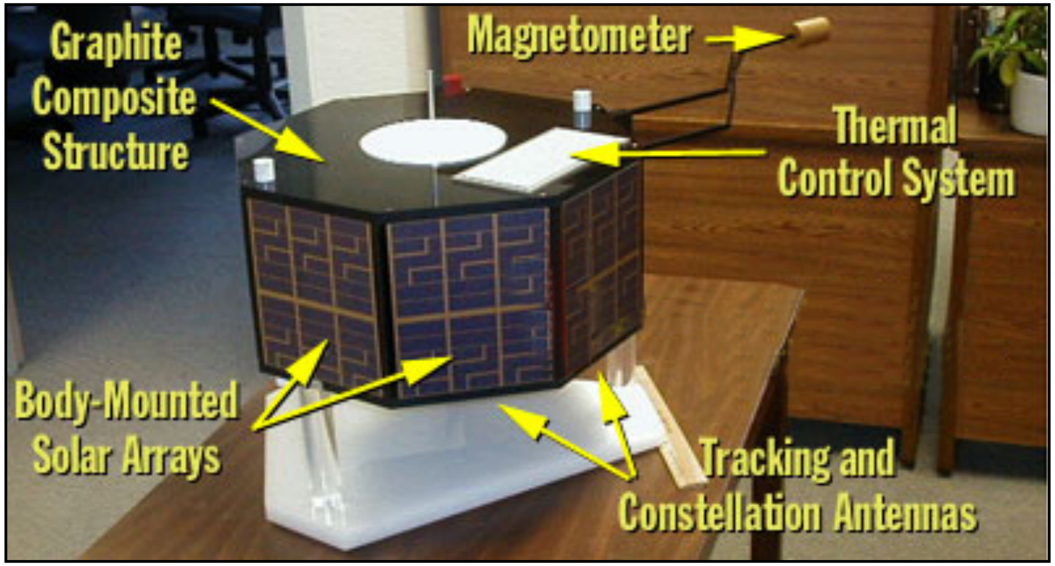
CRITERIA based on:

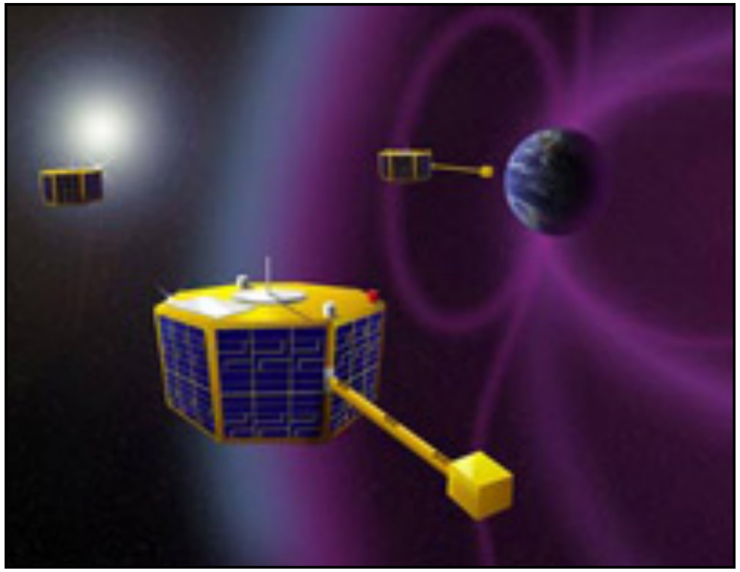
- Patentable or improves on existing patent
- Publishable in peer-reviewed journal
- Solution to long-standing hard problem
- Wins in direct competition with human players or other programs

Antenna for Nanosatellites

ST5 mission: Measure effect of solar activity on the Earth's magnetosphere
3 nanosatellites (50 cm)
Design of antenna to send data to ground station

[Lohn, Hornby, Linden, 2004]

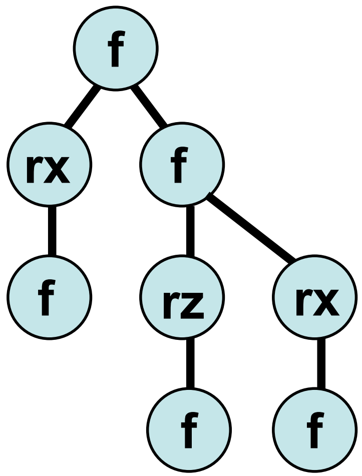


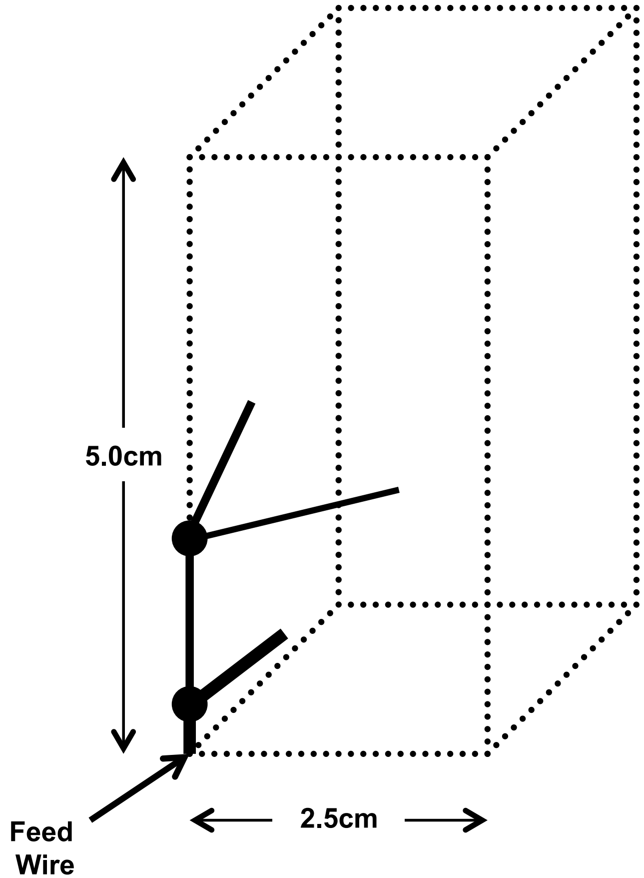


Genetic Representation

Tree-based Encoding
Execute instructions from root to leaves
Evaluate fitness according to specs in simulation
Build best and test in anechoic chamber

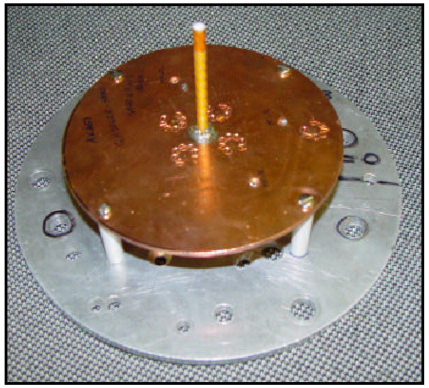
Function Set
f = forward (length, radius)
r x/y/z = rotate x/y/z
Terminal Set
Length, radius, x, y, z
Constraint:
max. 3 branches for each f node



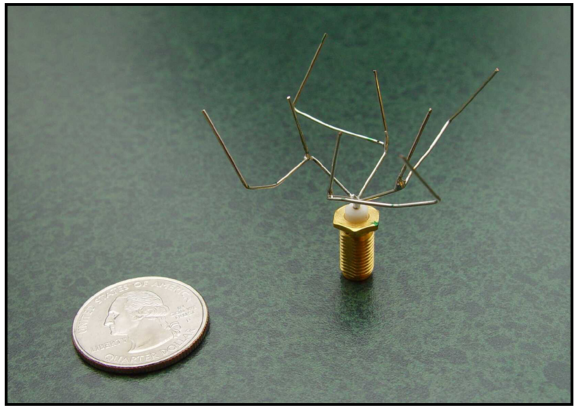


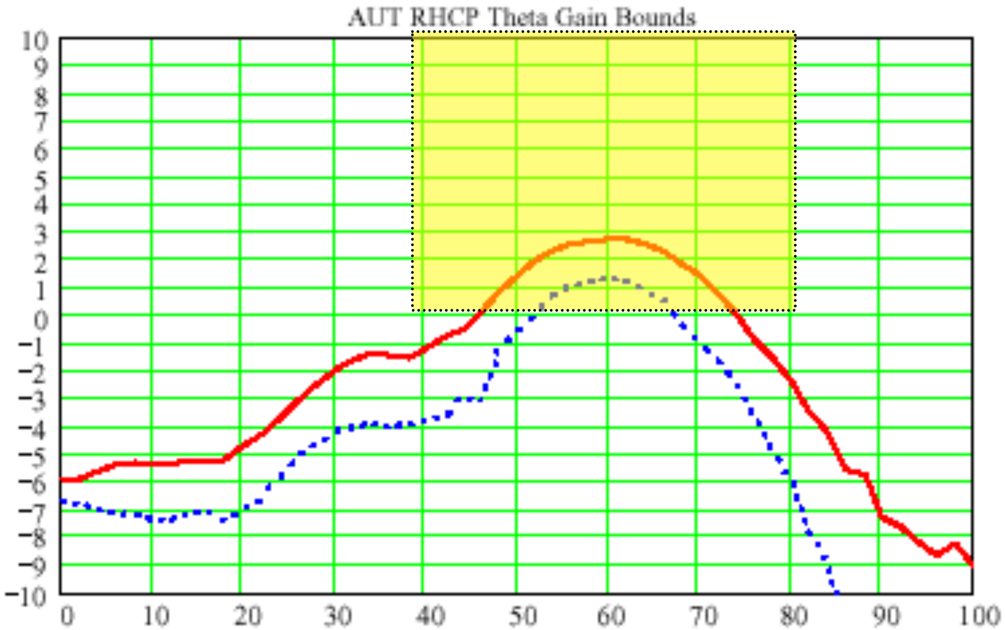
Comparison human/evolved

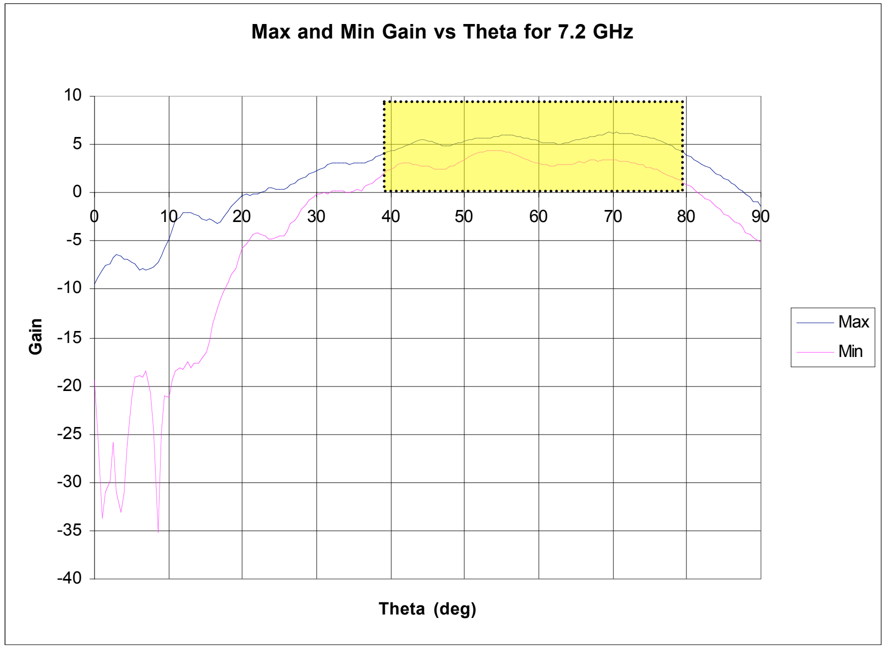
Human



Evolved

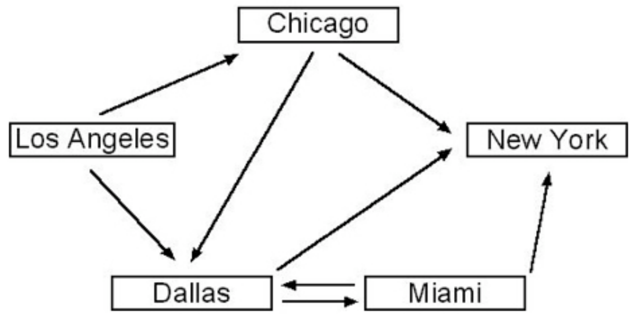




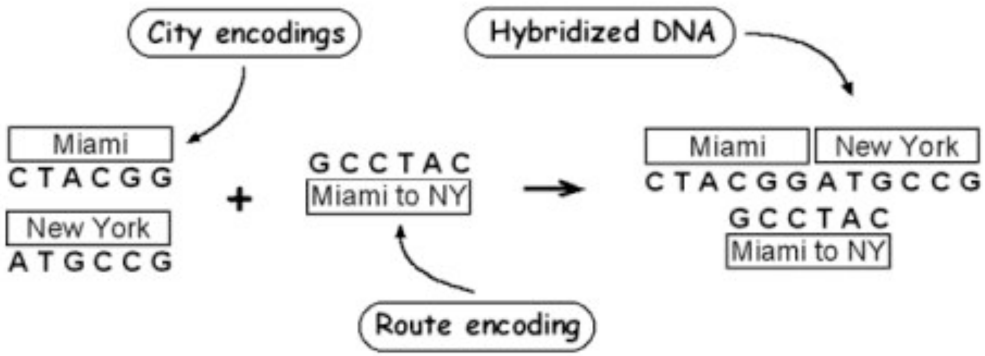


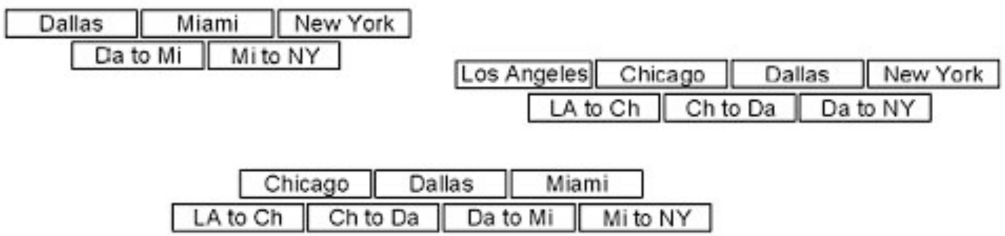
DNA Computing

See also *DNA primer* on
www.howstuffworks.com



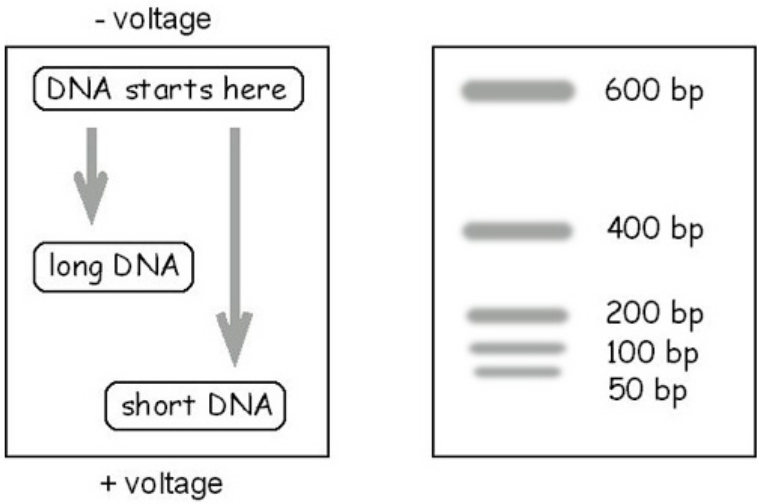
Exploit replication and binding of biological DNA to solve large combinatorial problems: e.g. TSP (Adleman, 1994)





- Create DNA cities
- Create DNA routes
- Make many copies with PCR
- Mix all DNA strings in test tube

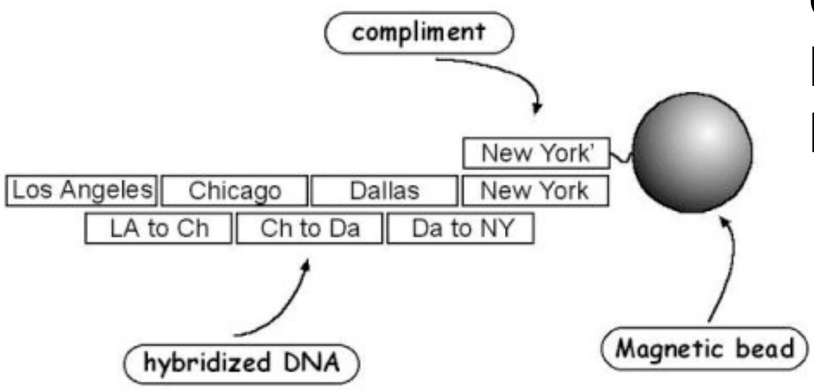
DNA Computing (cont'd)



Select only strands with good size
(e.g., 5 cities x 6 bases = 30 bases)

Create baits:
compliment of city + magnetic bead

Fish out all strands starting with 1st city
Out of those, fish out all strands with 2nd city
Out of those, fish out all strands with 3rd city
Out of those, fish out all strands with 4th city
Repeat for all cities until…
Remaining strands represent the best route



Problems of "DNA computers":
-number of strands
- laborious process
- restricted to few problems