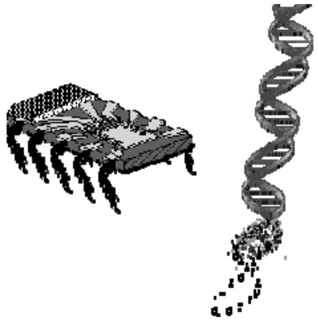


Genetic Algorithms

Jean-Louis Dessalles

ParisTech
Ecole Nationale Supérieure des Télécommunications

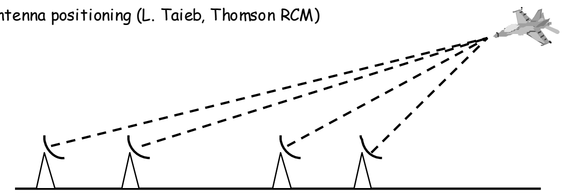
Genetic Algorithms

- **Concrete applications**
- **History**
  - Ch. Darwin, G. Mendel, J. Holland
- **Two didactic examples**
  - Binary sum
  - Labyrinth
- **Important concepts**
  - Selection, crossover, mutations, phenotype
  - Punctuated equilibria
  - Implicite parallelism
  - Schemata
- **Genetic programming**

Some applications of genetic algorithms

- **Optimisation problems**

Antenna positioning (L. Taieb, Thomson RCM)



- **Ill-defined problems**

ex: emergence of a communication code

W-D

Some applications of genetic algorithms

Oil flow optimization in a pipe-line (Goldberg 1989)

Jet engine turbine (General Electric)

Protein structure prediction (Schulze & Kremer 1992)

Message routing in telephone or data networks (Cox, Davis & Qiu)

Antenna positioning (L. Taieb, Thomson RCM)

Méga-Joule laser (CEA): phase blade profile optimization

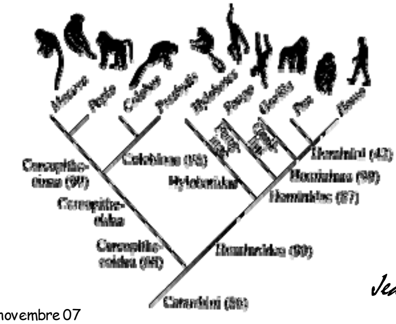
http://neo.lcc.uma.es/TutorialEA/semEC/cap03/cap_3.html

History

- **Evolution**
  - Jean-Baptiste Lamarck (1801): How species get transformed
  - Charles Darwin (1859): Natural selection and blind diversity generation (variations)
- **Genetics**
  - Gregor Mendel (1873): heredity is digital
- **Genetic Algorithms**
  - John Holland (1965): concept of schema and implicite parallelism
  - David Goldberg (1989): popularization

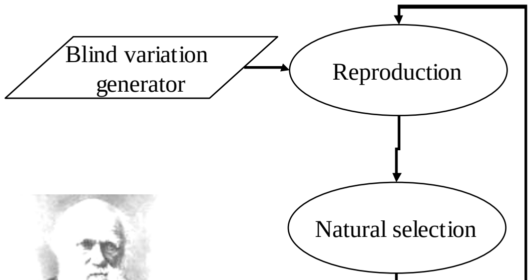
Evolution

1801



Jean-Baptiste Lamarck

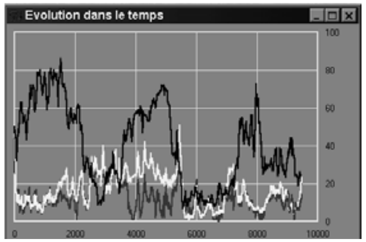
Evolution : Chance and necessity



Charles Darwin

1859

Fate of a mutation



favo

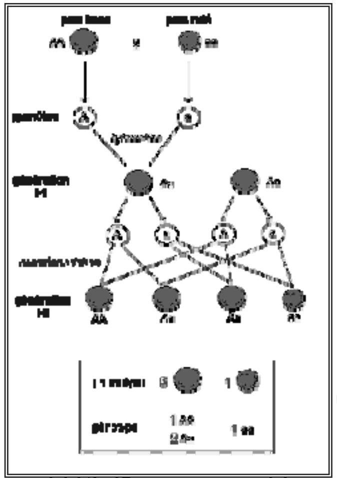
THE ORIGIN OF SPECIES

by Charles Darwin
1859

[…] we clearly see that the nature of the conditions is of subordinate importance in comparison with the nature of the organism in determining each particular form of variation;- perhaps of not more importance than the nature of the spark, by which a mass of combustible matter is ignited, has in determining the nature of the flames.

The evidence that accidental mutilations can be inherited is at present not decisive; but the remarkable cases observed by Brown-Sequard in guinea-pigs, of the inherited effects of operations, should make us cautious in denying this tendency.

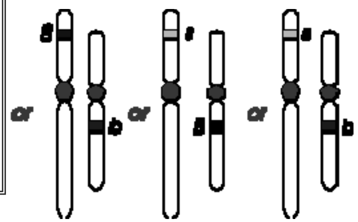
For peculiar habits confined to the workers or sterile females, however long they might be followed, could not possibly affect the males and fertile females, which alone leave descendants. I am surprised that no one has hitherto advanced this demonstrative case of neuter insects, against the well-known doctrine of inherited habit, as advanced by Lamarck.



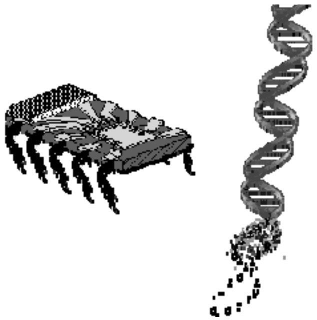
Heredity: a digital phenomenon

1873

Gregor Mendel



Genomes



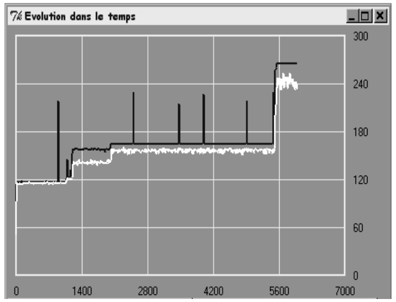
Aids virus :	9718 p.d.b.
Bacteria:	$5 \cdot 10^6$ p.d.b.
humans:	$3{,}3 \cdot 10^9$ p.d.b.

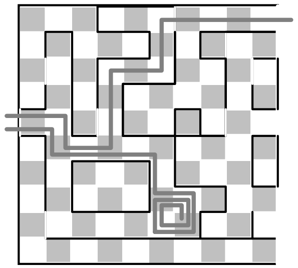
Genetic algorithms: a biological metaphore for engineers

- **Evolution through natural selection is an optimizer**
- **It is an efficient optimizer (rapidity)**
  - *Concept of schema: John Holland 1965*

John Holland

Punctuated equilibria

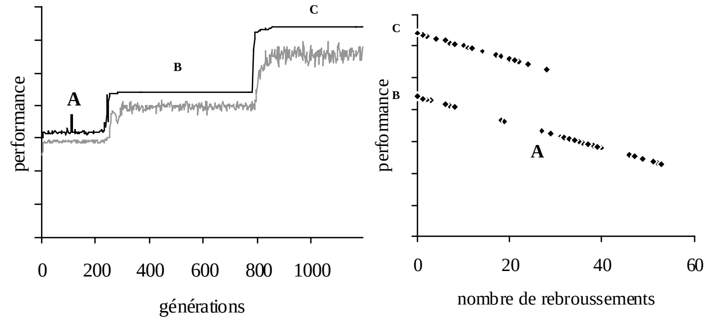




demo

1

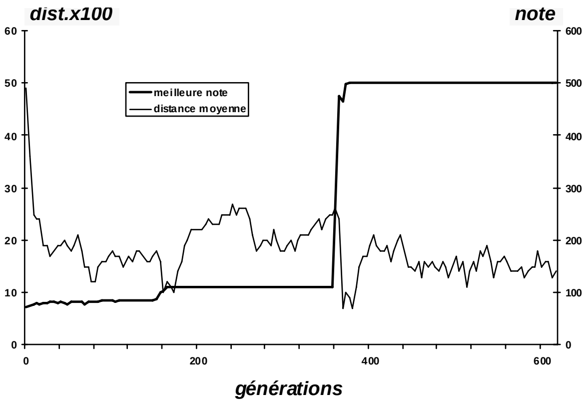
Punctuated equilibria



demo

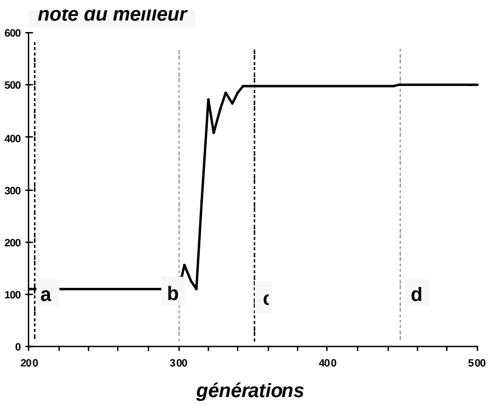
2

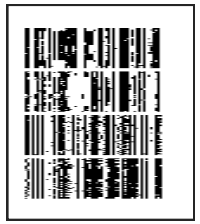
Punctuated equilibria



3

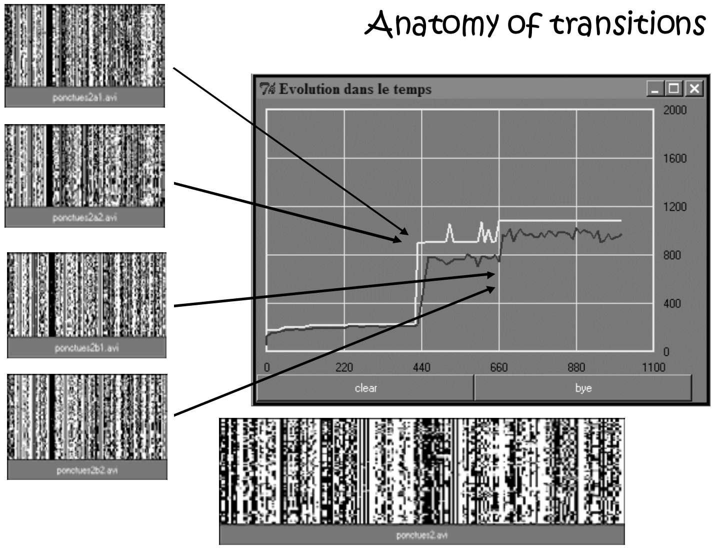
Punctuated equilibria





4

Anatomy of transitions



5

(Pagel, Venditti & Meade 2006)

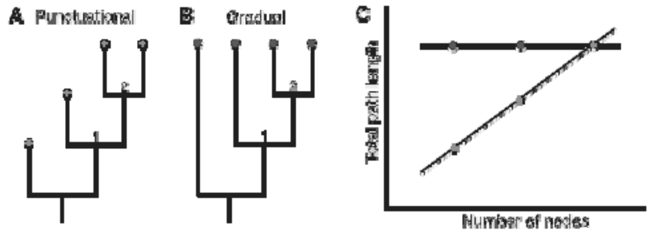
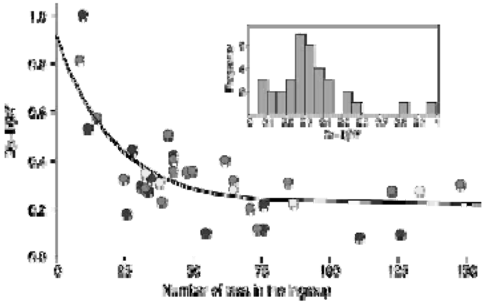


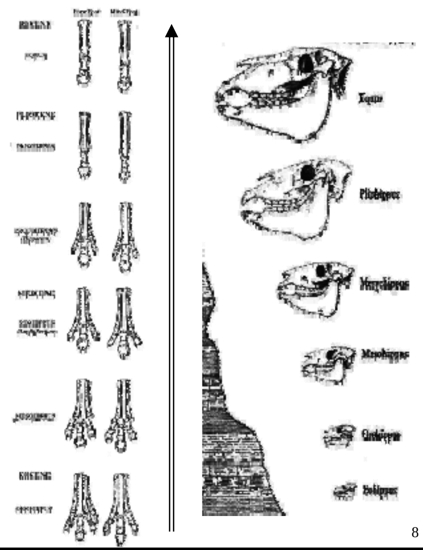
Fig. 1. Signatures of punctuational and gradual evolution on phylogenetic trees. (A) Punctuational evolution presumes a burst of evolution associated with each node of the tree. Path lengths, measured as the sum of branches along a path from the root to the tips of the tree, are proportional to the number of nodes along that path (C). Branches are assumed to be in units of nucleotide substitutions. (B) Gradual evolution presumes that change is independent of speciation events. Path lengths do not correlate with the number of nodes along a path (C). (C) Punctuational evolution predicts a positive relationship between path length and the number of nodes, whereas gradual evolution does not.

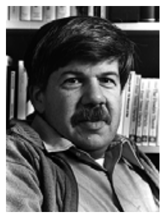


A bifurcating phylogenetic tree has 2(s – 1) branches where s is the number of species. The length of a phylogenetic tree, T, is the sum of the lengths of these branches, here measured in units of nucleotide substitutions.
The ratio 2(s – 1)β/T, then, measures the proportion of the tree length attributable to punctuational effects.

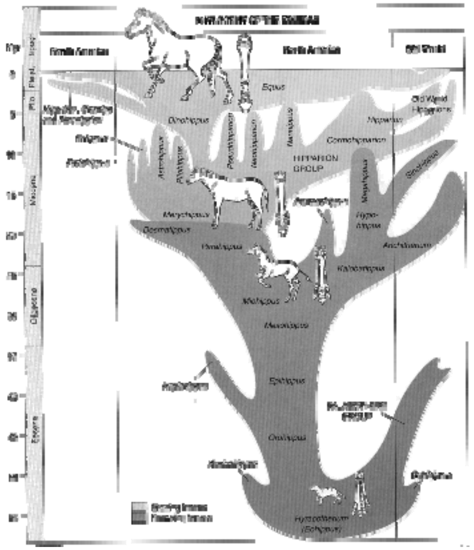
7

Evolution: a directed process ?

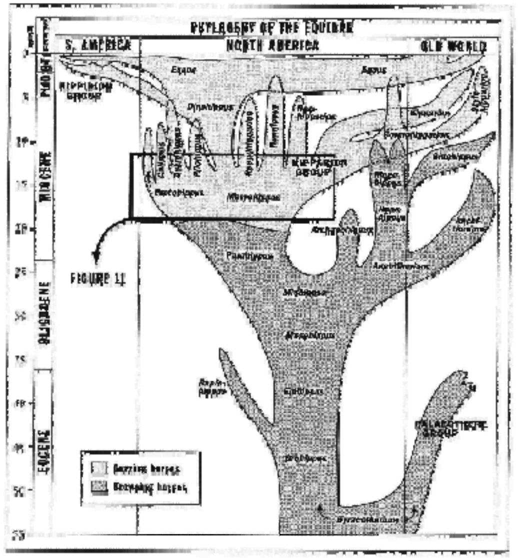




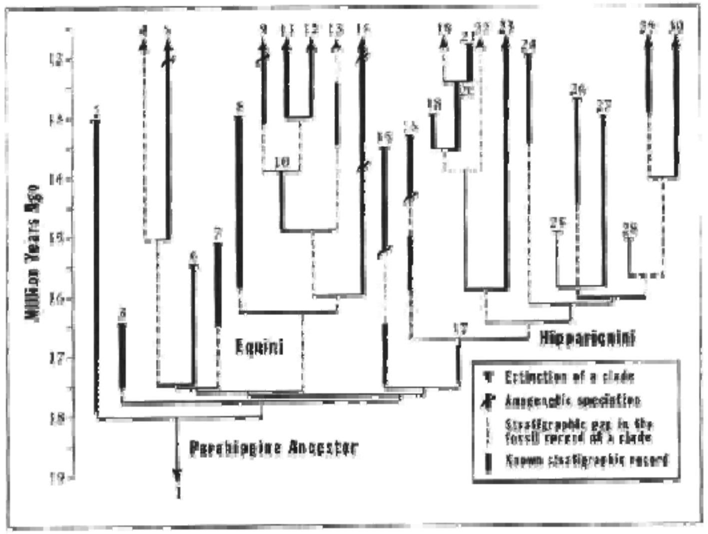
8



9

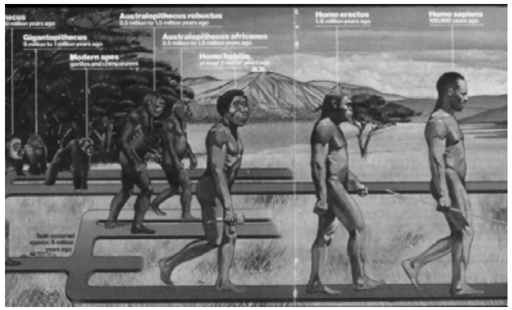


10



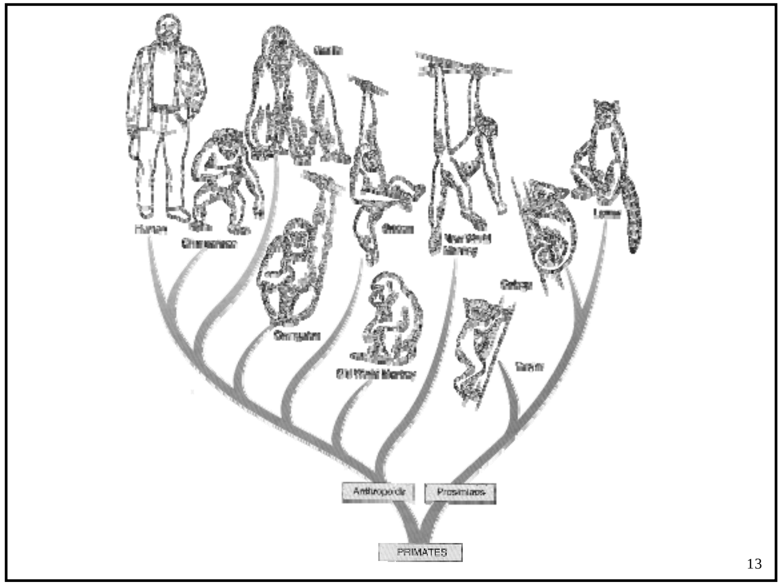
11

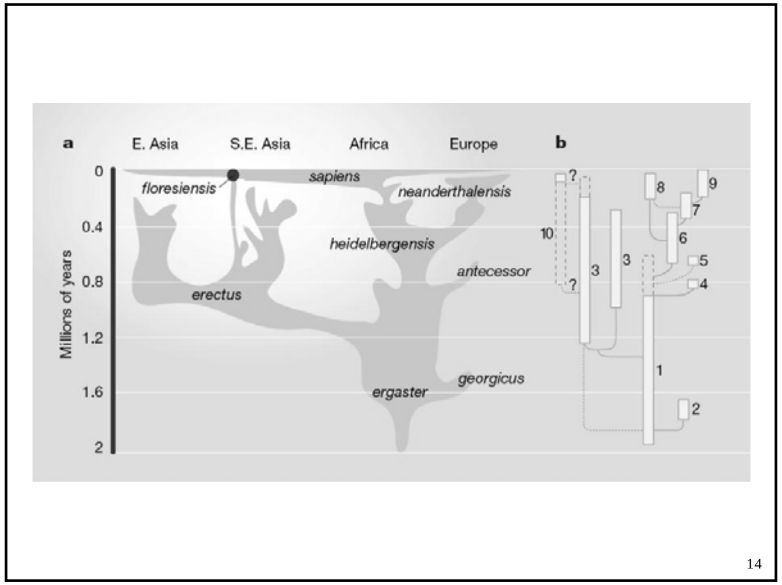
Evolution:
a directed process ?



Evolution through natural selection has no inertia

12


PRIMATES
13


a
E. Asia
S.E. Asia
Africa
Europe
b
Millions of years
0
0.4
0.8
1.2
1.6
2
floresiensis
sapiens
neanderthalensis
heidelbergensis
antecessor
erectus
ergaster
georgicus
14

7

The selfish schema

- **Prediction of evolutionary short-term future**
  - **Selection unit**

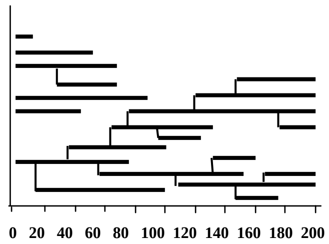
$$N^{t+1} = a \cdot N^t$$

- **Explain the rapidity of evolution**
  - **Implicit parallelism**

1

The unit of selection

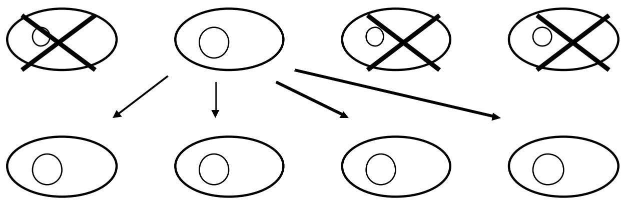
- **Group selection ?**



2

Group selection ?

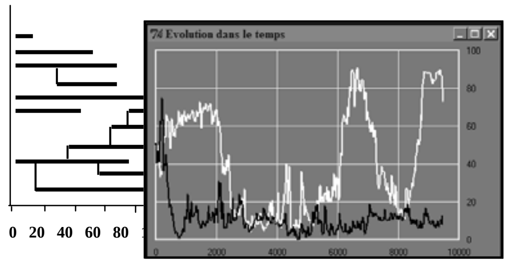
Ideal picture



Evolife

3

Group selection ?



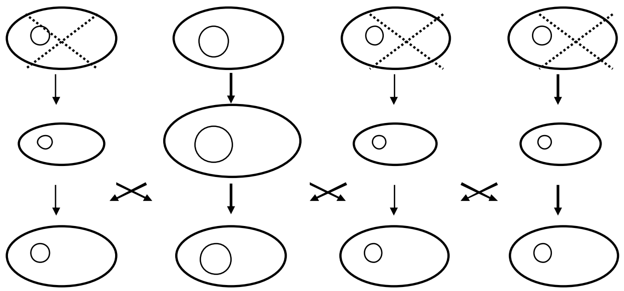
Taux de migration: 2

Evolife

4

Group selection ?

Actual picture

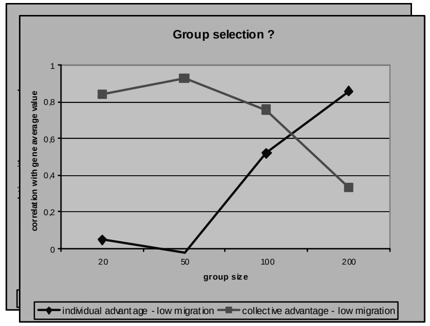


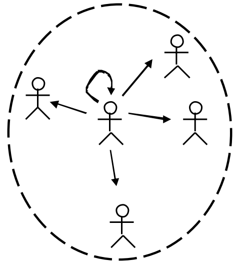
Evolife

Problem: correlation by descent

5

Group selection ?





Benefit of the next mutant:

$\alpha[p(G-1)+1]/G - \alpha pG/G = \alpha(1-p)/G$

Benefit of the group

αpG

α: group benefit per mutant
p: proportion of mutants in the population
G: groupe size

6

The unit of selection

Problem: predict evolution by observing a competition between *selection units*

1. Units U et V can be distinguished by incompatible properties u et v.
2. Units U and units V are in competition. A measure S_U of the average success of U in this competition is supposed to be available.
3. When S_U is greater (resp. smaller) than S_V, the proportion f_U of units U increases (resp. diminishes) systematically w.r.t. f_V.

$S_U > S_V \rightarrow f_U > f_V$

prédiction

7

The unit of selection

Do successful units propagate what makes their différence ?

Individual	≈	
Group	no	Only in case of kin-selection
Schema	yes	

8

The unit of selection

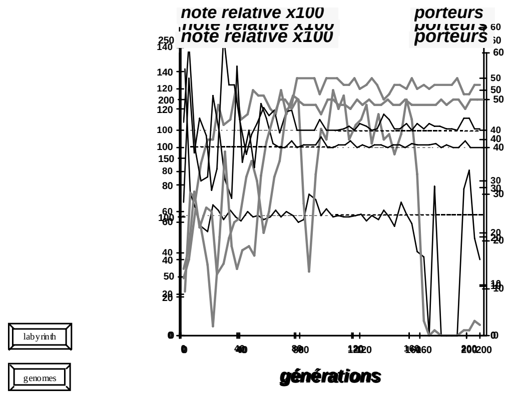
- **Competition between individuals ?**

- Two genes A and B on two different chromosomes
- with alleles A’ and B’
- Suppose that AB > A’B’ > AB’ > A’B
- "winners" AB are unlikely to propagate their différence

- Kin selection: animals such as social insects may not procreate and yet propagate their différence

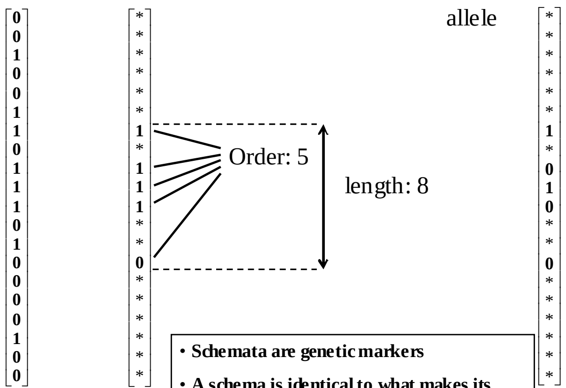
9

Schemata as selection units



10

Schema



- **Schemata are genetic markers**
- **A schema is identical to what makes its difference**

11

Schema

- **Schemata as selection units**
  - **Prediction of future evolution**

$$N_\sigma^{t+1} = a_\sigma \cdot N_\sigma^t$$

- **Schemata as "building blocks"**
  - **Explains why G.A. are so rapid: thanks to implicit parallelism**

12

Schemata as selection units

$$E(i) = K \cdot \left(\frac{A(i)}{\frac{1}{N}\sum_j A(j)} \right)$$

$$N_\sigma = \sum_i \delta_\sigma(i)$$

$$E_\sigma = \sum_i E(i) \cdot \delta_\sigma(i)$$

$$E_\sigma = K \cdot a_\sigma \cdot N_\sigma$$

$$a_\sigma = \frac{1}{N_\sigma} \cdot \sum_i A(i) \cdot \delta_\sigma(i) \Big/ \frac{1}{N} \cdot \sum_j A(j)$$

$$N_\sigma^{t+1} = a_\sigma \cdot N_\sigma^t$$

$$N_\sigma^{t+1} = E_\sigma / 2$$

13

Schemata as selection units

The genetic algorithm
sorts out schemata
in a quasi binary way

$$N_\sigma^{t+1} = a_\sigma \cdot N_\sigma^t$$

$$N_\sigma^{t+1} = E_\sigma / 2$$

14

Survival probability of a schema

- **Influence of crossover**

Chromosome of length $l \rightarrow (l\text{-}1)$ possible cutting points

Survival probability p_s :

$$p_s = \left(\frac{l - l_\sigma}{l - 1} \right)^n \approx 1 - n \frac{l_\sigma - 1}{l - 1}$$

- **Influence of mutations**

$$p_s = (1 - p_m)^{o_\sigma} \approx (1 - o_\sigma p_m)$$

$$N_\sigma^{t+1} = a_\sigma \cdot N_\sigma^t \left(1 - n \frac{l_\sigma - 1}{l - 1} - o_\sigma p_m \right)$$

15

Implicit parallelism

$$l_r \approx 1 + L \cdot (1 - p_r)/n$$

$$2^{l_r} + 2^{l_r - 1}(L - l_r) = 2^{l_r - 1}(L - l_r + 2)$$

$$N \cdot 2^{l_r - 1}(L - l_r + 2)$$

$$N \cdot 2^{l_r - 1}(L - l_r + 2)/2 \leq N_r \leq N \cdot 2^{l_r - 1}(L - l_r + 2)$$

$$N \approx 2^{l_r/2}$$

n individuals

n^3 schemata
processed in parallel

l_r

```
1001010001101010101010101111100001100011010
*0*10**0***********************************

1001010001101010101010101111100001100011010
*001***00**********************************
```

l_r

16