



# 第三章 遗传算法

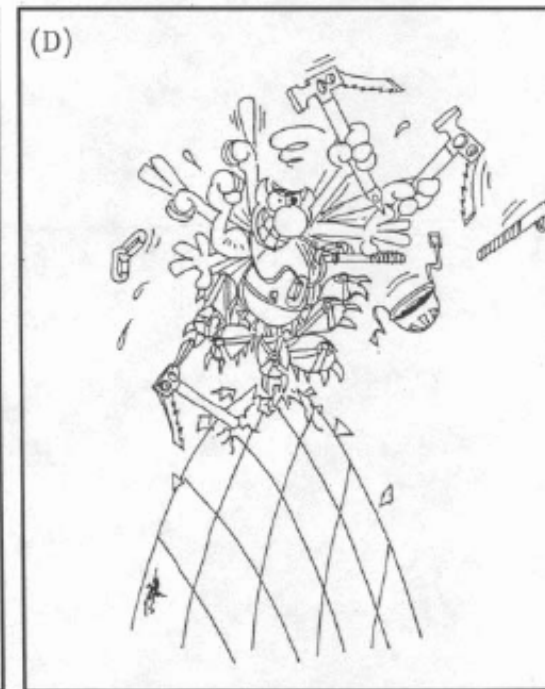
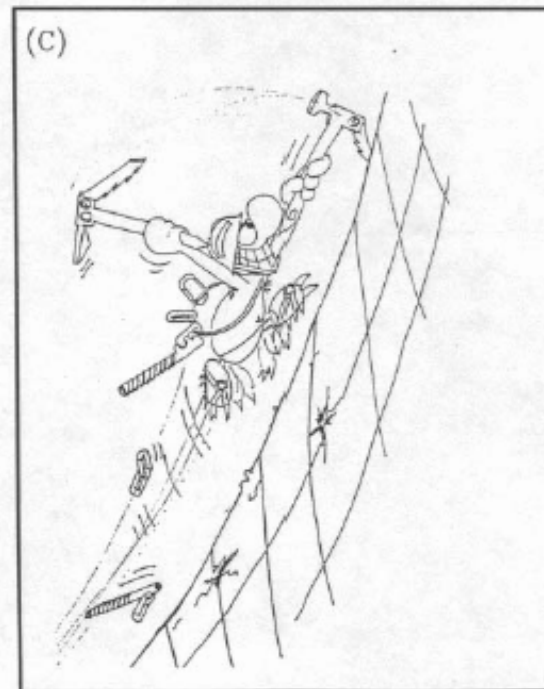
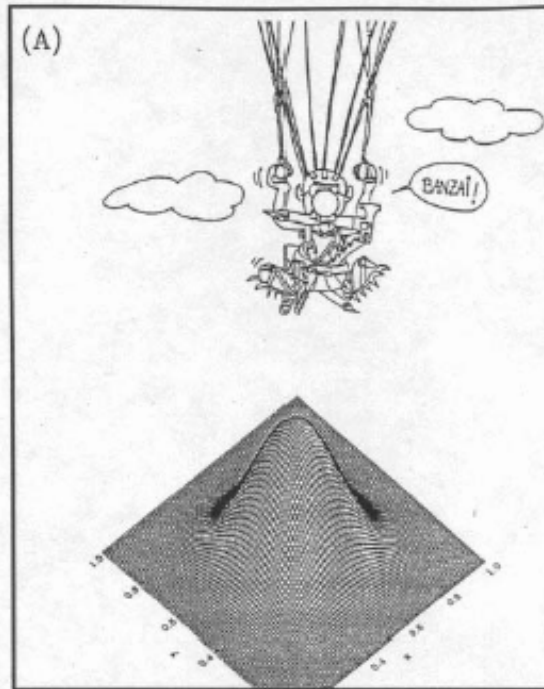
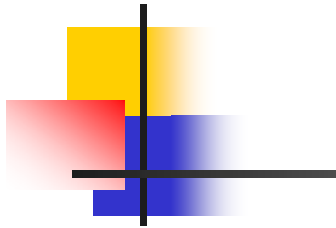
---

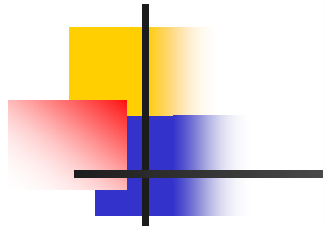


## 第三章 遗传算法

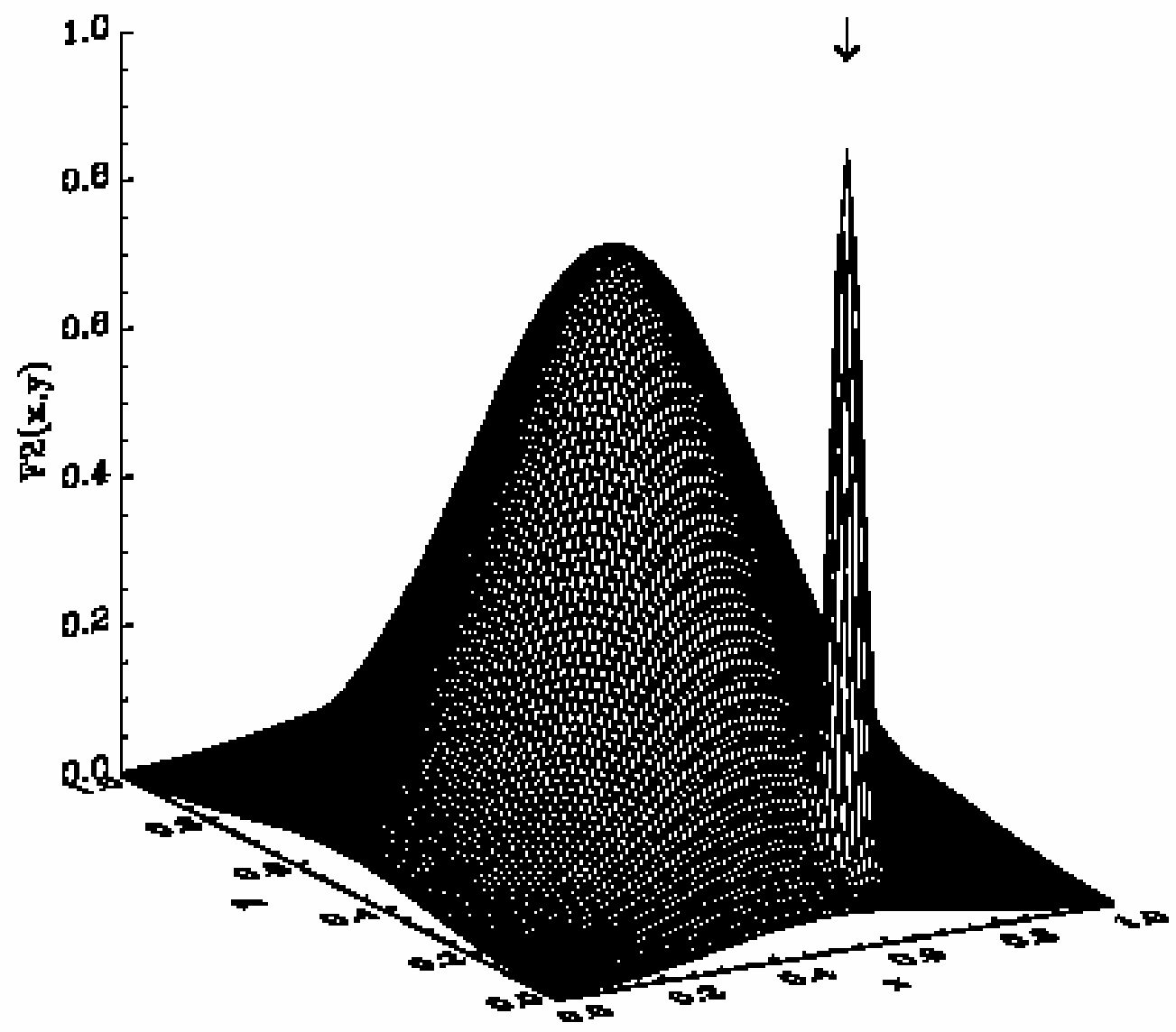
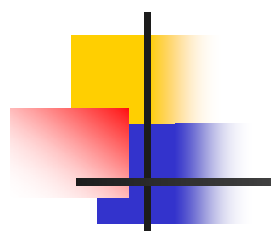
---

- 3. 1 从生物进化到进化计算
- 3. 2 标准遗传算法 (SGA)
- 3. 3 遗传算法的特点
- 3. 4 遗传算法理论研究
- 3. 5 遗传算法的应用
- 3. 6 遗传算法的基本术语

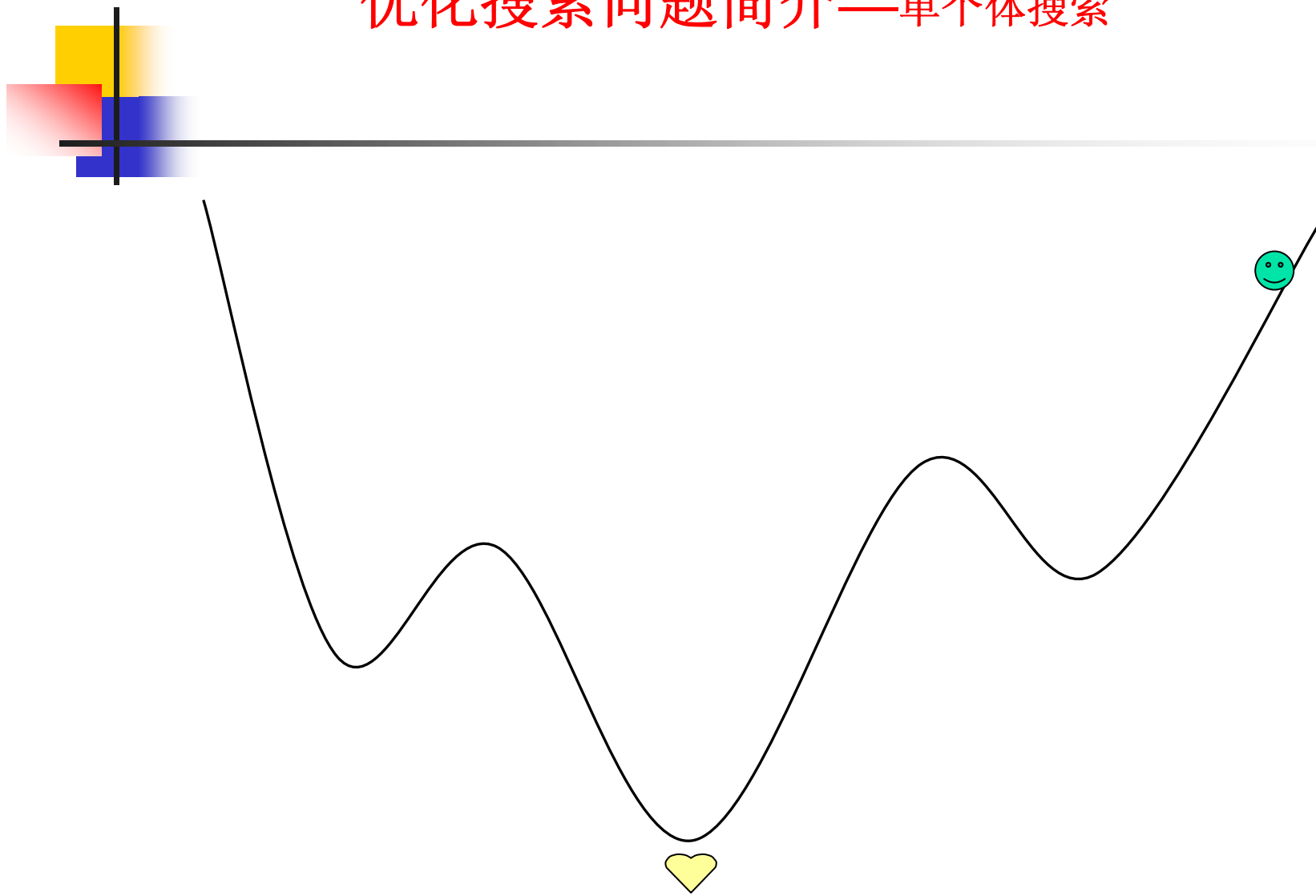




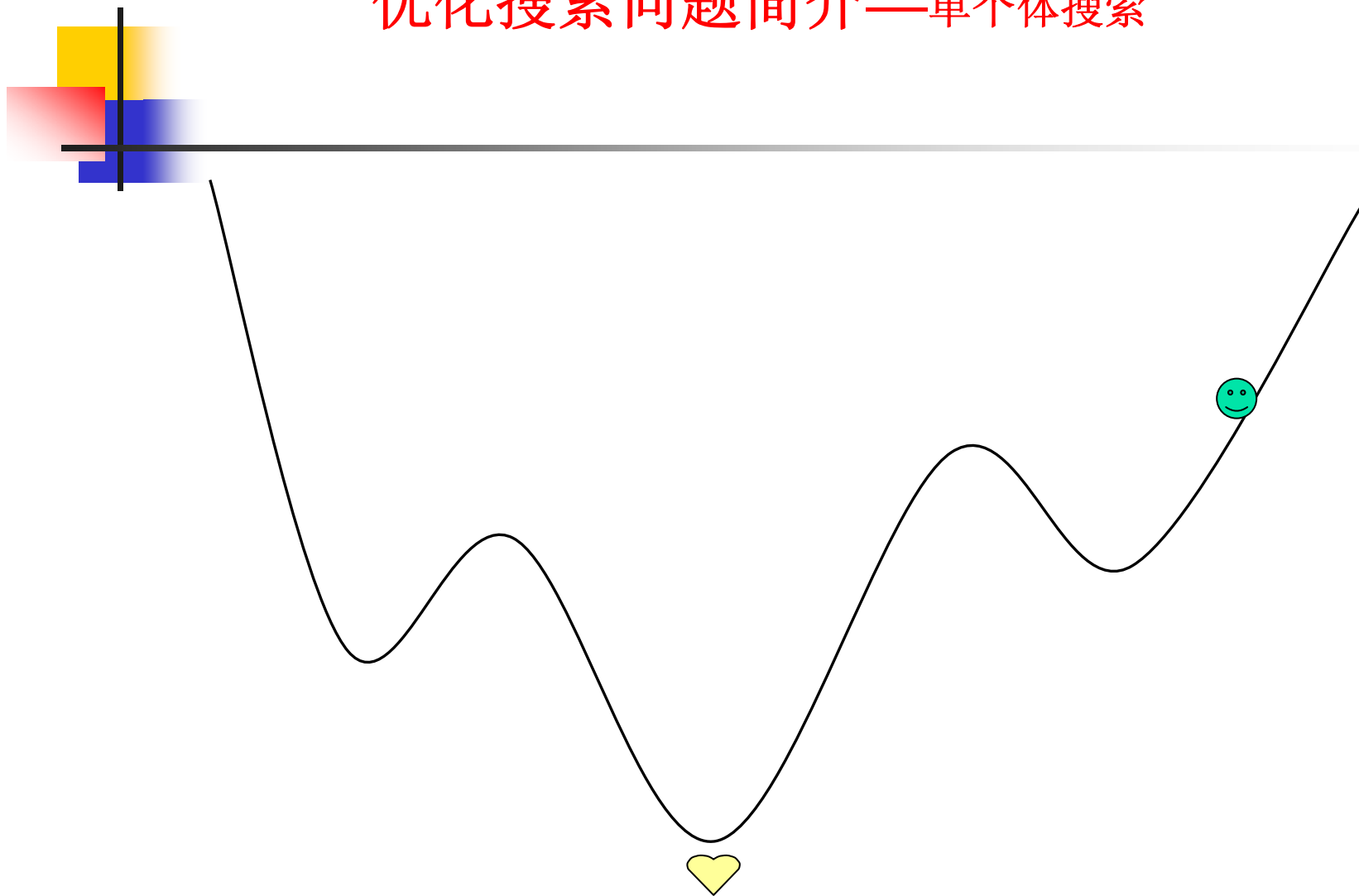
人多力量大  
吼吼吼



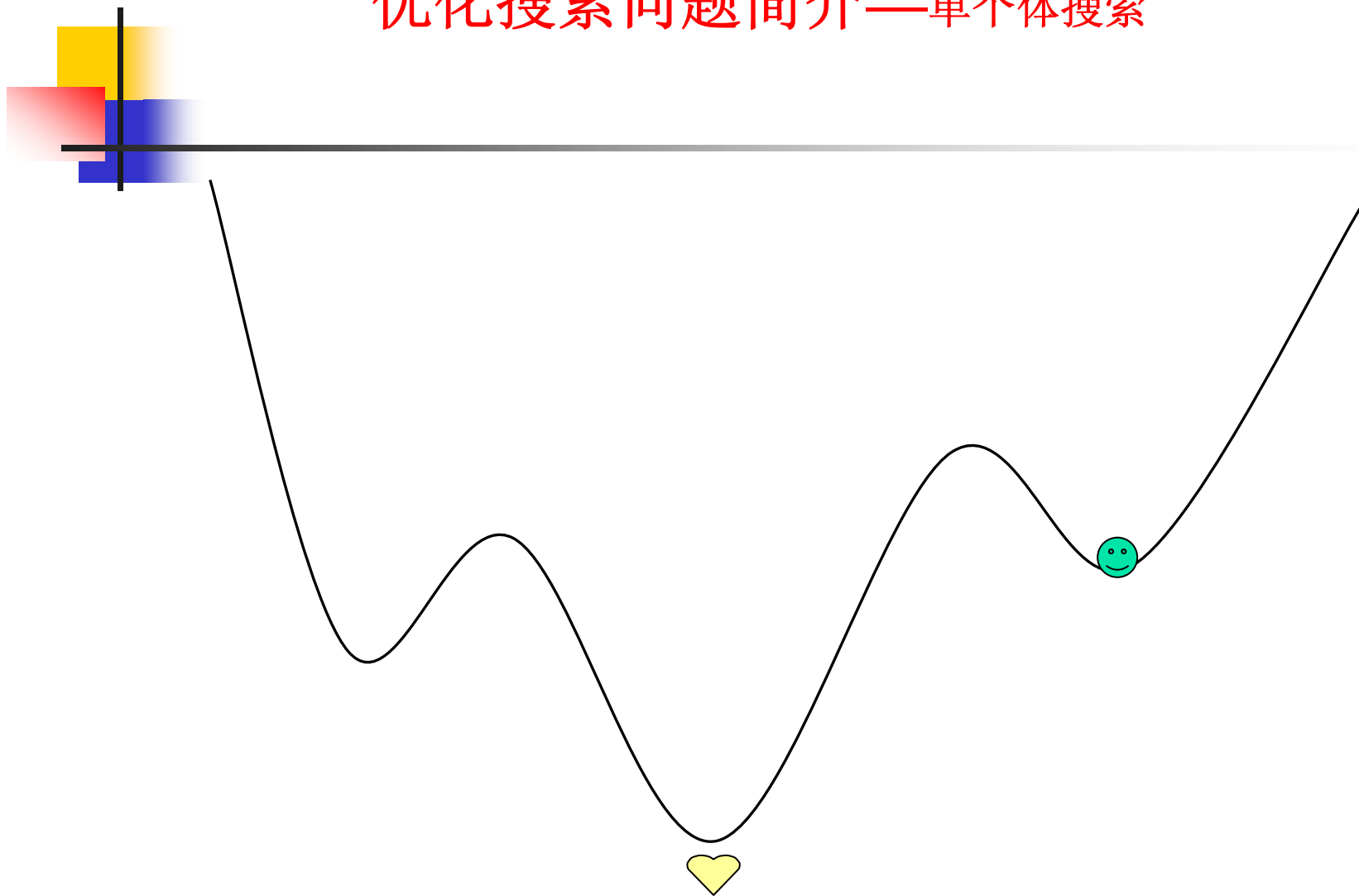
## 优化搜索问题简介—单个体搜索



## 优化搜索问题简介——单个体搜索

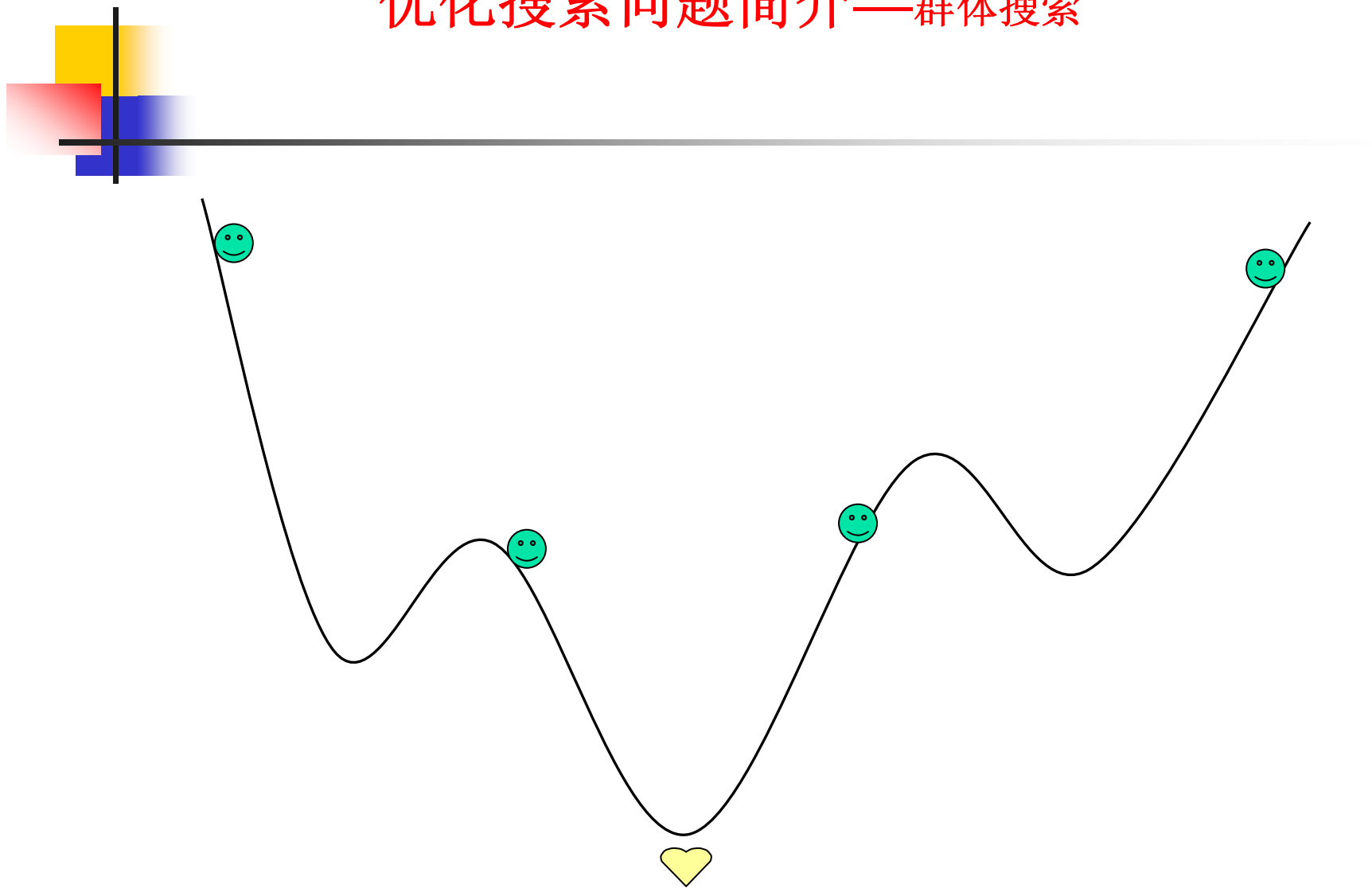


## 优化搜索问题简介—单个体搜索

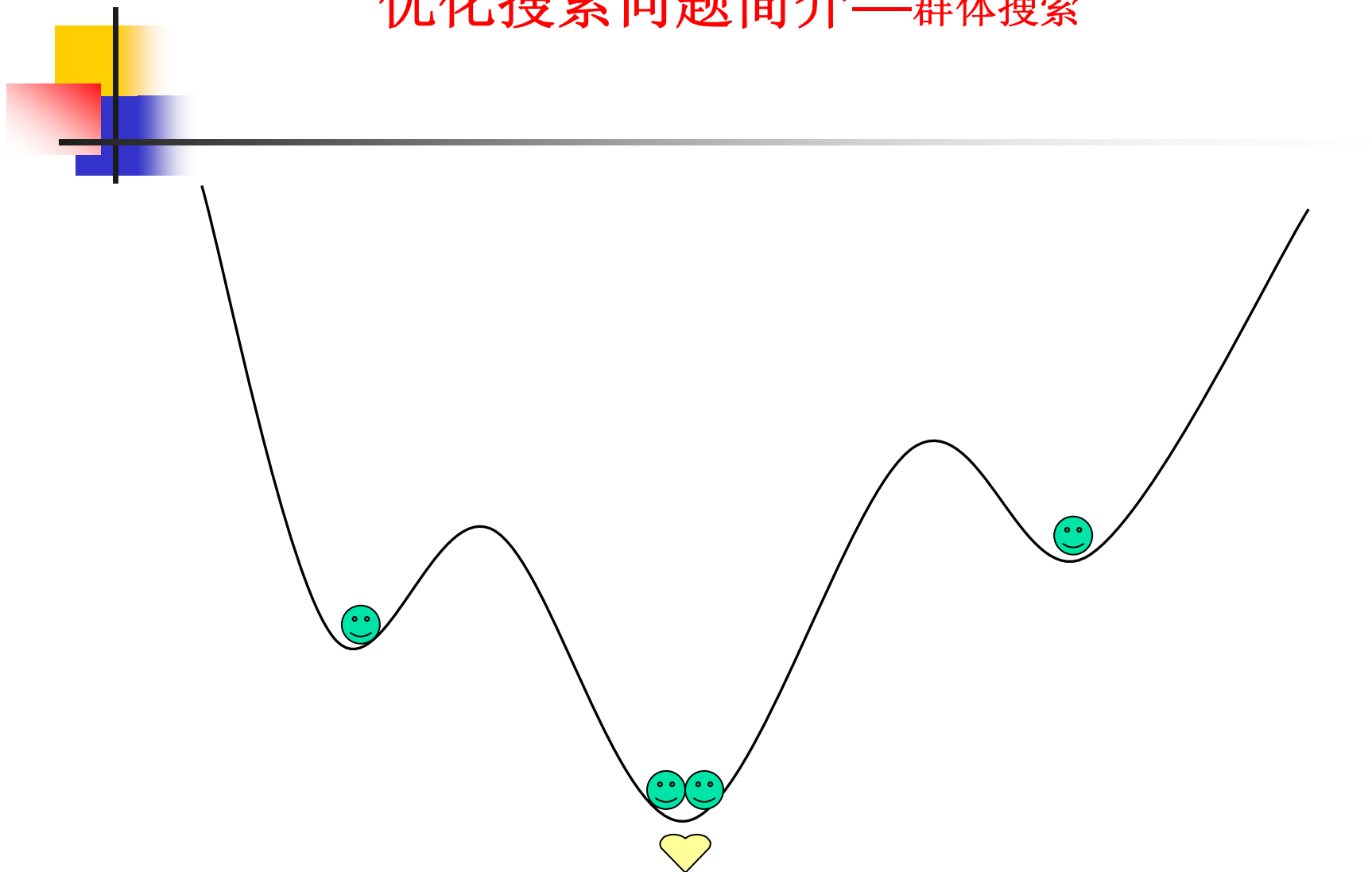




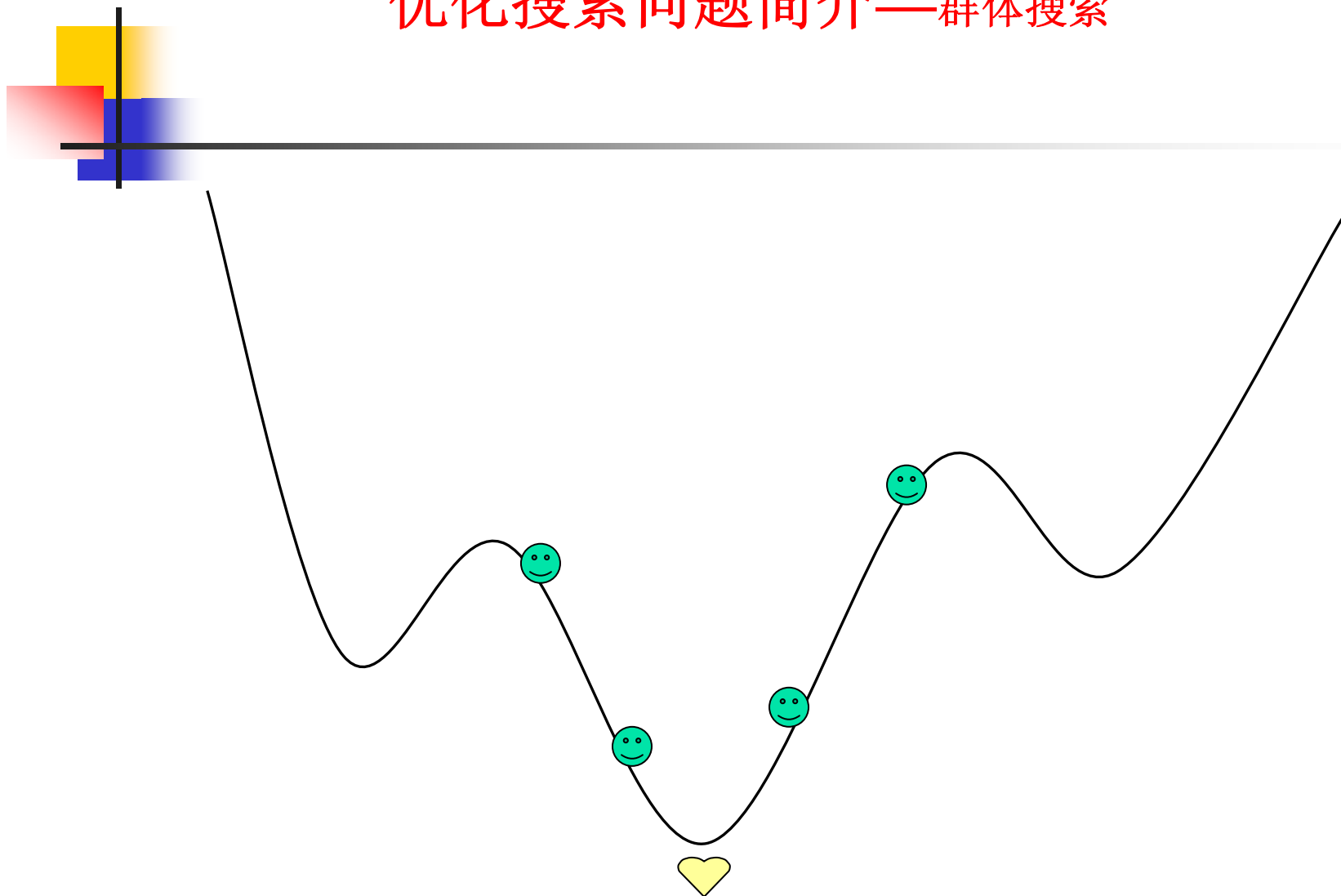
## 优化搜索问题简介——群体搜索



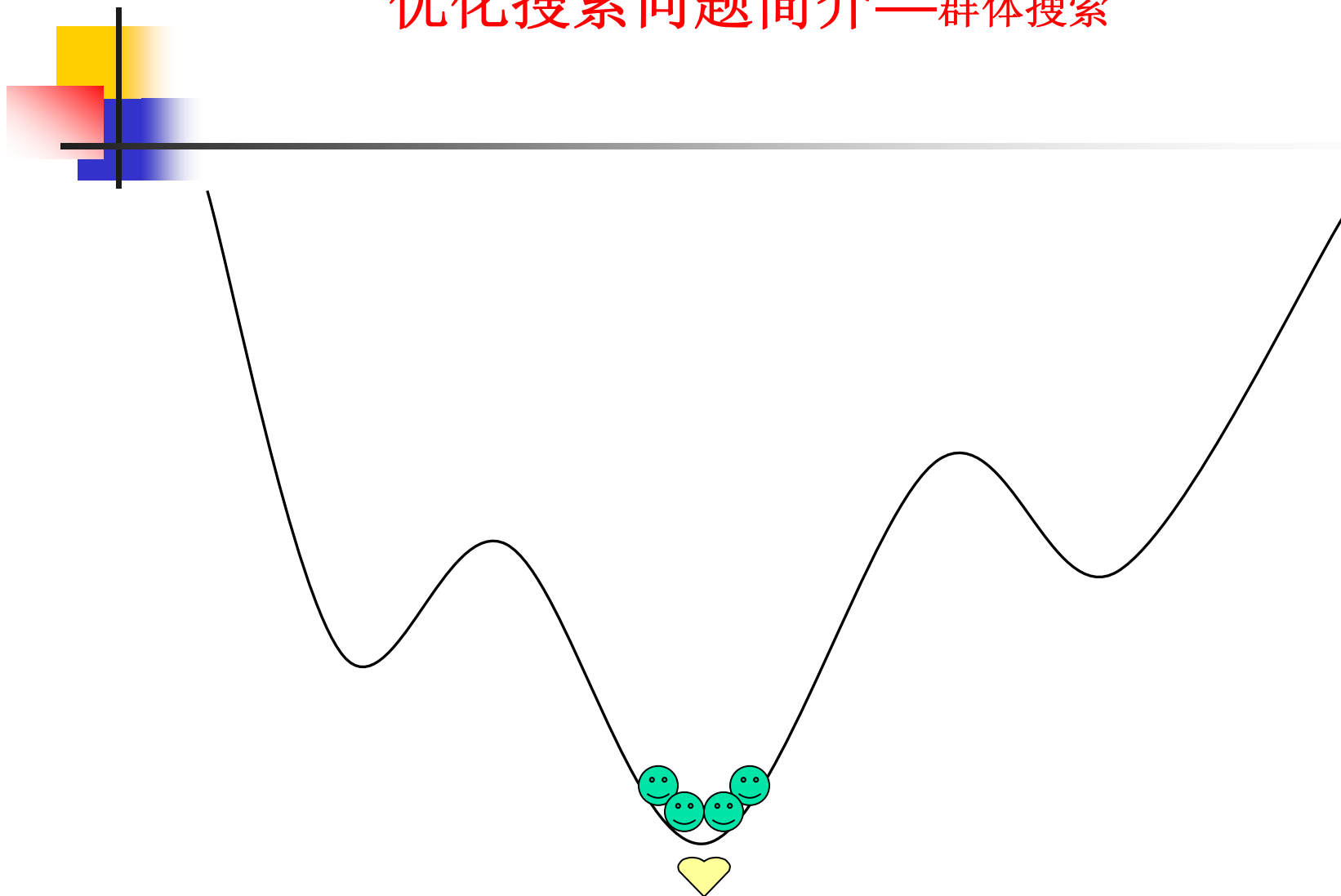
# 优化搜索问题简介——群体搜索



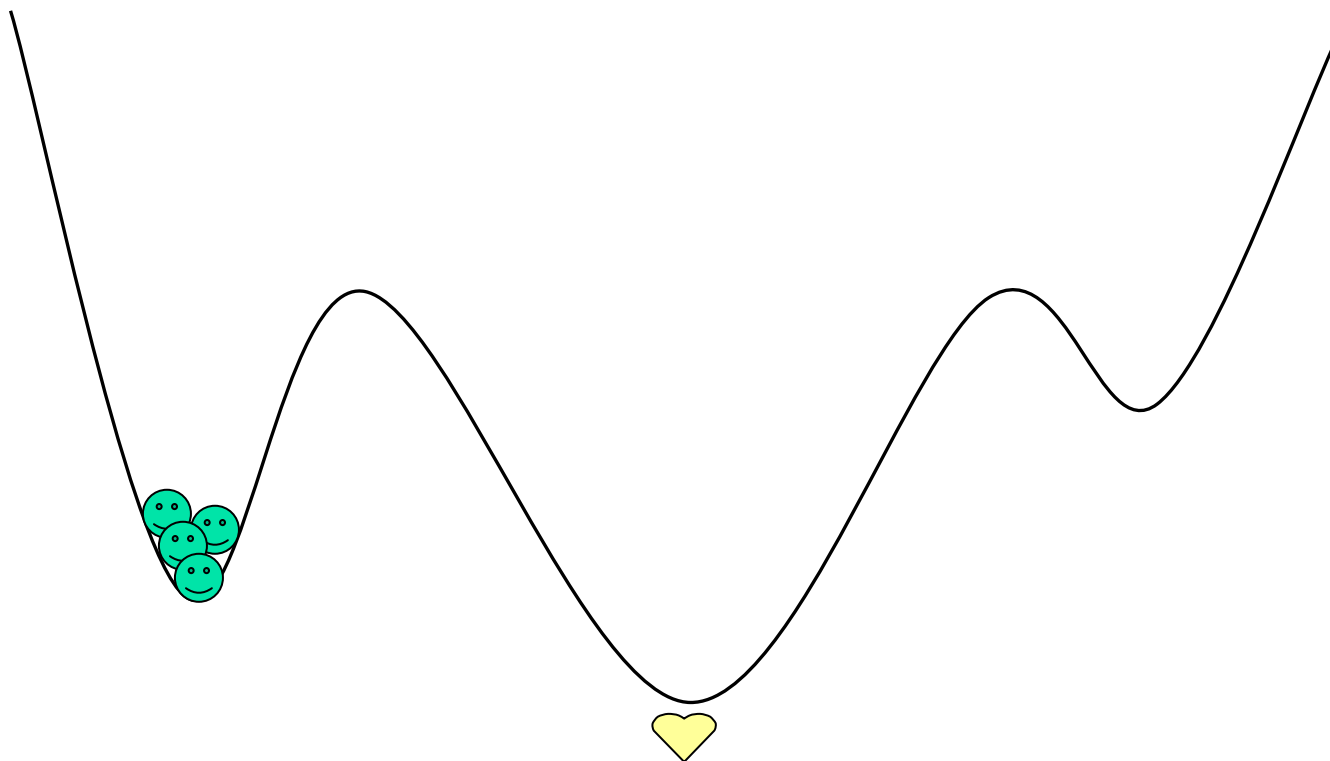
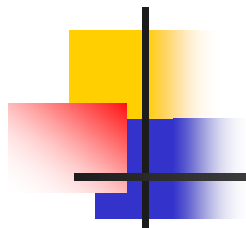
# 优化搜索问题简介——群体搜索



# 优化搜索问题简介——群体搜索



# 优化搜索问题简介——群体搜索



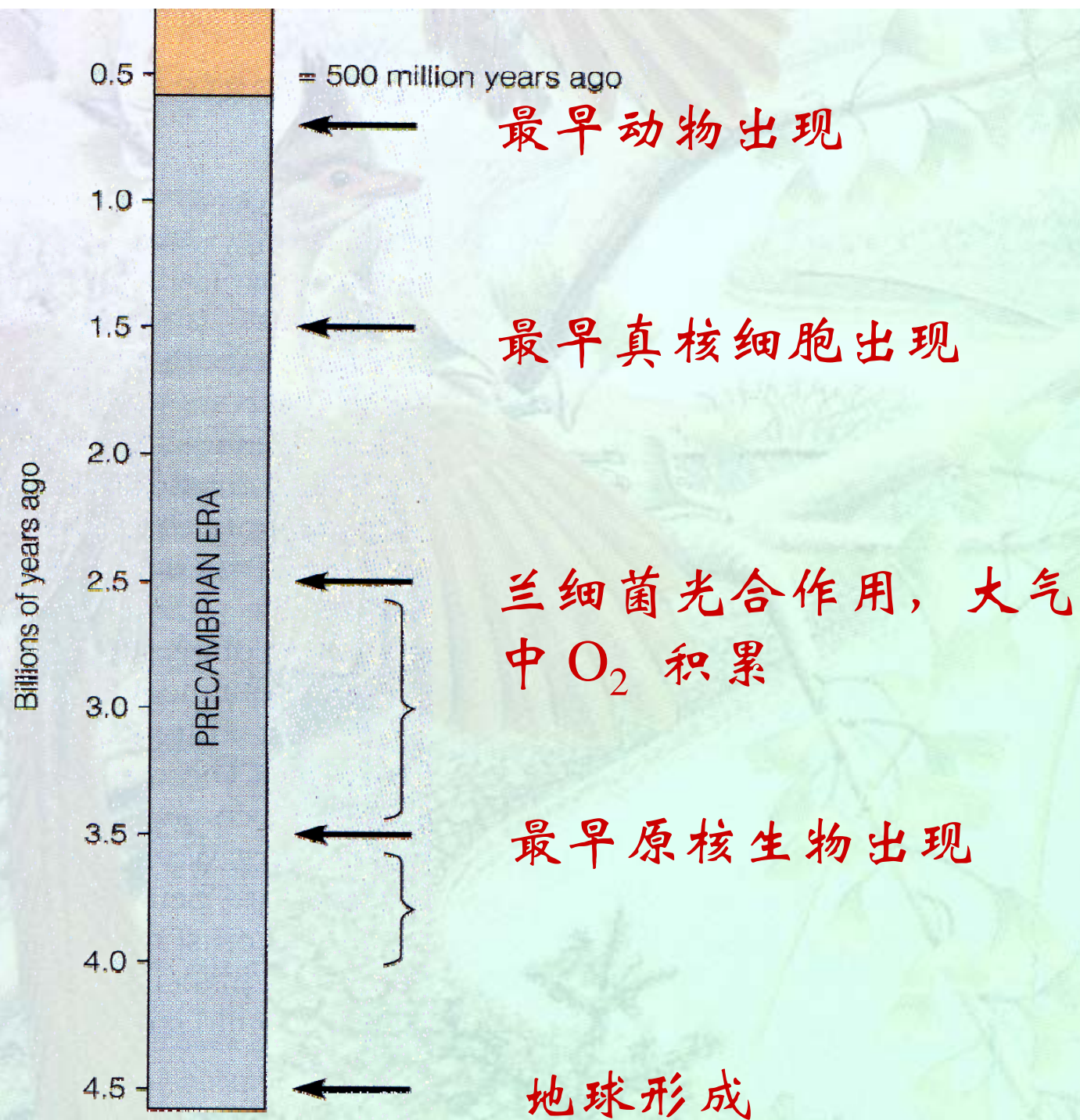


## 3.1 从生物进化到进化计算

---

生物界是一种自然进化（趋优）的系统。

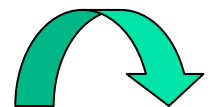
- 生物为什么会进化？（生物是如何进化的？）
- 只有生物界是自然进化的吗？
- 人工系统能否自然进化？如何实现？



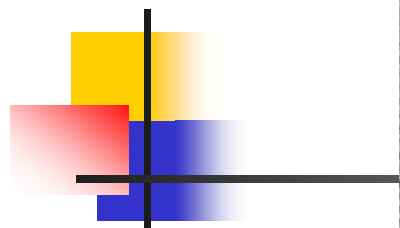




澳州西海岸的墊藻岩

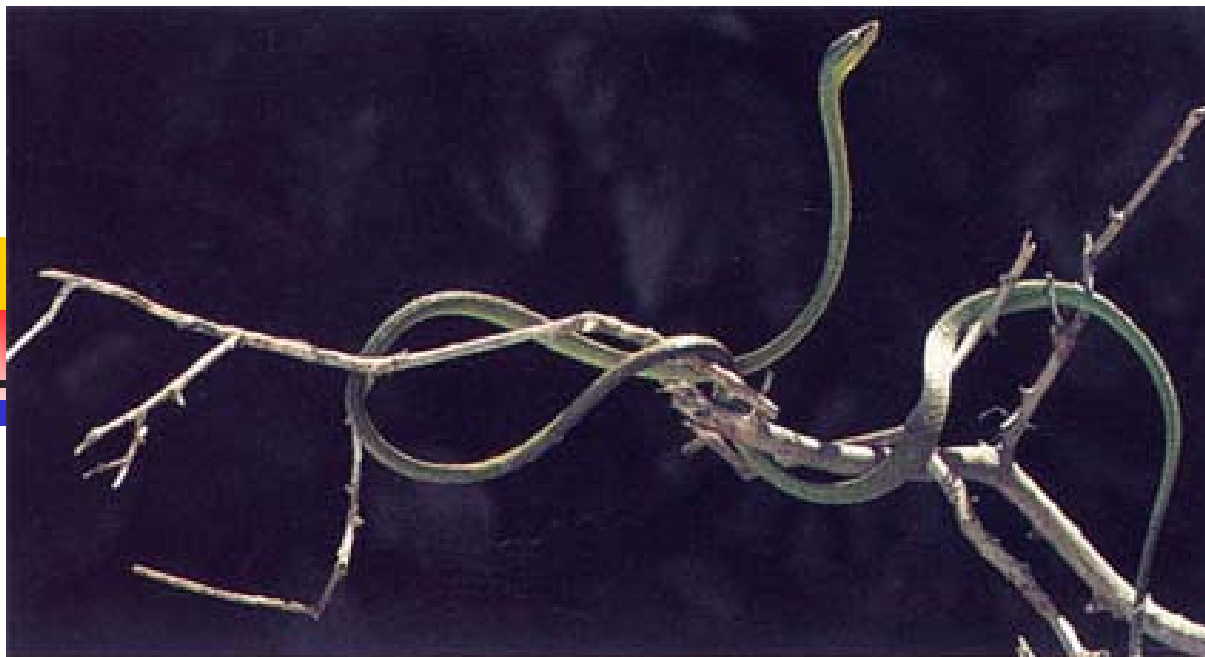






35  
亿  
年  
前  
的  
兰  
细  
菌  
化  
石



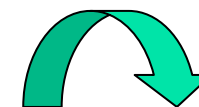


(a)

拟态



变色



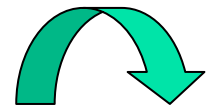


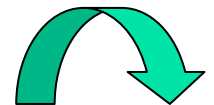
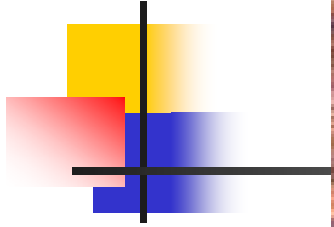


冬天白色



夏天黑色







西番莲是热带藤本植物，展足蛾产卵于叶面，  
西番莲的蜜腺象卵，起到保护叶子的作用

返回



## 3.1 从生物进化到进化计算

---

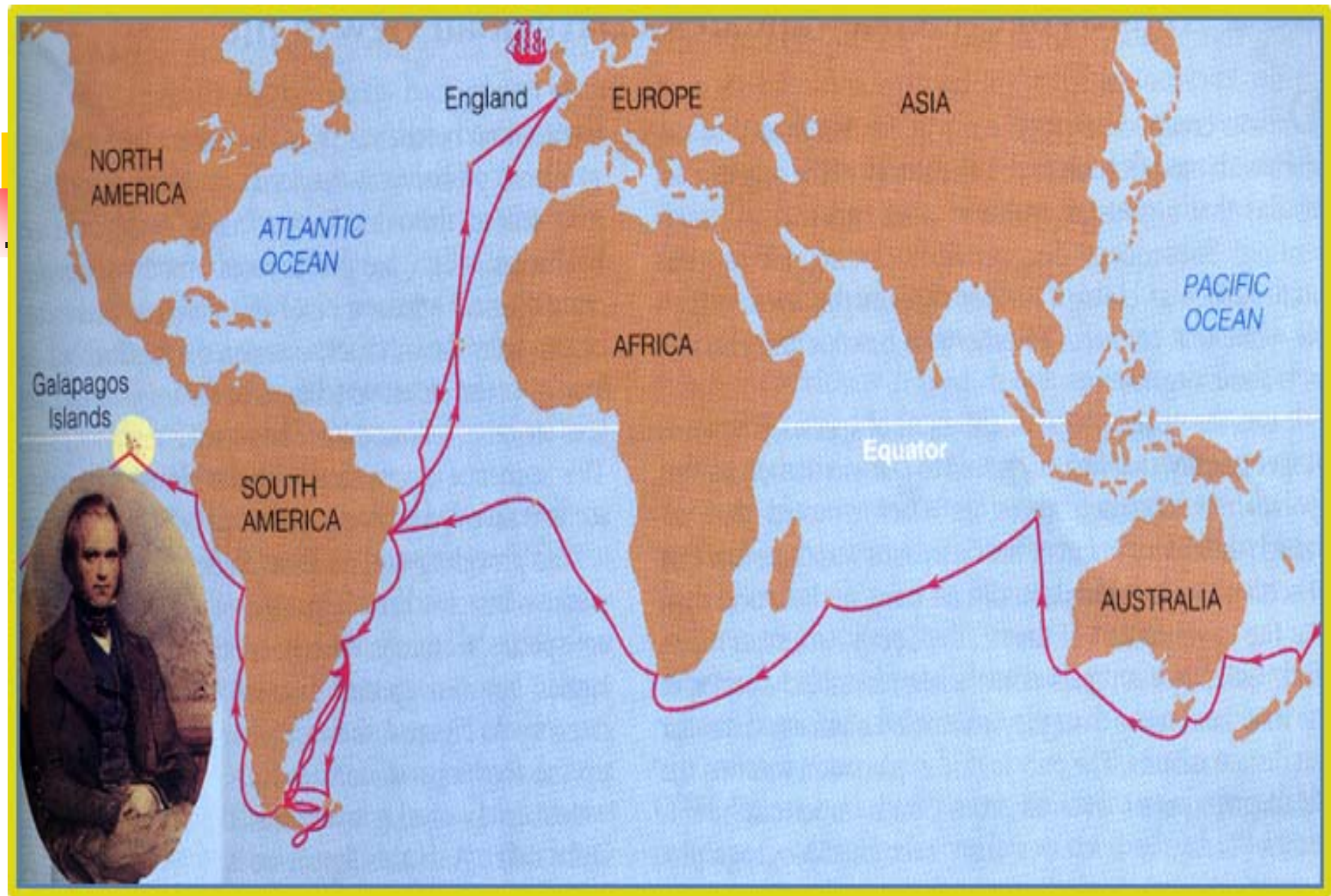
- 达尔文的进化论:

自然选择, 适者生存

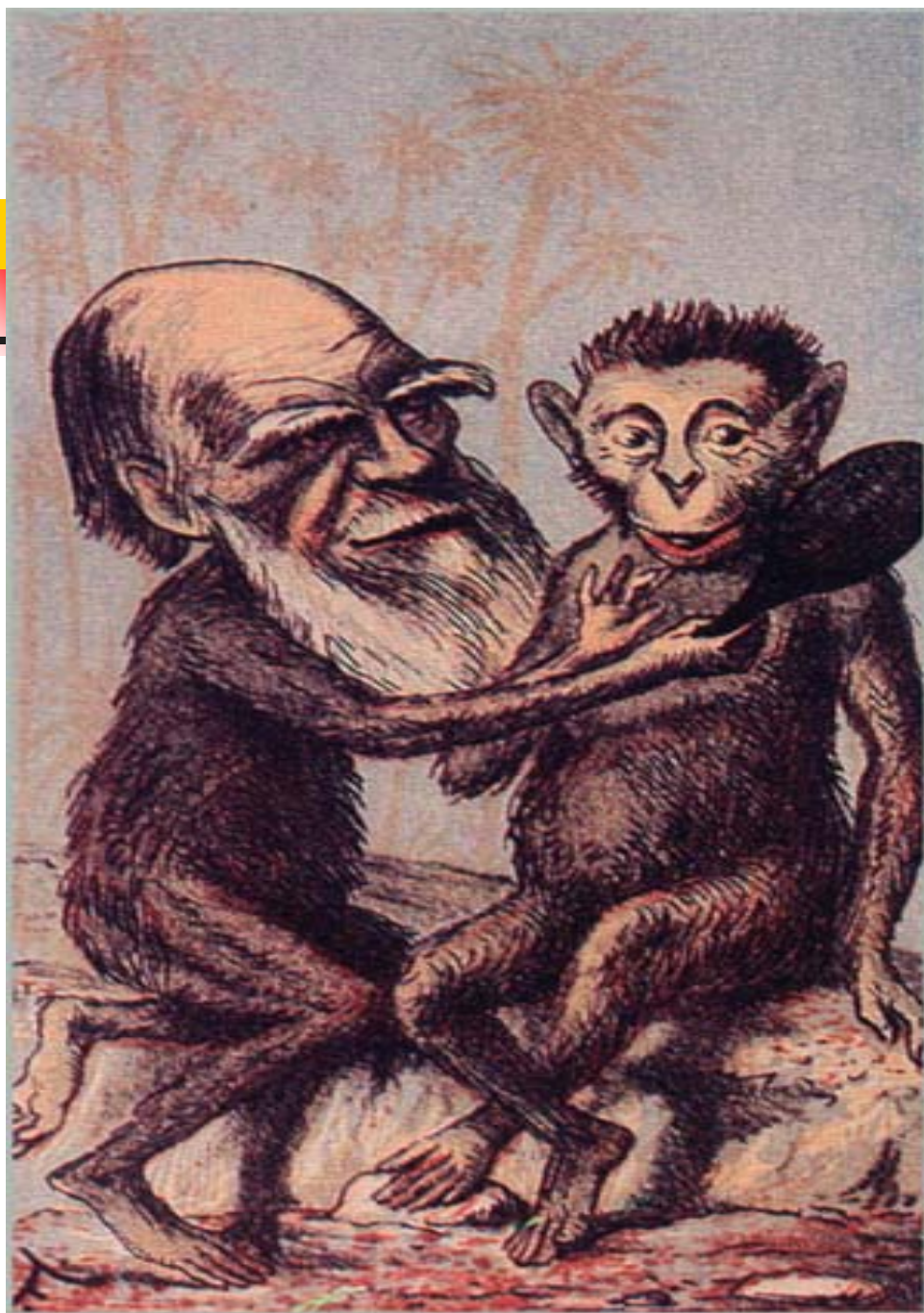
- 孟德尔与摩根的遗传学理论:

- 基因是决定生物特征的最基本的物质单元, 基因在染色体上以一定的顺序和结构排列, 每个基因有特殊的位置并控制生物的某些特性。
- 基因组合的特异性决定了生物体的多样性, 基因结构的稳定性保证了生物物种的稳定性, 而基因的杂交和变异使生物进化成为可能。





22岁的达尔文开始贝格尔舰航行



当时报纸上刊登的  
讽刺达尔文的漫画



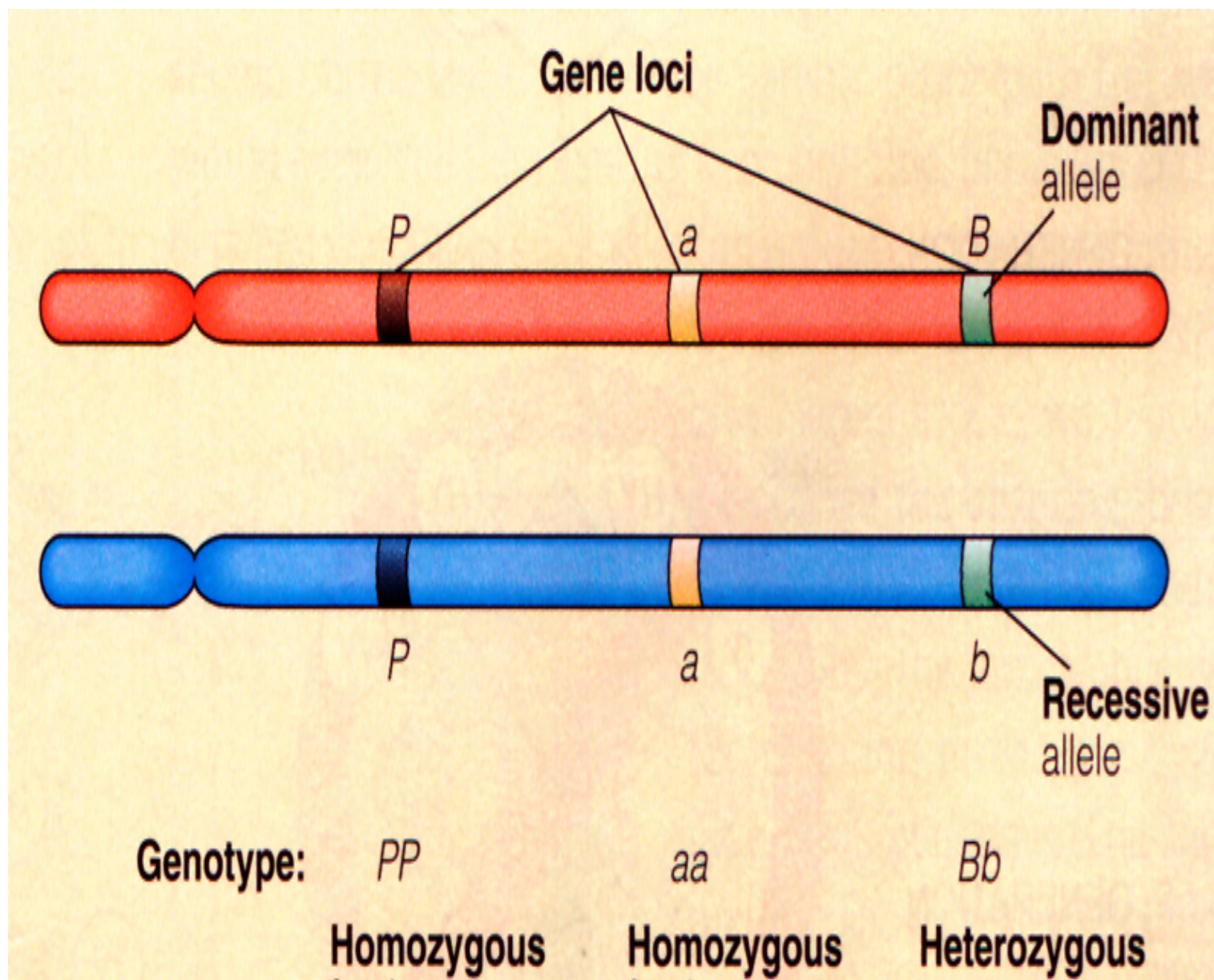


孟 德 尔 (1822-1884)



同源染色体分别带着控制

同一性状的两个等位基因



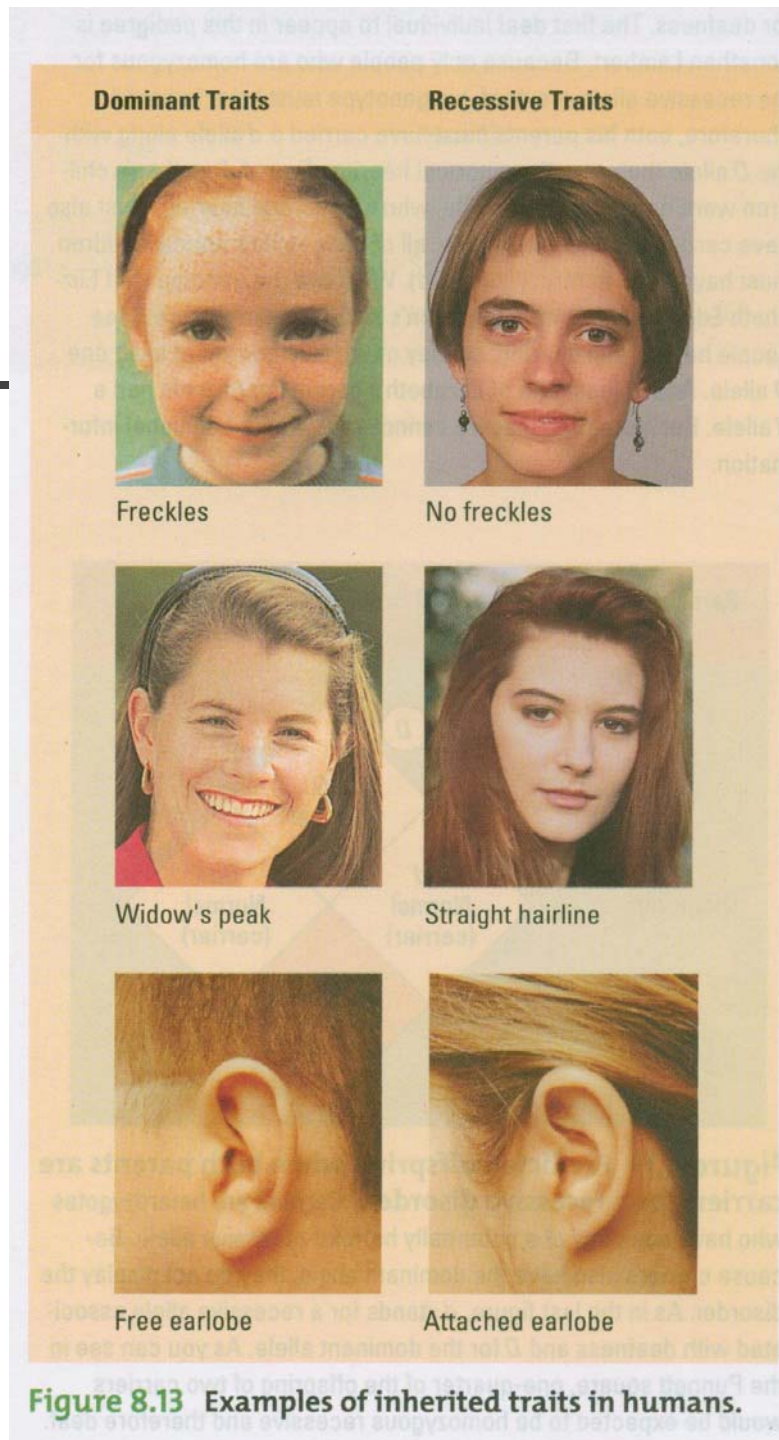
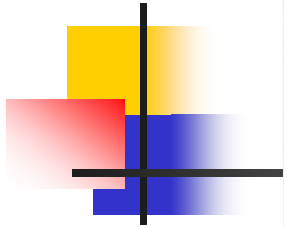
显性等位基因

纯合子

隐性等位基因

纯合子

杂合子



## 人类的显性与隐性基因举例

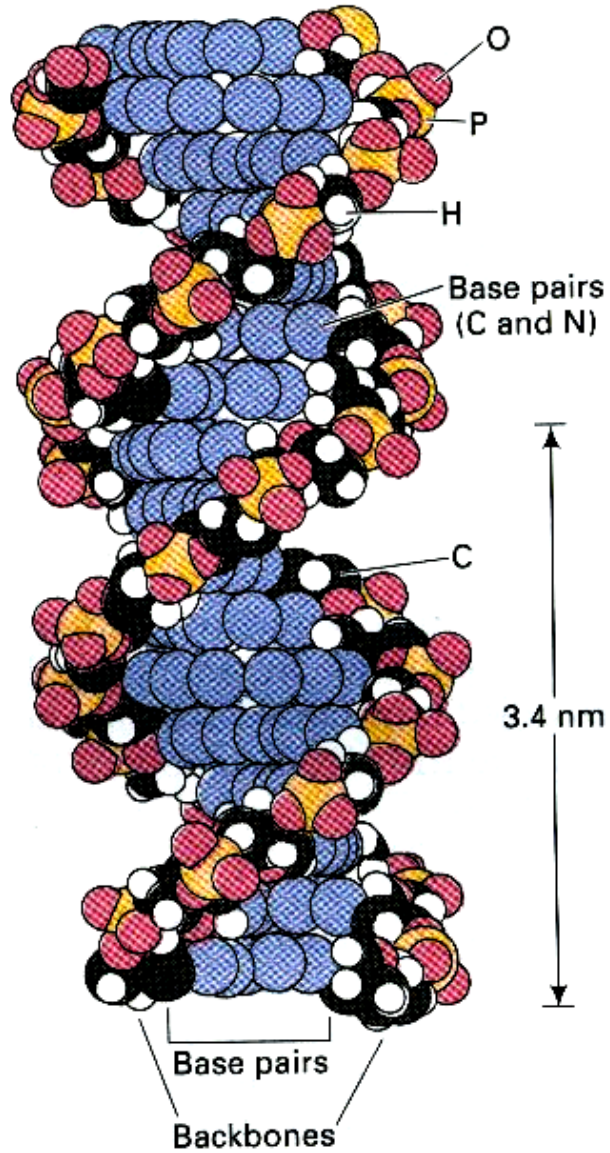
雀斑为显性

额头的V型发尖为显性

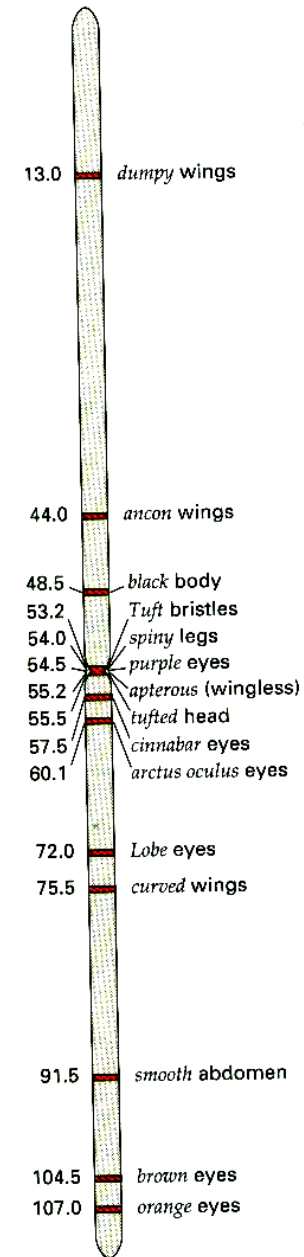
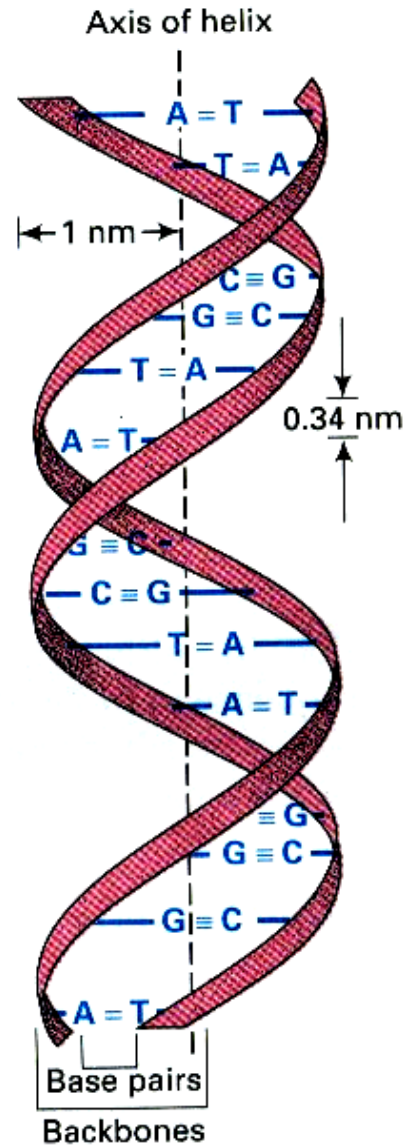
耳垂为显性

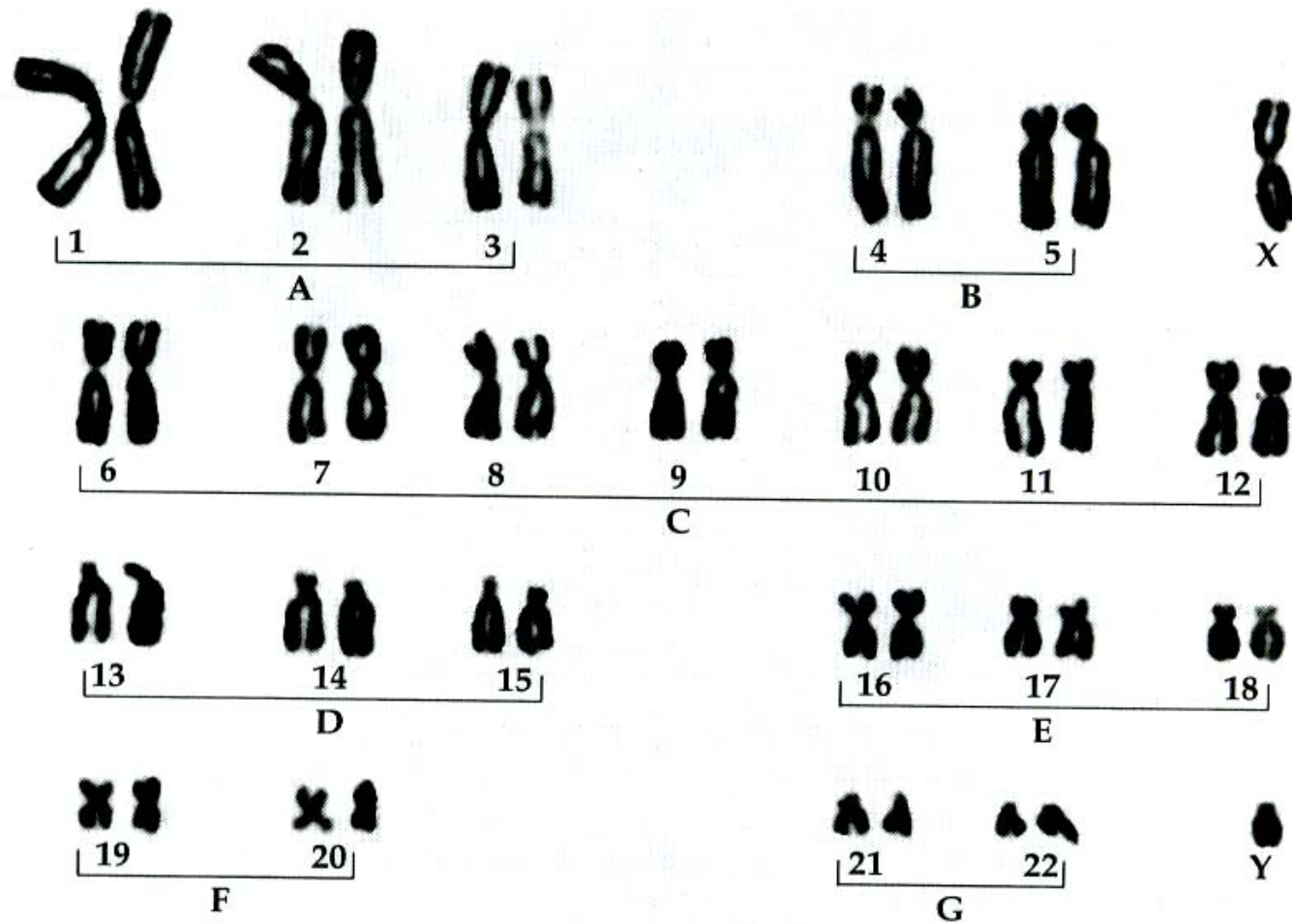
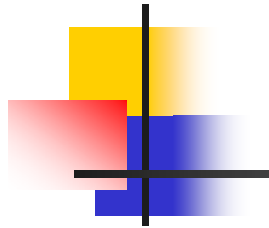


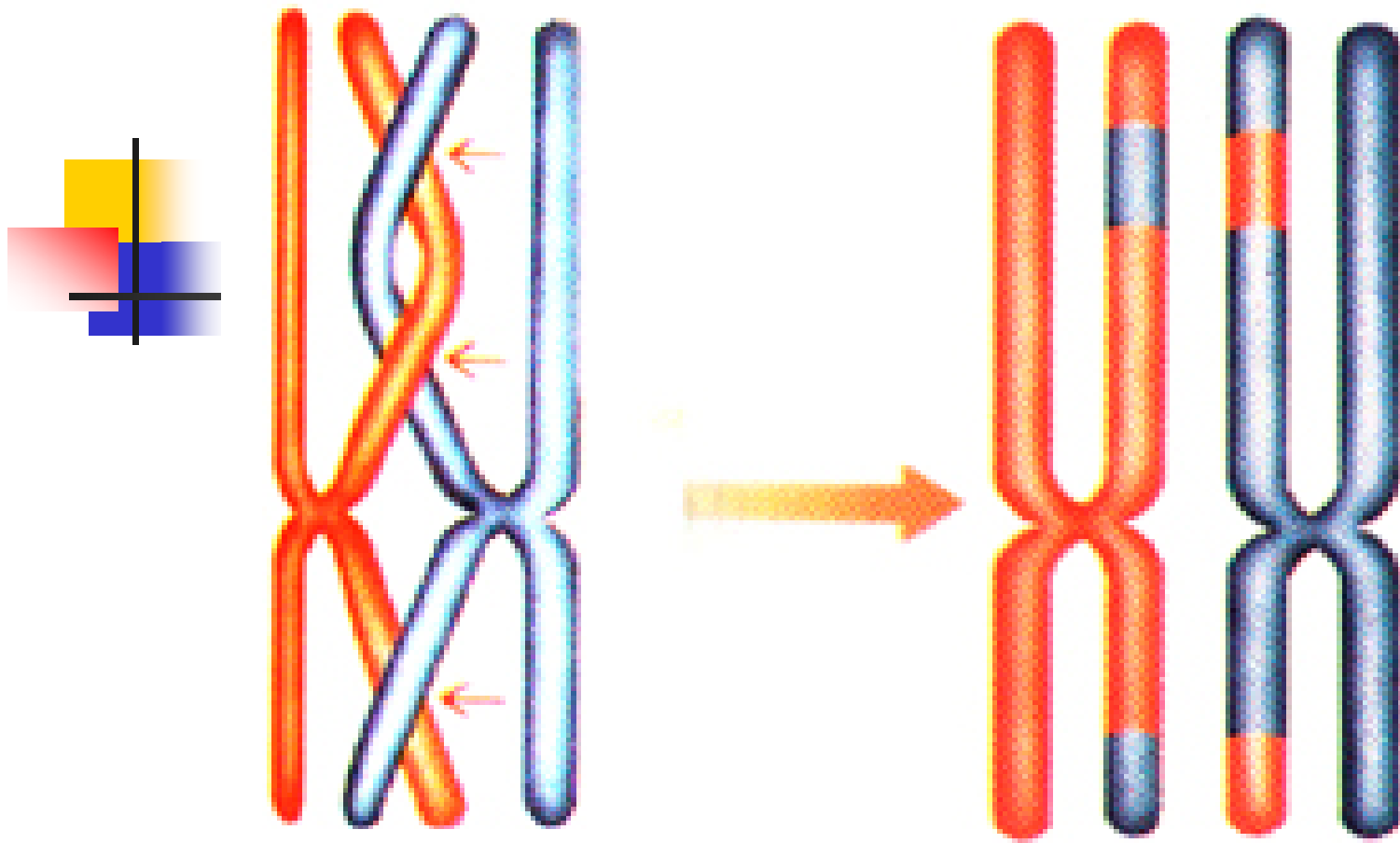
a) Molecular model



b) Stylized diagram







减数分裂时发生：染色体交叉/基因重组。



## 3.1 从生物进化到进化计算

---

生物进化过程的发生需要四个基本条件：

- 1) 存在由多个生物个体组成的种群;
- 2) 生物个体之间存在着差异，或群体具有多样性;
- 3) 生物能够自我繁殖;
- 4) 不同个体具有不同的环境生存能力，具有优良基因结构的个体繁殖能力强，反之则弱。
- 5) 存在竞争（优胜劣汰）。



## 3.1 从生物进化到进化计算

---

生物群体的进化机制包括三种基本形式:

- 1) 自然选择
- 2) 杂交
- 3) 突变

另外, 外界对生物的评价反映了生物的生存价值和机会。





## 3.1 从生物进化到进化计算

---

生物进化过程本质上是一种优化过程。

进化计算（**Evolutionary Computation, EC**）包括四个组成部分：

- 1) 遗传算法（**Genetic Algorithm, GA**），由**John H. Holland**教授等提出；
- 2) 进化规划（**Evolutionary Programming, EP**），由**Lawrence J. Fogel**等人提出；
- 3) 进化策略（**Evolutionary Strategies, ES**），由**Ingo Rechenberg**和**Hans-Paul Schwefel**提出。
- 4) 遗传规划（**Genetic Programming, GP**），由**John R. Koza**教授提出。



## 3.2 标准遗传算法

---

遗传算法的6个基本要素:

- 1) 参数编码;
- 2) 初始群体的设定;
- 3) 适应度函数的设计;
- 4) 遗传算子设计;
- 5) 控制参数设定
- 6) 迭代终止条件



## 3.2 标准遗传算法

---

标准遗传算法流程:

- 1. 编码
- 2. 初始群体的生成
- 3. 适应度评估检测
- 4. WHILE <未满足迭代终止条件> DO
  - 1. 选择
  - 2. 交叉
  - 3. 变异
  - 4. 适应度评估检测
- 5. END DO



## 3.2 标准遗传算法(示例)

---

问题: 求函数 $f(x) = x^2$ 的最大值,  $x \in [0, 31]$

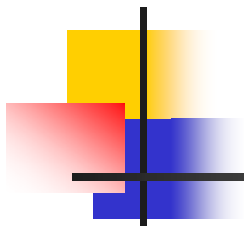
1. 编码 (定义问题参数空间到编码空间的映射)

$$13 \Rightarrow 0\ 1\ 1\ 0\ 1$$

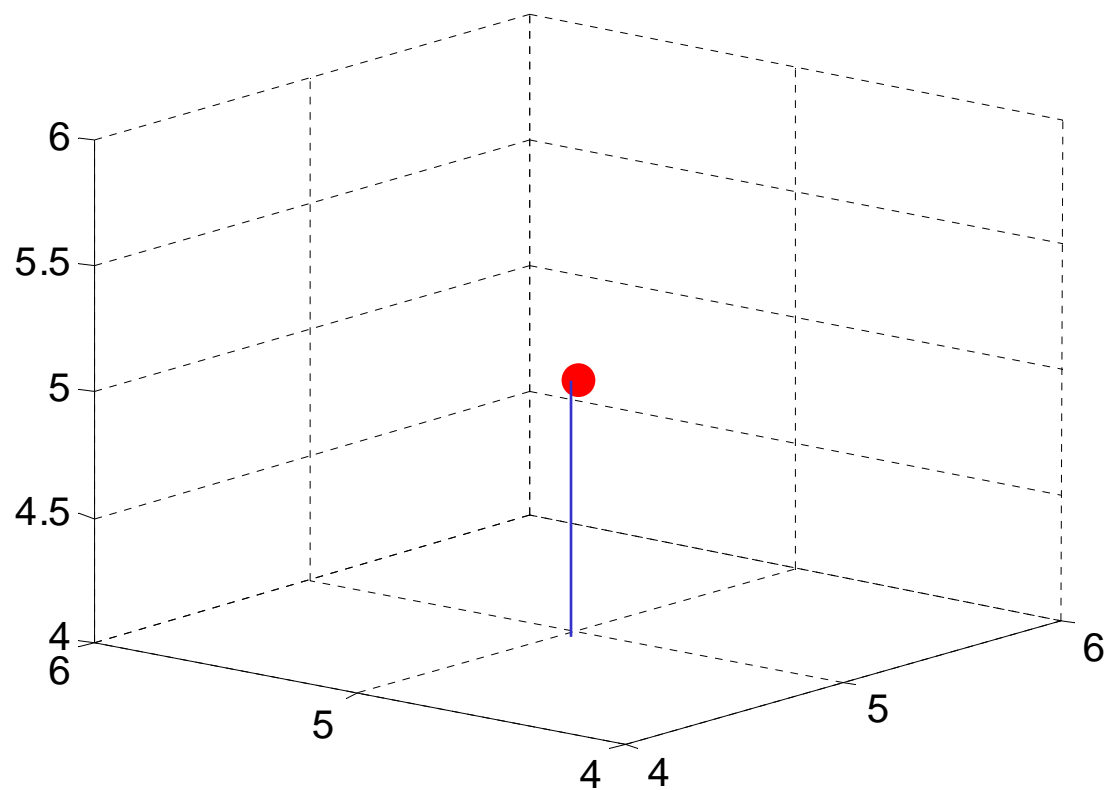
$$24 \Rightarrow 1\ 1\ 0\ 0\ 0$$

$$8 \Rightarrow 0\ 1\ 0\ 0\ 0$$

$$19 \Rightarrow 1\ 0\ 0\ 1\ 1$$



一个解对应于多维空间中的一个点





## 3.2 标准遗传算法(示例)

---

### 2. 初始群体的生成

$$\left\{ \begin{array}{ccccc} 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \end{array} \right\} \leftarrow \text{初始种群}$$



## 3.2 标准遗传算法(示例)

---

### 3. 适应度评估检测

$$f(x) = x^2$$

$$0\ 1\ 1\ 0\ 1 \Rightarrow x=13 \Rightarrow f(x) = 169$$

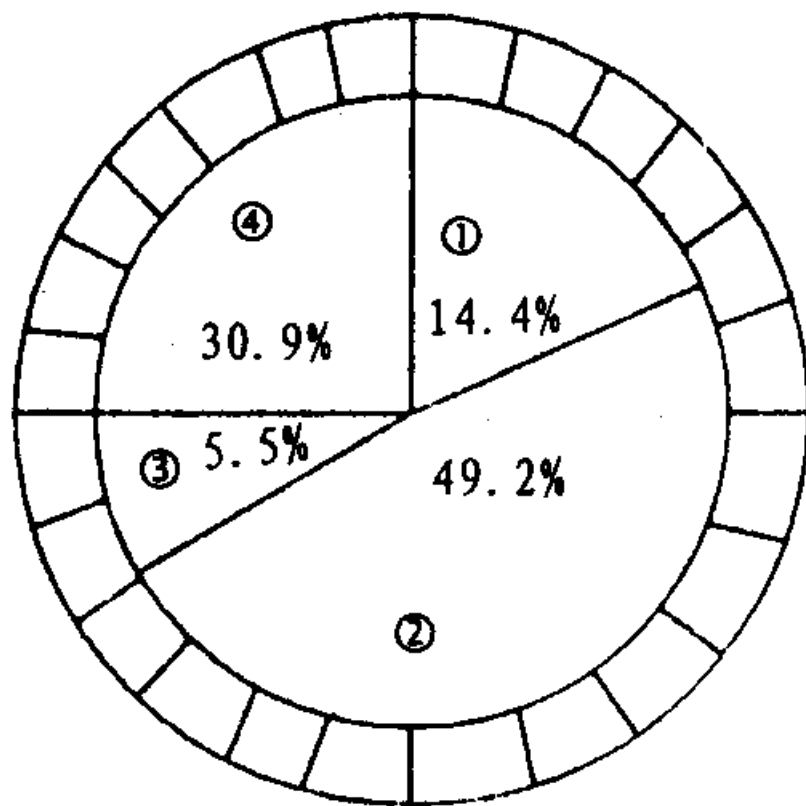
$$1\ 1\ 0\ 0\ 0 \Rightarrow x=24 \Rightarrow f(x) = 576$$

$$0\ 1\ 0\ 0\ 0 \Rightarrow x=8 \Rightarrow f(x) = 64$$

$$1\ 0\ 0\ 1\ 1 \Rightarrow x=19 \Rightarrow f(x) = 361$$

## 3.2 标准遗传算法(示例)——遗传算子

### 1) 选择（复制）



← 赌轮选择





## 3.2 标准遗传算法(示例)——遗传算子

---

### 2) 交叉

0 1 1 0   1	→	0 1 1 0 0
1 1 0 0   0		1 1 0 0 1



## 3.2 标准遗传算法(示例)——遗传算子

---

### 3) 变异

1 1 0 0 0  $\longrightarrow$  1 1 0 1 0



### 3.3 遗传算法的特点

---

1. 基于群体的搜索。
2. 直接处理的对象是参数的编码集而不是问题参数本身。
3. 搜索过程中使用的是基于目标函数值的评价信息，搜索过程既不受优化函数连续性的约束，也没有优化函数必须可导的要求。
4. 具有显著的隐并行性。（具有自学习能力）
5. 形式上简单明了。
6. 具有很强的鲁棒性。



### 3.4 遗传算法理论研究

---

遗传算法所追求的是当前群体产生比现有个体更好个体的能力，即遗传算法的**可进化性**或称**群体可进化性**。因此，遗传算法的理论和方法研究也就围绕着这一目标展开。



### 3.4 遗传算法理论研究

关于下面五个问题的回答，就成为**GA**理论研究的主要方向：

- 1) 遗传算法为什么会具有很好的优化的能力？遗传算法如何更好地模拟复杂系统的适应性过程和进化行为？
- 2) 遗传算法在优化问题求解中怎样才能具备全局收敛性？
- 3) 遗传算法的搜索效率如何评价？
- 4) 遗传策略的设计和参数控制的理论基础是什么？
- 5) 遗传算法与其他算法如何结合？





### 3.4.1 遗传算法的基础理论研究

---

遗传算法的基础理论主要以收敛性分析为主，即群体收敛到优化问题的全局最优解的概率。从整体上讲，可以分为基于随机过程的收敛性研究和基于模式理论的收敛性分析，我们将前者称为遗传算法的随机模型理论，后者称为遗传算法的进化动力学理论。



### 3.4.1 遗传算法的基础理论研究

---

#### 1. 随机模型理论

- 对于有限的编码空间和有限的群体，遗传算法的搜索过程可以表示为离散时间的马尔可夫链模型（Marko chain model），从而可以采用已有的随机过程理论进行严密分析。
- Rudolph用齐次有限马尔可夫链证明了标准遗传算法收敛不到全局最优解，但是如果采用精英保留策略，那么遗传算法是收敛的。



### 3.4.1 遗传算法的基础理论研究

#### 2. 进化动力学理论

- 基于某种具体运算形式的遗传算法的进化行为分析构成了进化动力学理论的基本内容。
- 由Holland提出的**模式定理**可以称为遗传算法进化动力学的基本定理。
- 积木块假设描述了遗传算法的重组功能。
- 模式定理和积木块假设构成了求解优化问题时遗传算法具备发现全局最优解的充分条件，也是分析遗传算法的进化行为的基本理论，统称为**模式理论**。



### 3.4.1 遗传算法的基础理论研究

---

#### 2. 进化动力学理论

- 通过基于模式理论的进化动力学分析，我们可以发现遗传算法所能求解的问题的范围，找出遗传算法的不足之处，进而可能对其进行改进。



## 3.4.2 遗传策略研究与设计

- 为了维持群体可进化性并最终搜索到问题的全局最优解，遗传算法必须采用合适的运算形式（计算流程、种群规模、种群初始化策略、GA算子、终止条件等），这就是遗传策略研究与设计的主要内容。
  - 在研究和设计遗传策略时，一般考虑遗传算法应具备两个方面的能力：
    - **探索**（exploration）能力：发现新模式的能力，GA探索未知解空间的能力。
    - **开发**（exploitation）能力：根据已有知识发现局部最优解的能力。
- 又称**求泛**（reforming）与**求精**（refining）能力。





## 3.4. 2 遗传策略研究与设计

- 遗传策略研究与设计可以分为**微观遗传策略**研究和**宏观遗传策略**研究两部分。
- **微观遗传策略**：主要讨论群体规模、遗传算子的形式和参数设计，及其对GA求解能力的影响。
- **宏观遗传策略**：主要讨论关于通过GA流程的再设计，改变GA的宏观特征，或者以GA流程为基础，引入其他算法构成混合GA，以期提高GA求解问题全局最优解的能力。



### 3.4.3 遗传算法的编码方式

---

- 编码问题：知识表示问题，经过编码的个体是GA操作的对象。
- 编码：由问题解空间到GA编码空间的映射，要满足完备性、健壮性、唯一性等一些基本要求。
- 编码还要满足模式定理和积木块假设。
- 编码与遗传算子的选择和设计关系密切。



### 3.4.4 遗传算法其他问题

---

- NFL (no free lunch) 定理
- 约束问题的求解 (抛弃、修正、惩罚非法解)
- 多目标优化问题的求解
- 并行遗传算法研究
- 遗传学习系统



## 3.5 遗传算法的应用

- 控制：瓦斯管道控制，防避导弹控制，机器人控制
- 规划：生产规划，并行机任务分配
- 设计：VLSI布局，通信网络设计，喷气发动机设计
- 组合优化：TSP问题，背包问题，图划分问题
- 图像处理：模式识别，特征抽取
- 信号处理：滤波器设计
- 机器人：路径规划
- 人工生命：生命的遗传进化
- 知识发现：规则提取，数据挖掘



## 3.6 遗传算法的基本术语

- 个体 (individual) : GA所处理的基本对象、结构。
- 群体 (population) : 个体的集合。
- 位串 (bit String) : 个体的表示形式。对应于遗传学中的染色体 (Chromosome)
- 基因 (gene) : 位串中的元素, 表示不同的特征。对应于生物学中的遗传物质单位, 以 DNA序列形式把遗传信息译成编码。
- 基因位 (Locus) : 某一基因在染色体中的位置。
- 等位基因 (allele) : 表示基因的特征值, 即相同基因位的基因取值。





## 3.6 遗传算法的基本术语

- 位串结构空间 (bit String Space) : 等位基因任意组合构成的位串集合, 基因操作在位串结构空间进行, 对应于遗传学中的基因型的集合。
- 参数空间 (Parameters Space) : 是位串空间在物理系统中的映射。对应于遗传学中的表现型的集合。
- 适应度值 (fitness) : 某一个体对于环境的适应程度, 或者在环境压力下的生存能力, 取决于遗传特性。
- 复制、选择 (reproduction or selection) : 在有限资源空间上的排他性竞争。



## 3.6 遗传算法的基本术语

- 交叉、交换、交配、重组 (crossover or recombination)：一组位串或者染色体上对应基因段的交换。
- 变异 (mutation)：位串或染色体水平上的基因变化，可以遗传给子代个体。
- 逆转或倒位 (inversion)：反转位串上的一段基因的排列顺序。对应于染色体上的一部分，在脱离之后反转  $180^\circ$  再连接起来。
- 单倍体 (haploid, monoploid)：细胞核中有  $n$  个正常的不配对染色体。



## 3.6 遗传算法的基本术语

- 二倍体 (diploid)、多倍体 (polyploid)：细胞核中有  $2n$  或更多个正常的配对染色体。
- 基因型 (genotype)：或称遗传型，指用基因组定义遗传特征和表现。对应于GA中的位串。
- 表现型 (phenotype)：生物体的基因型在特定环境下的表现特性。对应于GA中的位串解码后的参数。
- 基因连结 (linkage)：两个或更多的等位基因出现在同一个染色体上。对应于 GA 中的模式的概念。



## 3.6 遗传算法的基本术语

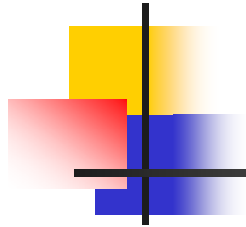
- 上位遗传或者基因关联（即epistasis）：两个非等位基因之间的相互作用，使得其中之一（上位基因）对另一个（下位基因）的表现型产生干扰或抑制。对应于优化函数的非线性特征（non-linearity）。
- 基因多效性（pleiotropy）：指单一基因对生物体多个物理性状的影响。对应于GA求解多目标优化问题中，某个变量对多个目标函数的影响。
- 多基因效性（polygeny）：指生物体某个物理性状由多个基因共同决定。对应于GA求解多目标优化问题中，某个目标函数的值由多个变量的状态所决定。



## 3.6 遗传算法的基本术语

- 遗传源变或遗传漂移 (genetic drift)：指群体的遗传组成的随机变化，不含自然选择的影响。
- 遗传 (heredity)：父代个体通过有性方式向子代个体的特征传递过程。
- 局部环境或者环境小生境 (environmental niches)：具有某种特征的子环境，其中生物体具有特定的染色体结构，表现出特定的物理性状。对应于多模态函数中局部极值点的邻域。





Parent 1: **011010**

Parent 2: **000110**

*Simple reproduction*

Offspring 1: **011010**

Offspring 2: **000110**

Parent 1: **011010**

Parent 2: **000110**

*Reproduction with  
crossover at locus 3*

Offspring 1: **011110**

Offspring 2: **000010**

Parent 1: **011010**

Parent 2: **000110**

*Simple reproduction with mutation  
at locus 3 for offspring 1*

Offspring 1: **010010**

Offspring 2: **000110**