Bioinformatique M1: Lecture 2

P. Derreumaux

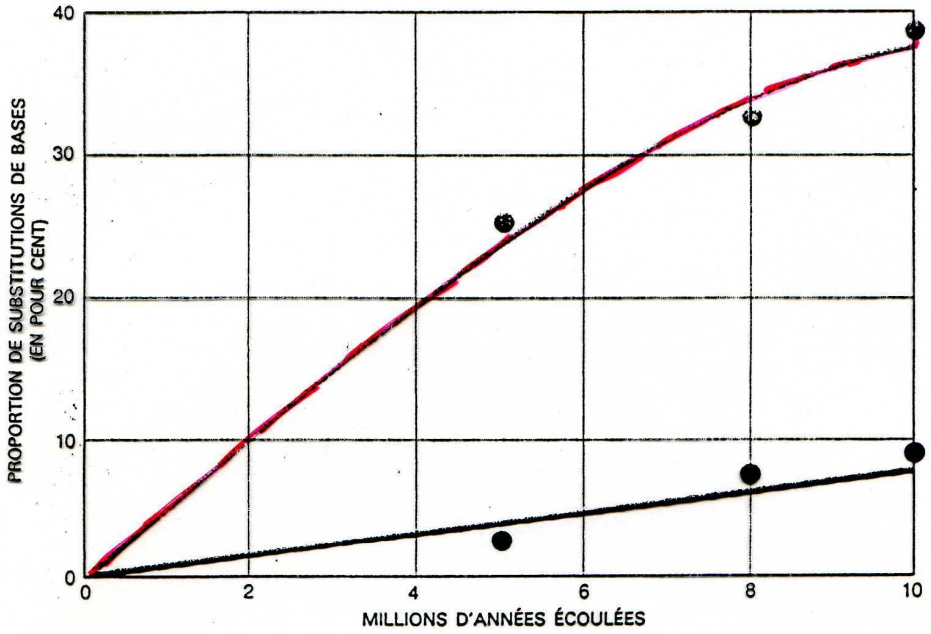
PROTEIN EVOLUTION

Definition: change in protein sequence (structure) over time.

They are several genetic mechanisms for altering DNA.

1. Point mutations :

→ single-nucleotide mutation.



5. LA NOTION DE CONTRAINTE FONCTIONNELLE apparaît quand on compare la fréquence de substitution des bases placées en deuxième position dans un codon (*courbe noire*) à celle des bases en troisième position (*courbe en couleur*). Un codon est un triplet de bases d'ADN codant un acide aminé. Les substitutions sont plus fréquentes en troisième position qu'en deuxième, car toutes les substitutions qui surviennent en deuxième position modifient l'acide aminé codé, alors que seule la moitié des autres provoque un changement. Les données utilisées proviennent de la comparaison de l'ADN de mitochondries humaines et de grands singes. L'évolution est plus lente dans les sites des gènes qui modifient directement la fonction

→ High frequency of simultaneous double-nucleotide substitution single-nucleotide mutation (Science 2000)

2. Insertions or deletions (multiple or not multiple of 3 nucleotides)

consequences on structures (loops, α-helix, β-hairpins)

adaptation is possible everywhere depending on the number of INDELS

consequences on functions vary with location of INDELS

3. Gene fusion or gene fission

- joining proteins to form a multienzyme polypeptide

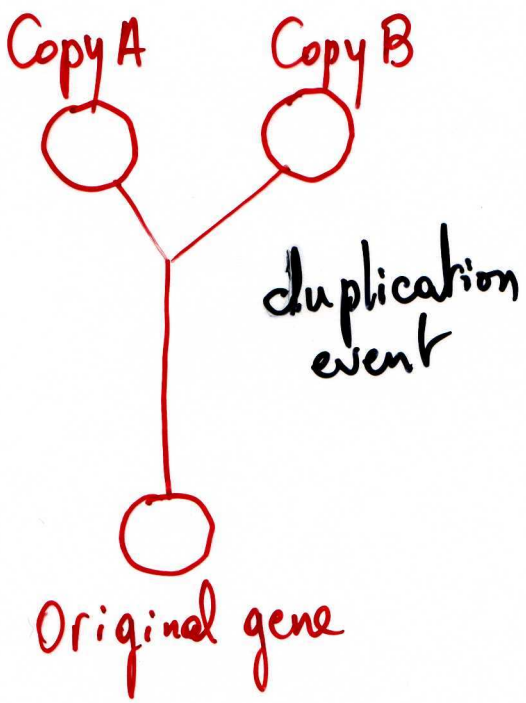
  e.g. 5 separate enzymes in E-coli

  V.S.

  5 enzyme activities in Neuropsora on one polypeptide chain

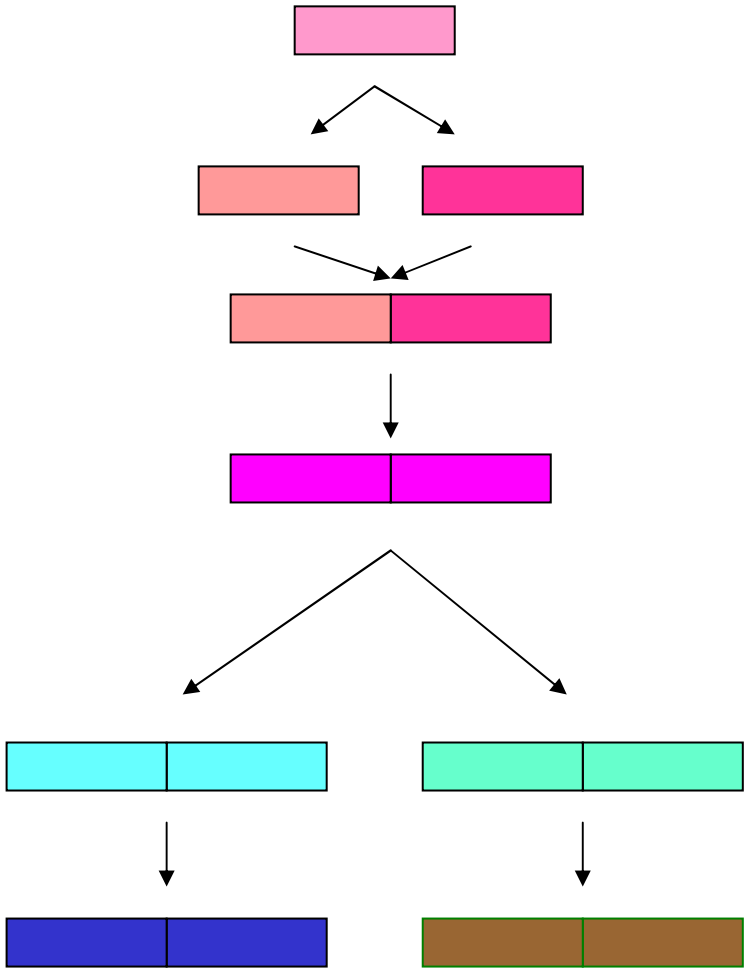
- joining domains

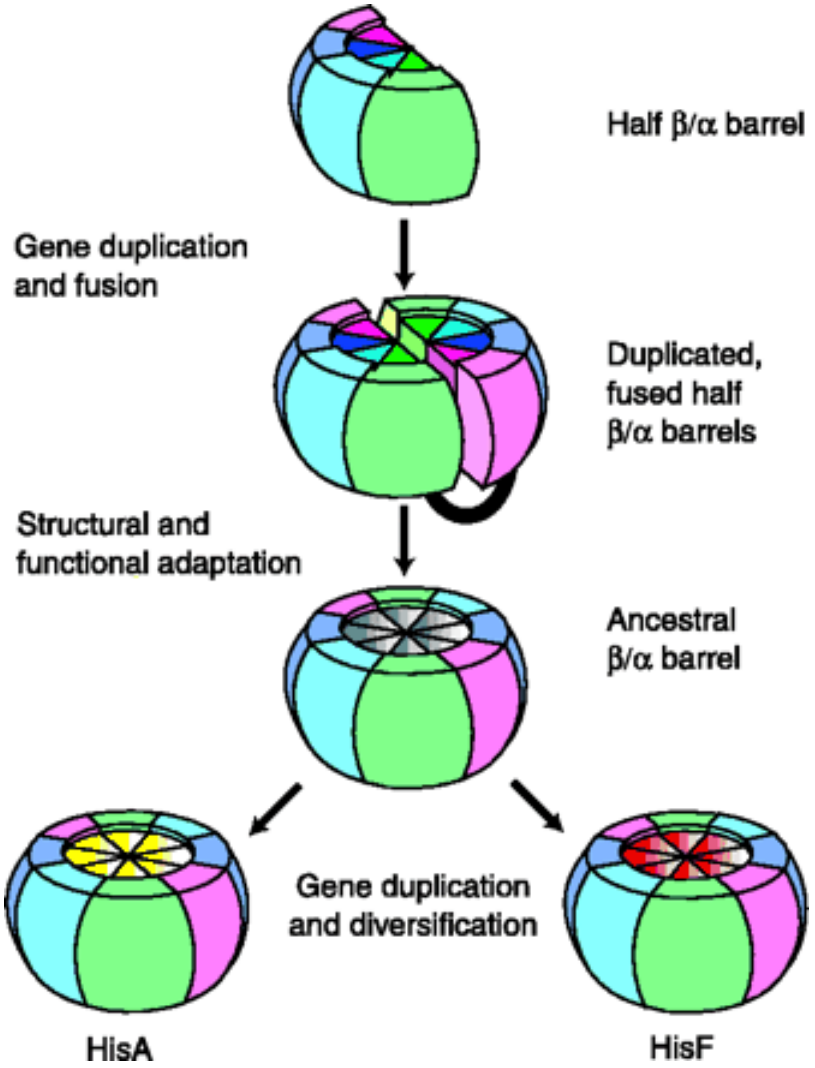
4. Gene duplication



- Two copies of original gene
- One copy may evolve new function (other retains original function)
- Both may persist relatively unchanged (provides redundancy)
- Copies A and B may partition the function of the original

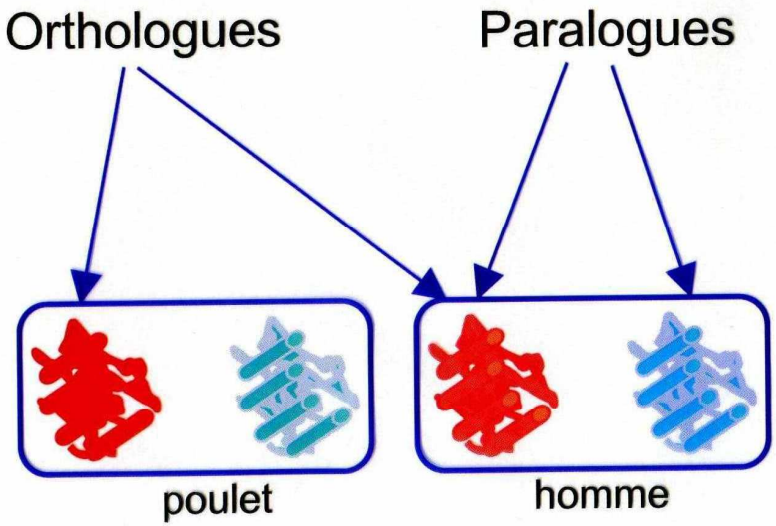
Mécanisme d'évolution des fonctions :
modèle pour HisA et HisF (biosynthèse His)





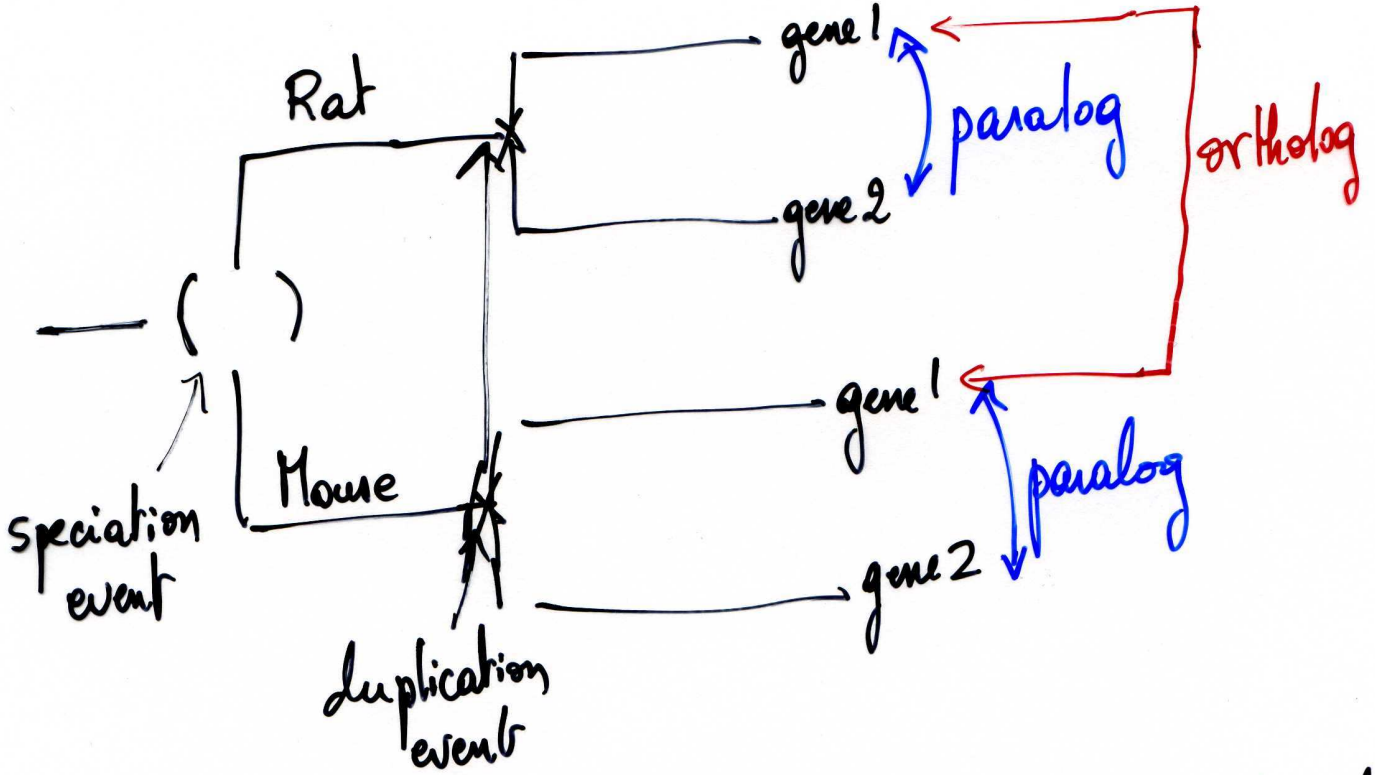
Paralogues et orthologues (Fitsh, 1970)

- ★ Homologues: gènes provenant d'un ancêtre commun
- ★ Paralogues: gènes homologues issus d'un phénomène de duplication
- ★ Orthologues: gènes homologues issus d'un phénomène de spéciation
- ★ Transfert horizontal: par endosymbiontes, etc. Fitsch a aussi introduit "xénologue" pour évoquer ce cas.



un résultat, pas à un mécanisme génétique spécifique

↘ échange de matériel génétique d'une espèce à une autre. Il est peut être révélé grâce au biais d'utilisation des codons [homogène chez certaines bactéries. Cet événement peut conduire à remettre en cause les arbres phylogénétiques issus des alignements de séquences



N.B. Do not confuse "orthology" with "functional equivalence"? Walker Fich (1970): While it is likely that two orthologs have similar functions, these functions are not necessarily identical.

A duplication event may involve

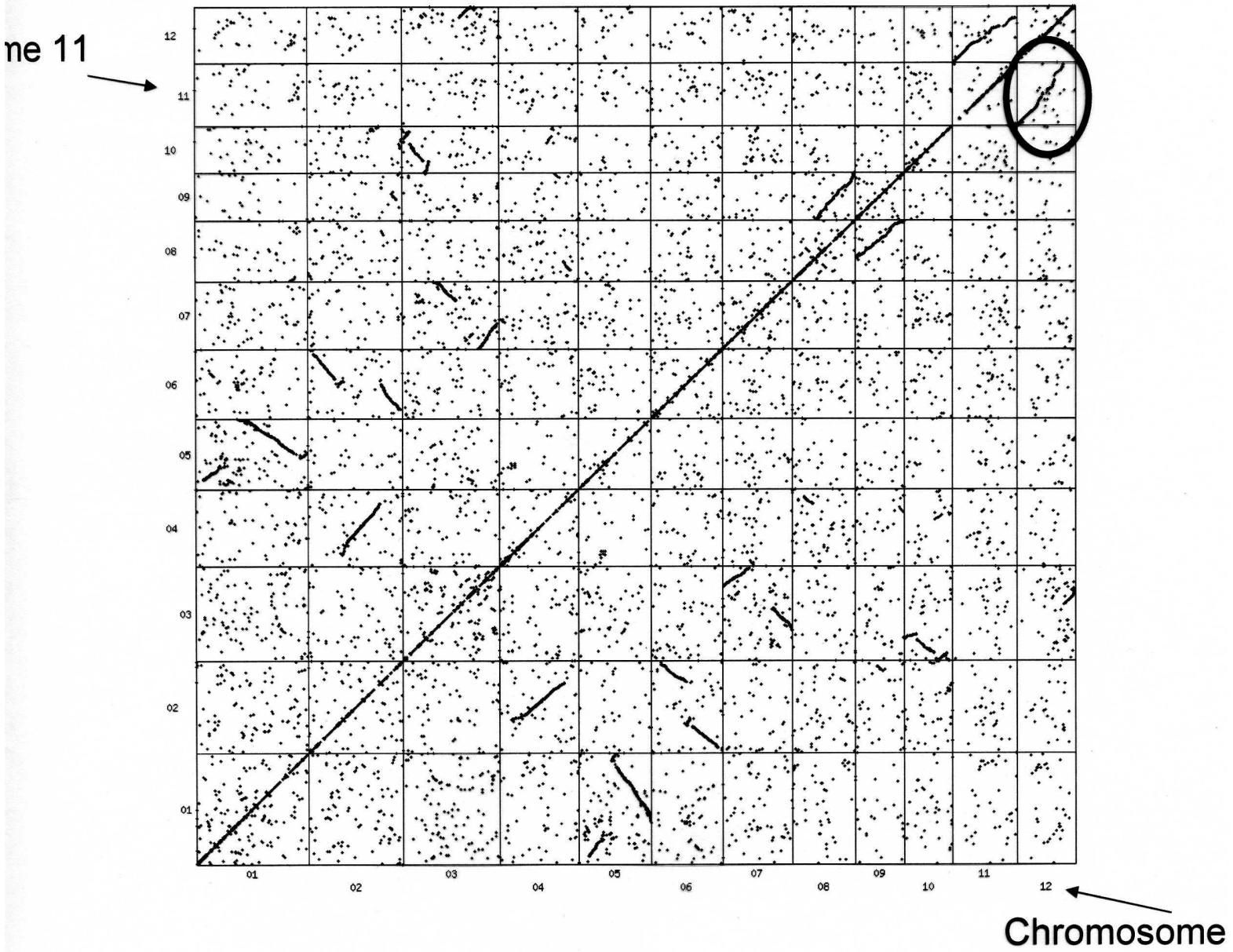
- a whole genome (polyploïdie)
- an entire chromosome (aneuploïdie ou polysomie)
  - trisomie 21 (fréquence 1/700 naissances)
  - trisomie 13 (—— 1/5000 ——)
  - trisomie 18 (—— 1/8000 ——)
- part of a chromosome
- a complete gene
- part of a gene

The proportion of sequences that has arisen by gene duplication is > 30%
(Teichman et al., Curr. Opin. in Struct. Biol. 9:390 (1999)

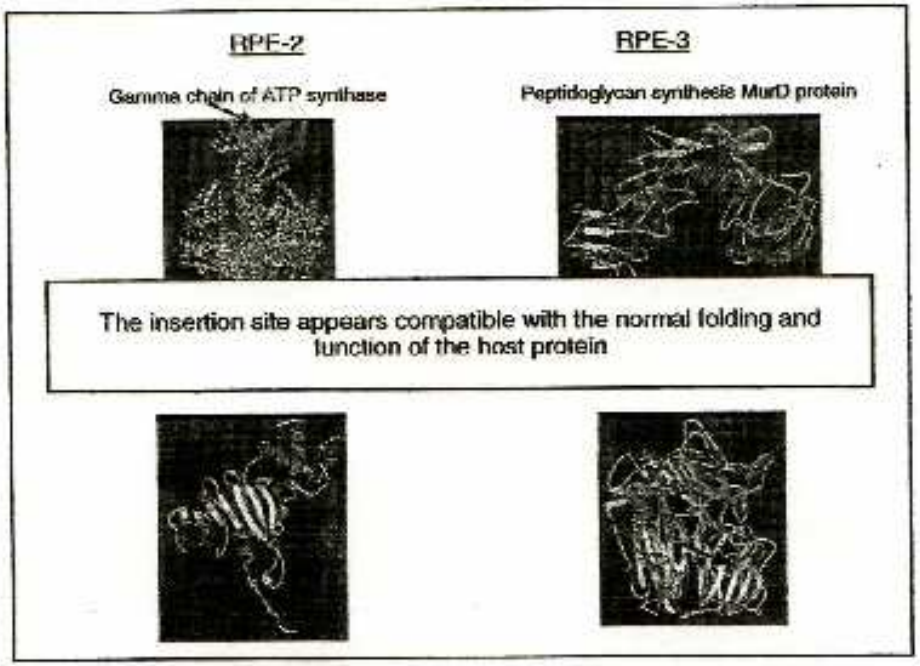
TABLE 10.1 **Examples of proteins with internal domain duplications**

Sequence (organism)	*Length of protein*[a]	*Length of repeat*	*Number of repeats*	*Percent repetition*
$\alpha_1\beta$-glycoprotein (human)	474	91	5	96
Angiotensin I-converting enzyme (human)	1306	357	2	55
Calbindin (human, bovine)	260	43	6	99
Calcium-dependent regulator protein (human)	148	74	2	100
Epidermal-growth-factor precursor (human)	1217	40	3	10
Ferredoxin (*Clostridium pasteurianum*)	55	28	2	100
Fibronectin (human)	2324	40	12	21
Guanosine cyclic 3′,5′-phosphate-dependent protein kinase (human)	670	120	2	36
Hemopexin (human)	439	207	2	94
Hexokinase (human)	917	447	2	97
Immunoglobulin γ chain C region (human)	329	108	3	98
Immunoglobulin ε chain C region (human)	423	108	4	100
Interleukin-2 receptor (human)	251	68	2	54
Interstitial retinol-binding protein (human)	1243	303	4	98
Lactase-phlorizin hydrolase (human)	1927	480	3[b]	79
Lymphocyte-activation gene 3 (LAG-3) protein (human)	470	138	2	59
Ovoinhibitor (chicken)	472	64	7	95
P-glycoprotein (human)	1280	609	2	95
Parvalbumin (human)	108	39	2	72
Plasminogen (human)	790	79	5	50
Proglucagon (rat)	161	36	3	67
Pro-von Willebrand factor (human)	2791	586	3	
		30	2	
		117	2	
		354	4[b]	86
Protease inhibitor, Bowman-Birk type	71	28	2	79

Comment visualiser une duplication génomique totale ?



5. Selfish ADN "ADN Egoiste"



See (Science 2001)

- The rate of mutation varies with external factors.

GENETICS

Can Organisms Speed Their Own Evolution?

Intriguing hints from cell and molecular biologists suggest that they might, but evolutionary biologists are not yet convinced

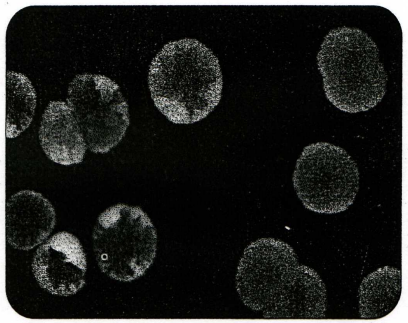
In November 1970, Miroslav Radman, a molecular geneticist now at the Université René Descartes in Paris, stunned his colleagues with a heretical proposal: that bacteria harbor a genetic program to make mutations. Through this program, Radman suspected, bacteria can crank up their mutation rates in stressful situations, helping accelerate their own evolution. Virtually no one believed him.

But 3 decades later, with the discovery of a new family of DNA-synthesizing enzymes, or polymerases, Radman feels vindicated. Unlike regular polymerases, this new family is prone to make mistakes. And recently, two independent groups led by molecular geneticist Susan Rosenberg of Baylor College of Medicine in Houston, Texas, and Patricia Foster of Indiana University, Bloomington, fingered one of them—polymerase IV—as a generator of mutations in times of stress. Says Radman: "These are the polymerases that I was dreaming about 30 years ago."

Based on these and other recent findings, the idea that organisms have ways of speeding their evolution by boosting their genetic variability is generating increasing excitement among a group of cell and molecular biologists. Within the past 2 years, for instance, researchers have unearthed molecular clues that could help explain apparent increases in genetic variability not only in bacteria but in eukaryotes as well. "I think it's very cool stuff that, amazingly, not enough people have really gotten the scoop on," says Rosenberg.

But a number of evolutionary biologists are trying to put the brakes on this mounting enthusiasm. Although the critics say the new molecular findings are intriguing, they question their origins and role in evolution. Specifically, these biologists say it is uncertain whether these processes were selected for their ability to generate variability in the first place. Nor is it clear whether they accelerate long-term evolution. Because most mutations are harmful, increased variability may often be costly to individuals and species. "It is hard to see how selection would directly favor a process that generated random variation or even one that just preserved it," says evolutionary biologist Jon Seger of the University of Utah in Salt Lake City.

Thus, a spirited tussle has ensued as researchers from these two camps put forth their interpretations of the new molecular findings. With roots dating back to Charles Darwin in the mid-1850s, the question of whether organisms harbor systems for adjusting their own rate of evolution remains open.



Bacterial lunch. Adaptive mutations allow a strain of *E. coli* to feed on lactose assessed by a blue indicator dye (solid blue colonies on right). Bacteria can also acquire this ability by amplifying the lactose-digesting genes, a temporary phenomenon (left).

CREDIT: SUSAN ROSENBERG

Error-prone enzymes

For decades, most biologists have worked under the assumption that mutation rates are constant and that individual organisms passively submit to the forces that shape evolution. Yet the idea that organisms may modulate their genetic variability has surfaced from time to time. Even Darwin suggested in *The Origin of Species* that environmental changes resulting from animal domestication affect variability.

But that didn't prepare members of the biological community for the jolt they received in 1988, when molecular biologist John Cairns and his colleagues at the Harvard School of Public Health published in *Nature* an even more shocking idea than Radman's. Cairns proposed that, depending on their environmental conditions, bacteria might be able to direct mutations to particular genes. The dogma-shattering idea that mutations might not be completely random "touched a raw nerve," says Cairns. It smacked of "Lamarckism"—a reference to Jean-Baptiste Lamarck's now-discredited theory that species evolve through the inheritance of characteristics acquired during an organism's lifetime. Outraged, a number of evolutionary biologists quickly embarked on their own studies to test the notion. The flurry of studies ultimately revealed that Cairns's original proposal was untenable, and the community, including Cairns, now at the Radcliffe Infirmary in Oxford, United Kingdom, discarded it.

Thanks to the commotion it ignited, however, Cairns's article prompted the study of a new phenomenon: the increased mutation rate observed in *Escherichia coli* during times of stress—in particular, starvation. Most of the evidence for this phenomenon comes from studies of a strain of *E. coli* that carries a mutation inactivating the lac operon, a group of genes that allow bacteria to digest lactose. When cells have plenty of food choices, they rarely acquire mutations that counteract the lactose deficiency. Yet, as Cairns and Foster described in the journal *Genetics* in 1991, when lactose is the only choice on the menu, rates of these compensating mutations skyrocket.

Over the past decade, researchers have been dissecting the molecular underpinnings of these so-called adaptive mutations. And within the last 2 years, they have made impressive strides. They have found, for example, that although these mutations are not directed to particular genes, as Cairns originally suggested, they don't uniformly pepper the bacterial genome either. "There are hot and cold regions for hypermutation," says Rosenberg, who is now working on defining these regions. "All regions are not equal."

One of the most exciting findings has been the discovery of the error-prone polymerases. "It's a great novelty," says Radman. "We knew of these *E. coli* genes for over 20 years but couldn't recognize them as polymerases." That changed in 1999, when researchers found that these proteins could copy lesioned DNA in a test tube.

Microbiologists already knew that when bacteria suffer DNA damage, they switch on a response, called SOS, that arrests the cell cycle and turns on genes that repair DNA and allow its duplication. They suspected that these genes might help regular polymerases avoid getting stuck when they run into a damaged stretch of DNA. But by monitoring the activities of the SOS-

- the rate of mutation is ≠ for different proteins

PAM = Percent of Accepted Mutations

if PAM corrected for multiple mutations

$$1u = \frac{1\ \text{PAM}}{108\ \text{years}}$$

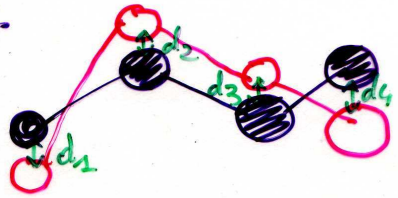
cytochrome c	5 u
fibrinopeptides	70 u
enzymes on average	5-50 u

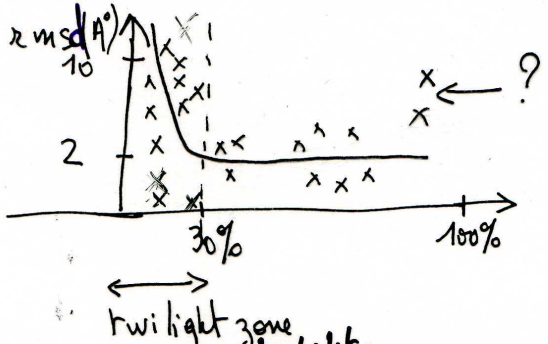
- the rate of mutation varies within a protein
  - catalytic sites
  - binding sites
  - folding sites
  - thermodynamic stability sites

consequences on structures

The level of structural change is directly related to the level of sequence change.

$$\text{rmsd} = f(\%\ \text{identity}) = \sqrt{\frac{\sum_{i=1}^{N} d_i^2}{N}}$$





→ New: Amyloid proteins

consequences on stability

→ stability essentially conserved

consequences on functions remain to be examined.
(see divergent evolution or convergent evolution)

- divergent evolution : identical ancestor but ≠ functions → significant sequence identity.
- convergent evolution : proteins start out with different sequences and remain ≠ in sequence, but they evolve to perform similar functions.