

1. Introduction

NOTRE SIÈCLE commence avec des avancées spectaculaires en génomique. Le cours examinera les méthodes informatiques qui supportent cette révolution. On vise une approche intégrative de biologie et d'informatique.

Organisation du cours

Style : séminaire interactif avec discussions

Sujets : études de génomes humaines 1

- \* technologies de séquençage
\* constitution du génome
\* diversité génétique
\* maladies humaines
\* histoire humaine

Sujets : études de génomes diverses

- \* testabilité de processus évolutifs : adaptation, spéciation
\* inférence de fonctionnalité d'éléments génomiques
\* organismes et environnements ancestraux
\* interactions entre organismes et diversité d'écosystèmes

Matériel du cours

- \* quelques livres de référence
[EG] Warren J. Ewens et Gregory R. Grant. Statistical Methods in Bioinformatics.2
[DEKM] Richard Durbin, Sean R. Eddy, Anders Krogh et Graeme Mitchison. Biological Sequence Analysis : Probabilistic Models of Proteins and Nucleic Acids.3
[Nielsen] Rasmus Nielsen. Statistical Methods in Molecular Evolution.4
\* 80 articles récents : 1-2 articles obligatoires à lire chaque semaine

Évaluation

- \* mini-questionnaires sur articles obligatoires : 10%
\* 3 devoirs (théorie, programmation, analyse de données actuelles) : 20-20-20%
\* présentation d'un sujet choisi : 30%

L'ADN

- \* composé de deux brins complémentaires => mécanisme de copiage du matériel génétique



1 Au début :

- 1. notion de homologie, contexte phylogénétique, évolution de répertoire de gènes
2. modèles statistiques de substitutions, trous, alignements
3. recherche de homologies entre génomes
4. séquençage de génomes
5. génomes personnels

2 Springer-Verlag, 2001. (2ème édition en 2005.) Texte surtout sur les aspects statistiques de la bioinformatique. Très utile pour la modélisation sensible de problèmes, ce qui est essentielle pour la conception d'algorithmes efficaces.

3 Cambridge University Press, 1988. Texte classique, avec un exposé de modèles de Markov caché, et les grammaires stochastiques.

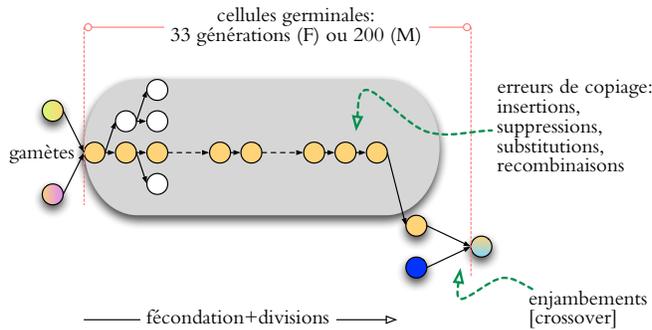
4 Springer-Verlag, 2005. Collection d'articles survolant des aspects avancés d'inférence avec des modèles probabilistes d'évolution.

Article snippet from Nature (April 25, 1953) titled 'MOLECULAR STRUCTURE OF NUCLEIC ACIDS' by Watson and Crick. It describes the discovery of the DNA double helix structure. A red box highlights a quote: 'It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material'.

★ la séquence (A,C,G,T) «se lit» dans le sens 5' → 3' sur un brin ou l'autre

### Descendance avec modification

Selon la théorie populaire de Charles Darwin, les organismes contemporains se descendent d'un ancêtre commun. Les enfants héritent les traits de leurs parents, mais *avec modifications*.



L'enfant hérite les allèles du parent et les nouvelles mutations dans la lignée de cellules germinales).

La probabilité de fixation dans la population entière dépend (a) de la taille de la population, et (b) l'effet de l'allèle (avantage ou désavantage possible).

### Taille de génomes

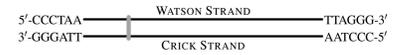
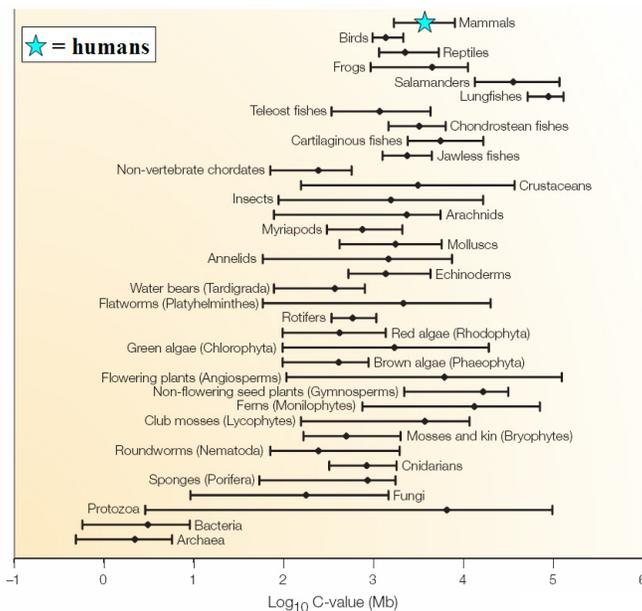


FIG. 1: Brins "Watson" et "Crick" (le brin Watson commence sur le petit bras, ou — dans d'autre contexte — il donne la séquence de référence dans la base de données)

FIG. 2: Les mutations lors d'héritage du matériel génétique entre générations introduisent de nouveaux allèles dans une population.

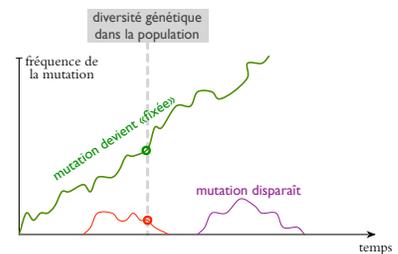


FIG. 3: Ultiment, l'allèle devient fixé dans la population entière, ou disparaît complètement.

FIG. 4: La taille du génome ne correspond pas à la «complexité» de l'organisme : influencée par taux d'erreurs de copiage + coût de camelote ADN + taille de population. [Gregory *Nat Rev Genet*, 6 :699-708, 2005]