

MCMC et analyse génétique



CM_pG

Pierre BERTHIER, Université de Berne, Suisse

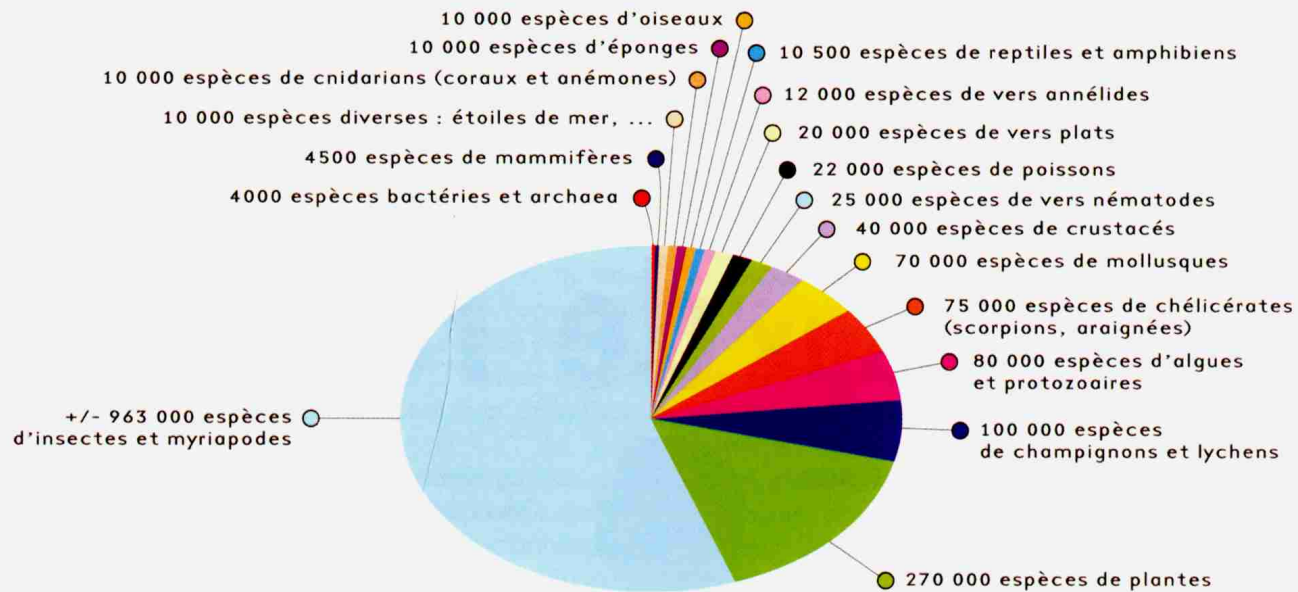
<http://cmpg.unibe.ch>

Plan

- La biologie évolutive
- Généalogies de gènes
- Applications

La biodiversité

les espèces en chiffre



Le règne animal compte +/- 1 386 000 espèces pour lesquelles on suppose plusieurs origines différentes. Les plantes quant à elles comptent +/- 270 000 espèces pour une seule origine connue.

L'arbre de la vie

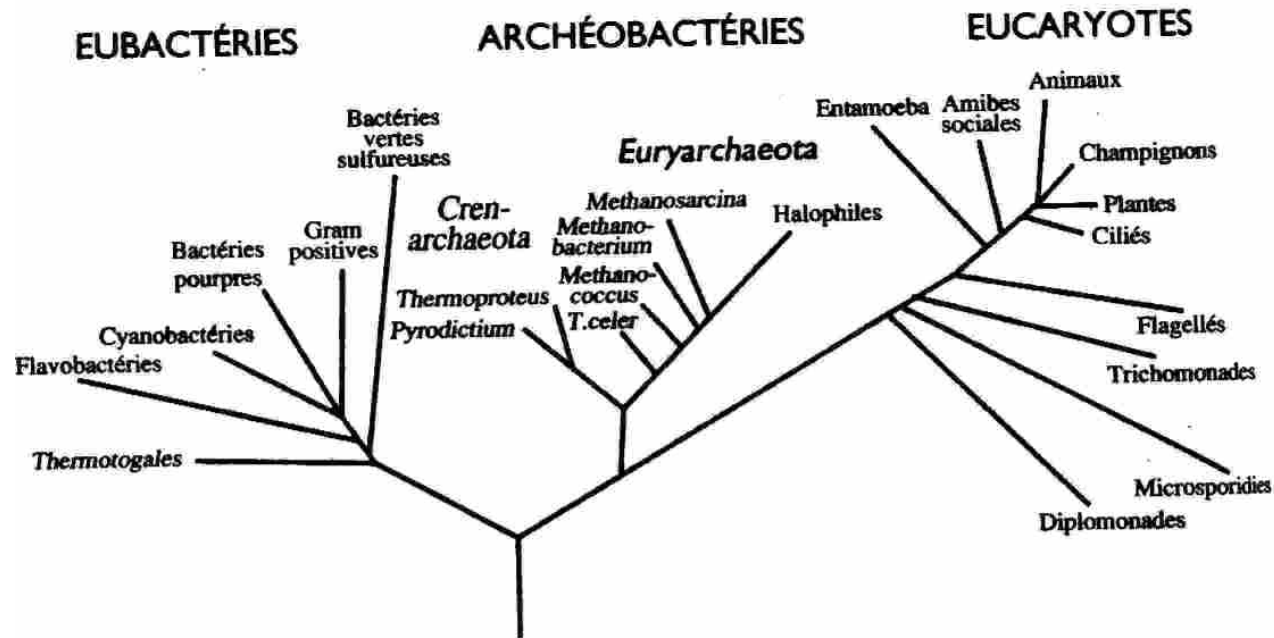
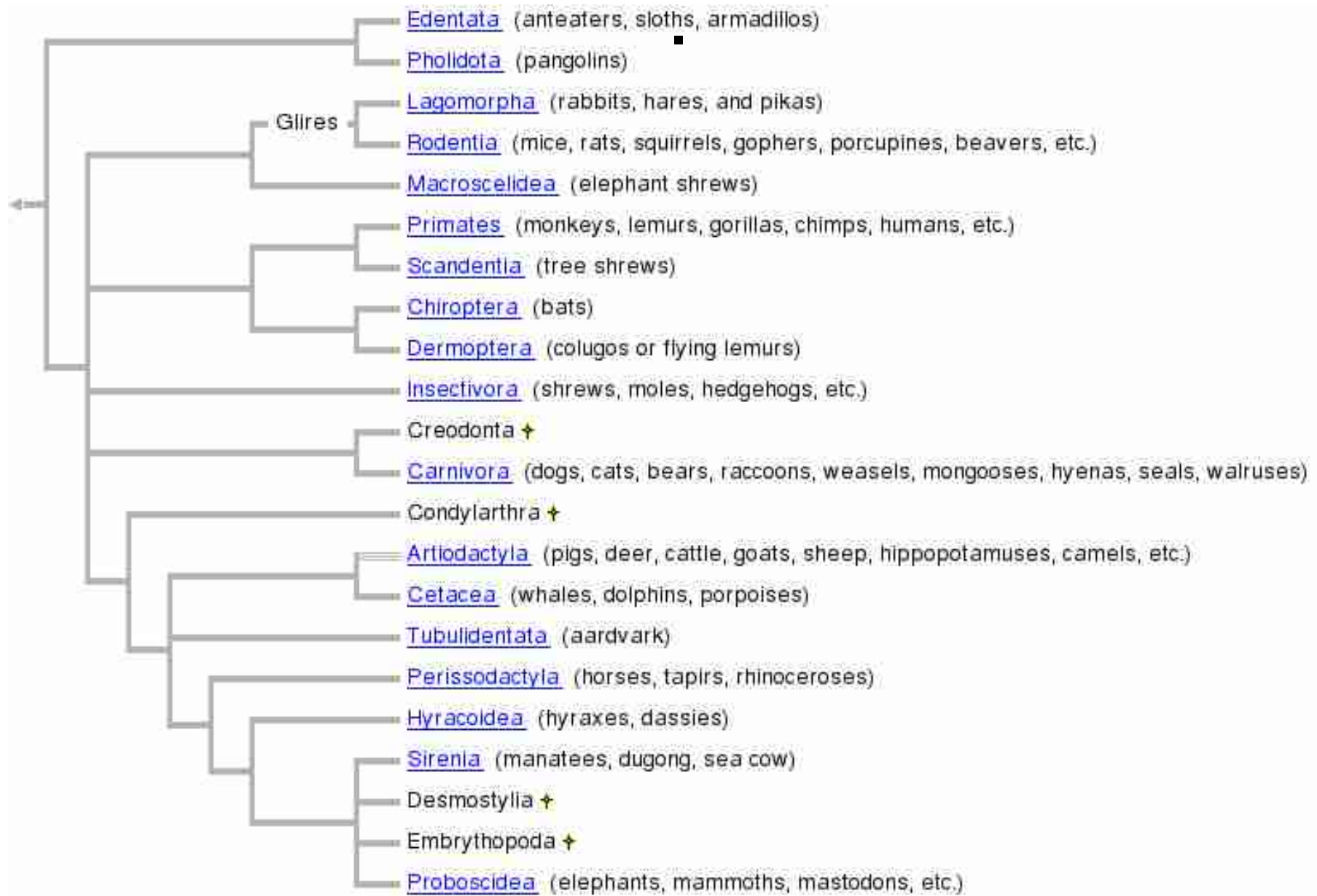


FIGURE 31

L'arbre évolutif de la vie comprend deux rameaux procaryotes et seulement un rameau eucaryote. Les plantes, les animaux et les champignons ne sont que de petites ramilles de ce dernier rameau.



Hérédité et variation

1735 Systématique, LINNÉ, *Systema Naturae*

1801 Le Transformisme, LAMARCK, *Philosophie zoologique*

1859 La Sélection Naturelle, DARWIN, *L'origine des espèces*

1865 Lois de l'héritabilité des caractères, MENDEL

1869 Découverte de l'« acide nucléique », MIESCHER

1900 Les lois de Mendel, DE VRIES, CORRENS, TCHERMAK

1944 ADN = support de l'hérédité, AVERY, MACLEOD, MCCARTY

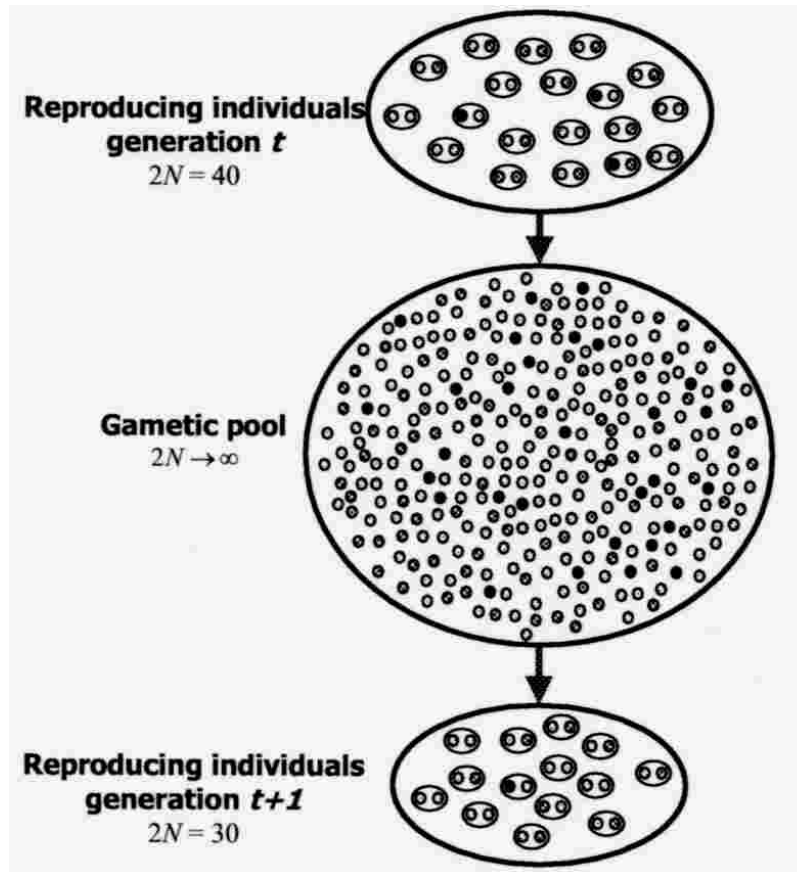
1953 La « double hélice » de l'ADN, WATSON ET CRICK

1961 Le code génétique, NIRENBERG ET MATTHAEI

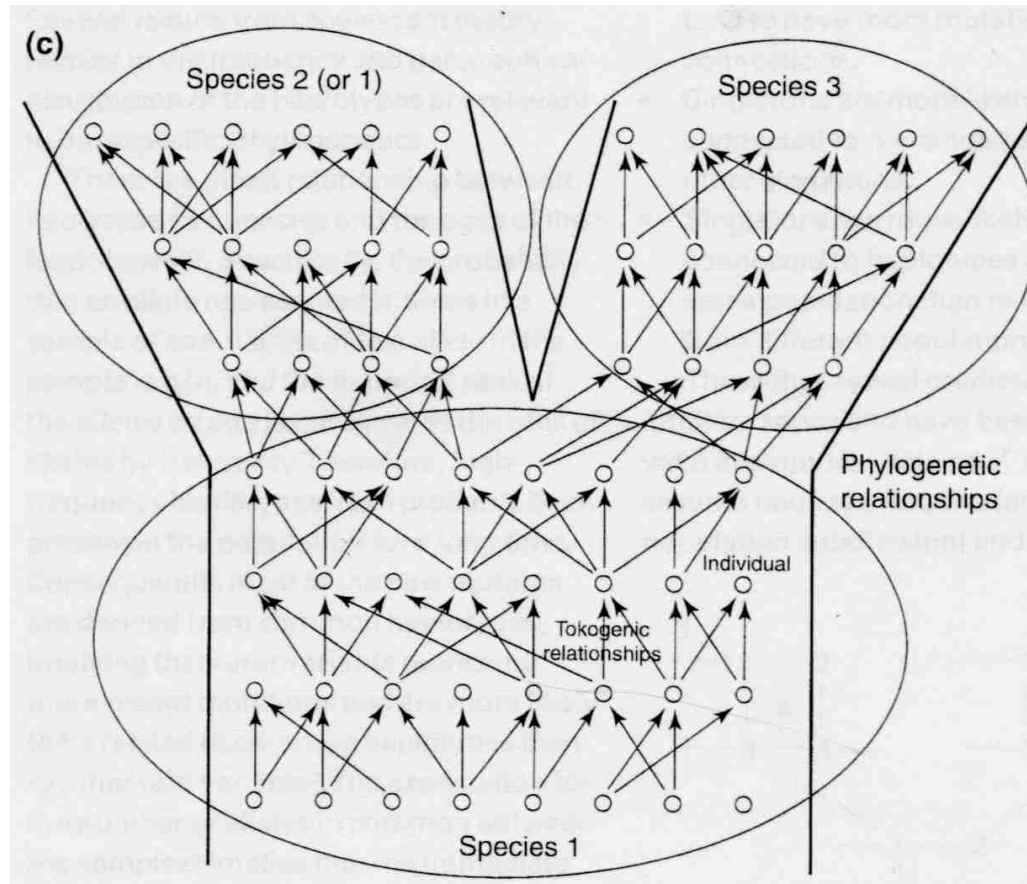
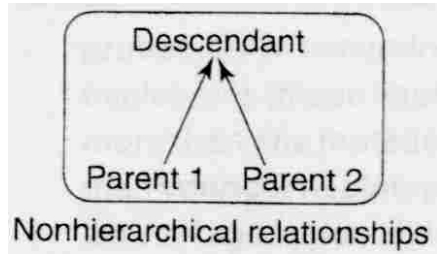
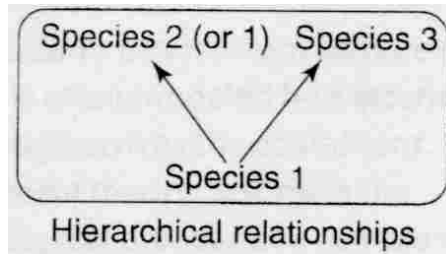
1970 Séquençage de l'ADN

1980 PCR

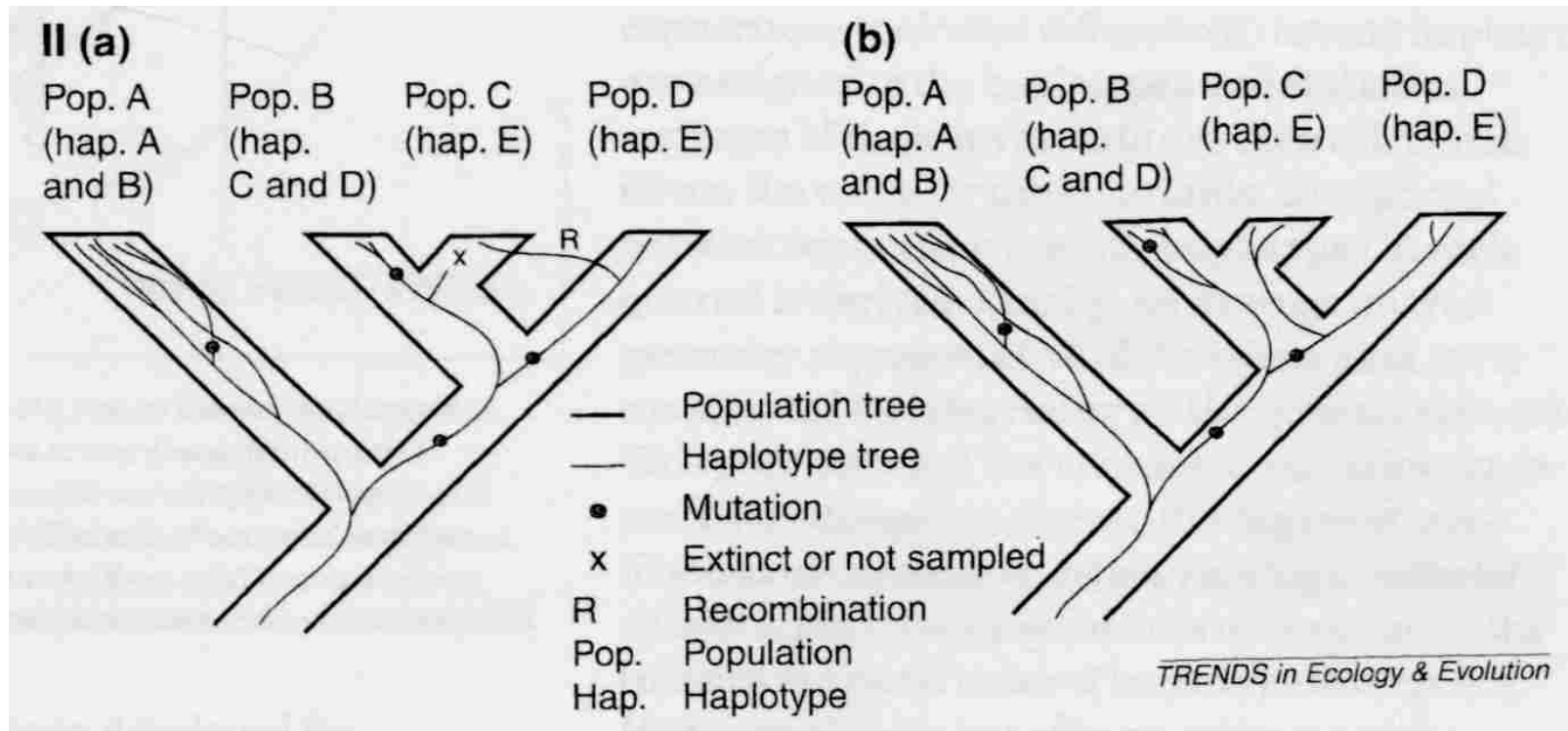
Une population



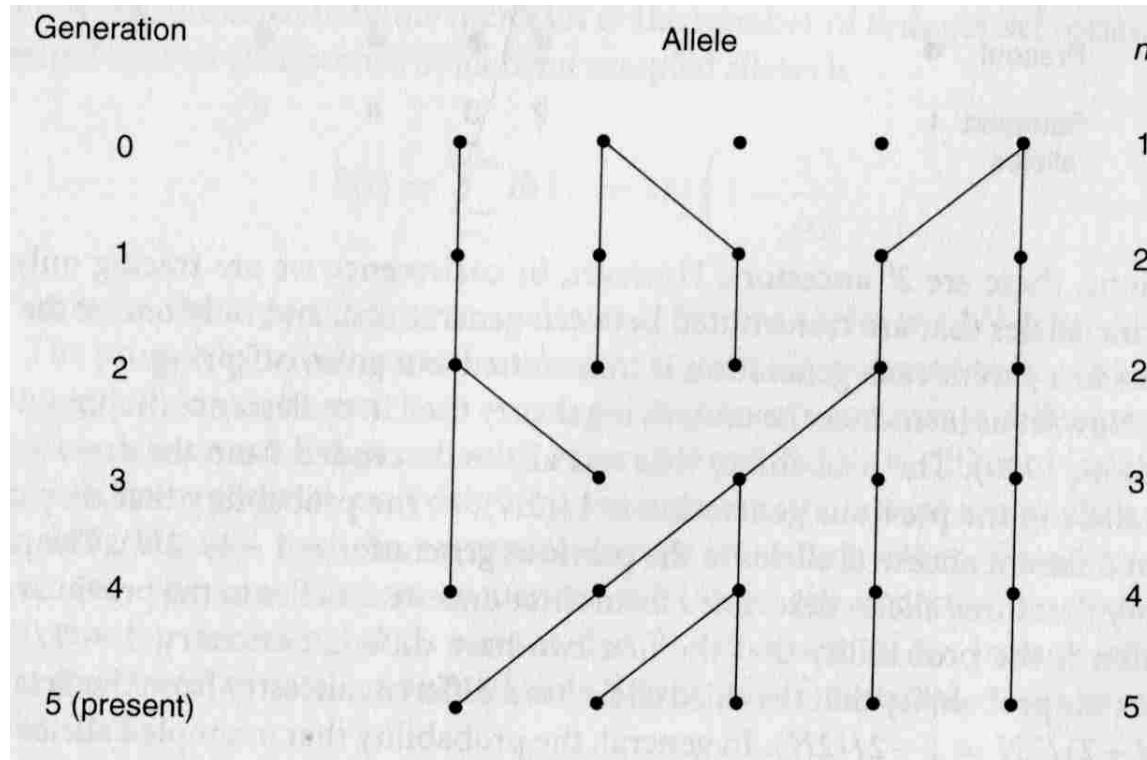
Généalogies



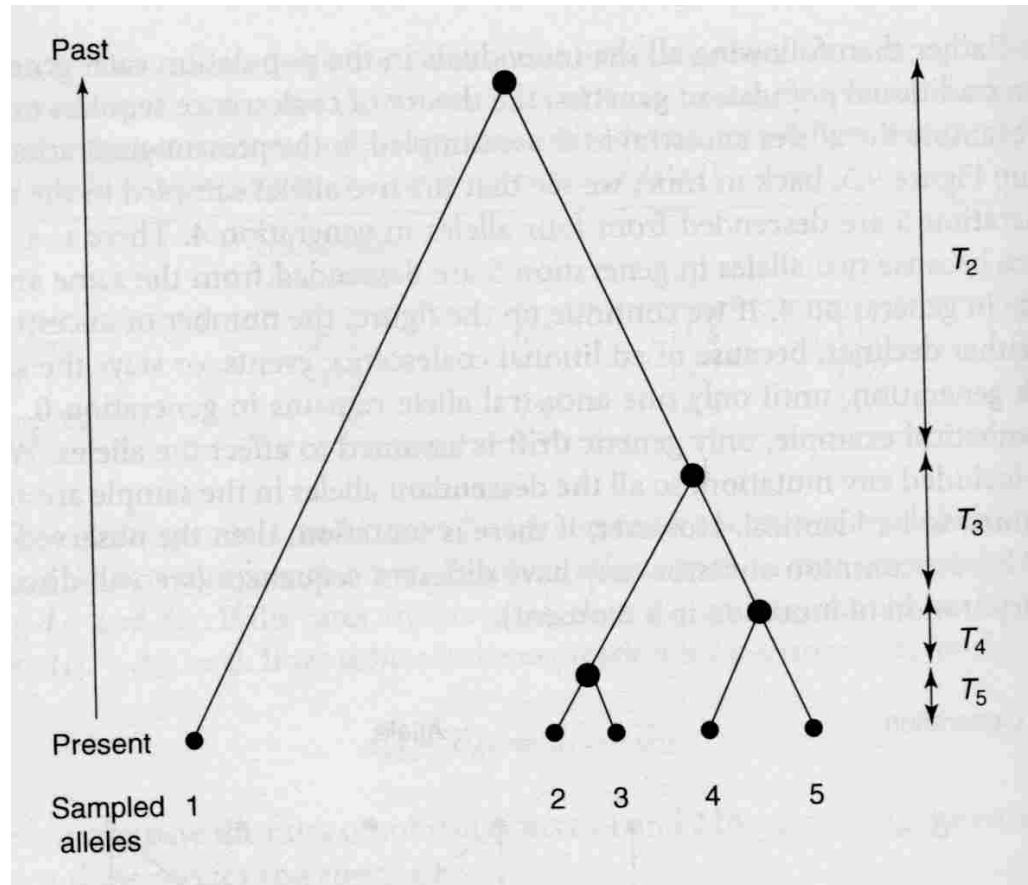
Généalogies



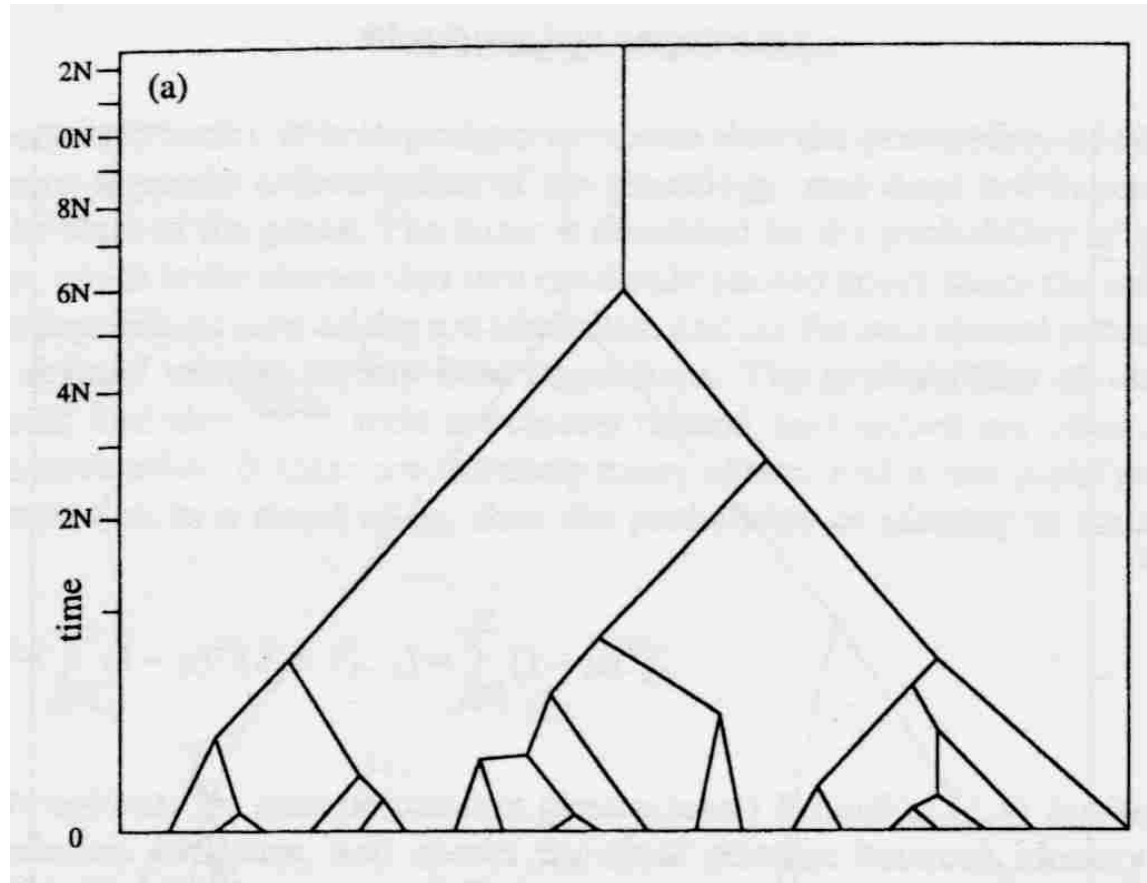
Généalogies



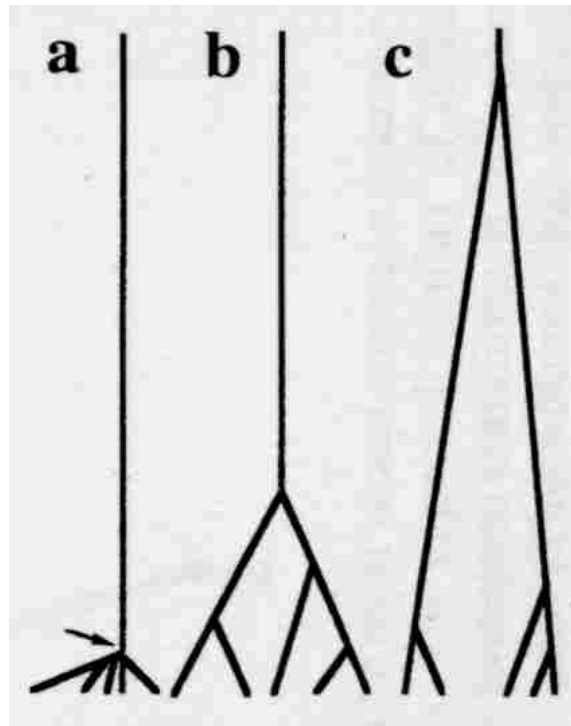
Généalogies



Généalogies



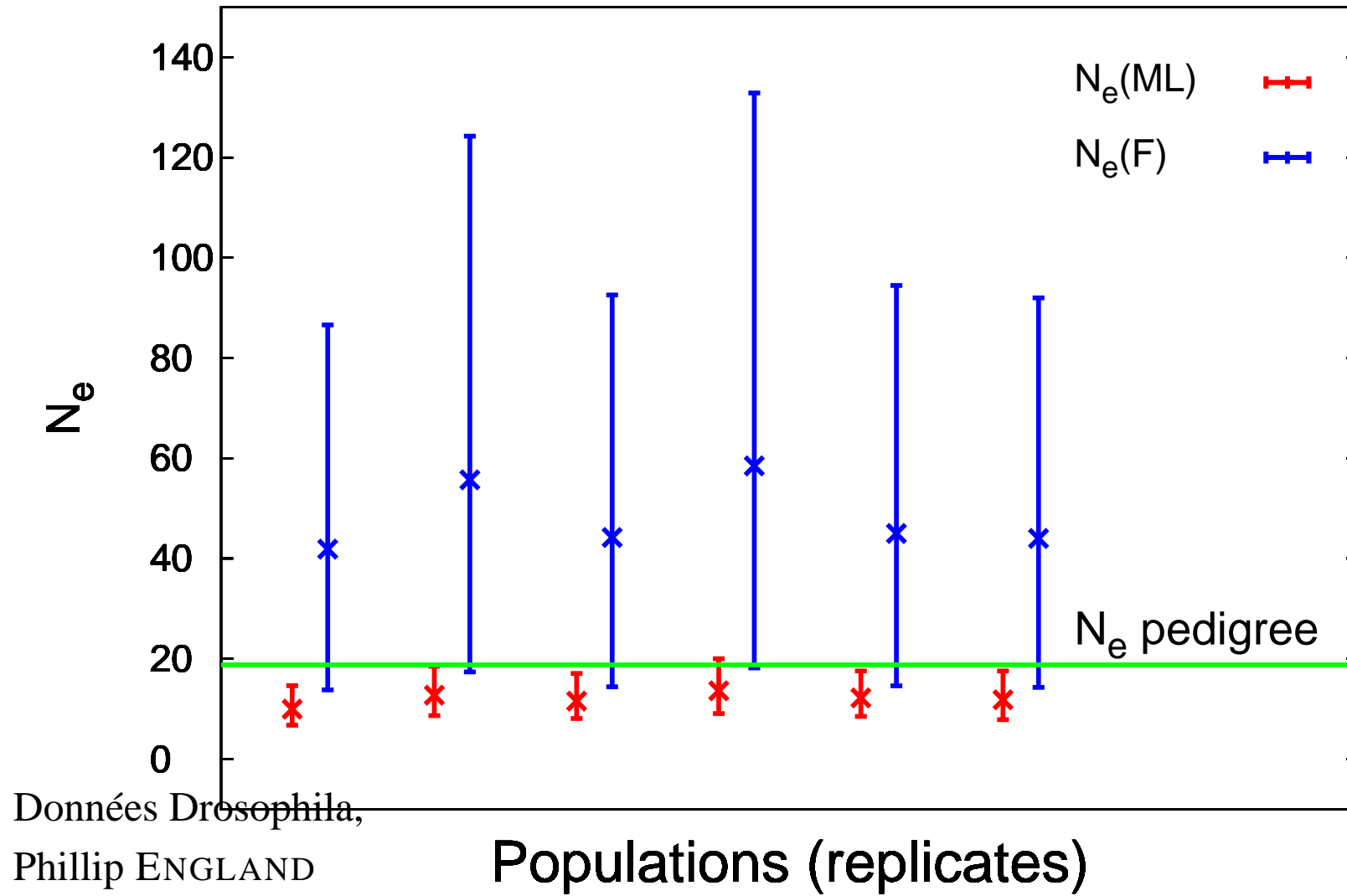
Généalogies



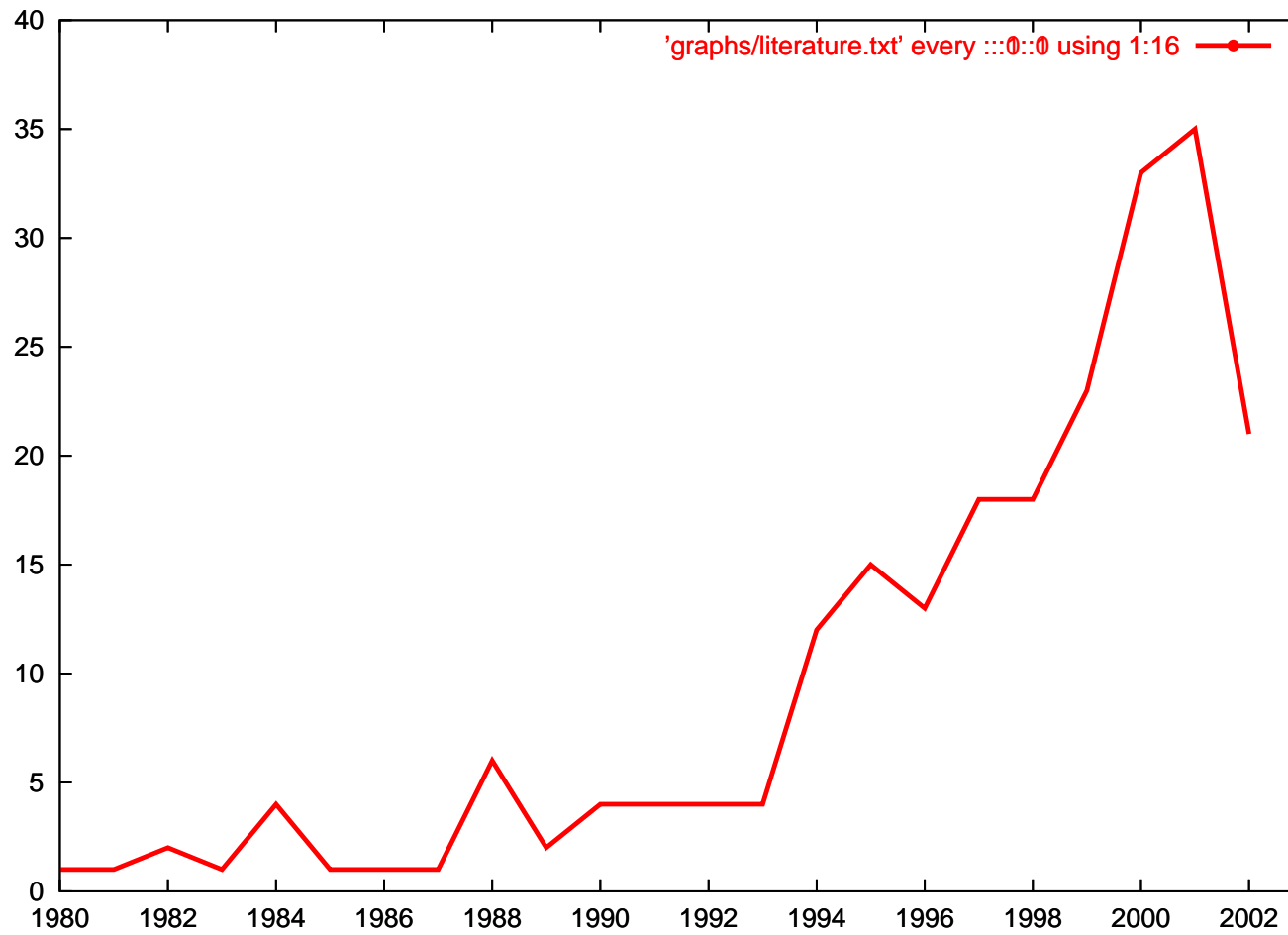
MCMC et Génétique des populations

- Modèles : depuis p.ex. FISHER, 1930 ; WRIGHT, 1931 ; KINGMAN, 1980
- Données : accessibles depuis 1980, produites en quantité depuis ~1990 (séquençage de l'ADN par PCR)
- Nouveaux types de données (SNPs, Génomes) → grand volumes
- Puissance de calcul pour MCMC, espace disque, RAM
- Évaluation des méthodes MCMC

Avantages du MCMC



MCMC Literature in *Genetics*



Évaluation des méthodes MCMC

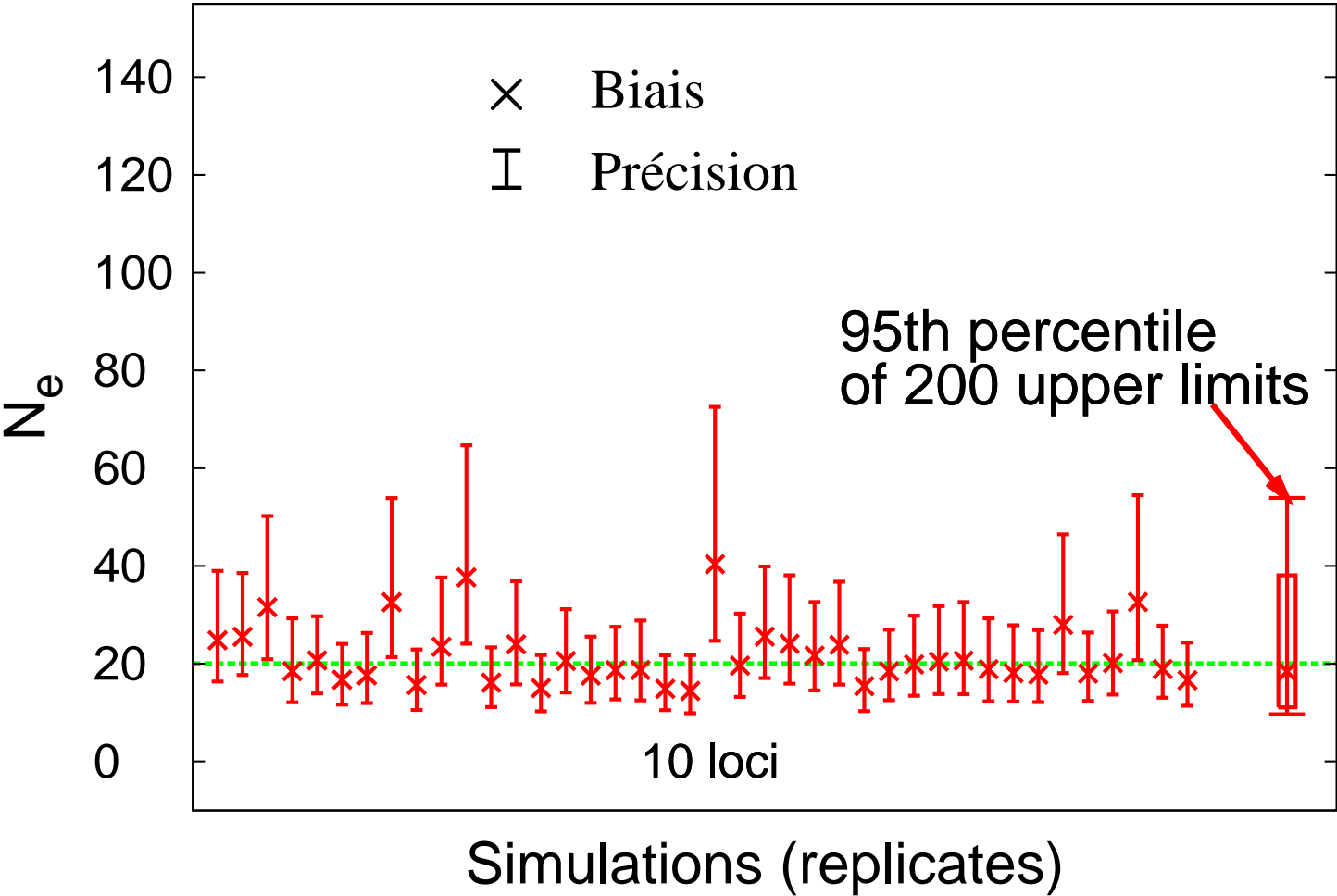
Les modèles en génétique des populations reposent sur certaines **hypothèses**, qui ne sont presque **jamais réalisées**, par exemple :

- pas de mutation
- marqueurs neutres (pas de sélection) et indépendants
- population(s) fermée(s)

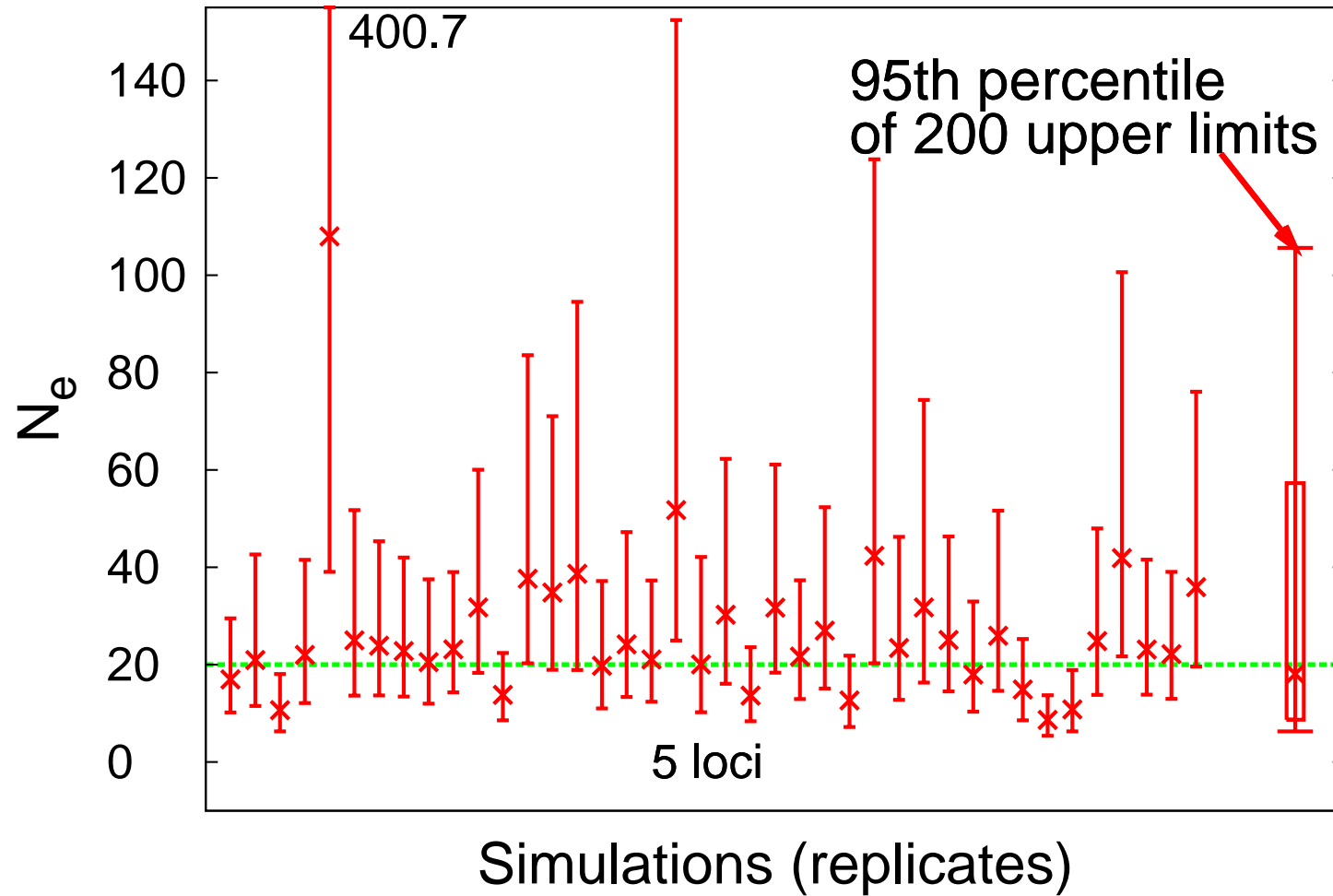
La question est : **dans quelles conditions** sont-ils **utilisables** ?

- quel type de marqueurs (données) génétiques ?
- quelle taille d'échantillon ?
- influence de l'histoire des populations ?

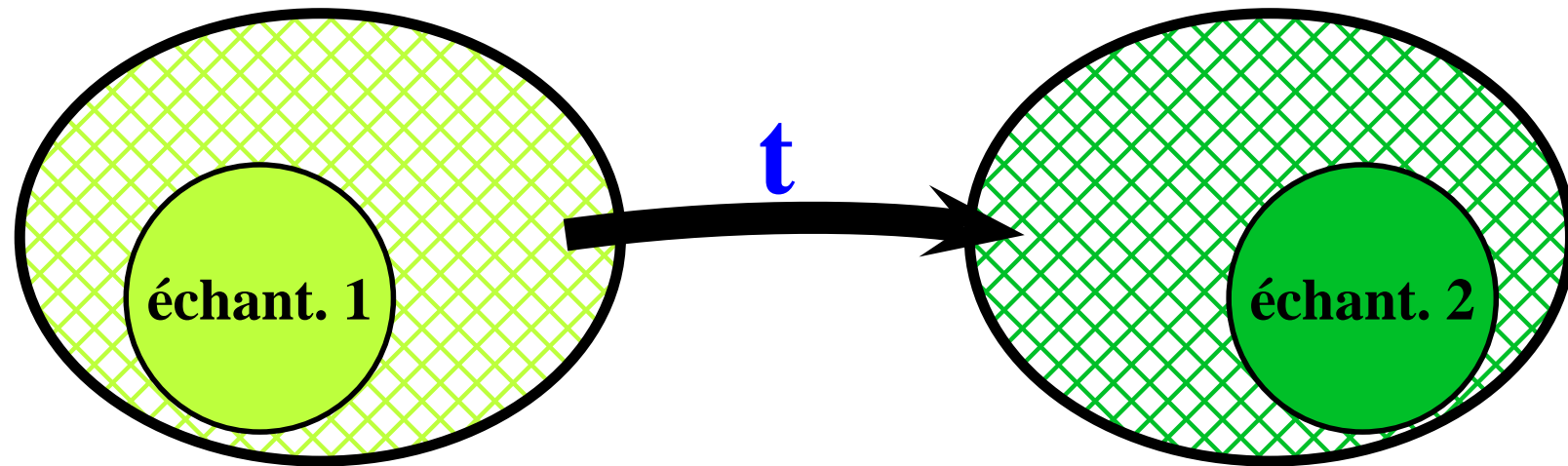
Exemple : un estimateur de l'effectif efficace



Exemple d'évaluation



Un estimateur classique de N_e , $N_{e_{F_k}}$

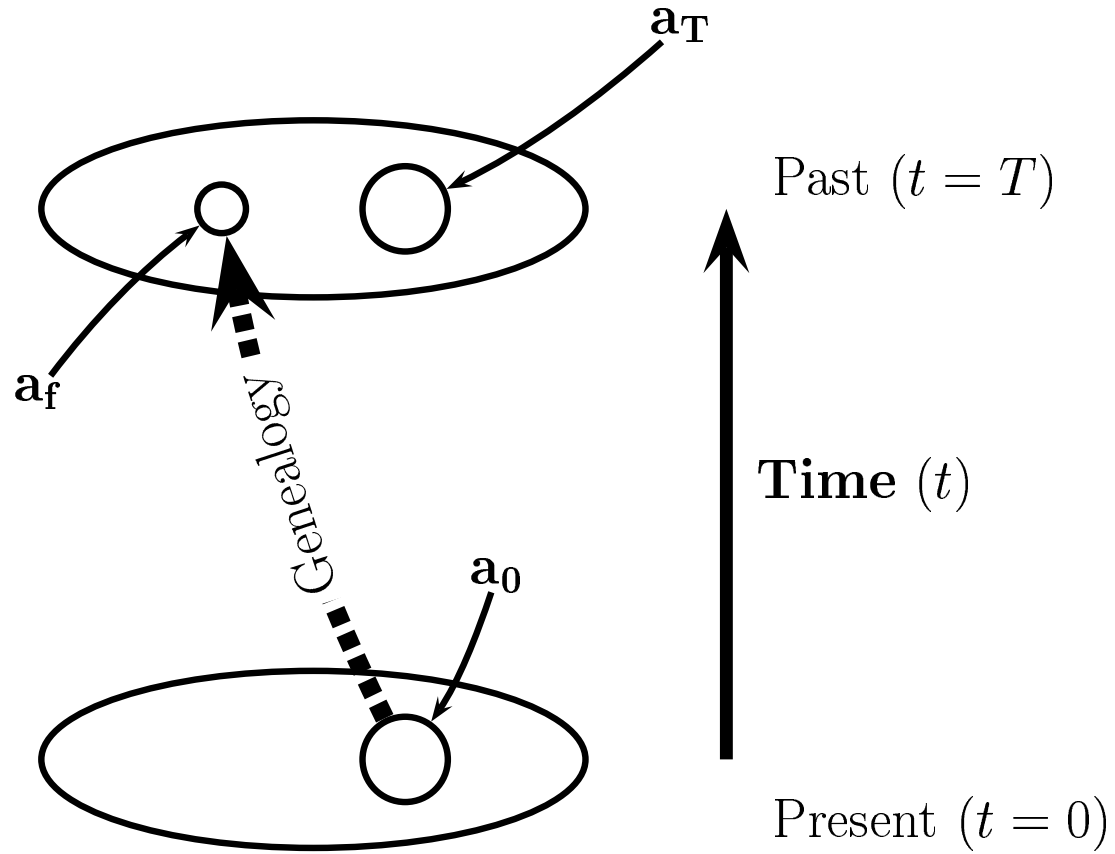


F_{ST} - "temporel"

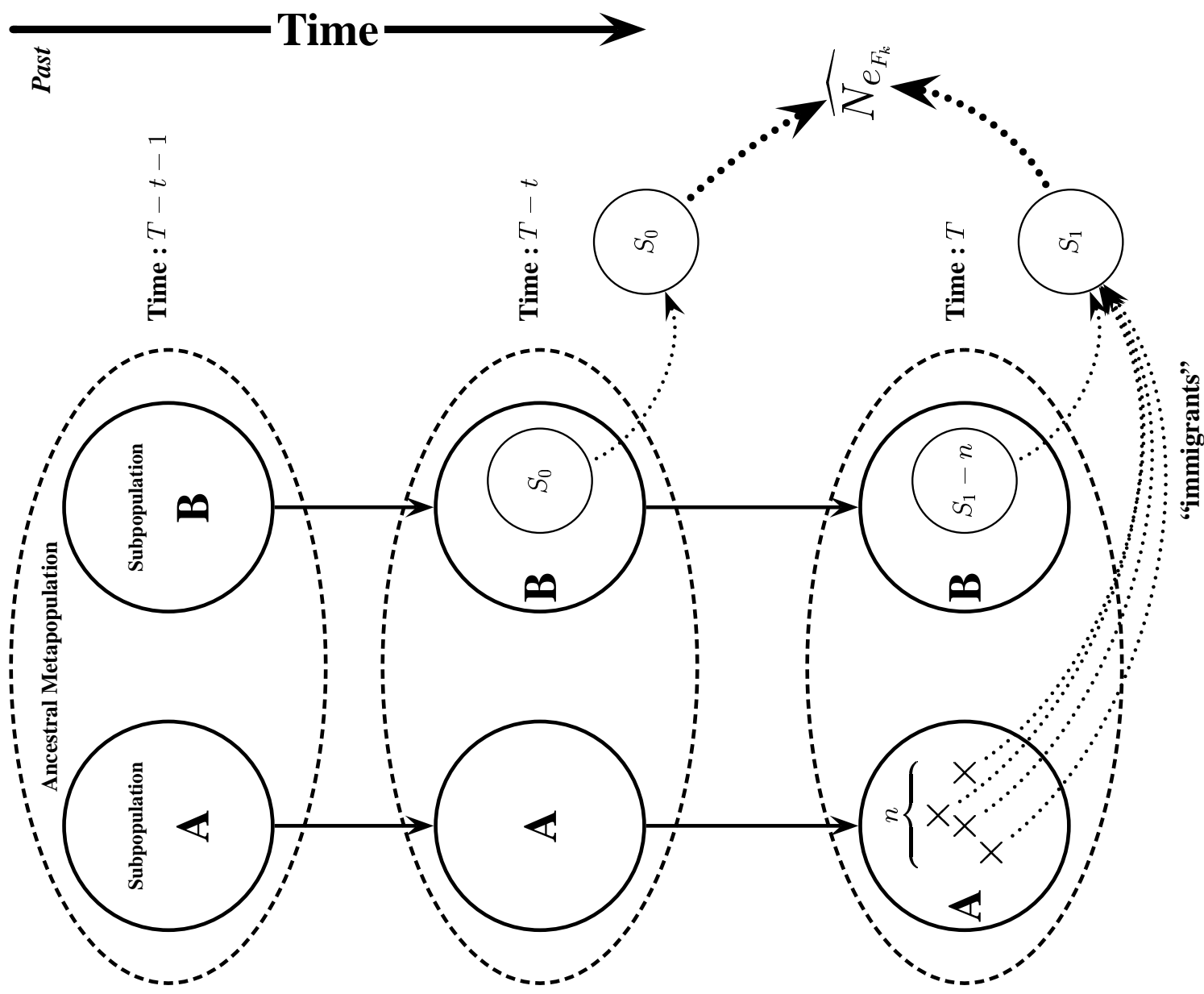
↓

$$N_e \text{ "variance" } (\widehat{N_{e_{F_k}}}) \sim \frac{t}{2 \times F_{ST}}$$

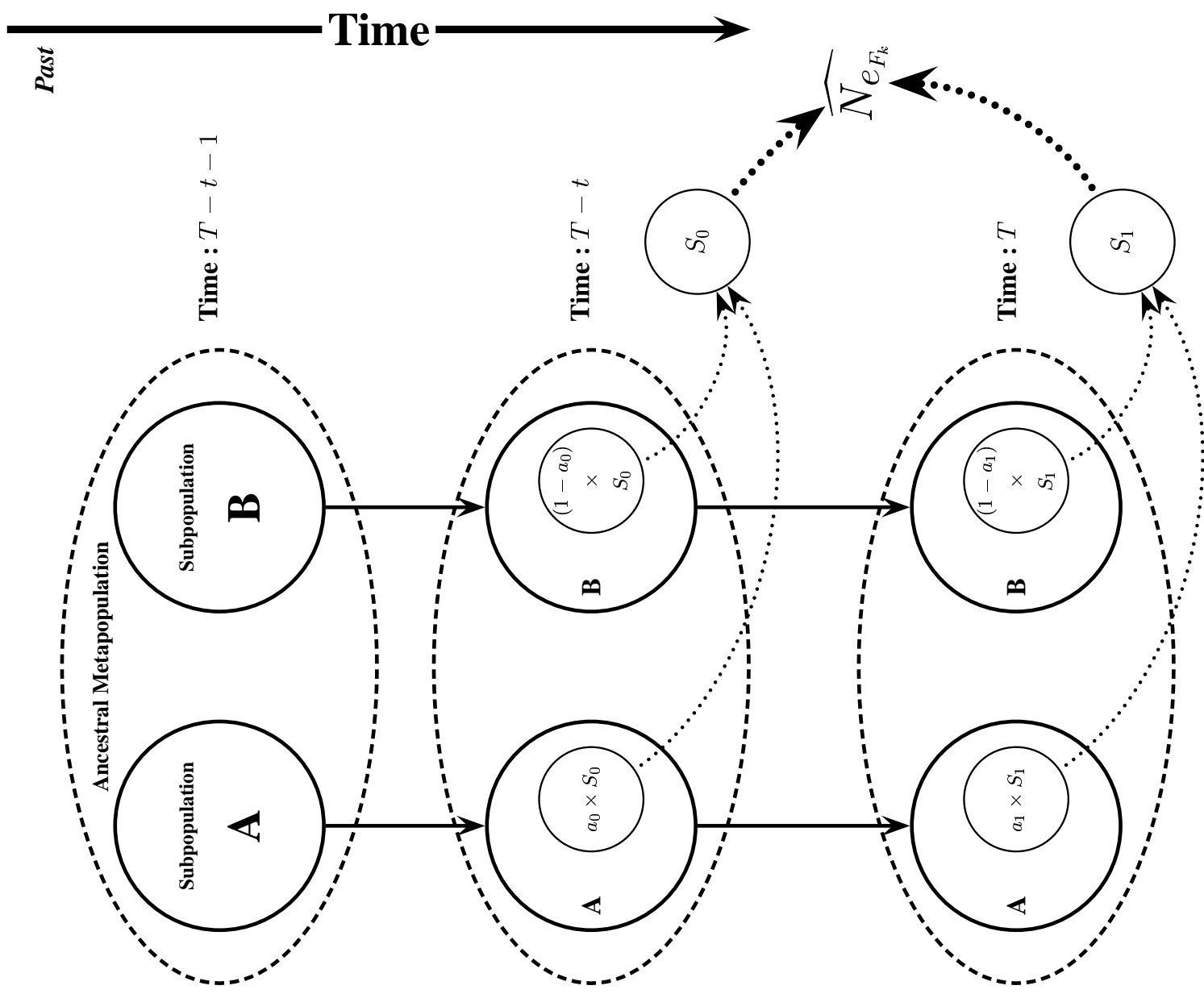
$N_{e_{ML}}$ model



Migration model



Substructure model



Usage des MCMC en génétique des pops

1. Taille efficace (N_e)
2. Consanguinité
3. Relations de parenté
4. Phase haplotypique
5. Proportions d'admixture de deux populations
6. Structuration de la variance génétique
7. Locus sous sélection
8. Cartographie des gènes
9. Âge d'une mutation
10. Tests d'assignation
11. Détection des hybrides
12. Paramètres de Migration
13. Démographie passée
14. Déséquilibre de liaison
15. Phylogenies
16. Horloge moléculaire (datation dans les phylogenies)

Applications

1. Médecine
2. Médecine légale (forensics)
3. Biologie de la conservation
4. Gestion de stocks (agriculture, élevage, populations naturelles)
5. Systématique
6. Histoire des populations