



HAL
open science

Inférer l'histoire démographique et sélective de populations à partir de leur diversité génétique: apports de la génomique. Application au cas des populations de poissons migrateurs

Sylvie Oddou-Muratorio

► To cite this version:

Sylvie Oddou-Muratorio. Inférer l'histoire démographique et sélective de populations à partir de leur diversité génétique: apports de la génomique. Application au cas des populations de poissons migrateurs: Application au cas des populations de poissons migrateurs. Master. UPPA campus d'Anglet, France. 2024. hal-04542693

HAL Id: hal-04542693

<https://hal.inrae.fr/hal-04542693>

Submitted on 11 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Inférer l'histoire démographique et sélective de populations à partir de leur diversité génétique: apports de la génomique

Application au cas des populations de poissons migrateurs
Master 2 Dynamique des Ecosystèmes Aquatiques

Sylvie MURATORIO, UMR ECOBIOP

INRAE



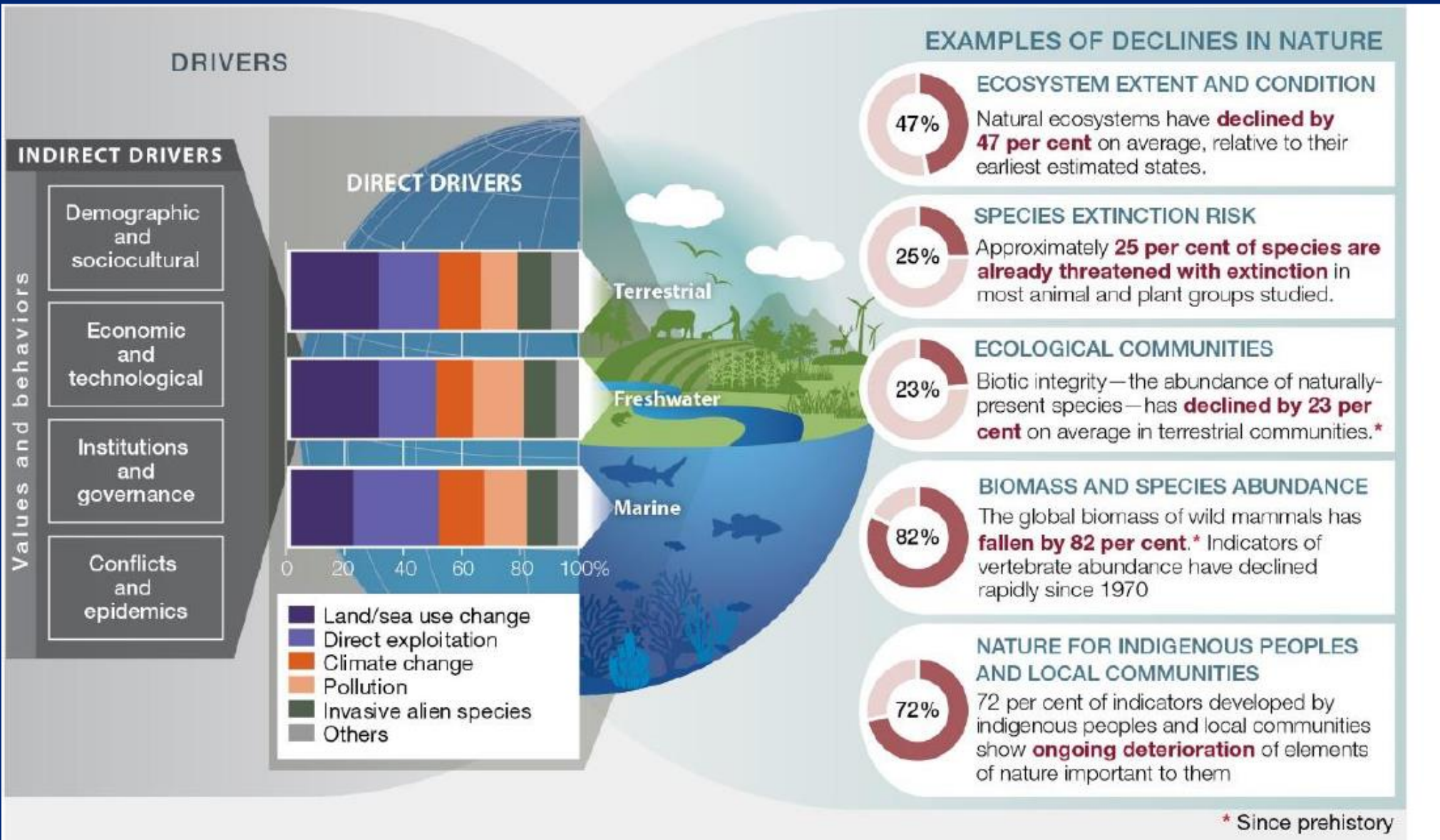
Plan du cours

- I. Introduction
- II. Rappels de base en génétique des populations
- III. Estimation de la démographie historique : principes fondamentaux, méthodes courantes, étude de cas
- IV. Estimation de l'histoire sélective et de l'adaptation génétique: principes fondamentaux, méthodes courantes, étude de cas
- V. Vers l'inférence conjoint de processus démographiques et sélectifs

Plan du cours

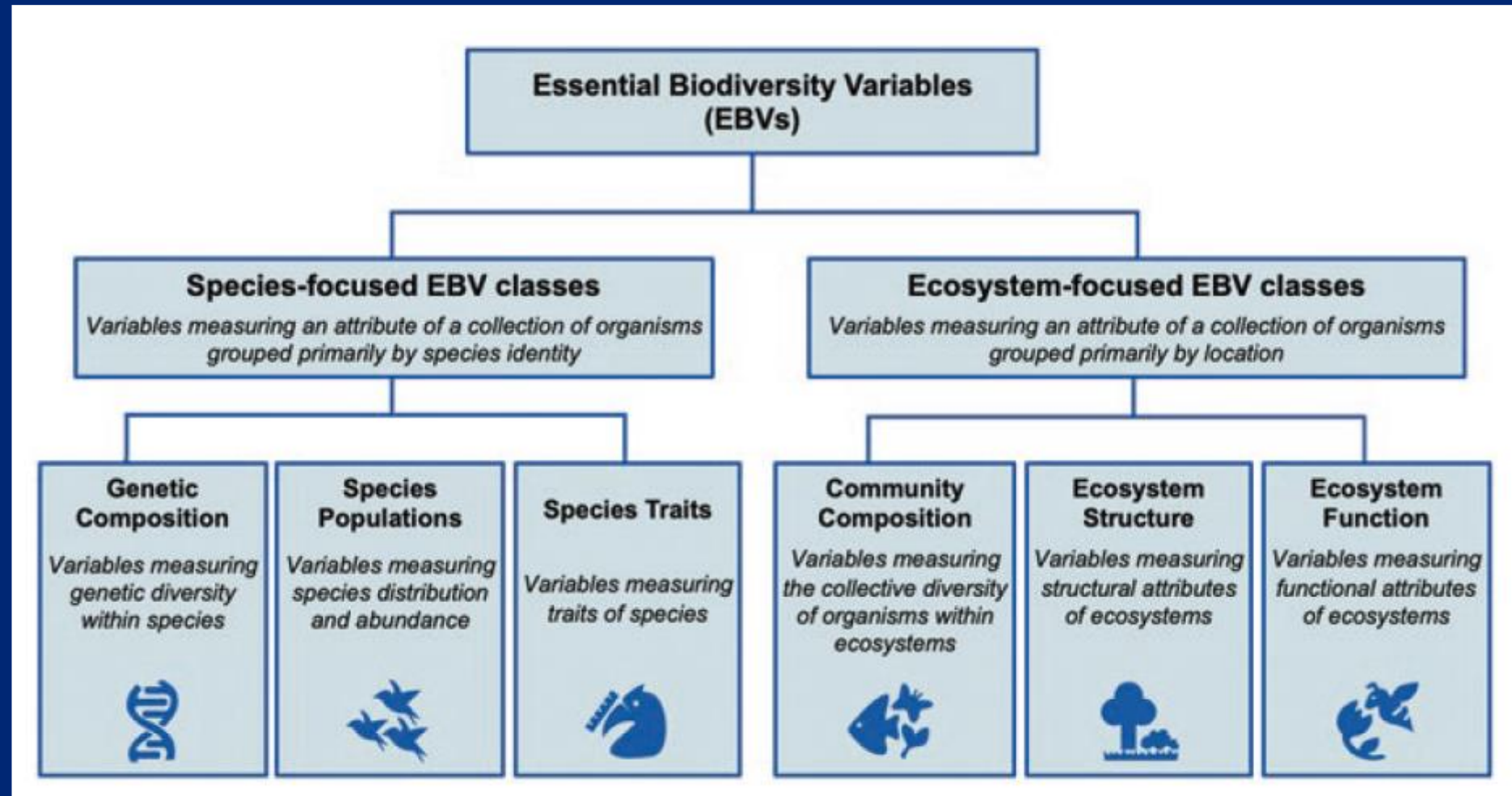
- I. Introduction
- II. Rappels de base en génétique des populations
- III. Estimation de la démographie historique : principes fondamentaux, méthodes courantes, étude de cas
- IV. Estimation de l'histoire sélective et de l'adaptation génétique: principes fondamentaux, méthodes courantes, étude de cas
- V. Vers l'inférence conjoint de processus démographiques et sélectifs

Le déclin de la biodiversité

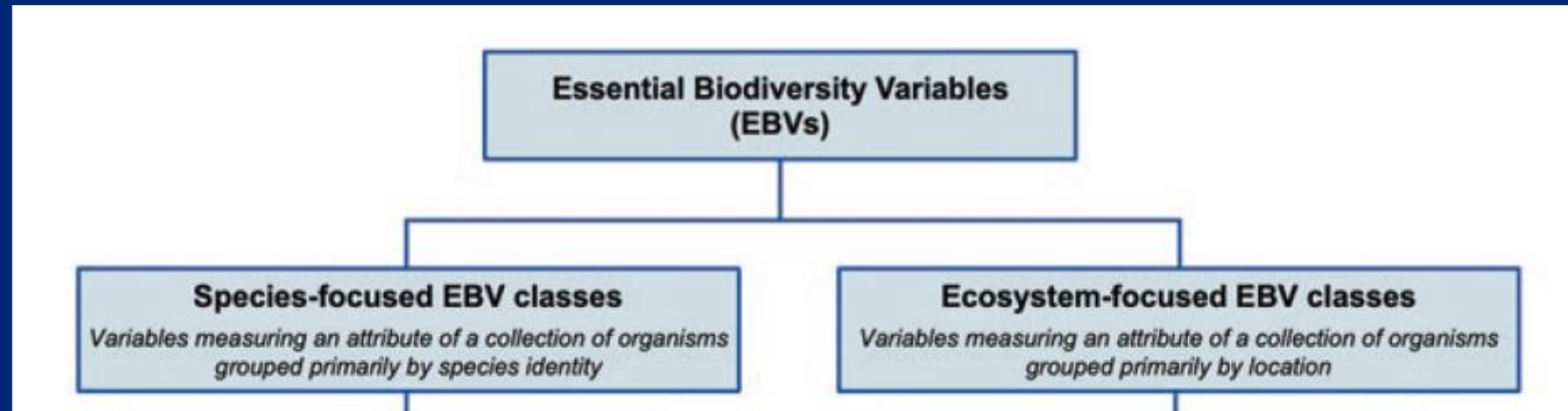


IPBES 2019
Global assessment
report of the
intergovernmental
science-policy
platform on
biodiversity
and ecosystem
services

Les variables essentielles de biodiversité (GEO BON)



Les variables essentielles de biodiversité (GEO BON)



Variation génétique

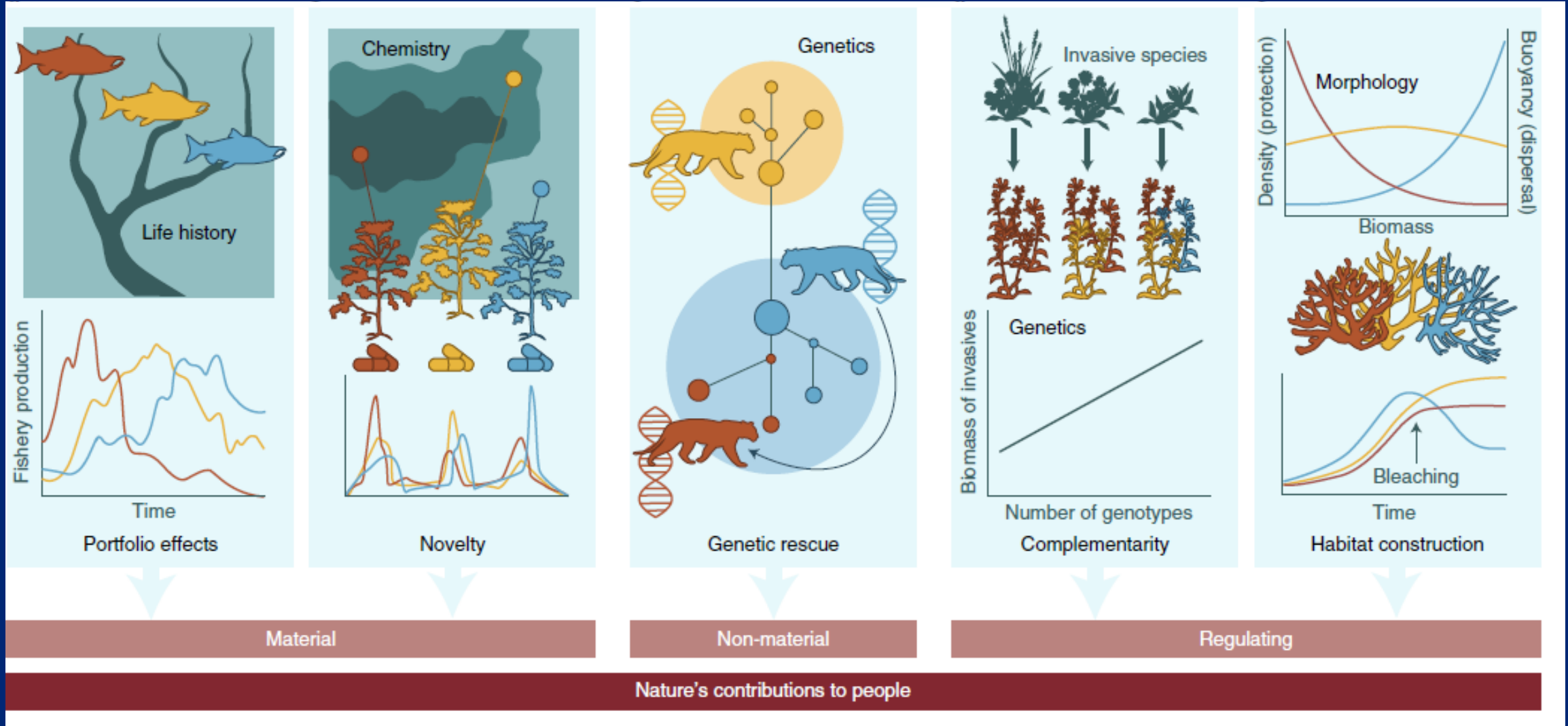
- Diversité génétique
 - Différenciation génétique
- Taille efficace de population
- Consanguinité

Conséquences d'une baisse de diversité génétique

- À court terme : la dépression de consanguinité
- À moyen terme : diminution de l'efficacité de la sélection
- À long terme : une baisse du potentiel adaptatif des populations

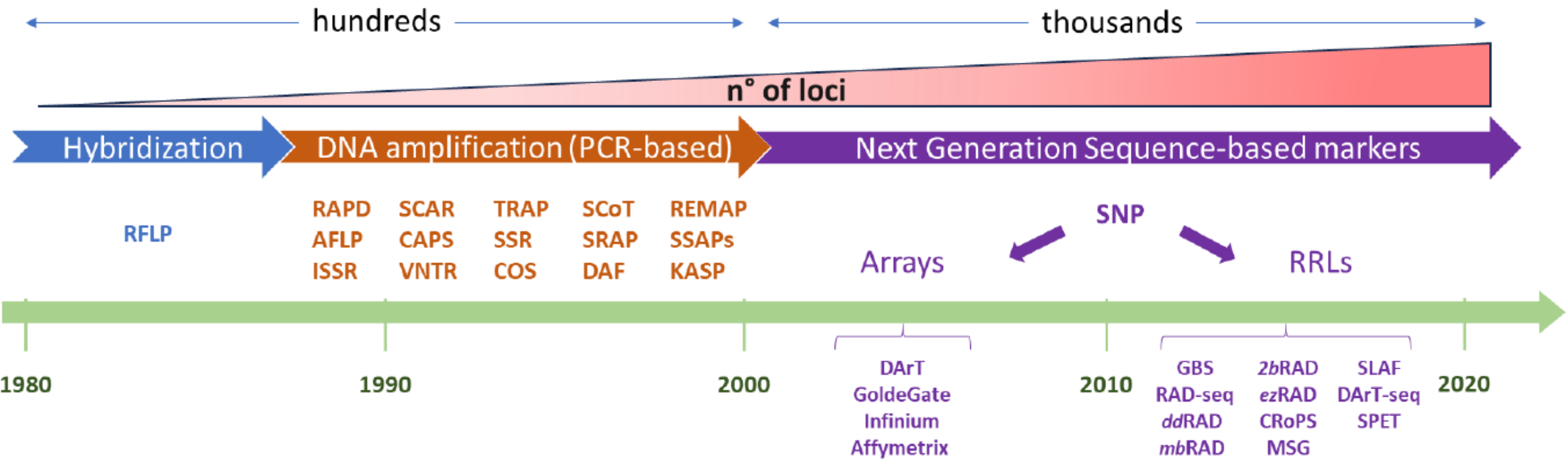
Importance de la diversité intraspécifique

Carburant de l'adaptation via la réponse à la sélection naturelle

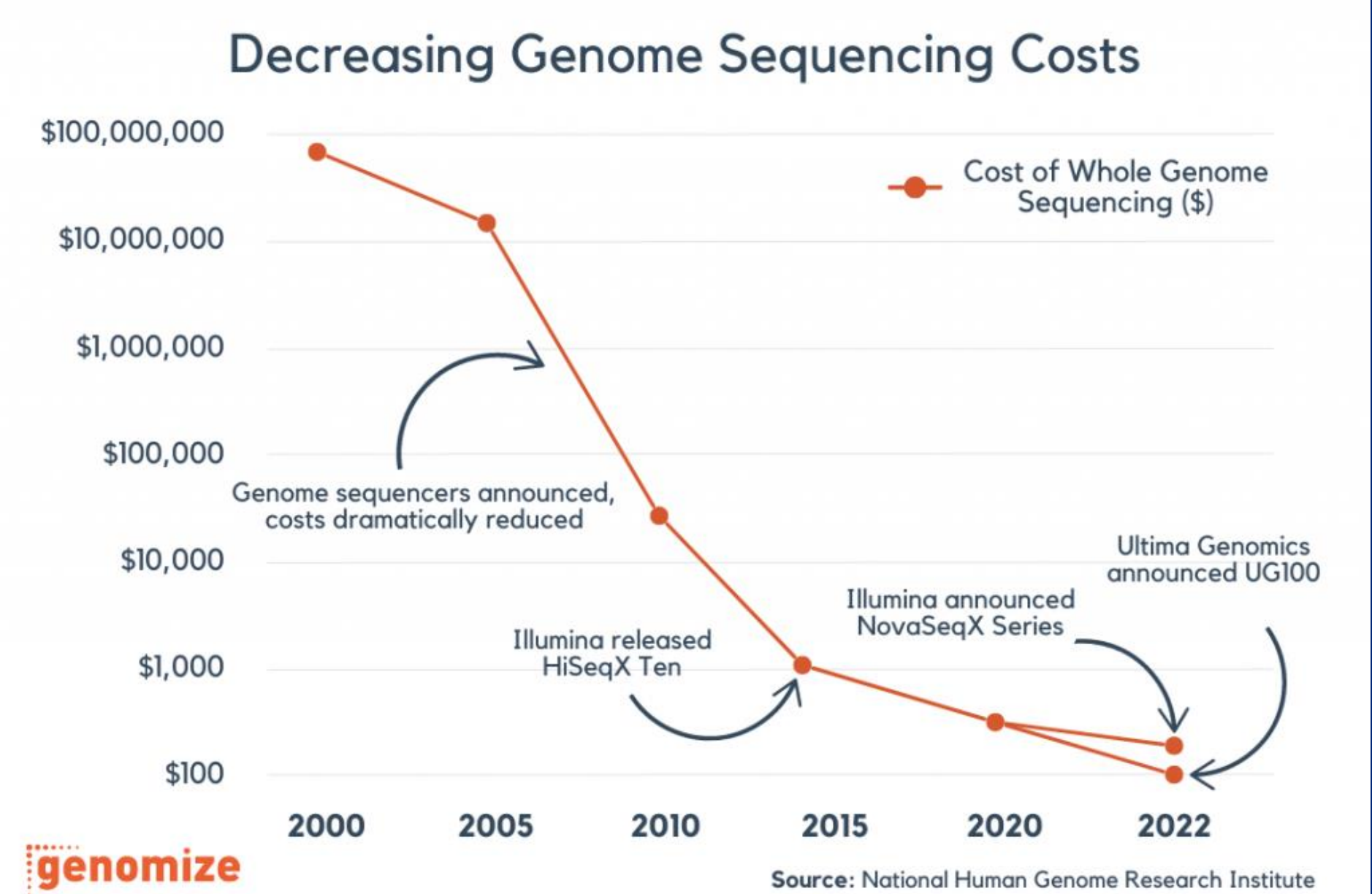


Comment caractériser la diversité génétique ?

Evolution des marqueurs ADN disponibles



Comment caractériser la diversité génétique ?



Comment les données génomiques peuvent-elles contribuer à la conservation de la biodiversité ?

Genomics and the future of conservation genetics

2010

Fred W. Allendorf*, Paul A. Hohenlohe^{§||} and Gordon Luikart*[#]

Abstract | We will soon have complete genome sequences from thousands of species, as well as from many individuals within species. This coming explosion of information will transform our understanding of the amount, distribution and functional significance of genetic variation in natural populations. Now is a crucial time to explore the potential implications of this information revolution for conservation genetics and to recognize limitations in applying genomic tools to conservation issues. We identify and discuss those problems for which genomics will be most valuable for curbing the accelerating worldwide loss of biodiversity. We also provide guidance on which genomics tools and approaches will be most appropriate to use for different aspects of conservation.

nature reviews genetics

2023

Perspective

Translating genomic advances into biodiversity conservation

Review

2010

Conservation genetics in transition to conservation genomics

N. Joop Ouborg¹, Cino Pertoldi², Volker Loeschcke², R. (Kuke) Bijlsma^{3,4} & Phil W. Hedrick⁵

Opinion

INTRODUCTION

Genotyping-by-sequencing in ecological and conservation genomics

2013

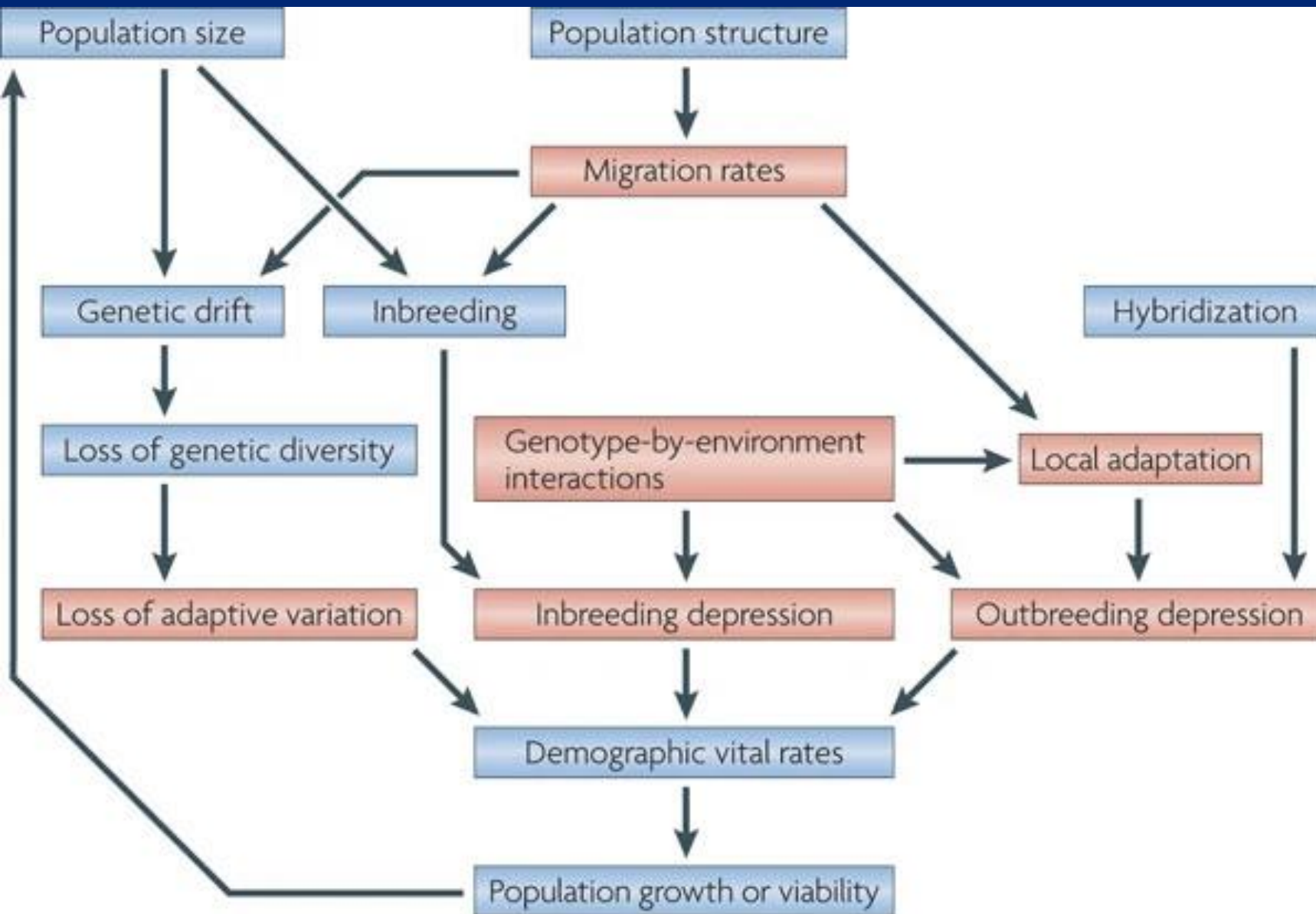
SHAWN R. NARUM,*† C. ALEX BUERKLE,‡ JOHN W. DAVEY,§ MICHAEL R. MILLER¶ and PAUL A. HOHENLOHE**

CellPress

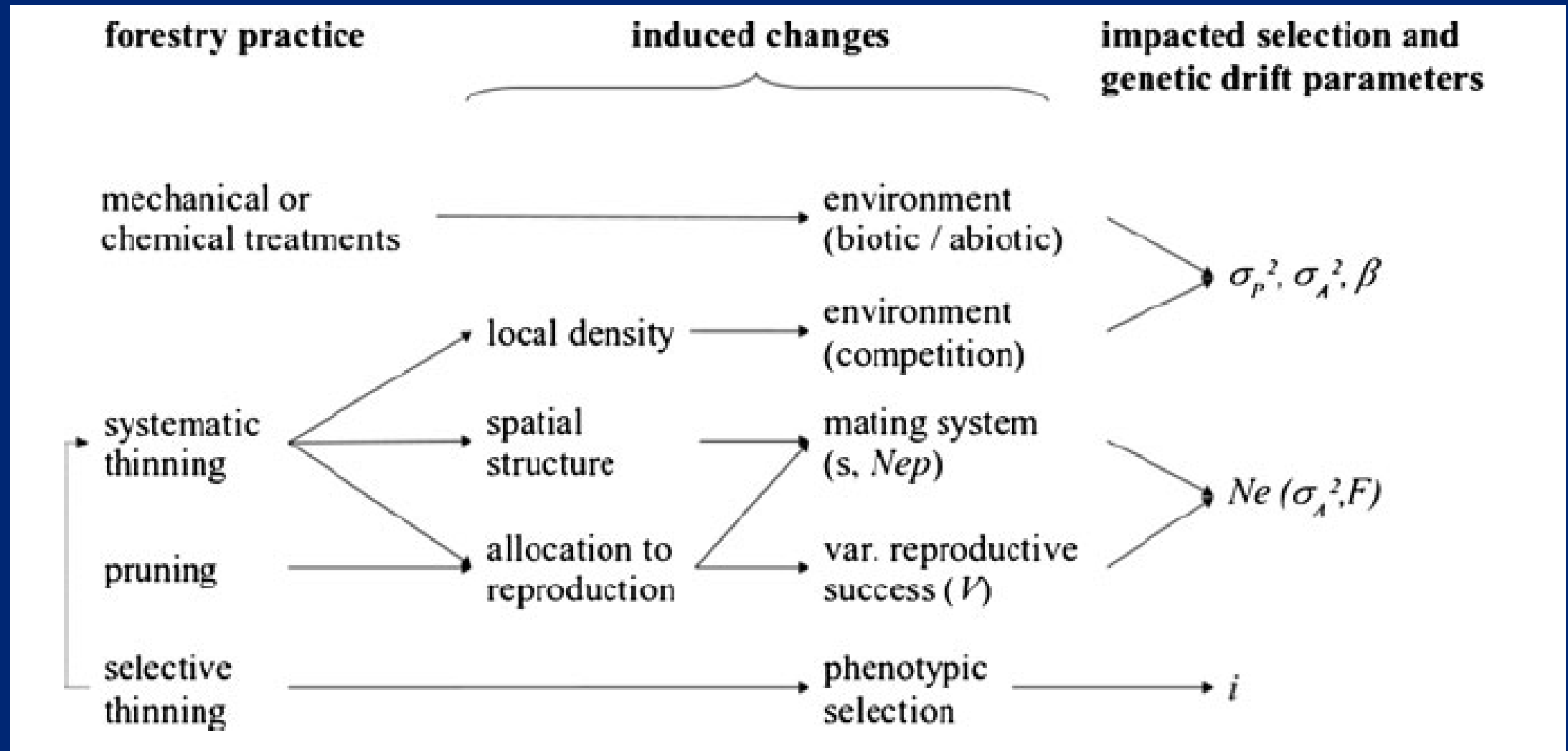
Genomics and the challenging translation into conservation practice

Aaron B.A. Shafer¹, Jochen B.W. Wolf¹, Paulo C. Alves², Linnea Bergström¹, Michael W. Bruford³, Ioana Brännström¹, Guy Colling⁴, Love Dalén⁵, Luc De Meester⁶,

Comment les données génomiques peuvent-elles contribuer à la conservation de la biodiversité ?

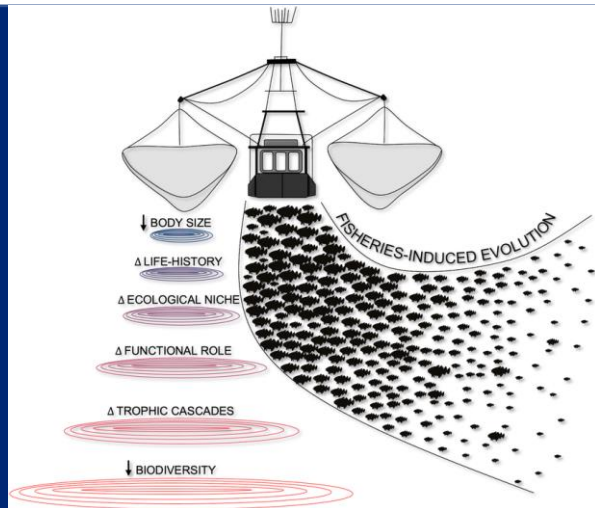
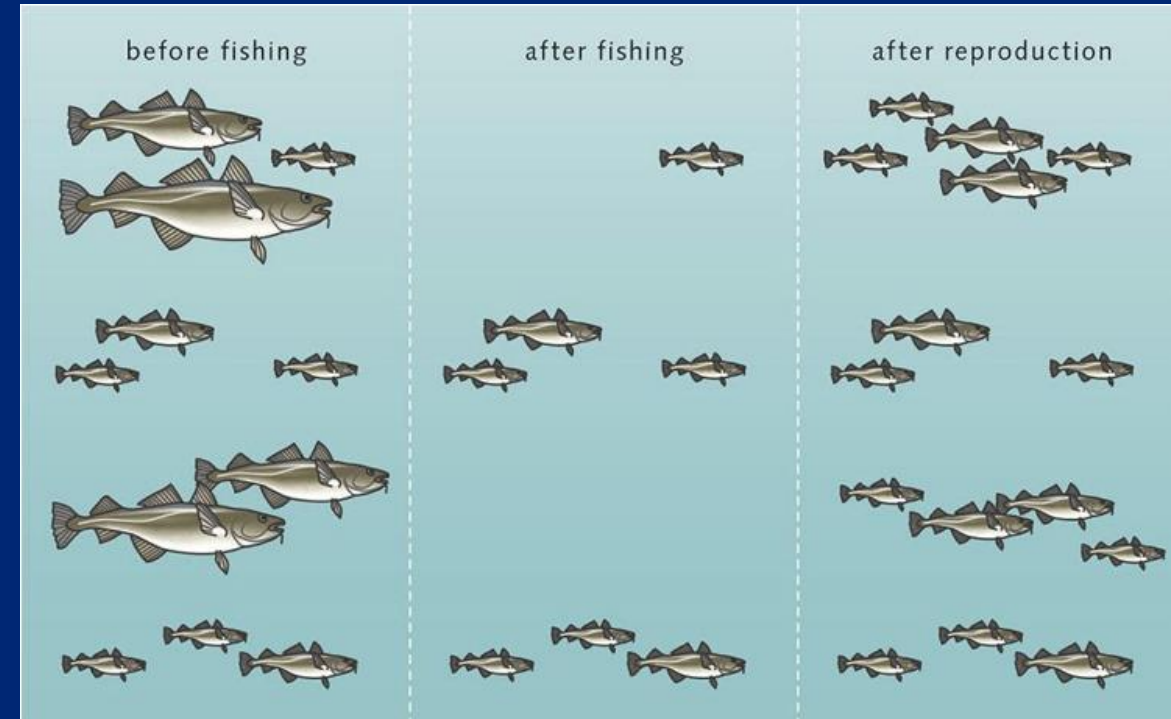
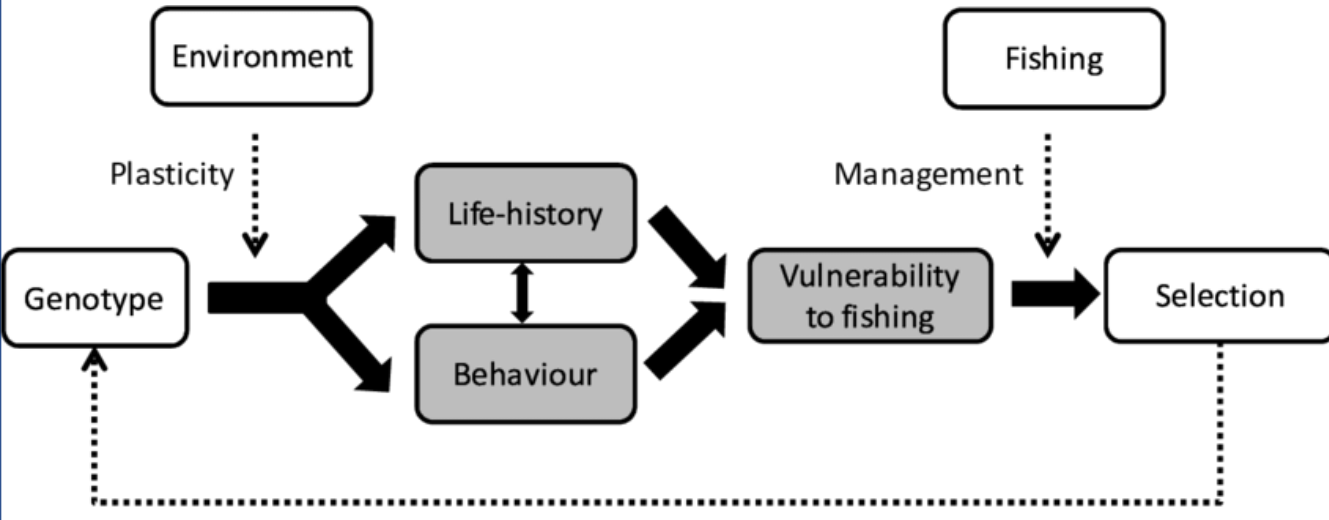


Comment les données génomiques peuvent-elles contribuer à la gestion adaptative des populations?



Comment les données génomiques peuvent-elles contribuer à la gestion adaptative des populations?

Fisheries induced evolution



Kuparinen & Merila TREE 2007

Heino et al. Ann Review Ecol Syst Evol 2015

L'essentiel à retenir

- La diversité génétique est cruciale pour la survie et l'adaptation des populations d'organismes vivants dans un environnement en constante évolution. Elle joue un rôle fondamental dans la résilience des écosystèmes et dans la capacité des espèces à s'adapter aux défis changeants de leur environnement
- Avec les nouvelles technologies de séquençage, caractériser la diversité génétique à un grand nombre de marqueurs (voire sur le génome entier) est devenu accessible pour un grand nombre d'espèces (même non modèles)
- Ces données génomiques permettent d'inférer de nouveaux paramètres importants pour la conservation des populations, et aussi d'estimer plus précisément des paramètres classiques. Pas encore de recul sur comment elles peuvent contribuer à guider la gestion adaptative des populations.

Plan du cours

- I. Introduction
- II. Rappels de base en génétique des populations
- III. Estimation de la démographie historique : principes fondamentaux, méthodes courantes, étude de cas
- IV. Estimation de l'histoire sélective et de l'adaptation génétique: principes fondamentaux, méthodes courantes, étude de cas
- V. Vers l'inférence conjoint de processus démographiques et sélectifs

Qu'est-ce que la génétique des populations ?

Historiquement une synthèse entre

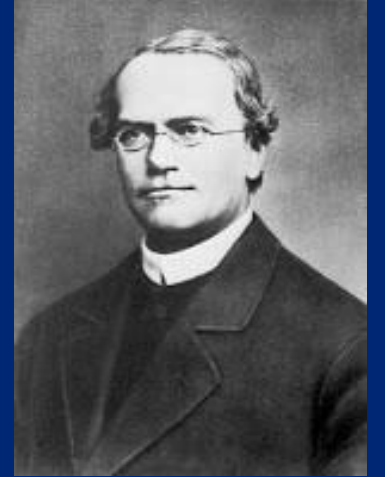
1. Application des lois de la génétique de la transmission

- Caractères à déterminisme simple

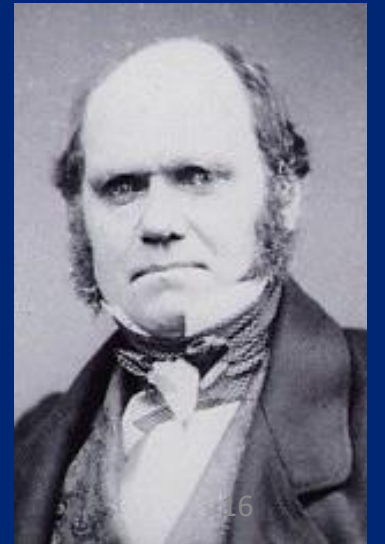
2. Application de la théorie de l'évolution et de la sélection naturelle

- Evolution biologique
- Evolution par sélection des individus les plus adaptés

G. Mendel



C. Darwin



Quelques définitions

- Information génétique → **locus** → plusieurs états alléliques distincts
- Variation génétique s'exprime par les **fréquences** (proportions) relatives des différents allèles ; VARIATION= **POLYMORPHISME**
- **génotype** = chez un organisme diploïde, composition allélique d'un individu à un locus donné (ex : AA, Aa, aa); même état allélique (**homozygote**) ou non (**hétérozygote**).
- **marqueur moléculaire** : locus à déterminisme mendélien, polymorphe, codominant (ex : microsatellites, RFLP, AFLP, SNPs)
- L'évolution agit sur des **populations** d'individus (groupes d'individus susceptibles de se reproduire entre eux à court terme)
- L'évolution se traduit par une variation des fréquences alléliques dans les populations au cours du temps
- **Forces évolutives** = les processus qui agissent sur les fréquences alléliques

Facteurs influençant la diversité génétique

- Le modèle 'nul' : population de Hardy-Weinberg: croisements aléatoires (panmixie), => pas d'évolution génotypique, allélique.
- Impact du système de reproduction : ouvert/fermé (hétéro/homogamie, consanguinité, autofécondation) => modifie la distribution des différents génotypes
- Forces évolutives : => modifient la fréquence des allèles.

Citez les forces évolutives

Facteurs influençant la diversité génétique

- Le modèle 'nul' : population de Hardy-Weinberg: croisements aléatoires (panmixie), => pas d'évolution génotypique, allélique.
- Impact du système de reproduction : ouvert/fermé (hétéro/homogamie, consanguinité, autofécondation) => modifie la distribution des différents génotypes
- Forces évolutives : => modifient la fréquence des allèles.
 - Mutation
 - Dérive génétique
 - Sélection
 - Migration

Les forces évolutives

A- Mutation

B- Dérive génétique

C- Sélection

D- Migration

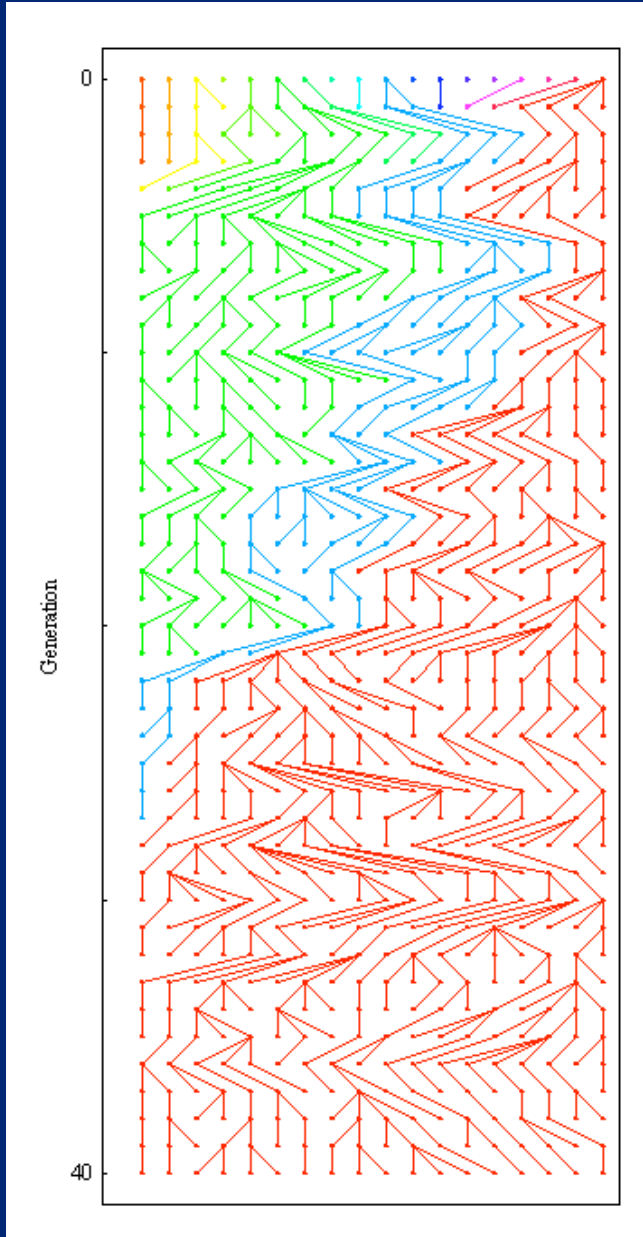
- fluctuations aléatoires des fréquences alléliques résultant d'un échantillonnage aléatoire parmi les gamètes, pour générer la génération suivante
- introduction d'allèles (nouveaux) dans une population via le mouvement des individus ou des gamètes
- source fondamentale de variation génétique
- liée à l'existence d'un avantage ou désavantage adaptatif d'un allèle, favorisant sa survie ou sa reproduction (fitness); peut être naturelle ou artificielle

Mutation

Définition: Changement héréditaire dans le matériel génétique, source fondamentale de variation génétique

- Événement rare... : substitutions : 10^{-8} à 10^{-9} nouvelles mutations par nucléotide (base) par génération
 - si $2Nu$ mutations apparaissent à un nucléotide donné par génération (dans une population diploïde de taille N , pour un taux de mutation génique u) ...
 - alors si $N = 10000$ et $u = 10^{-9}$, il y a 0.00002 mutation par nucléotide dans la population..
 - et donc si l'on considère un génome de taille 100 Mpb chez *Arabidopsis thaliana* (ou 3400 Mpb chez *Homo sapiens*).on obtient 2000 (68000) nouvelles substitutions dans la population en une génération !
 - soit 0.2 (6.8) substitutions dans un zygote formé
- Devenir d'une mutation EN PRATIQUE
 - mutation **neutre** : le plus souvent éliminée de la population, mais peut aussi se substituer à l'allèle sauvage, à cause des effets aléatoires de la dérive génétique dans les petites populations.
 - mutation **défavorable** : diminue en fréquence (sélection négative).
 - mutation **favorable** : augmente en fréquence (sélection positive).

Dérive génétique

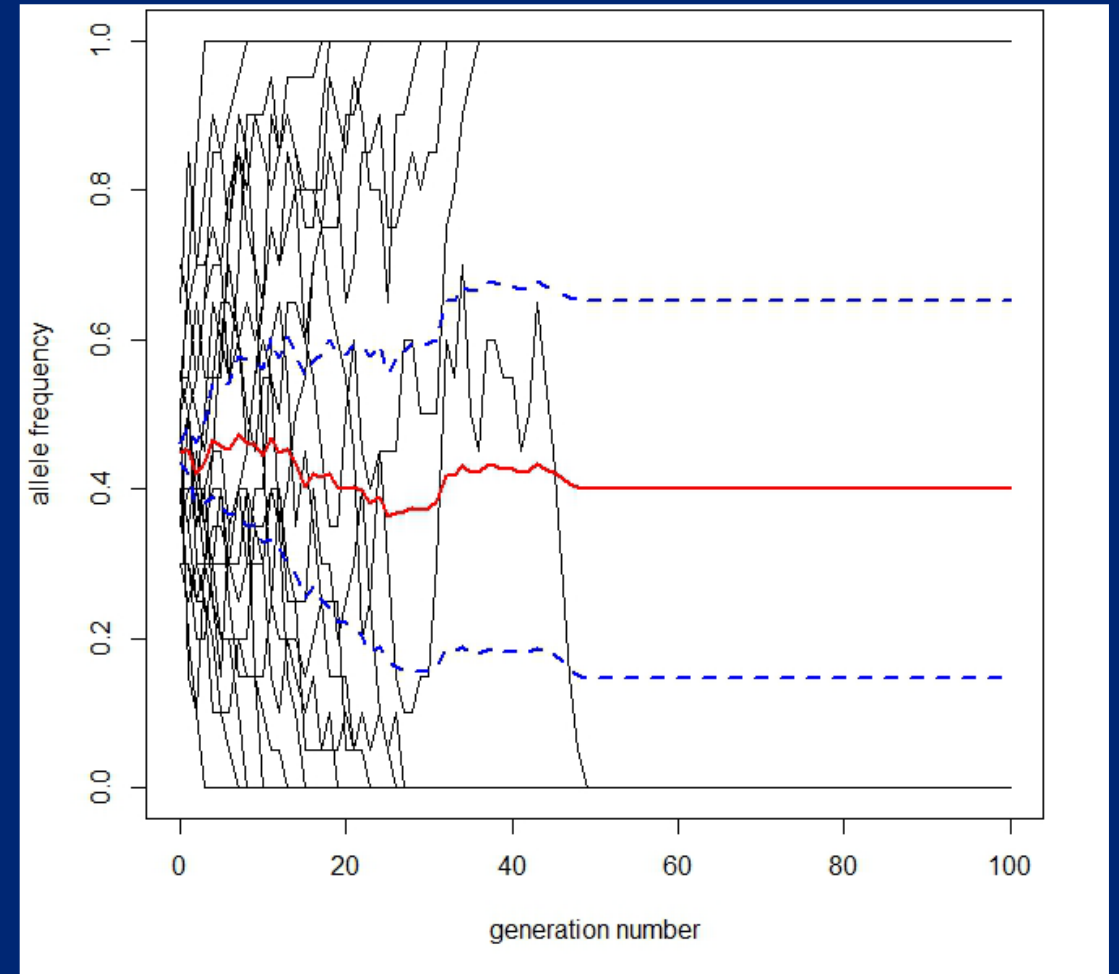
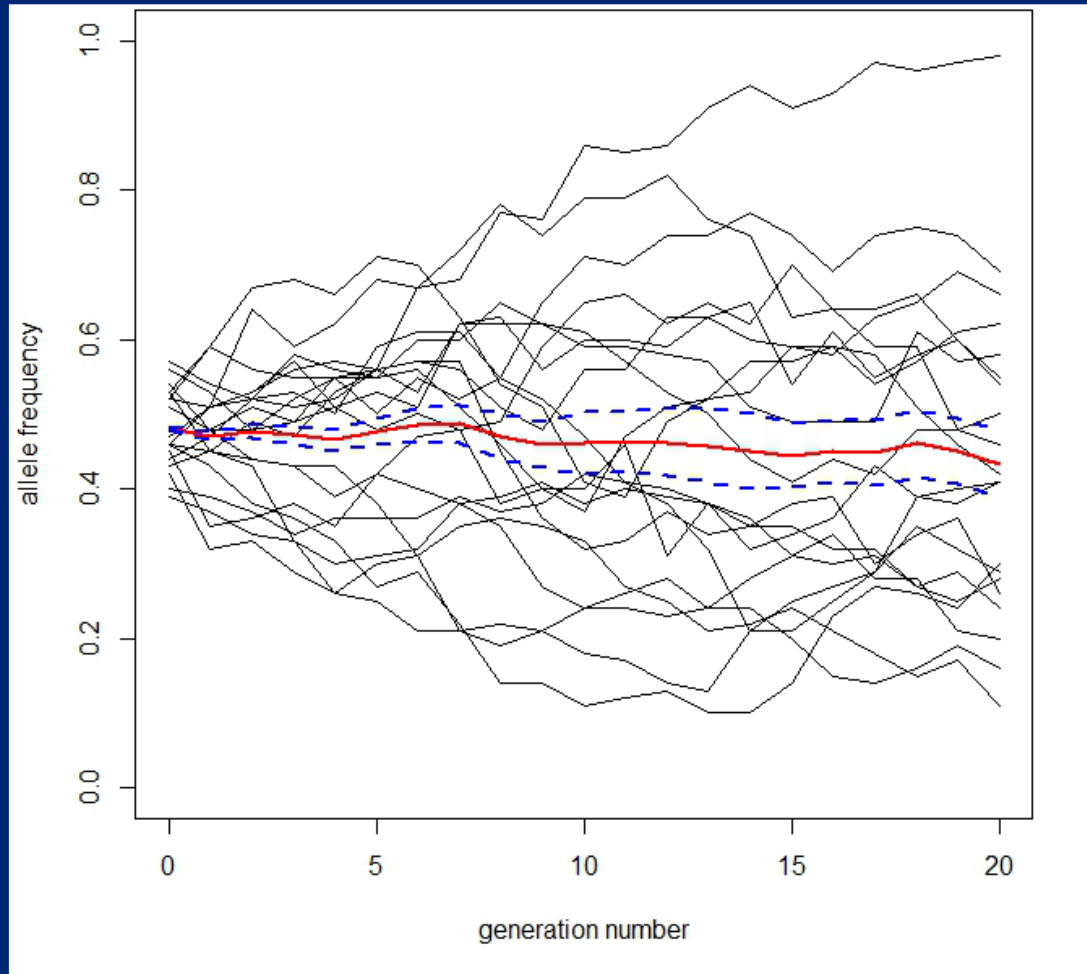


Définition: fluctuations aléatoires des fréquences alléliques < échantillonnage aléatoire parmi les gamètes

Modèle de Wright-Fisher

- Population de taille finie et constante, où les individus se reproduisent une fois et au même moment (générations non chevauchantes)
- Chaque gène à une génération est la copie d'un gène d'une génération précédente
- En l'absence de mutation et de sélection, les fréquences alléliques dérivent (augmentent ou diminuent aléatoirement) inévitablement jusqu'à la fixation d'un allèle
- La dérive conduit donc à une perte de diversité à l'intérieur des populations

Dérive génétique

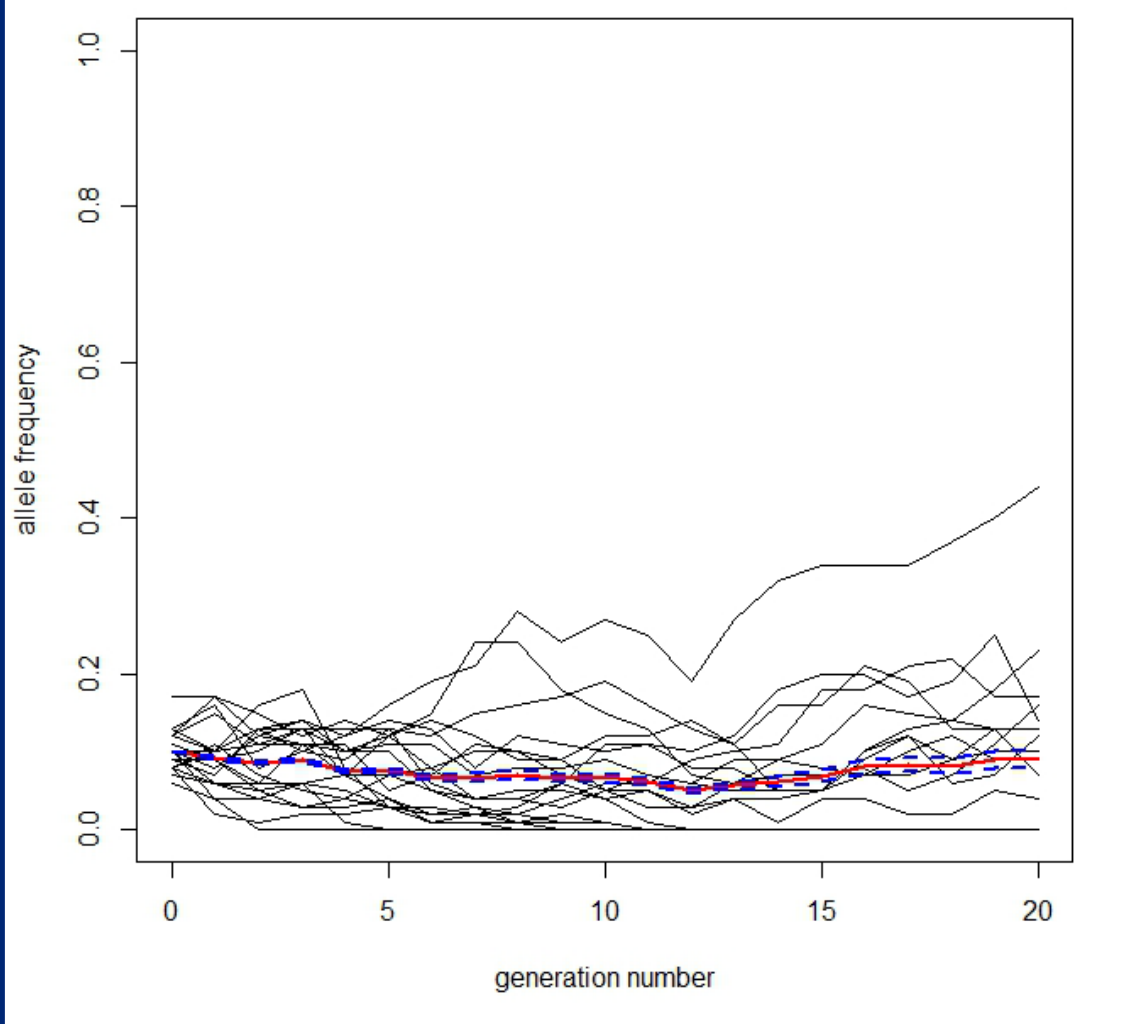


Taille de population = 100

Fréquence initiale = 0.5, 20 simulations (rouge = moyenne, bleu = variance).

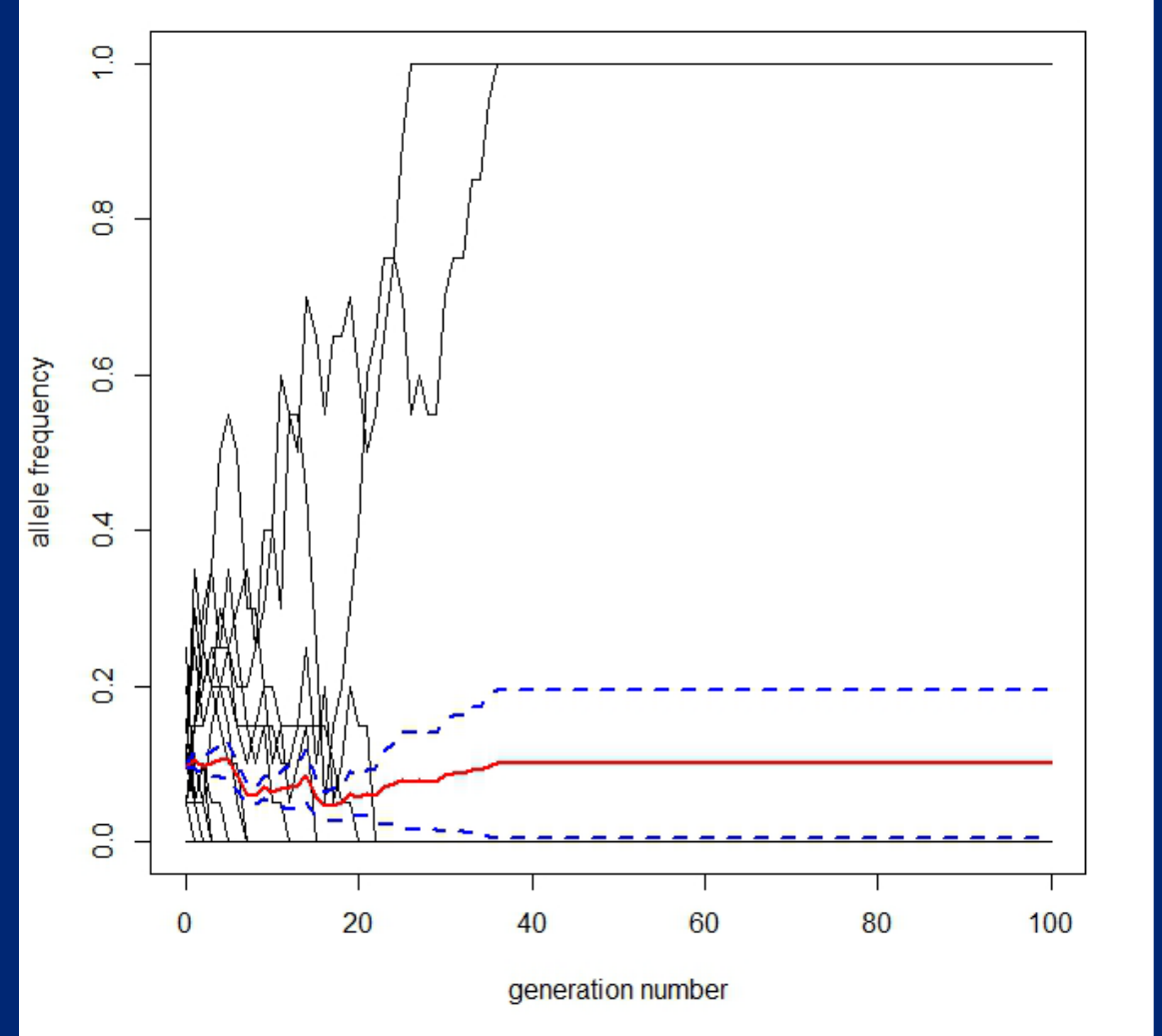
Taille de population = 20

Dérive génétique



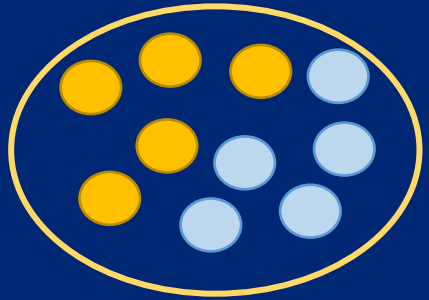
Taille de population = 100

Fréquence initiale = 0.1, 20 simulations (rouge = moyenne, bleu = variance).

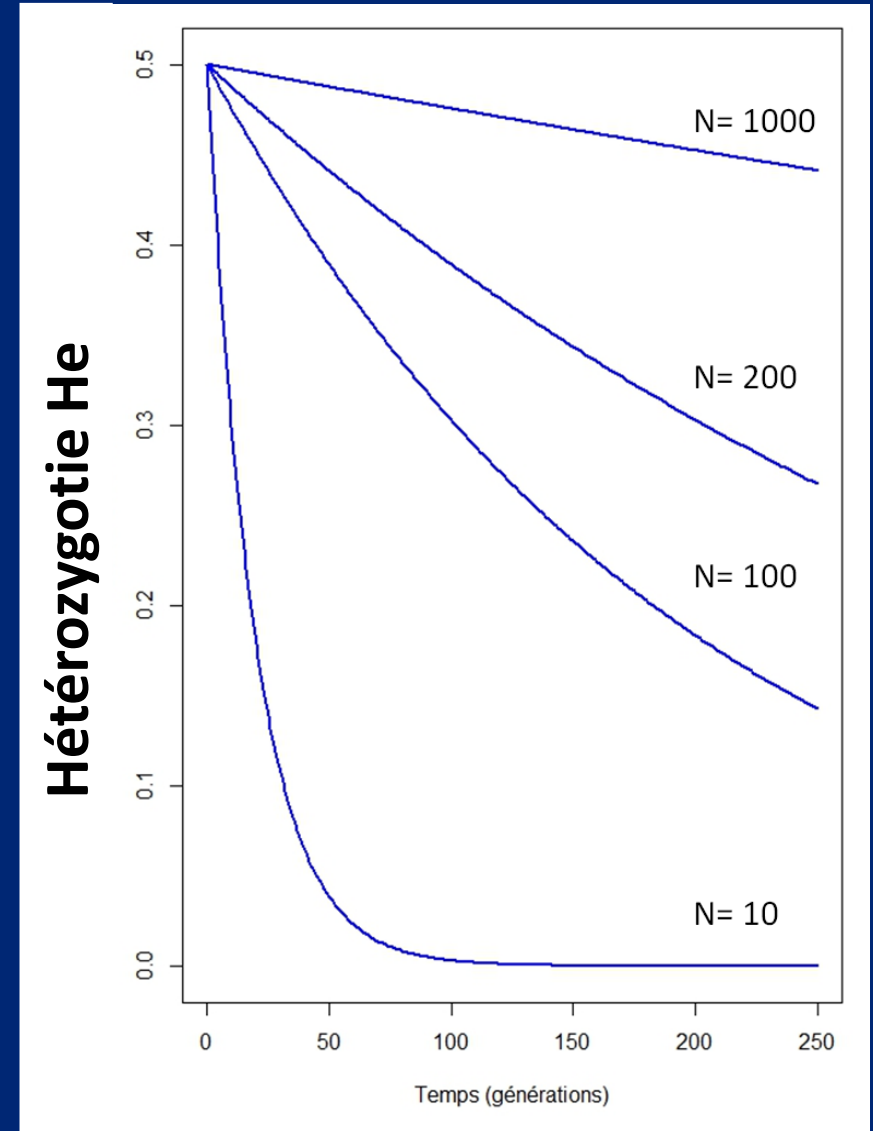
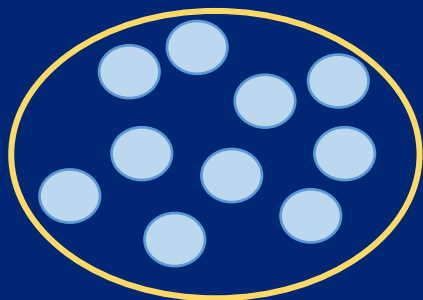
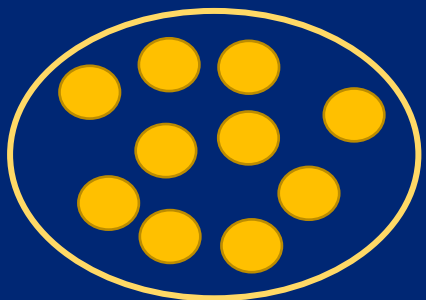


Taille de population = 20

Conséquences de la dérive sur la diversité intrapopulation

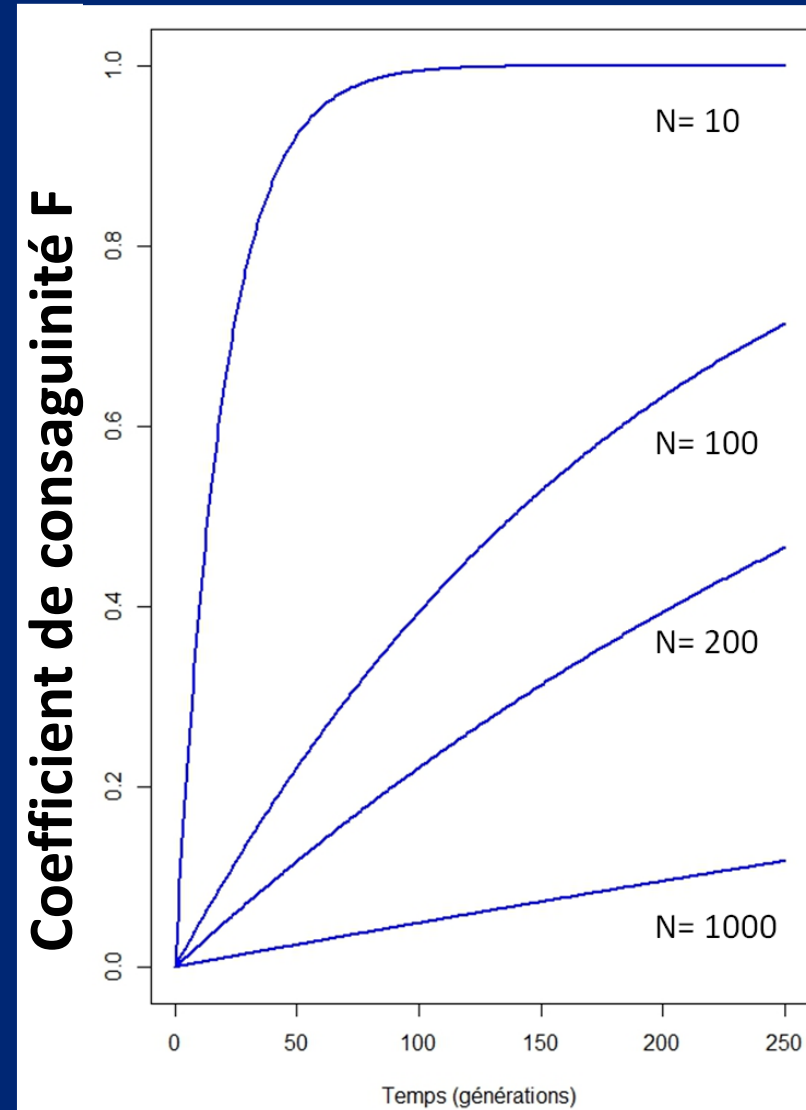
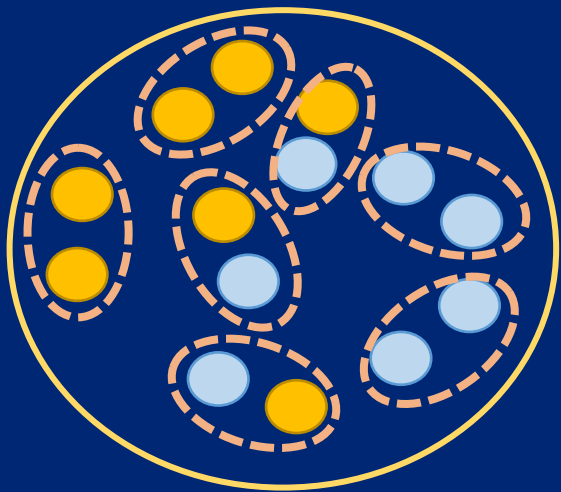
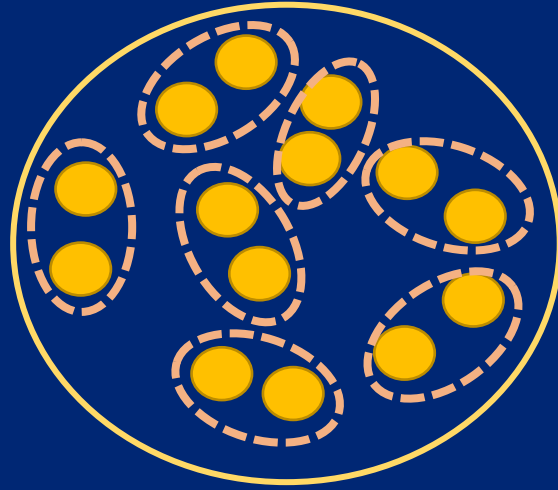


$$H_e = 1 - \sum_{i=1}^2 p_i^2 = 2p(1-p)$$

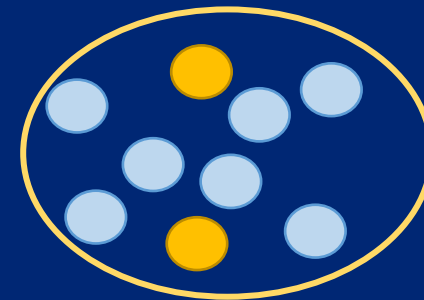
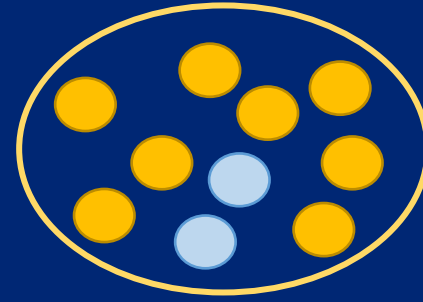
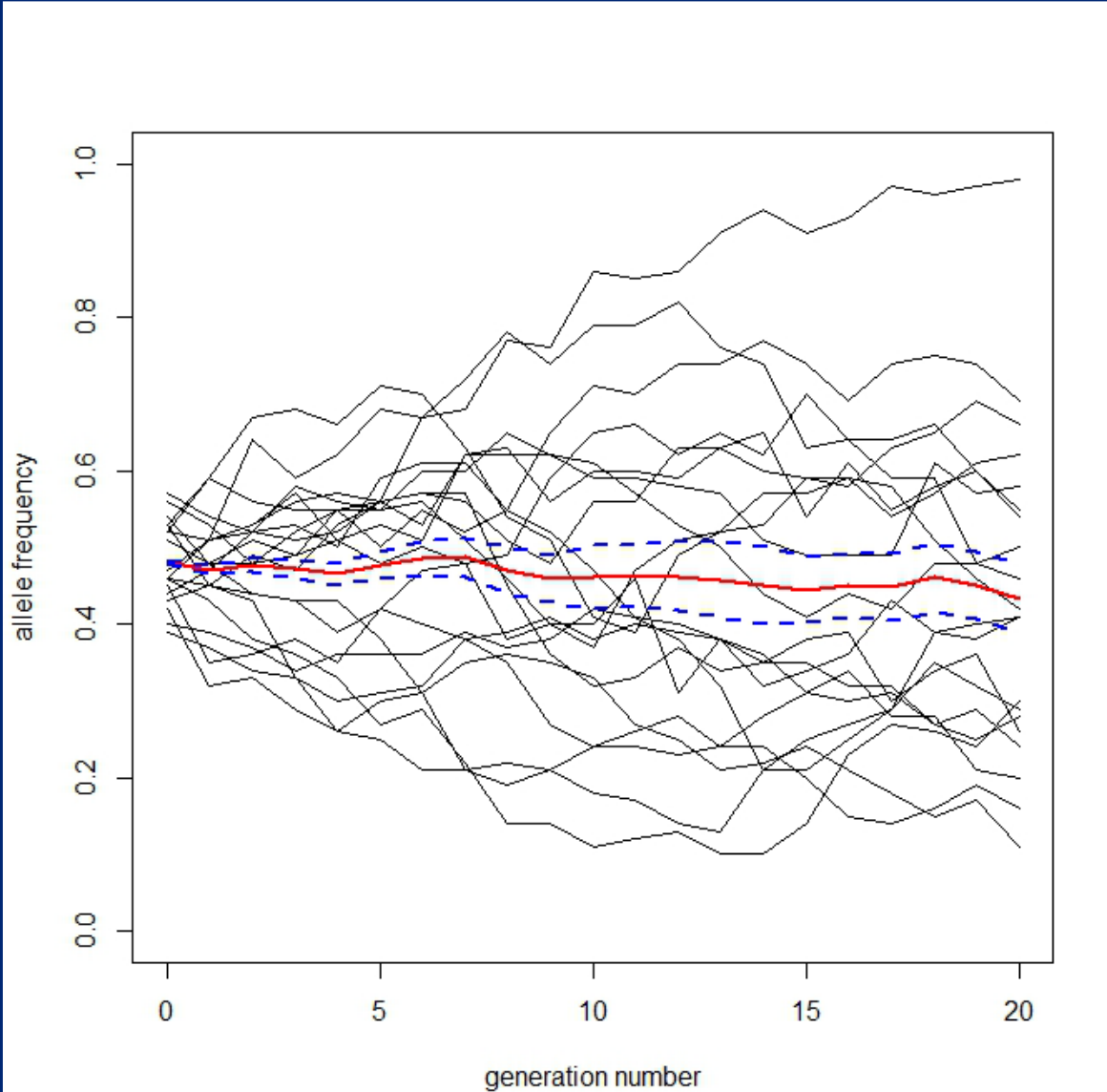


Conséquences de la dérive sur la diversité intrapopulation

$$F = 1 - \frac{H_{obs}}{2p(1-p)}$$



Conséquences de la dérive sur la différenciation interpopulation



Variance des fréquences alléliques entre populations

$$F_{ST} = \frac{\sigma_S^2}{\sigma_T^2} = \frac{\sigma_S^2}{\bar{p}(1 - \bar{p})}$$

Variance des fréquences alléliques dans la pop totale

Fréquence moyenne dans la pop totale

Dérive génétique: conclusion et notion de taille efficace

- par dérive génétique on perd la diversité allélique au locus considéré...
- ... mais en plus tous les gènes de la population au locus considéré finissent par être des copies d'un seul des gènes de la population de départ !
- si l'on remonte dans le temps, on trouvera un seul gène ancêtre de tous les gènes actuellement présents au locus dans une population (notion de coalescence).
- il en est de même **pour tous les locus du génome !**

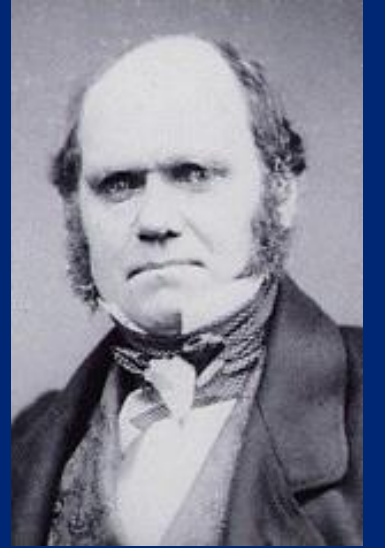
Dérive génétique: conclusion et notion de taille efficace

- par dérive génétique on perd la diversité allélique au locus considéré...
- ... mais en plus tous les gènes de la population au locus considéré finissent par être des copies d'un seul des gènes de la population de départ !
- si l'on remonte dans le temps, on trouvera un seul gène ancêtre de tous les gènes actuellement présents au locus dans une population (notion de coalescence).
- il en est de même **pour tous les locus du génome !**

- La taille (ou l'effectif) efficace N_e de la population est rarement égale à celle de la population réelle N que l'on pourrait recenser.
- On définit la taille efficace N_e d'une population comme le nombre d'individus d'une population "idéale" de type Wright-Fisher où la dérive génétique aurait la même intensité que dans la population qui nous intéresse.
- $N_e < N$

Sélection

C. Darwin

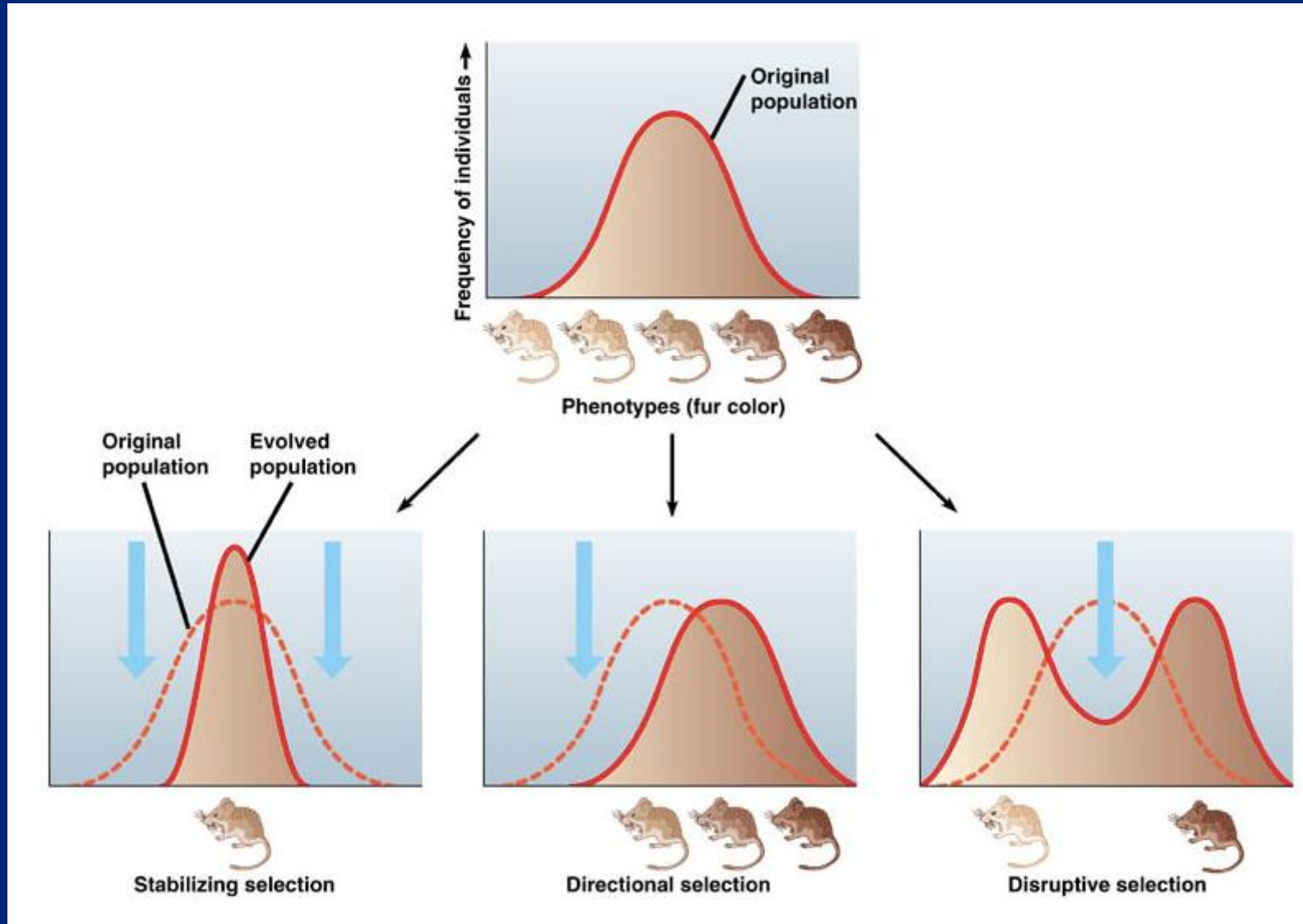


Définition: Processus par lequel les individus les mieux adaptés à leur environnement ont une meilleure chance de survie et de reproduction, transmettant ainsi leurs caractères héréditaires à la génération suivante.

- **Principe 1 = variabilité** : les individus d'une espèce diffèrent les uns des autres pour certains caractères (phénotypes et donc génotypes).
- **Principe 2 = adaptation / survie** : les individus les plus adaptés au milieu survivent et se reproduisent davantage.
- **Principe 3 = transmission des caractères avantageux** : les caractéristiques avantageuses doivent être héréditaires.

Sélection naturelle dans une population

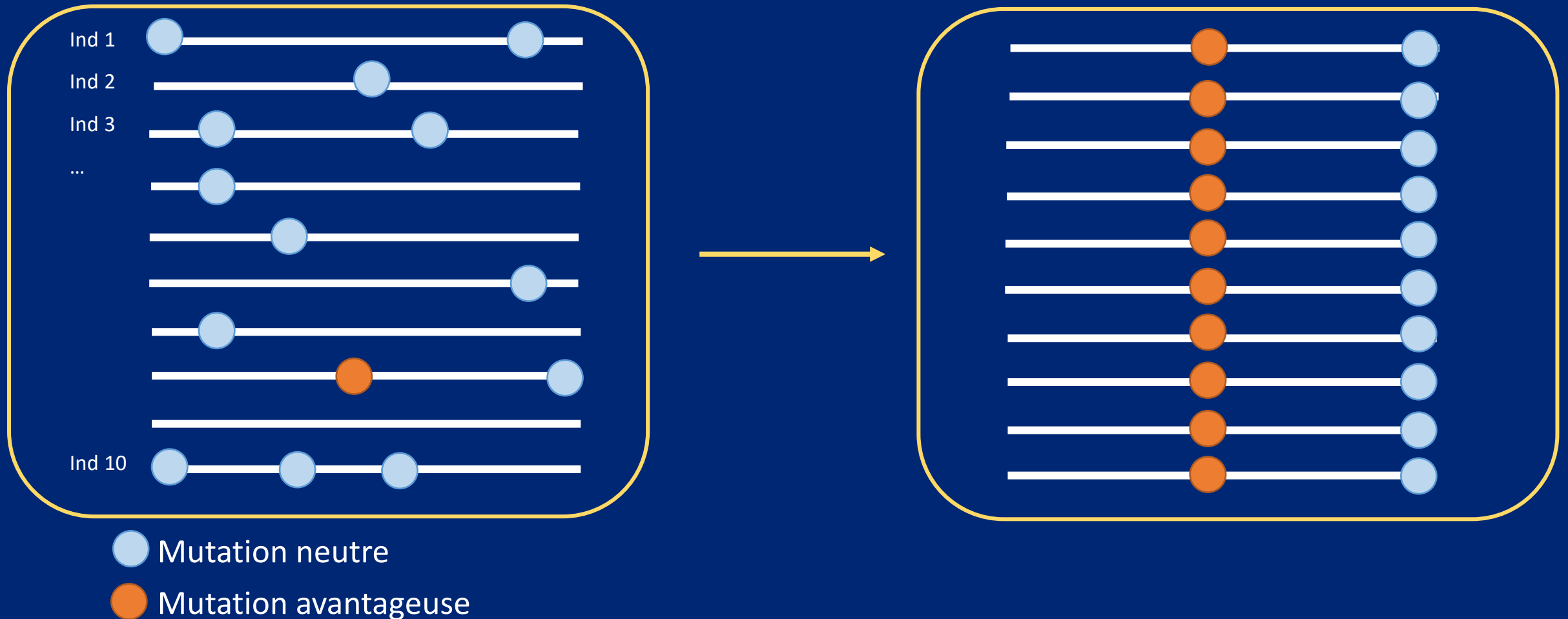
Ce qu'on observe au niveau phénotypique



Sélection naturelle dans une population

Ce qu'on observe au niveau moléculaire

Sélection positive (darwinienne, directionnelle)

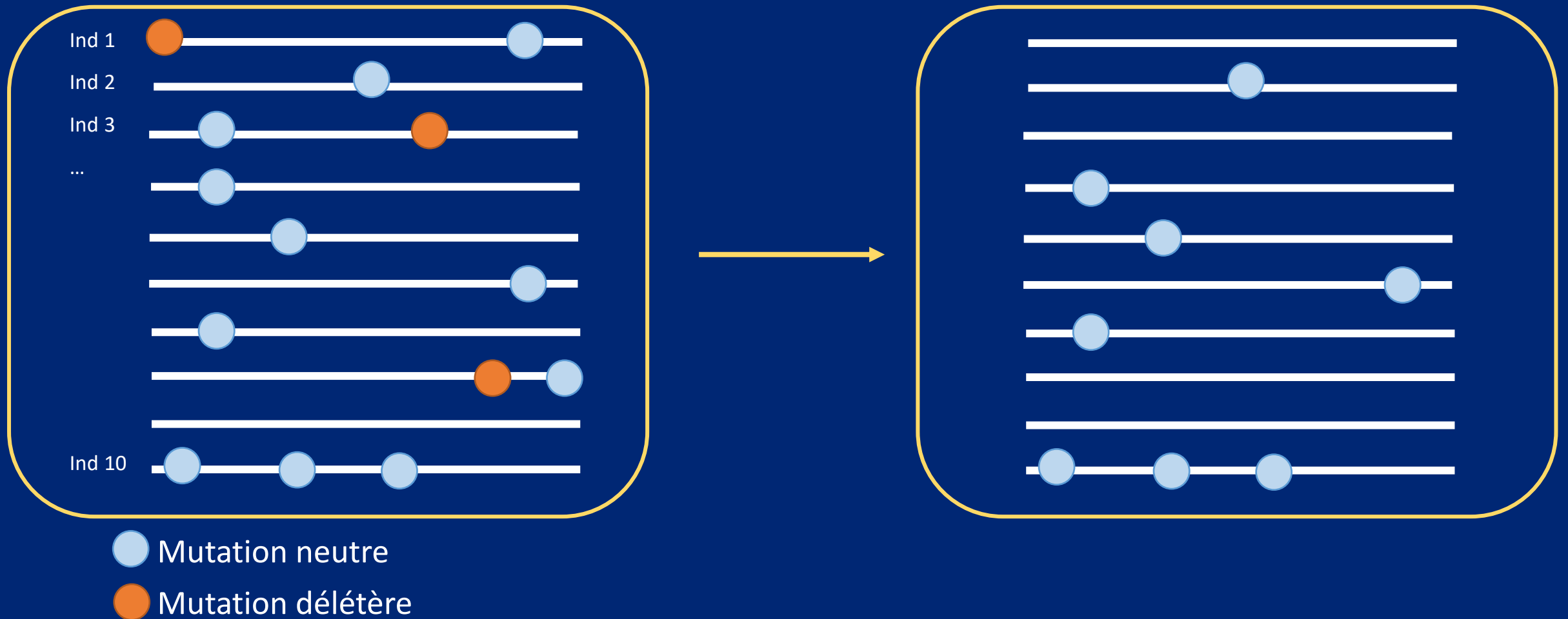


Gènes de l'adaptation (ex: résistance aux insecticides/pesticides, à la sécheresse)

Sélection naturelle dans une population

Ce qu'on observe au niveau moléculaire

Sélection purifiante (stabilisante, background selection)

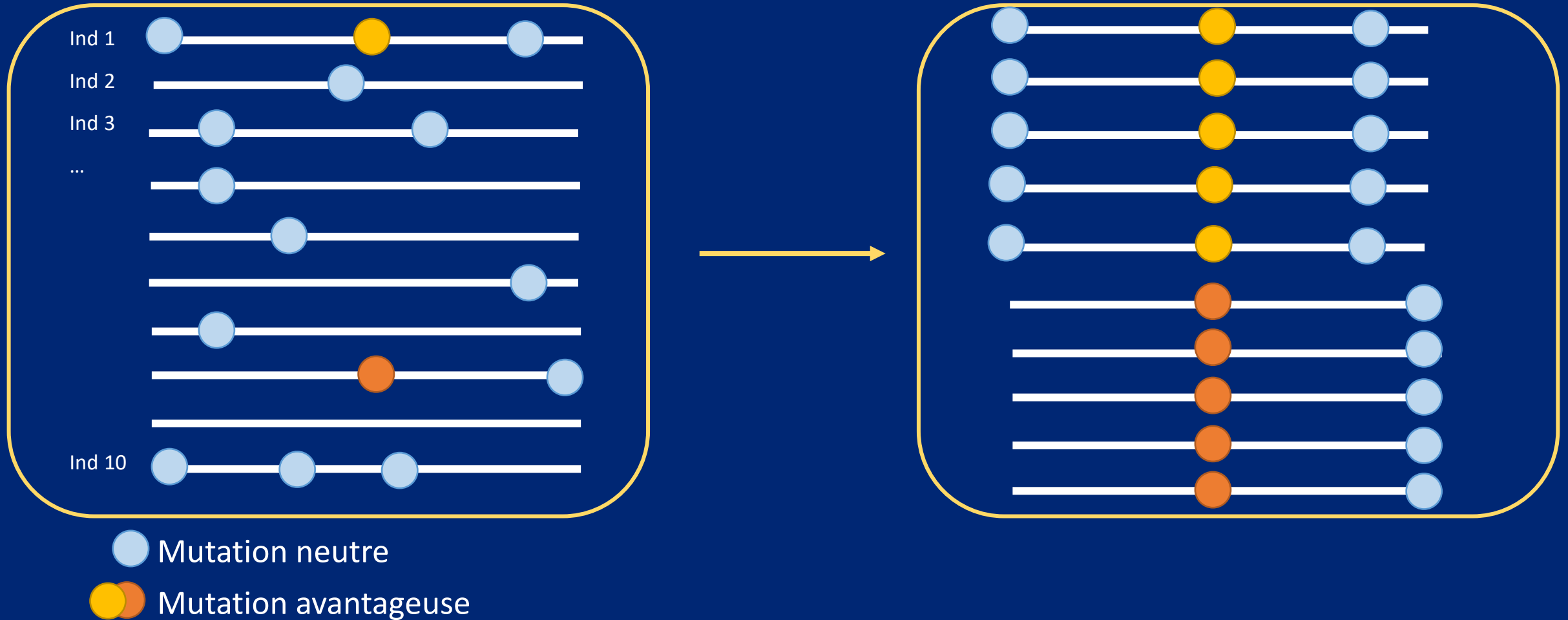


Gènes du développement (housekeeping genes), changements contre-sélectionnés

Sélection naturelle dans une population

Ce qu'on observe au niveau moléculaire

Sélection balancée (diversifiante)



Avantage à l'hétérozygote, sélection fréquence-dépendante (ex: anémie à hématie falciforme, gènes de l'immunité)

Sélection naturelle dans une population

Ce qu'on observe au niveau moléculaire

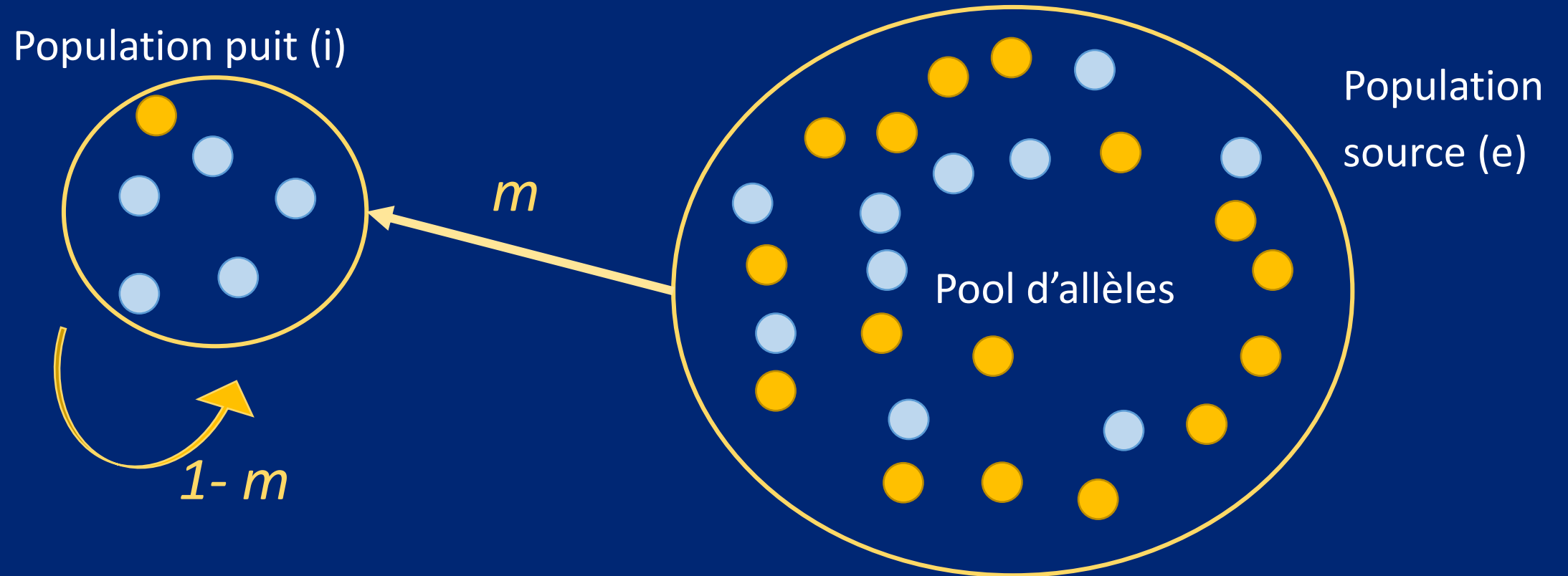
Sélection positive	Sélection purifiante	Sélection balancée
Réduction ou augmentation de la diversité globale	Réduction de la diversité globale	Augmentation de la diversité
Augmentation de la fréquence des variants rares	Augmentation de la fréquence des variants rares	Augmentation de la fréquence des variants communs
Augmentation de la différenciation entre populations		Réduction de la différenciation entre populations

Sélection

- la sélection naturelle conduit à une évolution déterministe des fréquences alléliques, à l'opposé de la dérive (aléatoire).
- la sélection naturelle agit sur certains gènes du génome, pas tout le génome à l'opposé de la dérive.
- la sélection naturelle peut réduire la diversité allélique d'un gène (sélection positive, négative) mais aussi maintenir activement le polymorphisme (sélection balancée, par exemple pour les gènes des récepteurs de l'immunité).
- la sélection artificielle sur des espèces domestiquées (cultures, bétail) a des effets similaires à la sélection positive (directionnelle) et/ou stabilisante.
- la sélection sexuelle peut avoir les mêmes signatures génomiques que la sélection naturelle

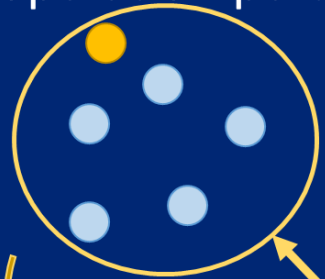
Migration

Définition: Flux de gènes entre populations résultant de l'échange d'individus (ou de gamètes)



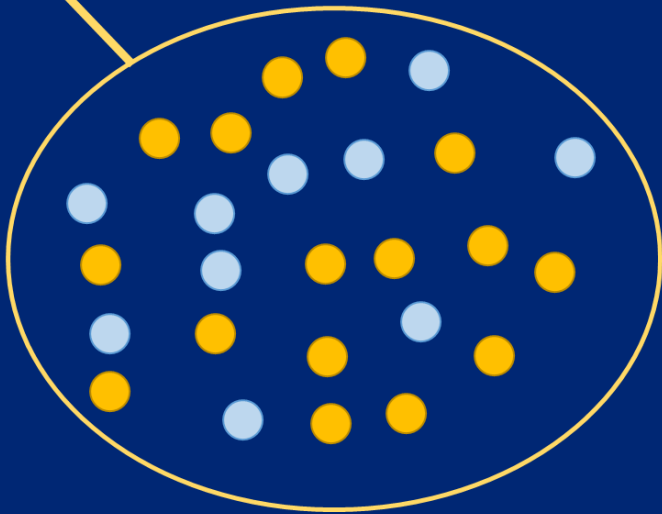
Migration

Population puit (i)

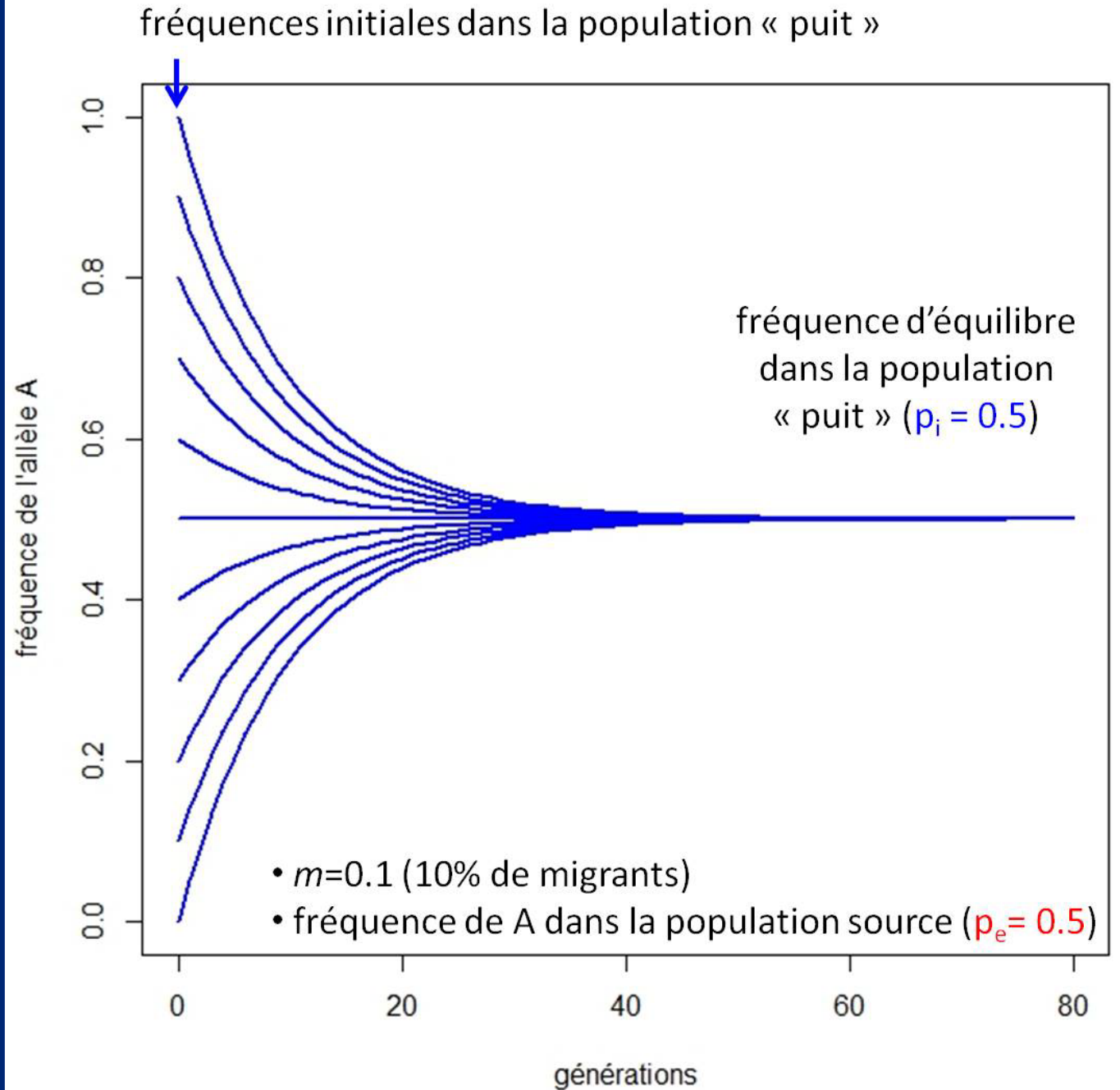


m

Population source (e)



$1-m$



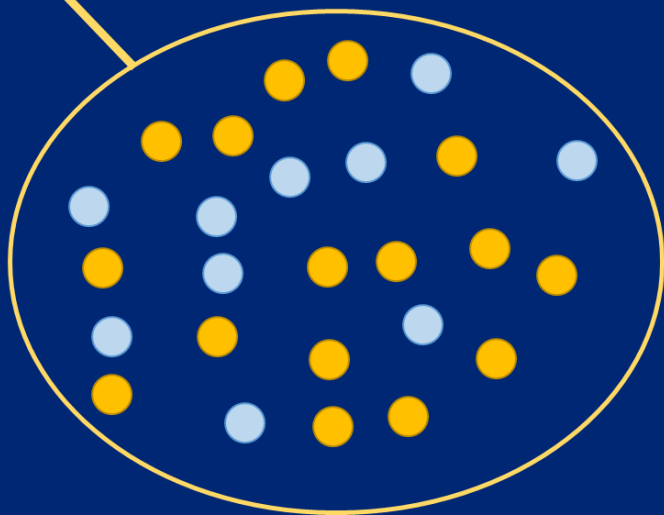
Migration

Population puit (i)

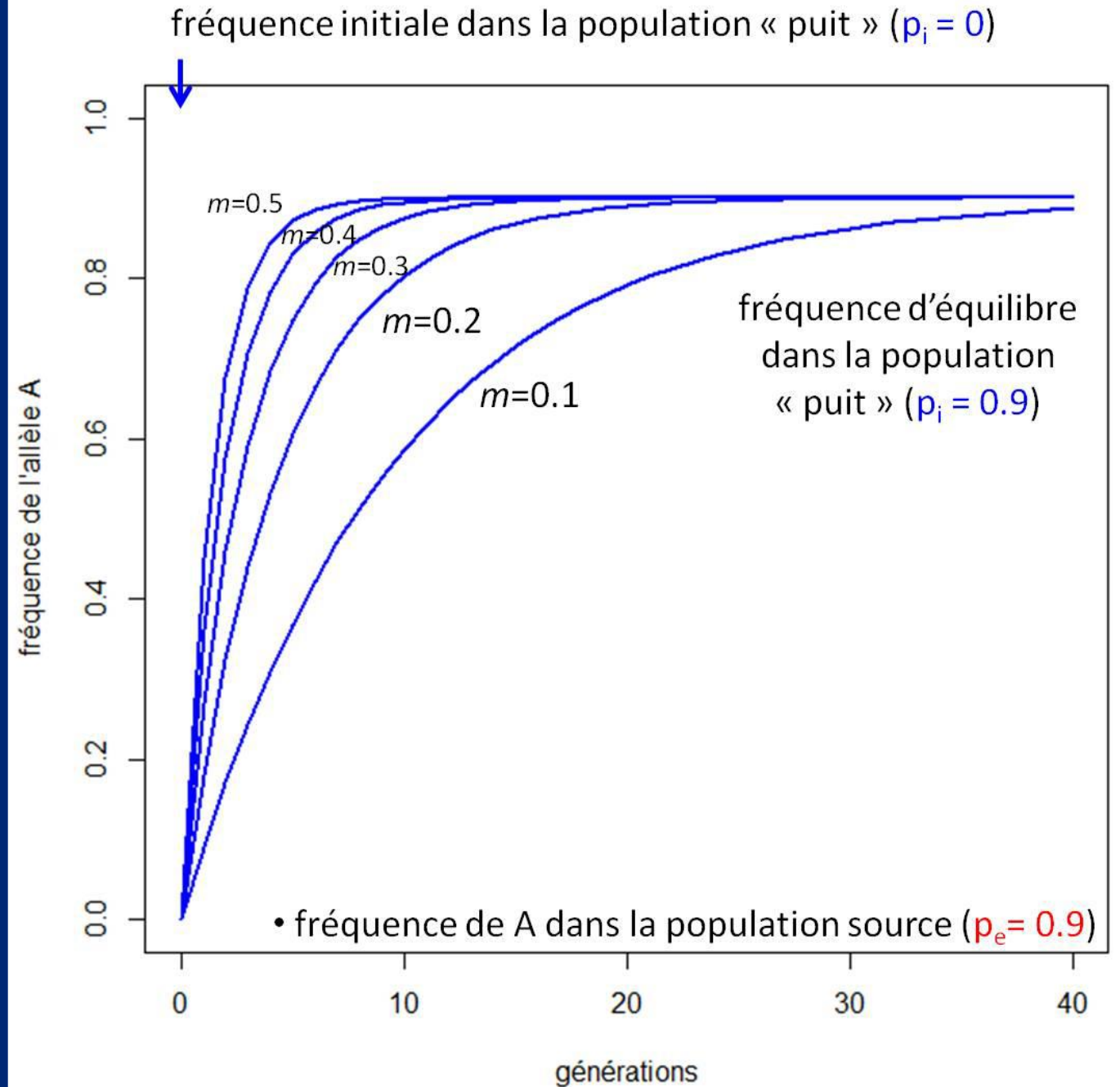


m

Population source (e)

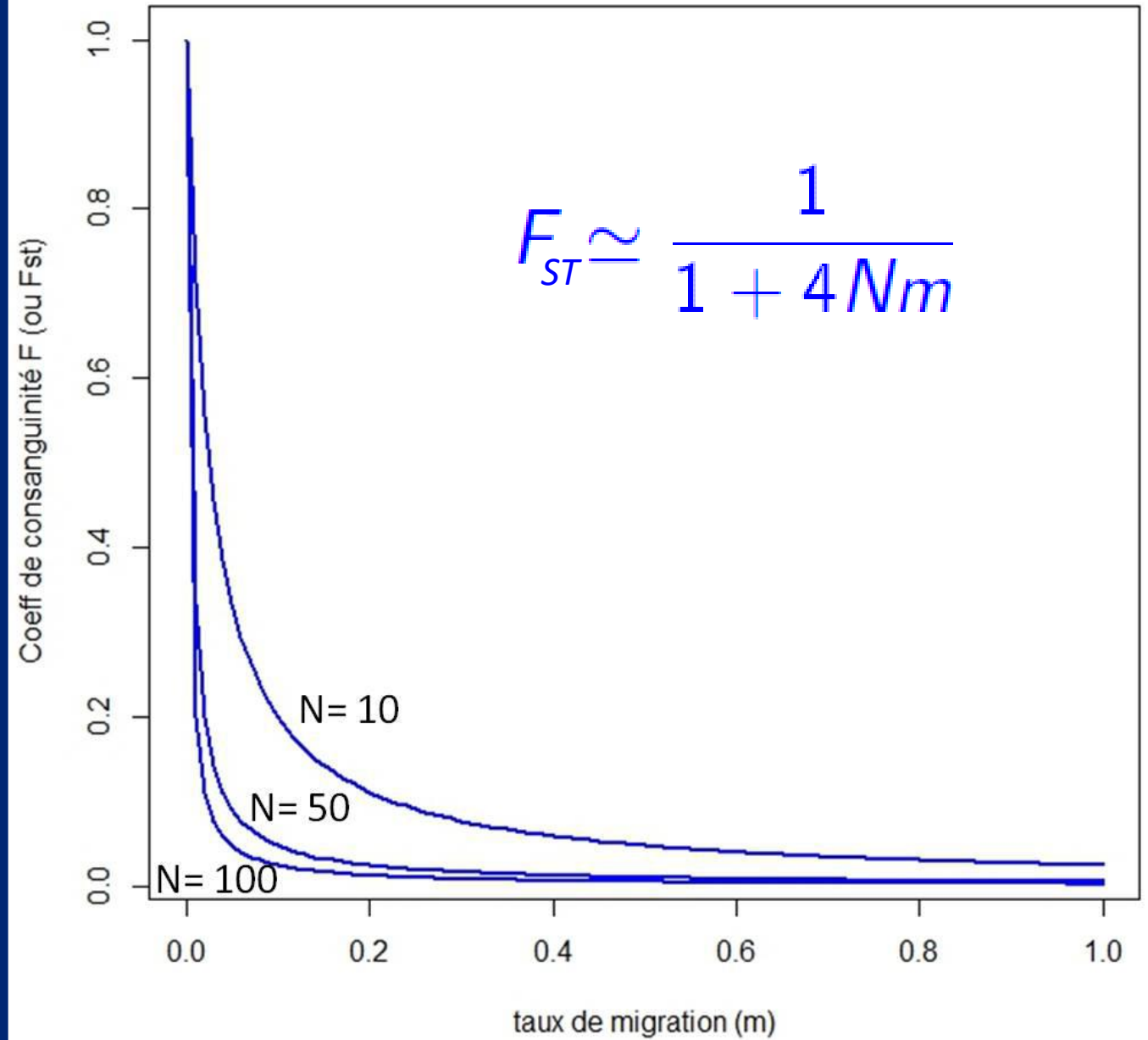
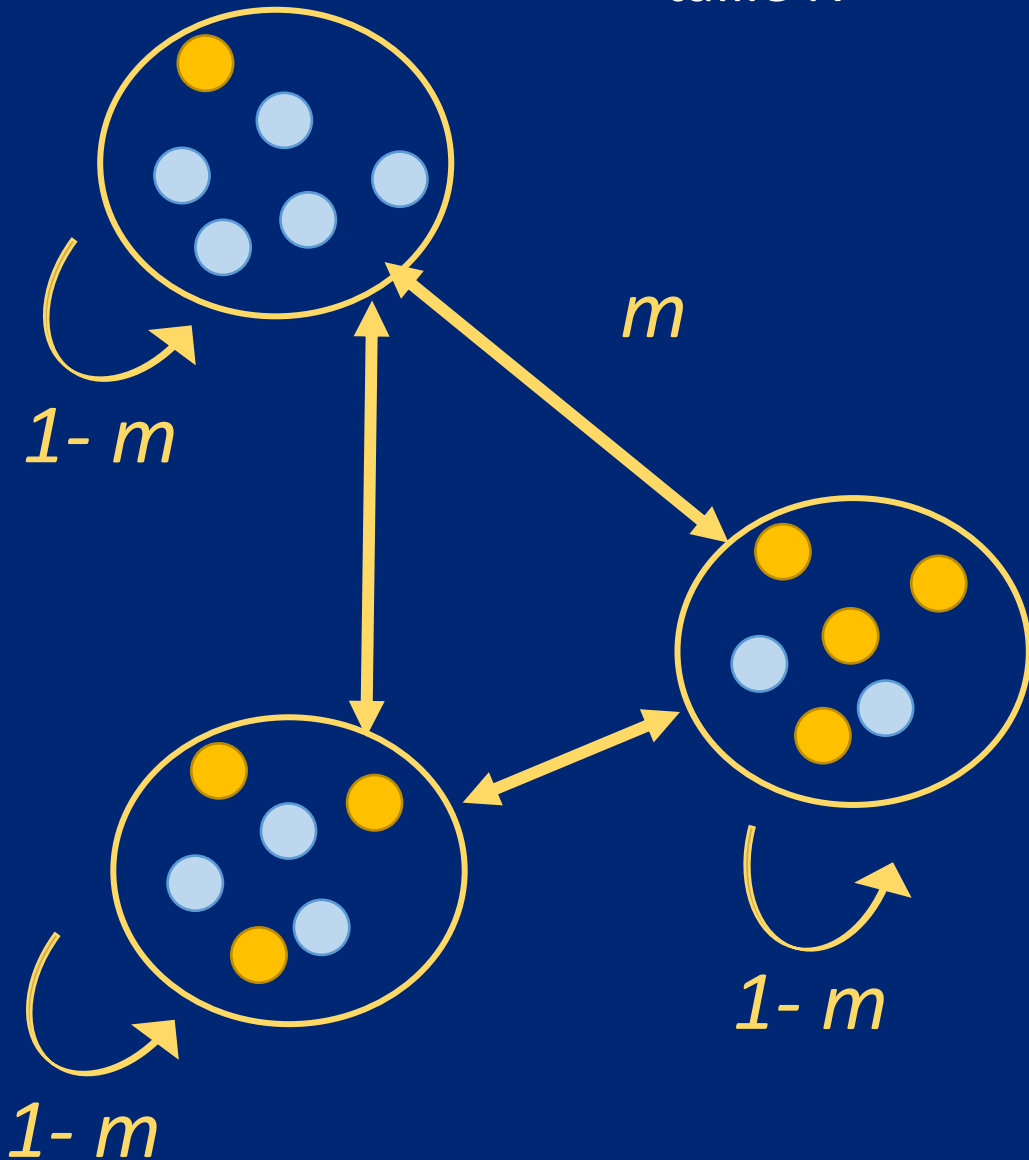


$1-m$



Migration

P populations de
taille N



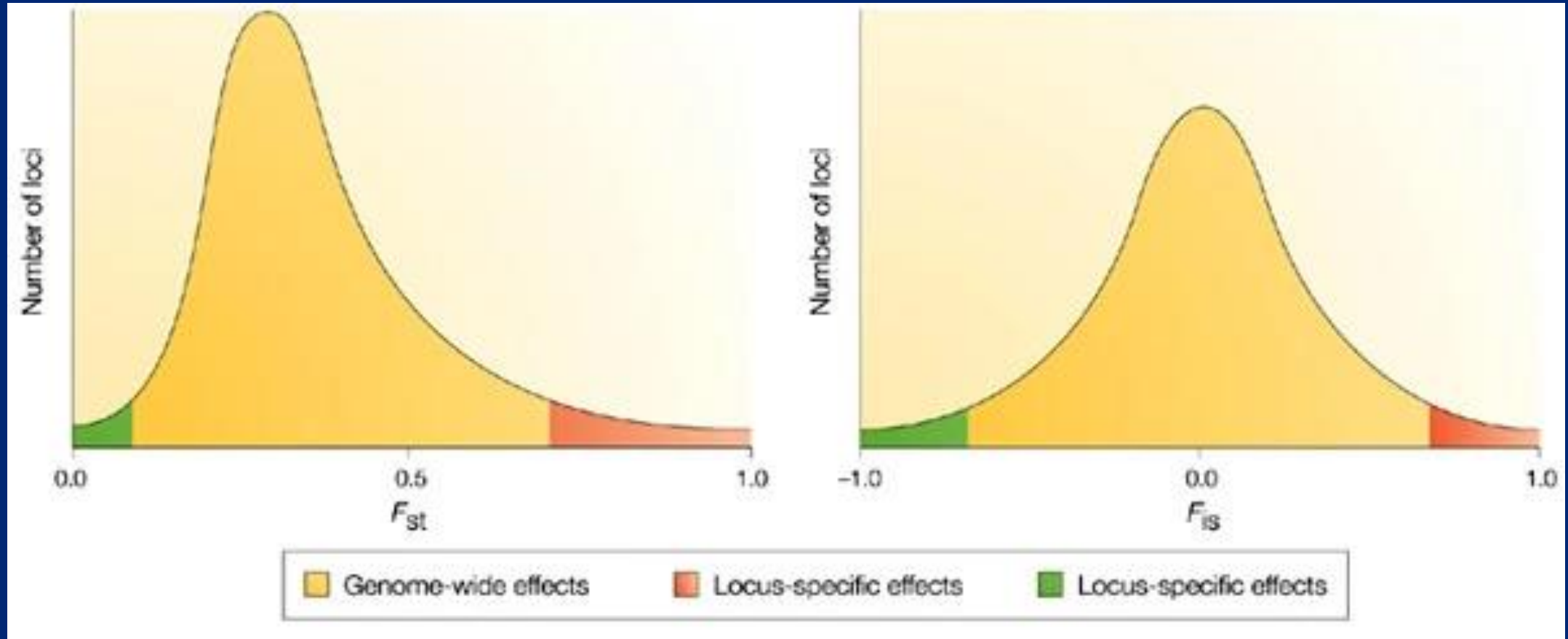
Migration

- comme la mutation, la migration apporte de **nouveaux allèles** dans une population, c'est une source d'innovation génétique.
- mais la migration est une force de bien **plus forte intensité que la mutation**.
- En l'absence d'autres forces évolutives, la migration tend à **homogénéiser les fréquences alléliques** entre populations qui échangent des gènes (= mélange).
- Elle s'oppose donc à la différenciation génétique qui tend à s'installer entre des populations sous l'effet de la dérive notamment.
- Le modèle de migration (la façon dont les individus se déplacent dans l'espace, la dispersion en groupe ou solitaire, etc..) influence beaucoup la distribution spatiale du polymorphisme.
- d'autres modèles que le modèle en îles sont étudiés (isolement par la distance, populations subdivisées (structurées), métapopulations, admixture, hétérogénéité des taux de migration etc...).

L'essentiel à retenir

- La diversité peut être étudiée/mesurée à **différents niveaux**: intra-population (H_e , F) - entre populations (F_{ST})
- La diversité génétique va **naturellement s'éroder dans une population fermée** (sous l'effet de la dérive génétique et/ou de la sélection positive/purifiante). Ce phénomène peut être accentué si les effectifs des populations sont faibles ou si la sélection est forte.
- La sélection balancée et/ou la migration entre populations permettent au contraire de **maintenir la variabilité génétique**
- La mutation est la source fondamentale de variation à l'échelle de la population totale, mais elle joue peu sur la variations des fréquences alléliques au fil des générations

L'essentiel à retenir

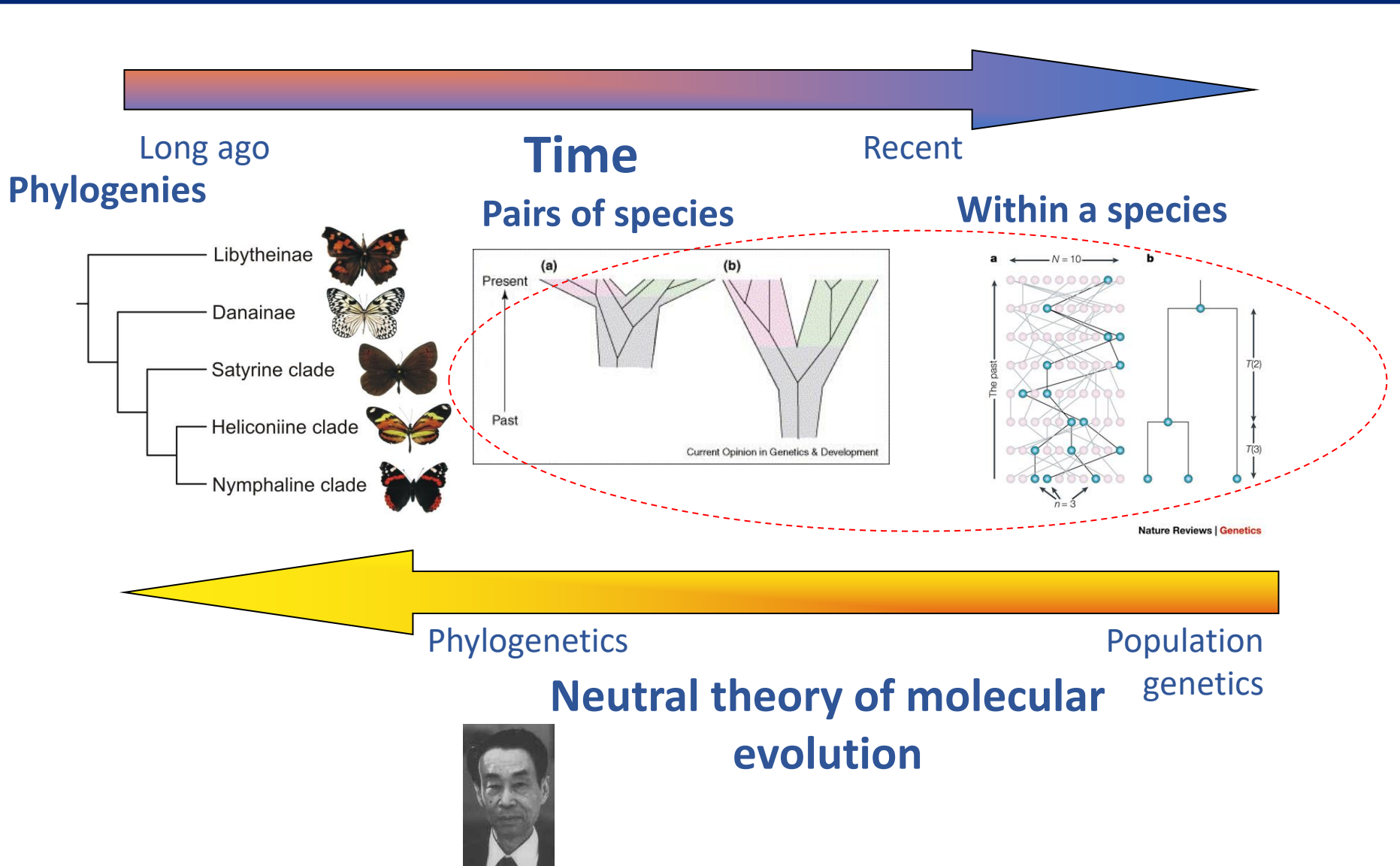


- Les processus démographiques (dérive génétique, migration) affectent l'ensemble des locus alors que la sélection cible des locus spécifiques

Plan du cours

- I. Introduction
- II. Rappels de base en génétique des populations
- III. Estimation de la démographie historique : principes fondamentaux, méthodes courantes, étude de cas
- IV. Estimation de l'histoire sélective et de l'adaptation génétique: principes fondamentaux, méthodes courantes, étude de cas
- V. Vers l'inférence conjoint de processus démographiques et sélectifs

L'histoire démographique peut être étudiée à différentes échelles de temps



- Fluctuations démographiques (expansion, contraction bottleneck)
- Taille de population
- Migration
- Hybridation, admixture entre lignées ou espèces proches

Méthodes courantes d'inférence démographique

- Méthodes basées sur le SFS (Site Frequency Spectrum) et la théorie du **coalescent**
- Méthodes basées sur le **coalescent** séquentiel markovien par paire
- Méthodes basées sur l'**ABC** (Approximate Bayesian Computation) avec des statistiques résumées

Introduction à la théorie du coalescent

- La théorie classique de la génétique des populations tente de prédire ce qui se passera à l'avenir dans une population donnée. Il s'agit d'une **approche prospective**.
- La théorie du coalescent est une approche **rétrospective** de la génétique des populations basée sur la **généalogie des copies de gènes**. Cette généalogie est représentée sous la forme d'un **arbre**.
- La théorie du coalescent utilise les mathématiques pour décrire les caractéristiques de la jonction des lignées dans le temps jusqu'à un ancêtre commun. **Cette jonction de lignées est appelée coalescence**.

Un exemple de généalogie

Présent



20 individus

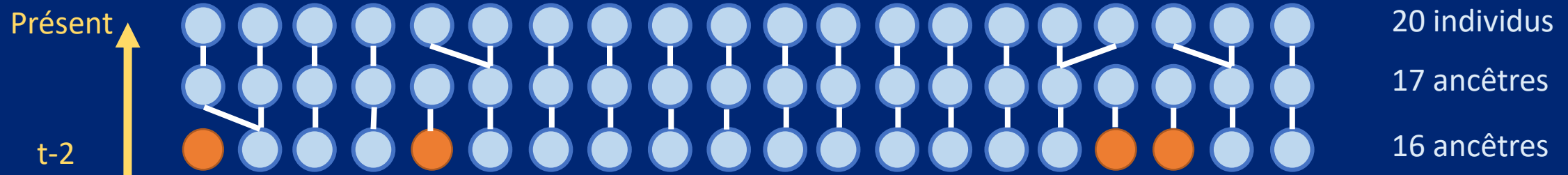
Temps



Un exemple de généalogie



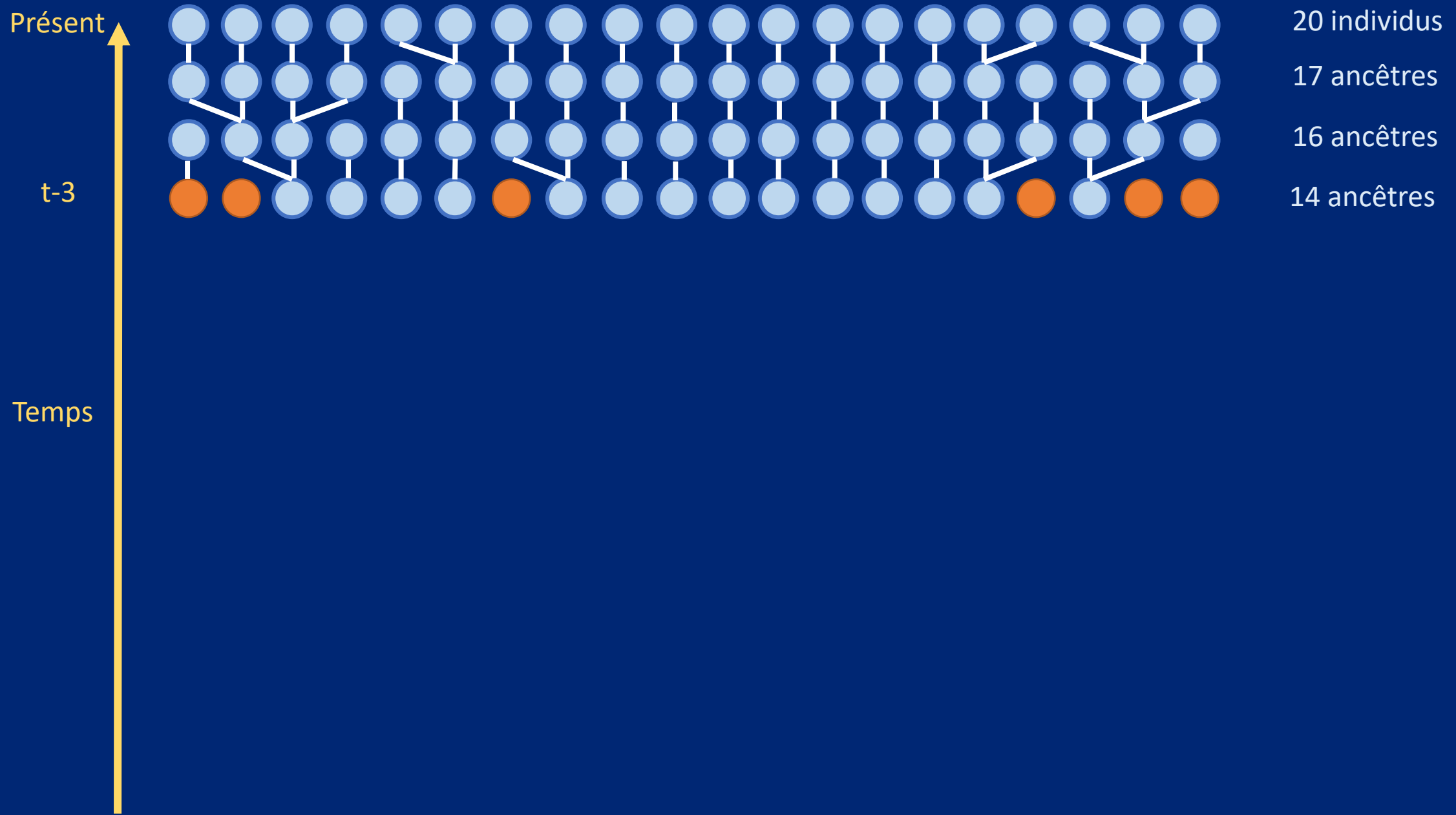
Un exemple de généalogie



Temps

1 ancêtres

Un exemple de généalogie



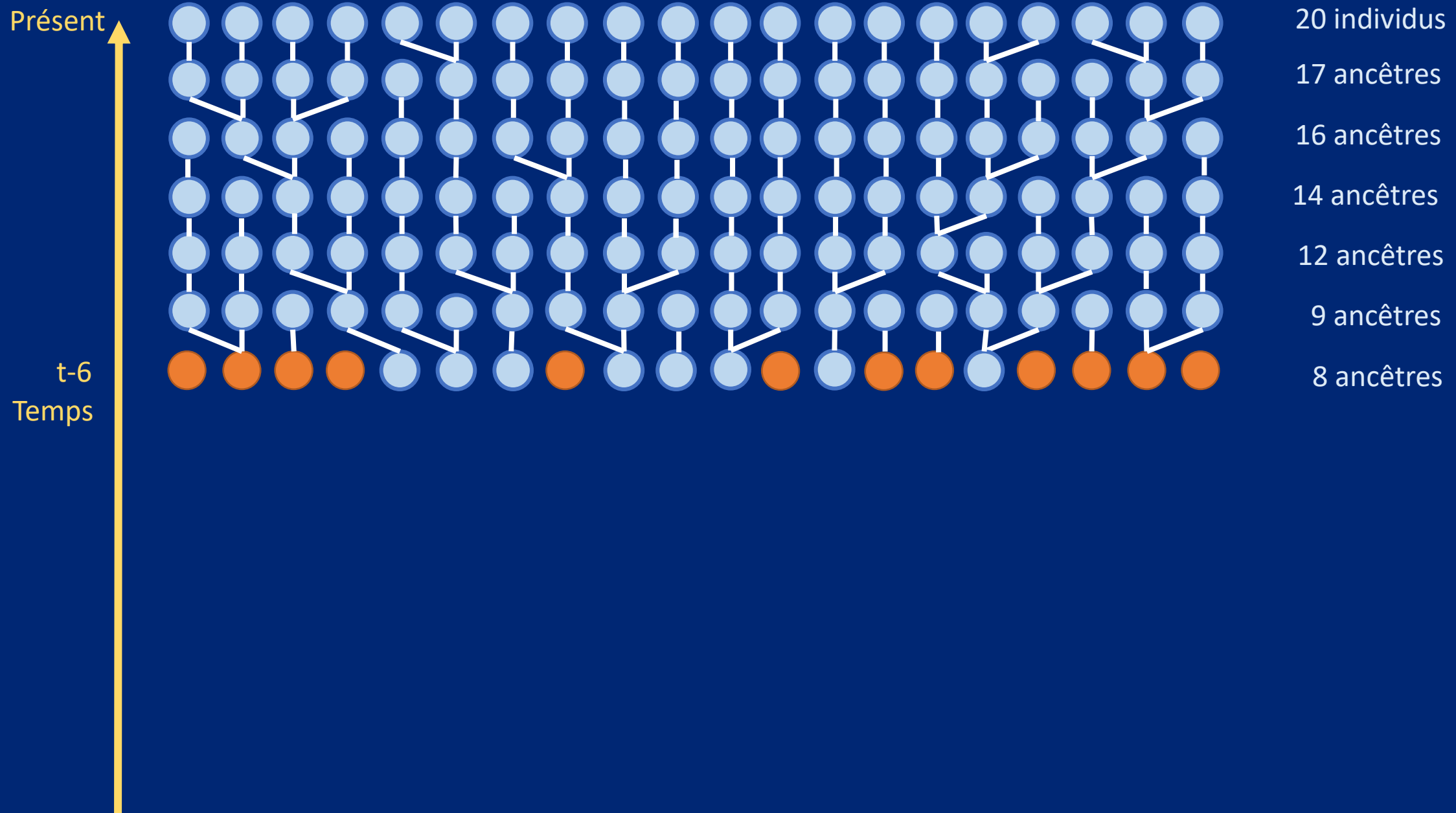
Un exemple de généalogie



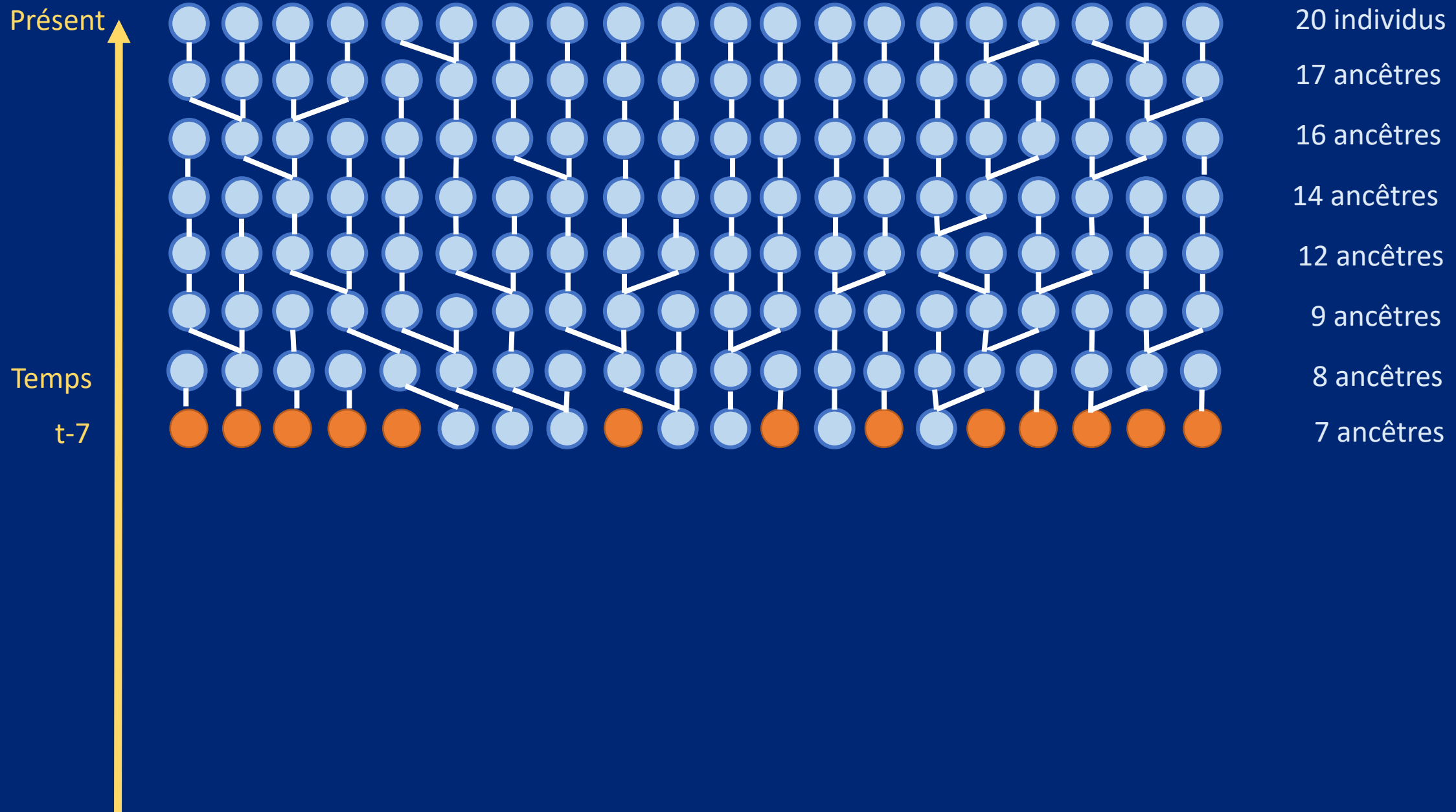
Un exemple de généalogie



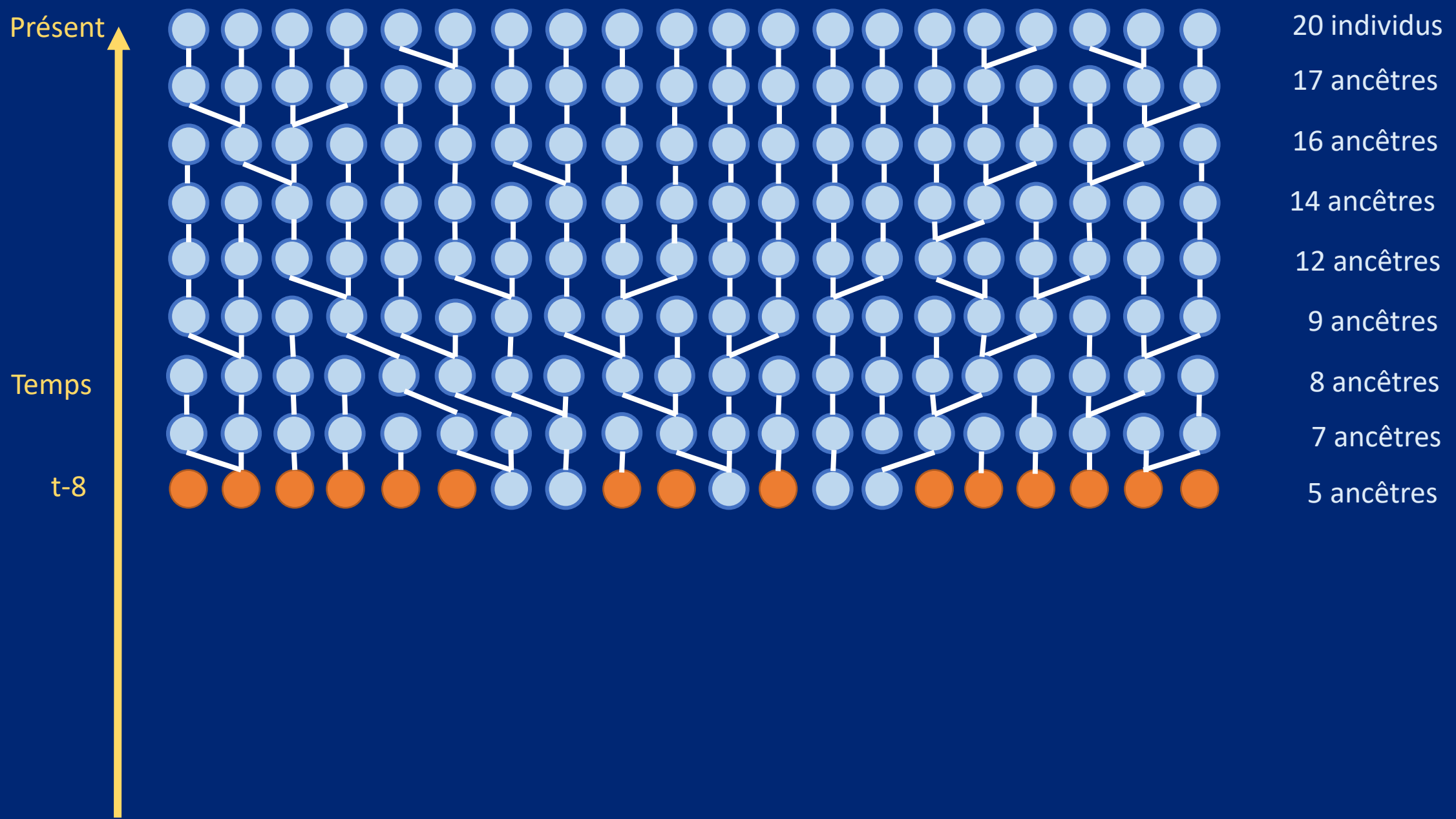
Un exemple de généalogie



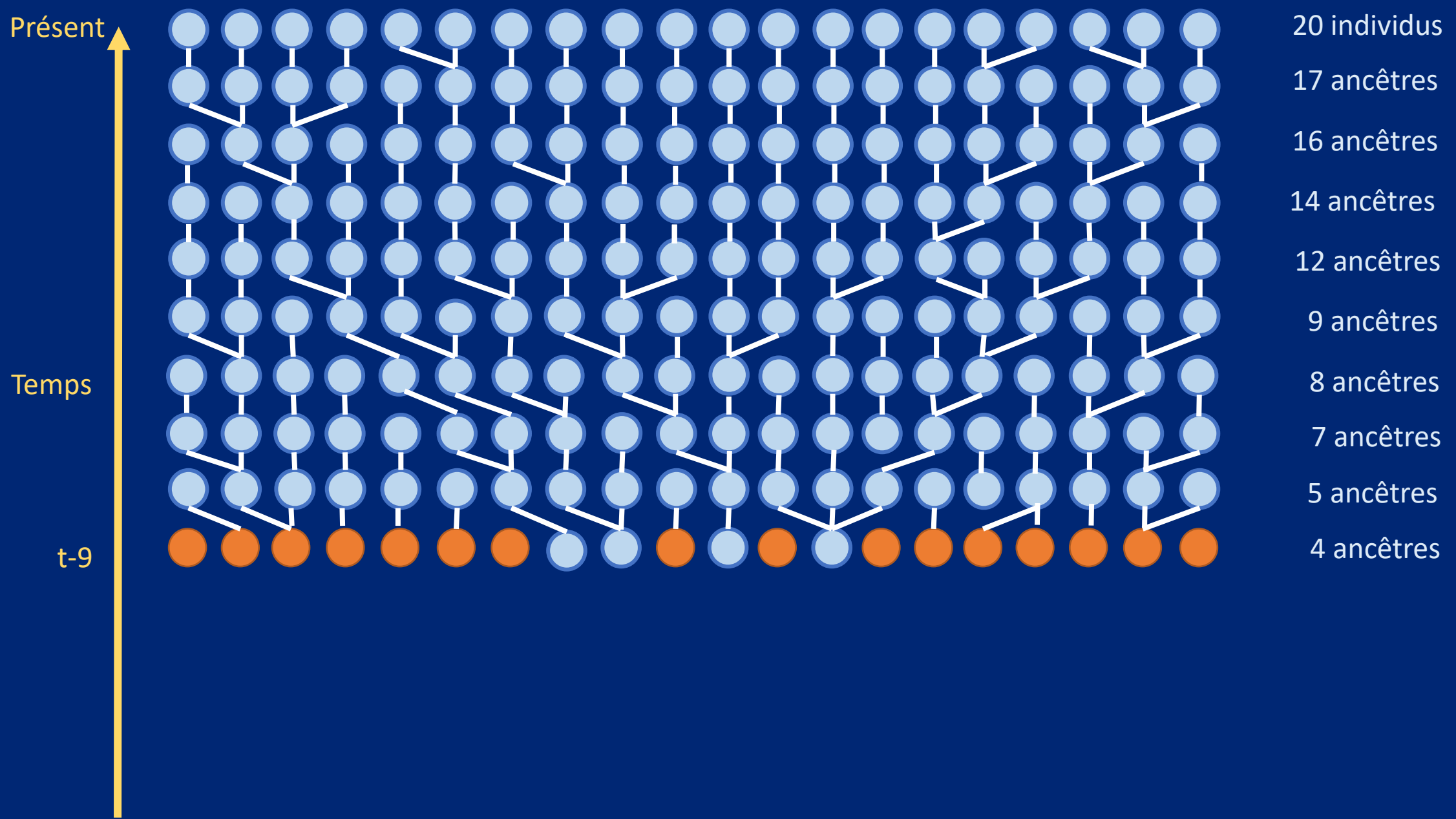
Un exemple de généalogie



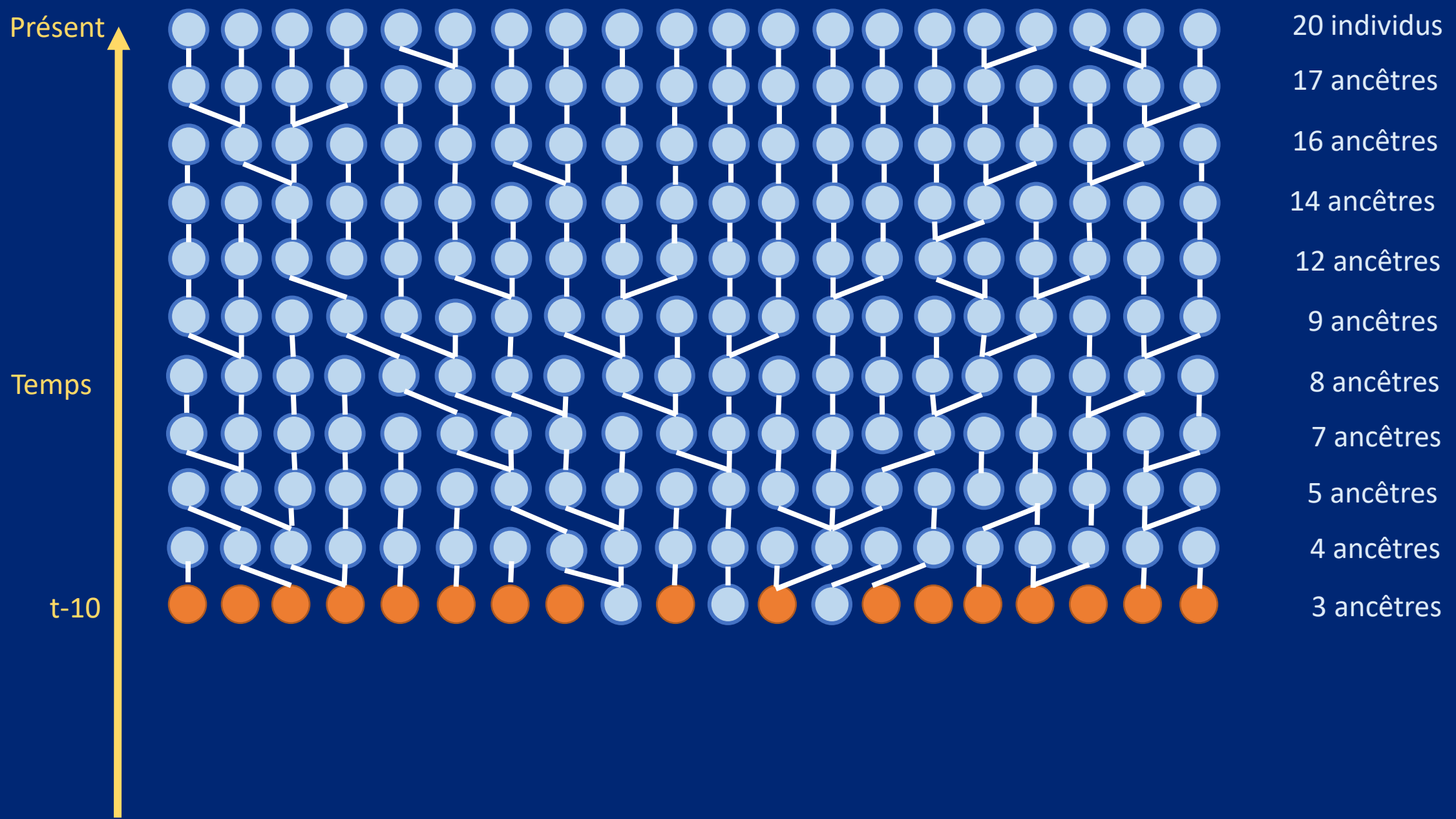
Un exemple de généalogie



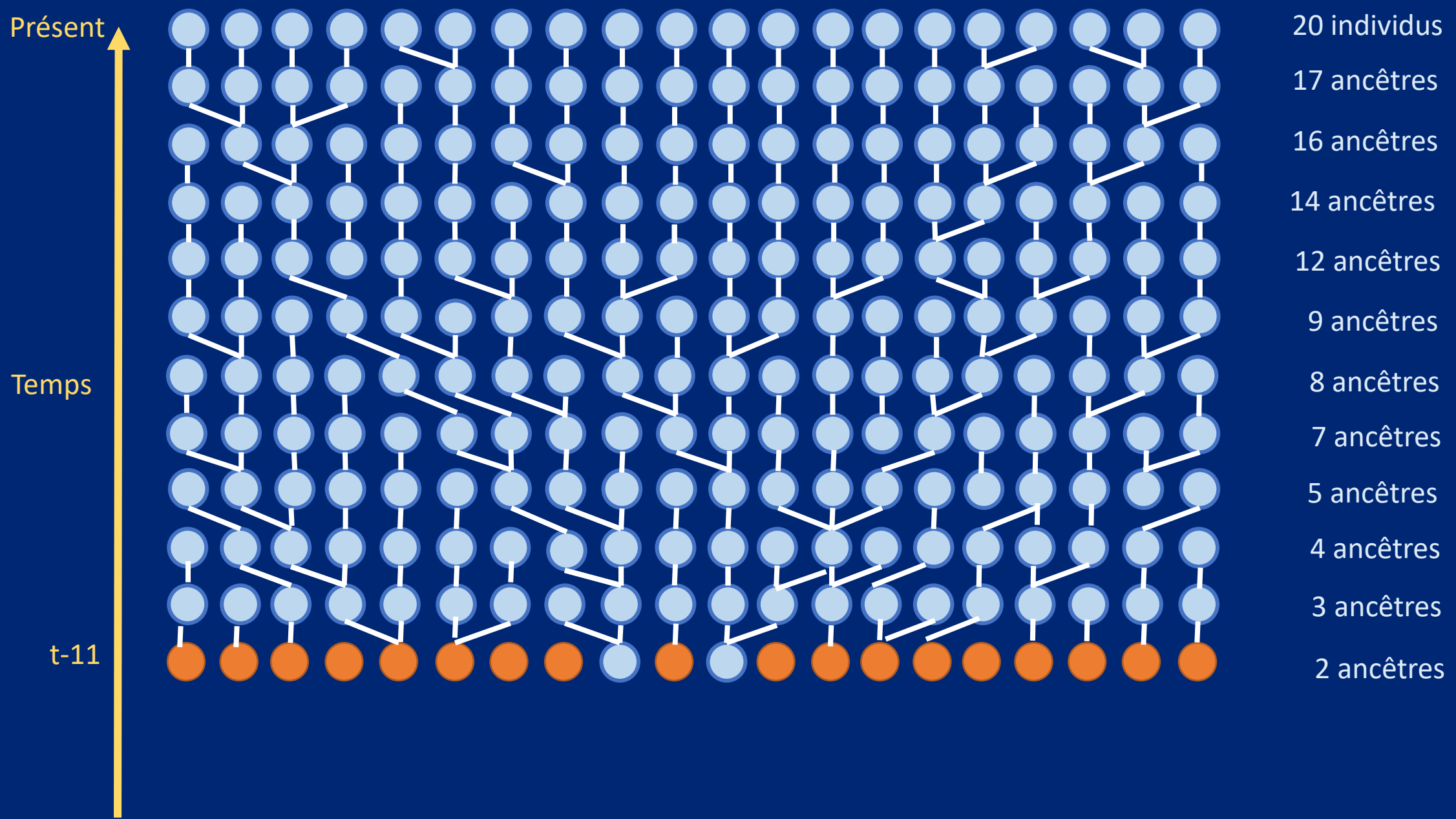
Un exemple de généalogie



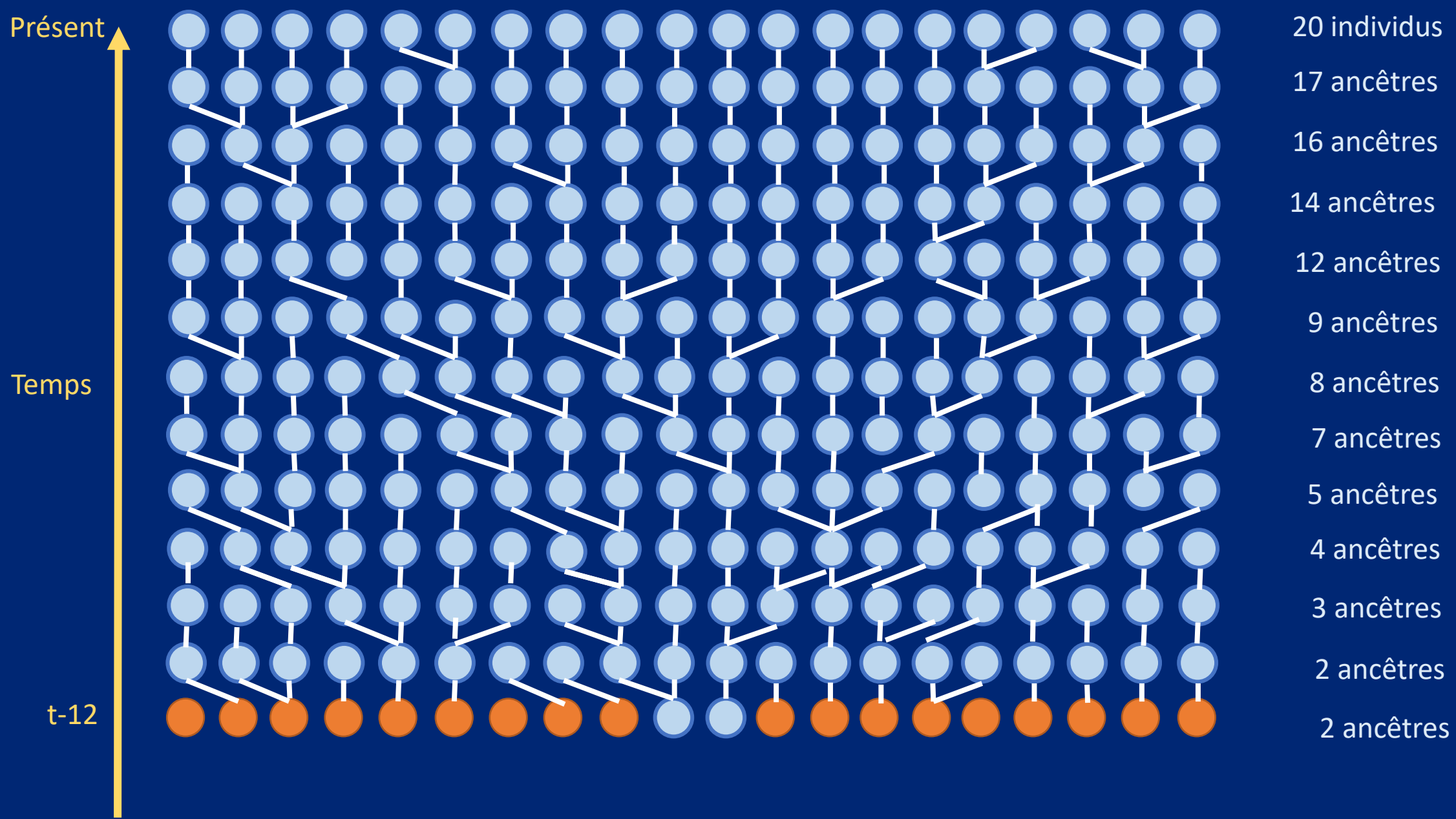
Un exemple de généalogie



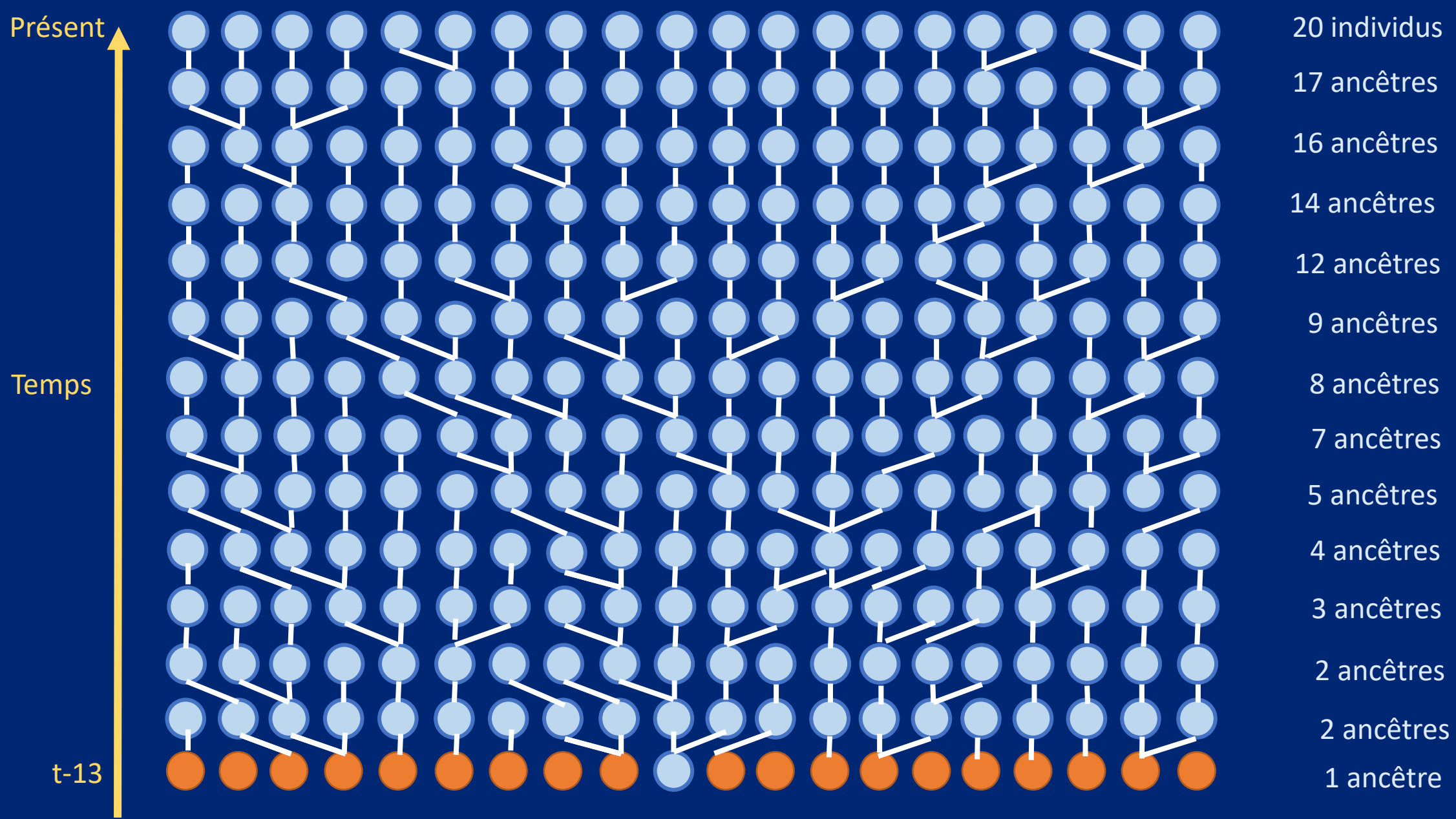
Un exemple de généalogie



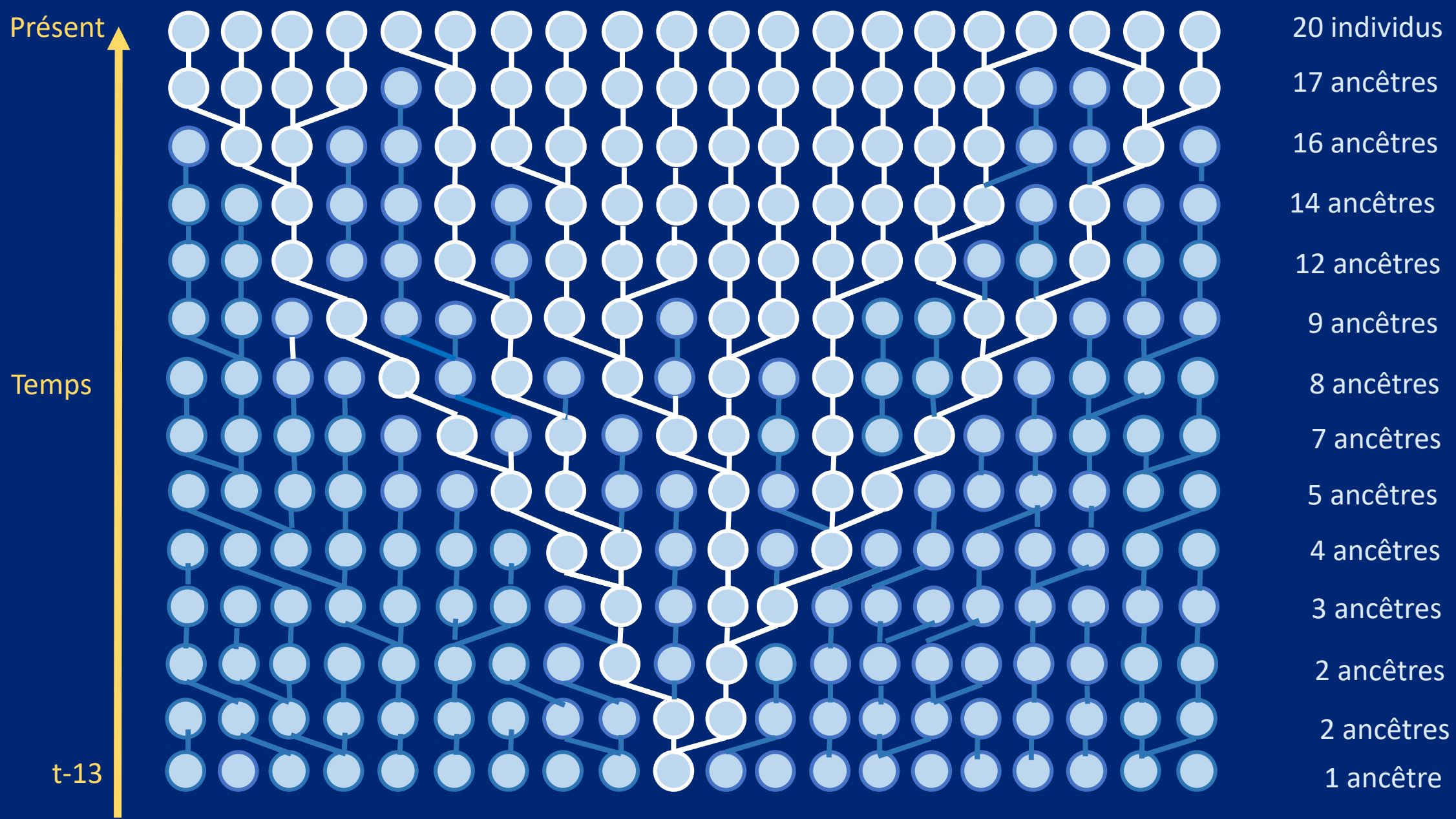
Un exemple de généalogie



Un exemple de généalogie



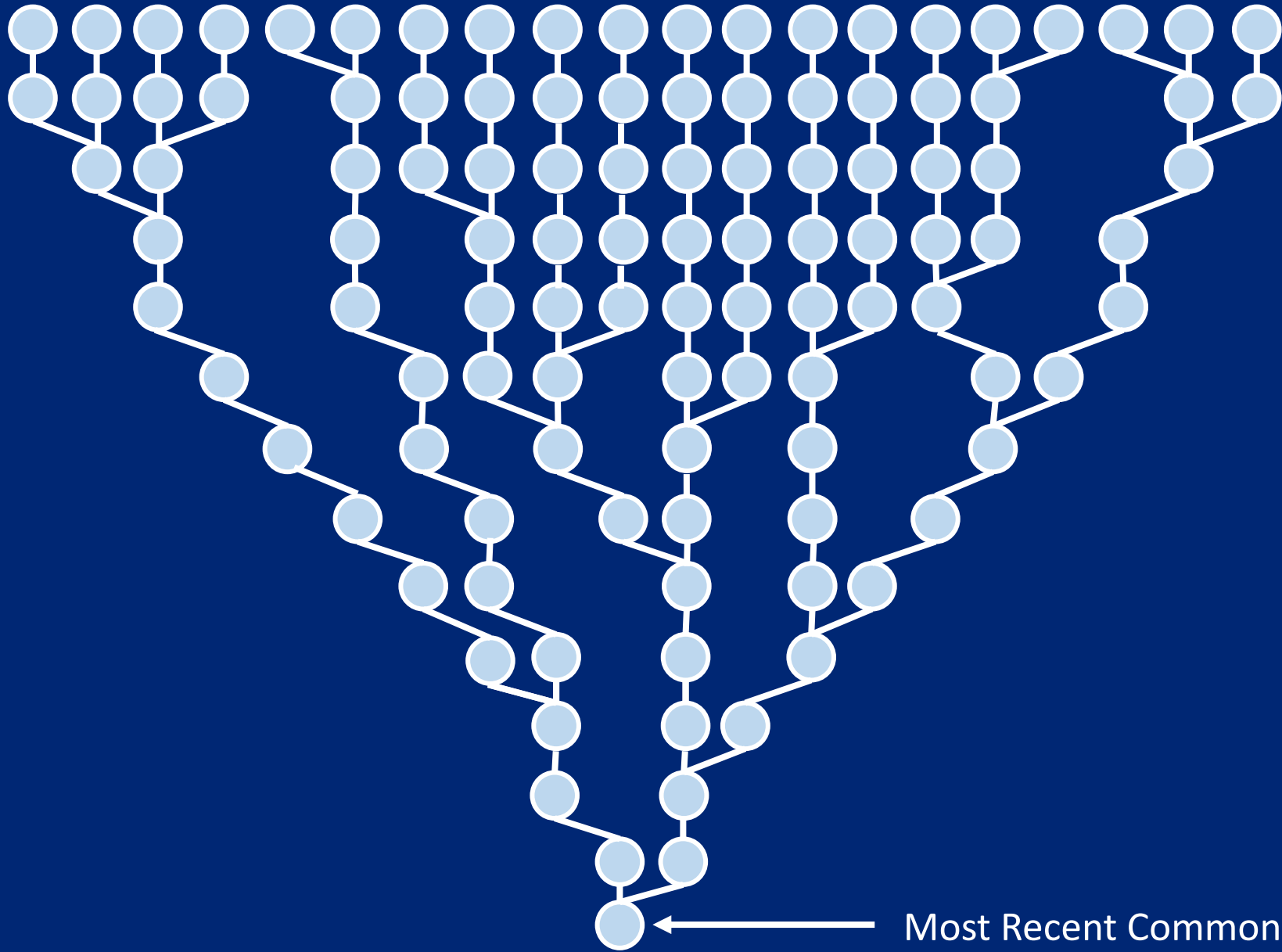
Un exemple de généalogie



Notion d'ancêtre commun le plus récent

Présent

Temps

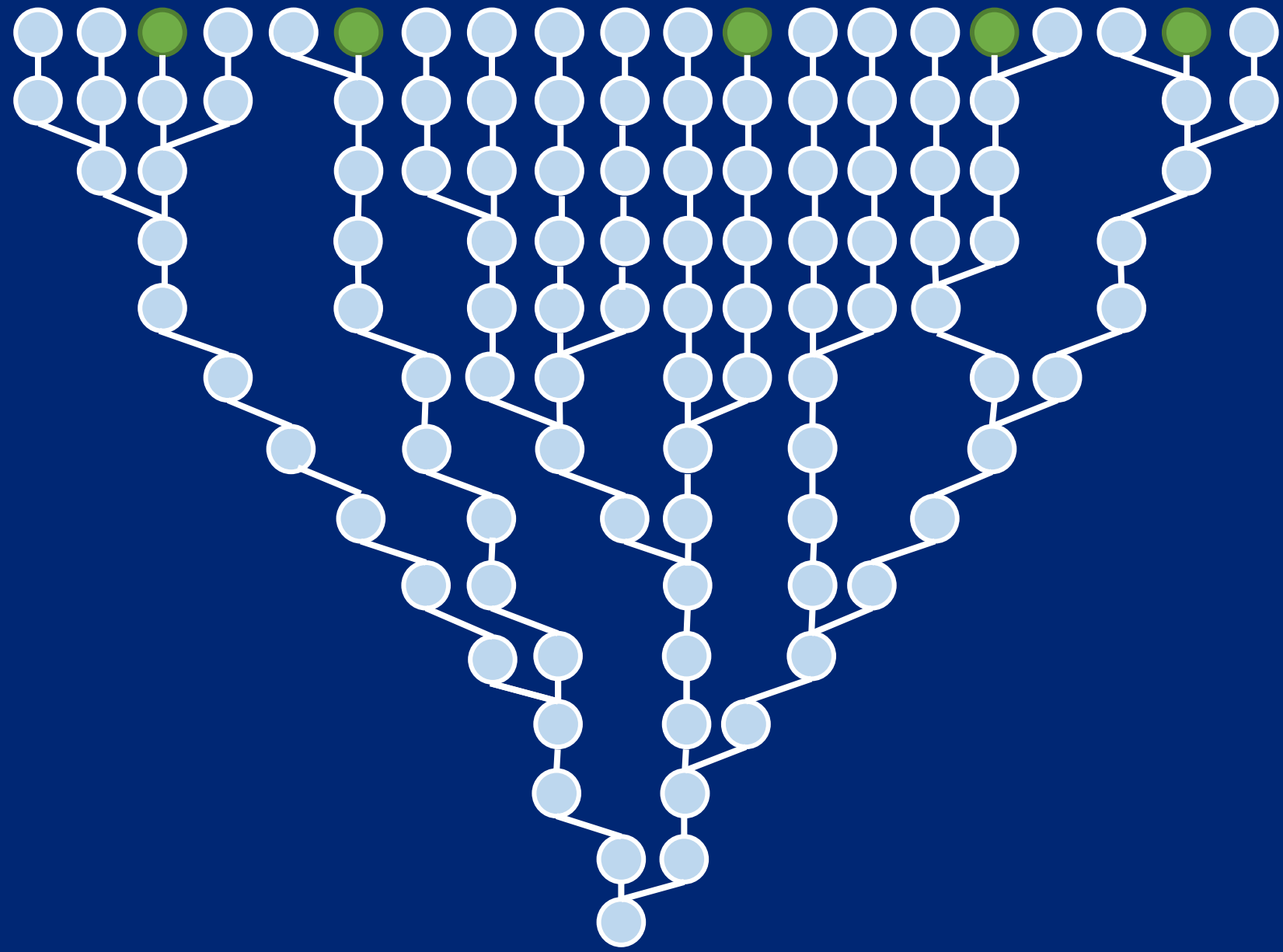


← Most Recent Common Ancestor (MRCA)

Vers la théorie du coalescent

Présent

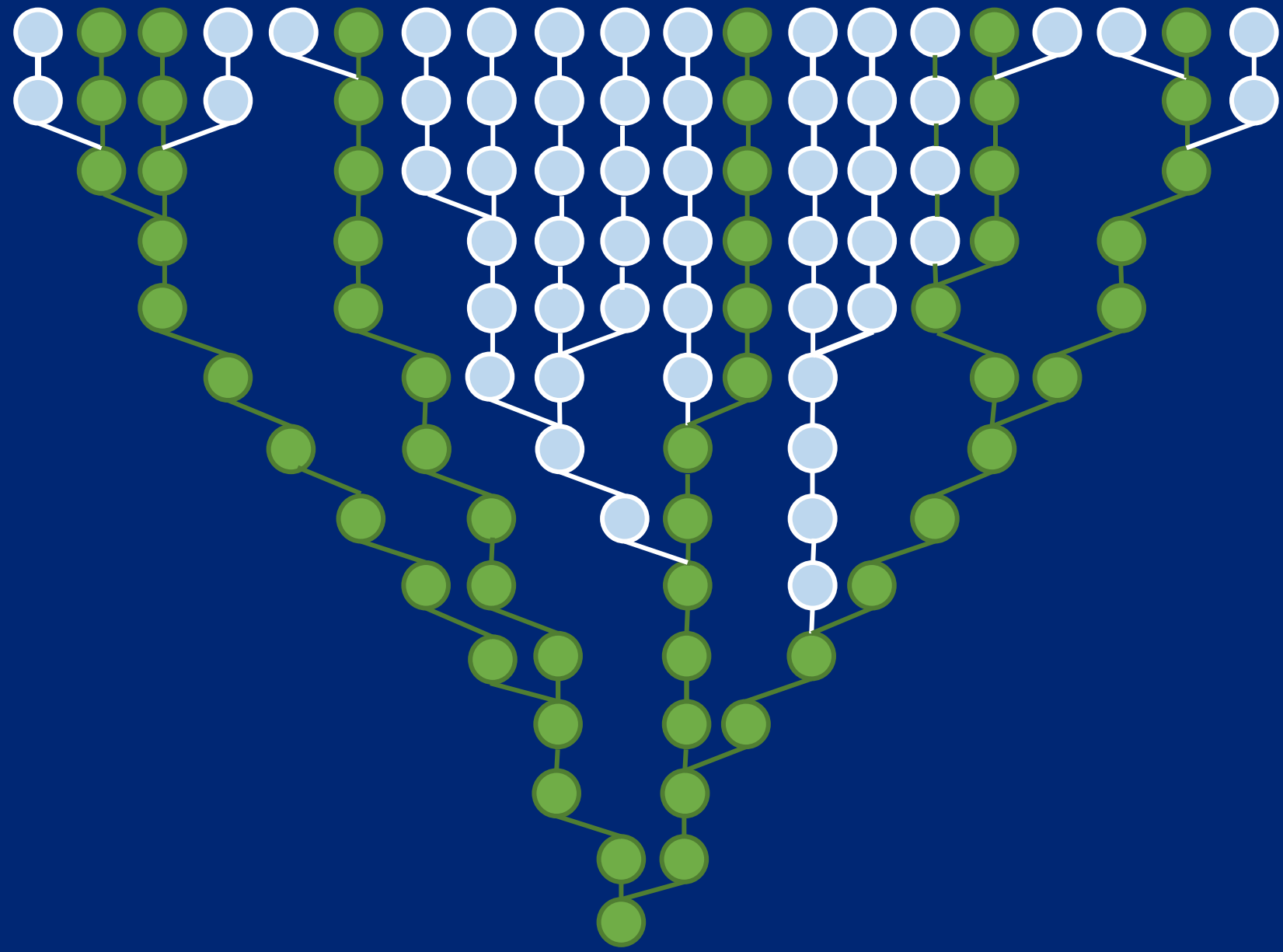
Temps



Vers la théorie du coalescent

Présent

Temps



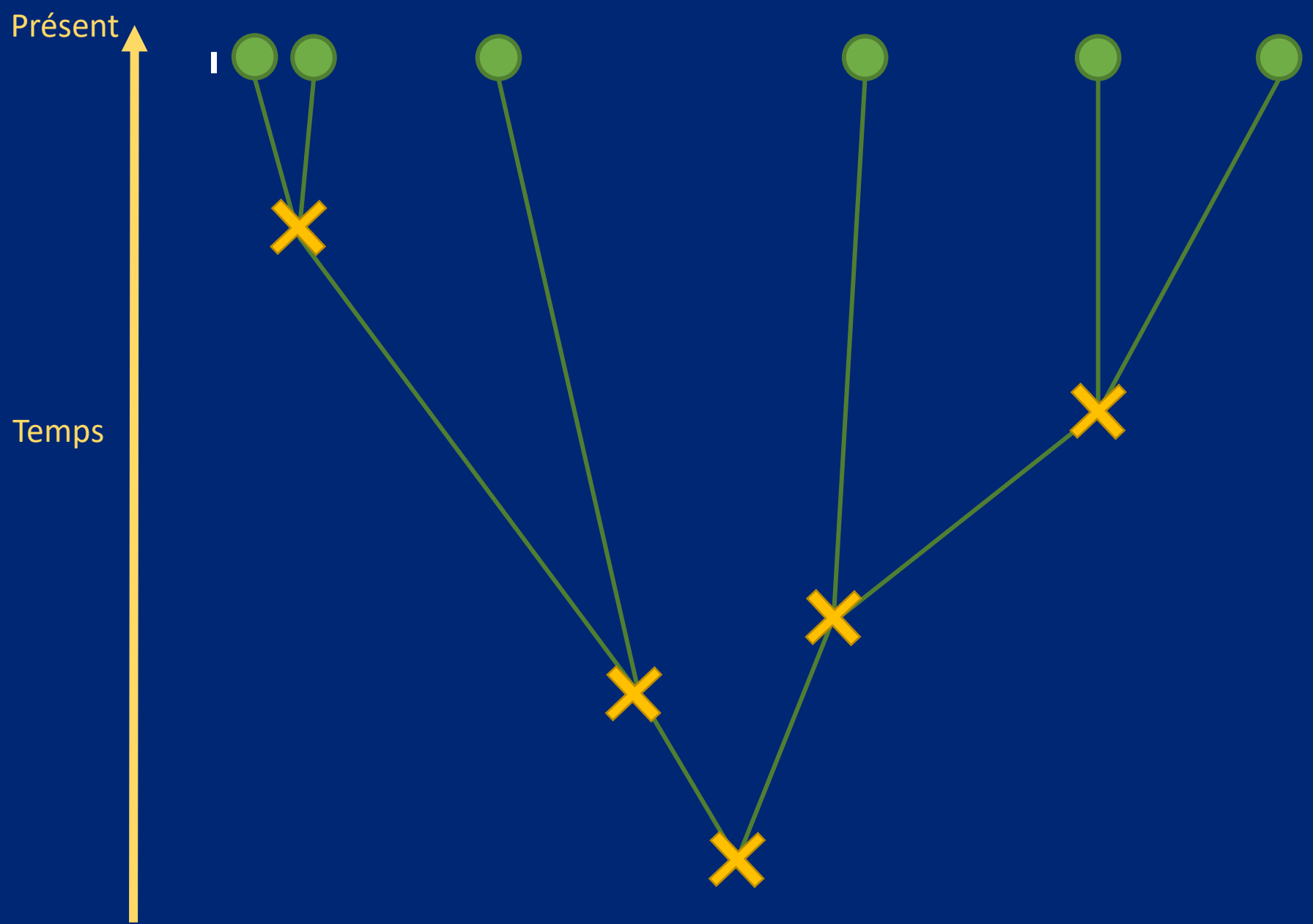
Vers la théorie du coalescent

Présent

Temps



Vers la théorie du coalescent

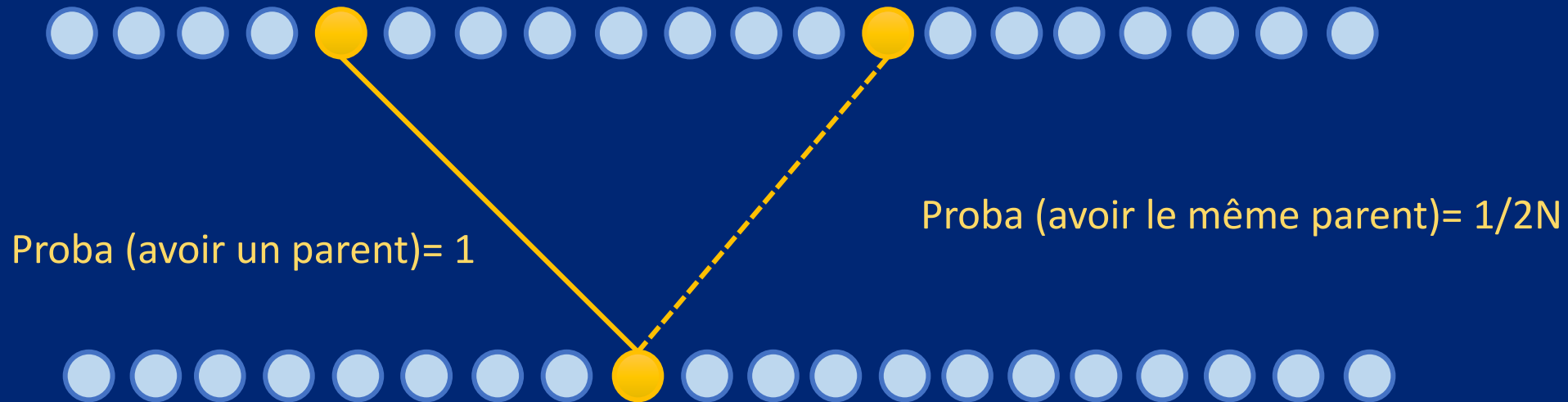


Esprit et spécificité de l'approche coalescence

- Questions typiques: quel est l'ensemble des arbres généalogiques possibles pour un modèle de génétique des populations donné ? Si on connaît tous les arbres possibles pour un modèle donné, quels sont les paramètres de ce modèle le plus compatible avec l'arbre observé ? Les paramètres sont typiquement des valeurs pertinentes en génétique des populations (taille de la population, taux de migration, taux de croissance de la population, force de la sélection etc.)
- Les notions d'arbres "possibles", paramètres "compatibles" sont formalisables, pour cela, recours à des notions plus ou moins simples de probabilité. On ne cherche pas si un arbre est "possible", mais quelle est la probabilité associée à cet arbre sous un certain modèle.

Formulation probabiliste

Dans un modèle de Wright-Fisher (population de taille constante, croisements aléatoires, en génération discrètes non chevauchantes)

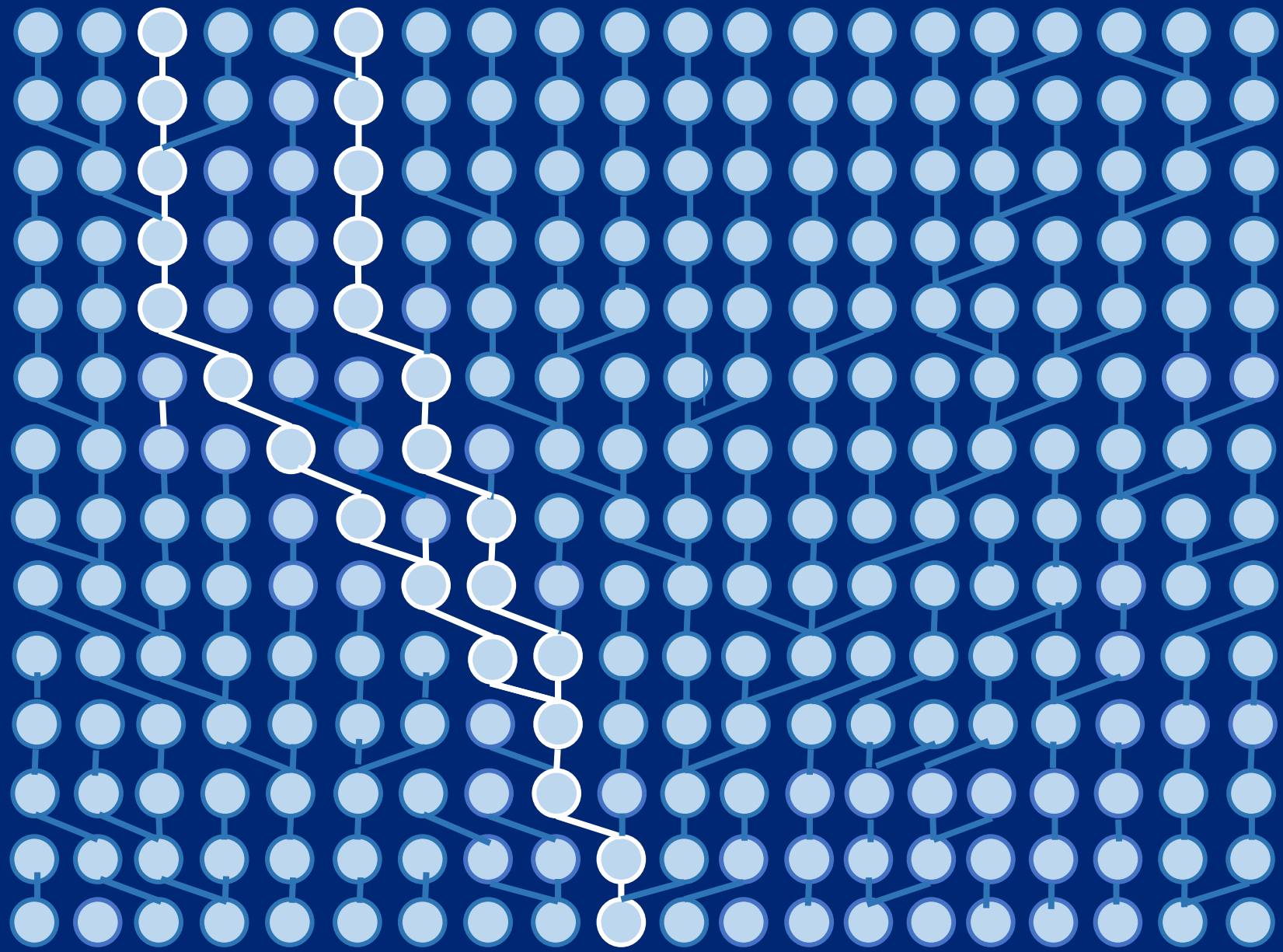


Formulation probabiliste pour 2 gènes/lignées

Présent

Temps

t-10



$$p = 1 - \frac{1}{2N}$$



$$p = 1 - \frac{1}{2N}$$
$$p = \frac{1}{2N}$$



$$p = \left(1 - \frac{1}{2N}\right)^9 \times \frac{1}{2N}$$

Formulation probabiliste pour 2 gènes/lignées

- Le temps de coalescence de deux lignées/ gènes suit une loi géométrique de paramètre $1/2N$.
- La probabilité que deux gènes coalescent t générations auparavant (ni avant ni après) est de :

$$\left(1 - \frac{1}{2N}\right)^{t-1} \times \frac{1}{2N}$$

- L'espérance du temps de coalescence de deux lignées/ gènes est **$2N$**
- Mais la variance est très grande:

$$2N \times (2N - 1) \sim N^2$$

Le n-coalescent de Kingman

- On considère k gènes
- Il a y plus de chance d'observer un événement de coalescence: $1/2N$ pour chaque paire possible de lignées parmi les k

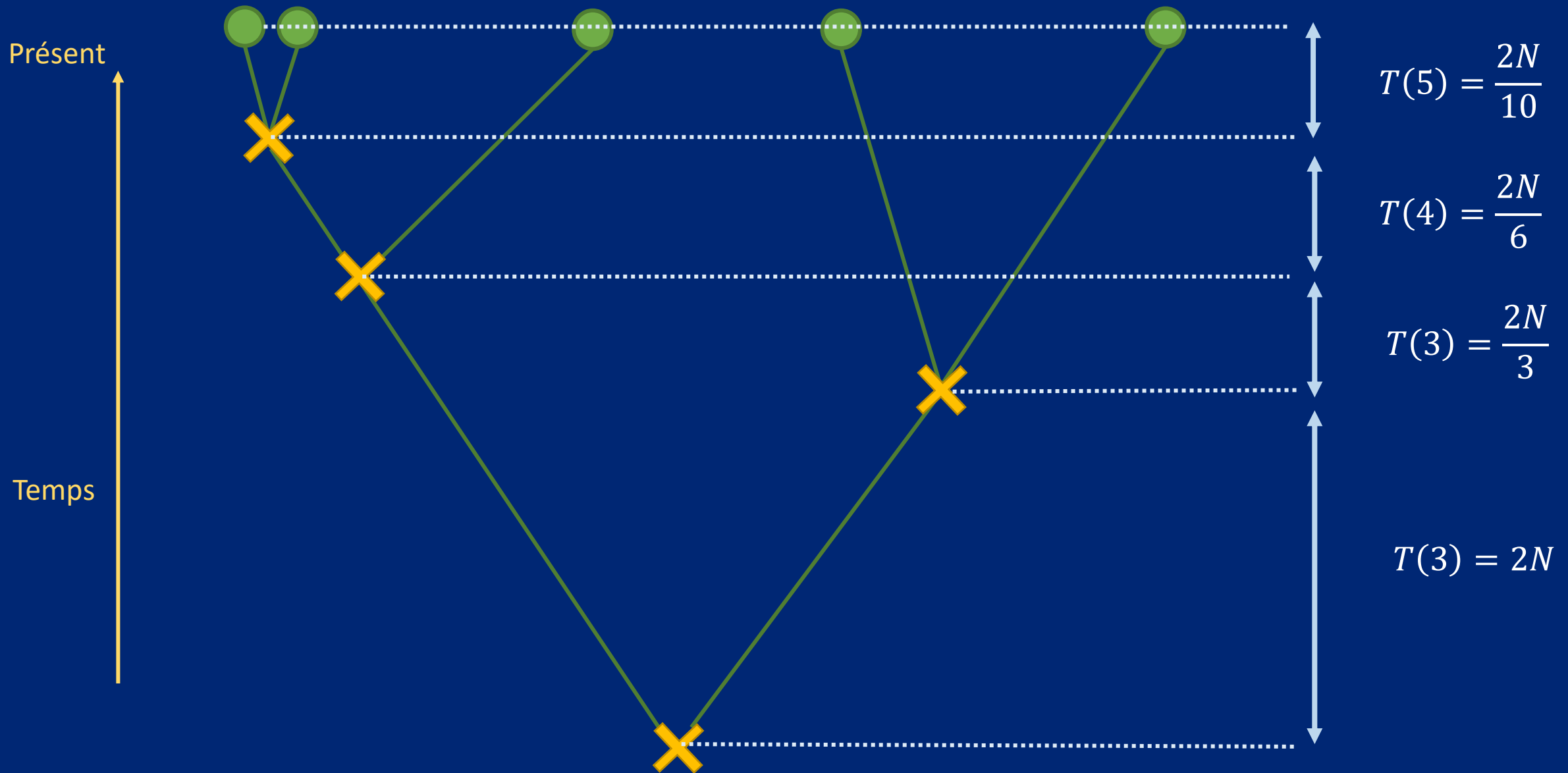
- Nombre de paires possibles:
$$\binom{k}{2} = \frac{k \times (k - 1)}{2}$$

- La probabilité totale pour n'importe quelle paire de coalescer à la génération précédente est alors :
(dépend de k)

$$p = \frac{k \times (k - 1)}{4N}$$

- Le temps de coalescence suit toujours une loi géométrique, mais de paramètre $\frac{k \times (k - 1)}{4N}$
- L'espérance du temps de coalescence de 2 lignées parmi k : $T(k) = \frac{4N}{k \times (k - 1)}$
- On a toujours $T(2)=2N$

Vers la théorie du coalescent



Quelques idées générales

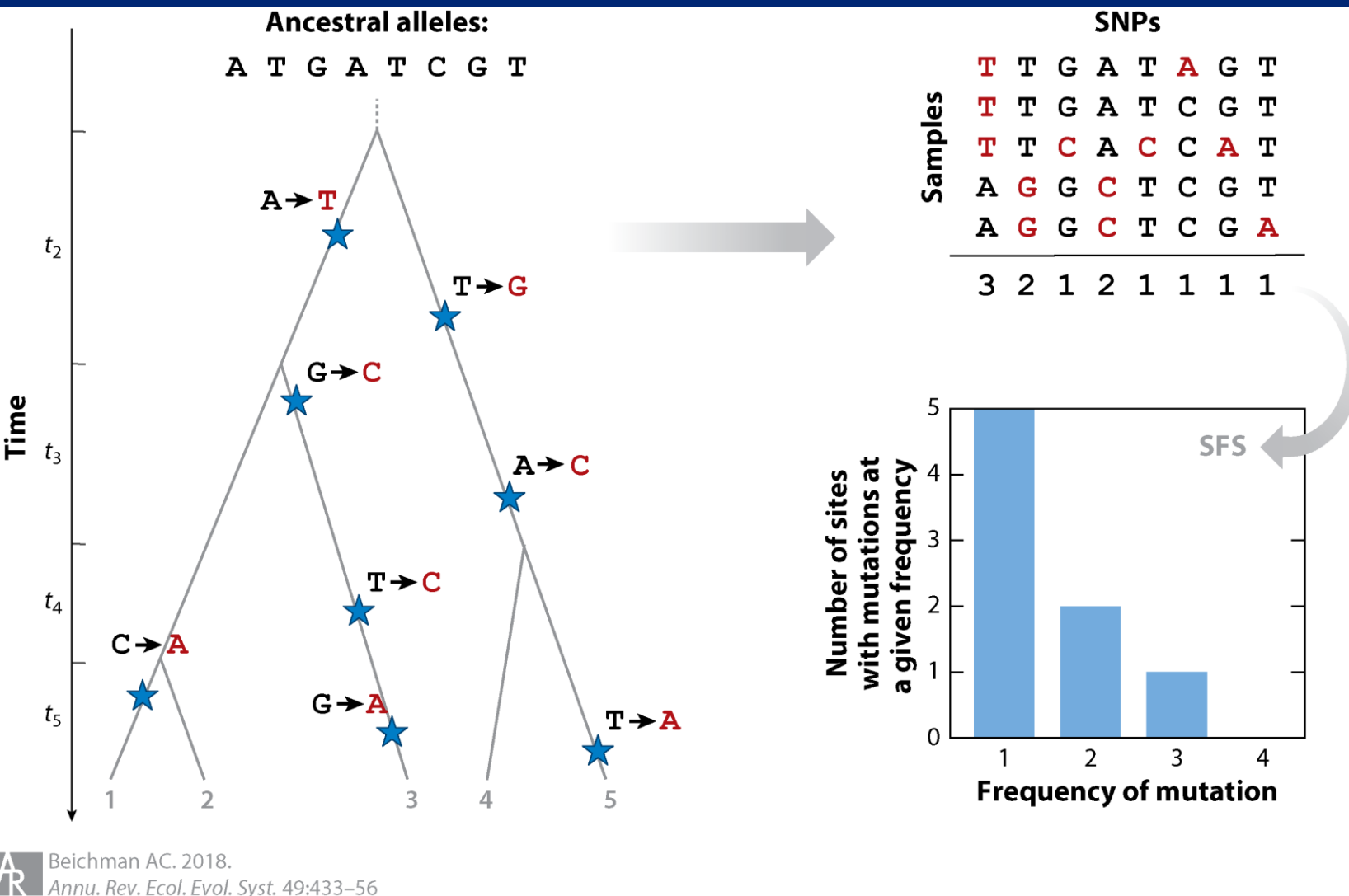


- Les généalogies de gènes peuvent être très variables dans des populations de taille constante, à la fois en terme de topologie et de longueur de branche.
- Généralement on a des longues branches internes et des branches externes courtes
- On peut calculer divers paramètres pour caractériser la topologie de l'arbre: longueur totales des branches, temps jusqu'au MRCA...
- Le coalescent est un cadre conceptuel très puissant pour simuler des généalogies, car on ne simule que ce dont on a besoin (les branches conduisant au MRCA)
- La complexité augmente avec le nombre de lignées.

Quelques caractéristiques

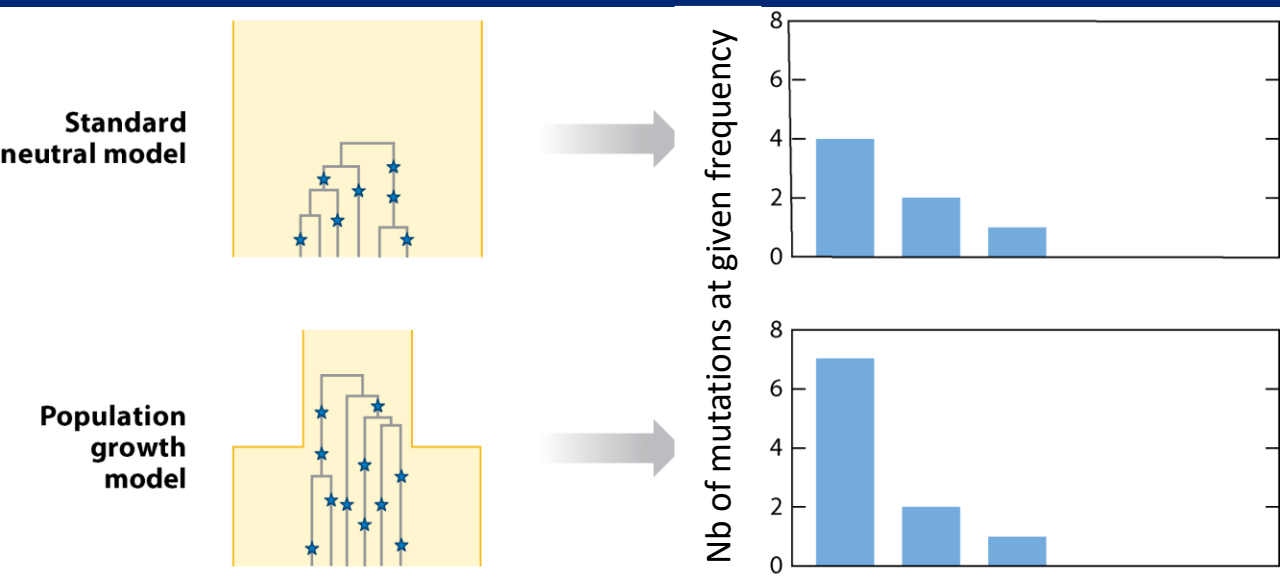
- Le coalescent capture bien l'effet du processus de dérive génétique sur la diversité génétique au temps présent; la forme de l'arbre ne dépend que de la démographie de la population et de l'échantillonnage.
- Et les autres forces évolutives ?
 - ✓ Le processus de mutation peut être modélisé comme un processus indépendant qui se superpose à l'arbre de coalescence réalisé -> approches basées sur le SFS
 - ✓ On peut ajouter de la migration, de l'admixture...
 - ✓ (On peut ajouter de la sélection)

Méthodes d'inférence démographique basée sur le SFS



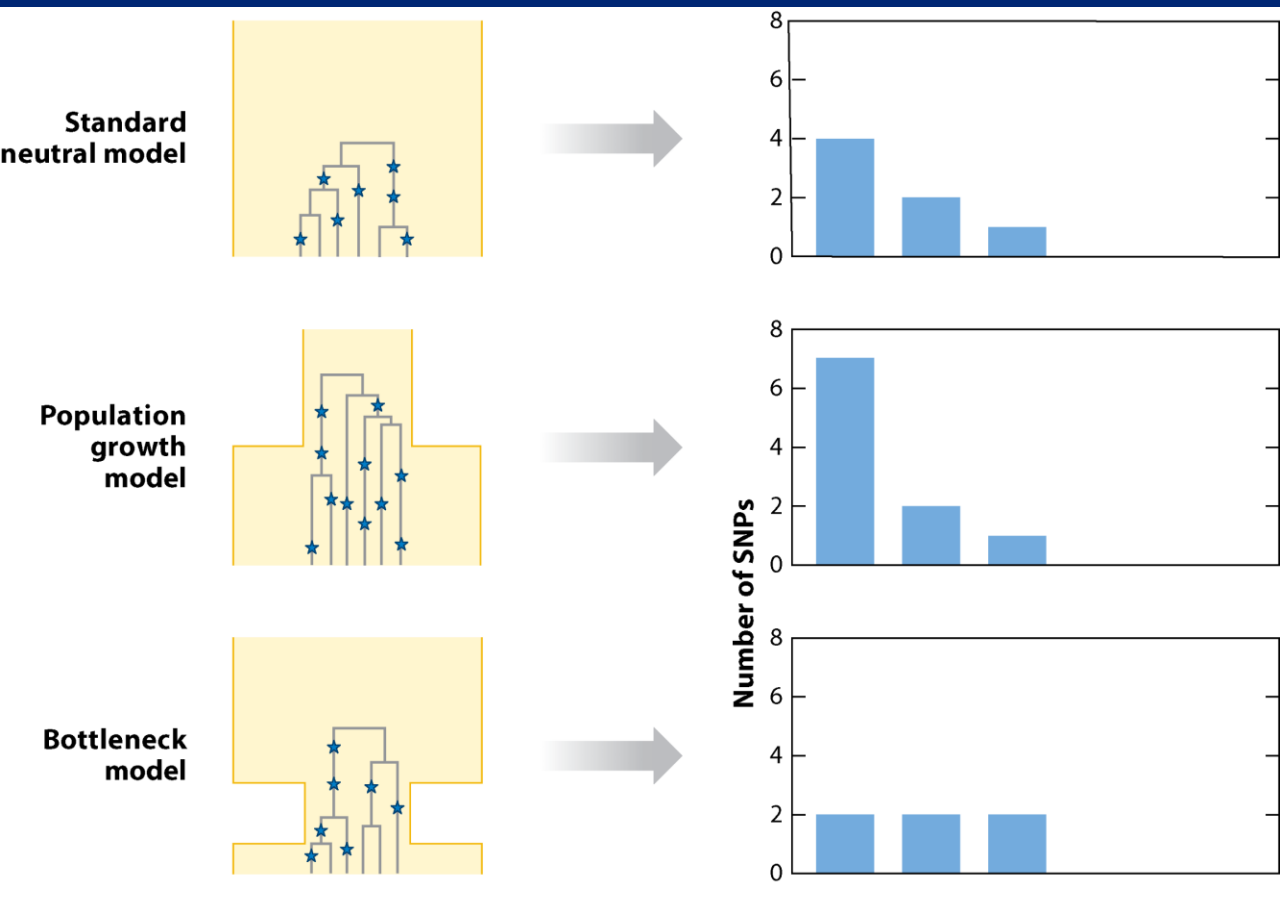
- SFS généralement construit à partir de données de séquence plutôt que des SNPs
- Nombre d'individus requis: pas moins de 10 (événement anciens) et jusqu'à 100 ou 1000
- Le SFS est influencé par l'histoire démographique de la population

Méthodes d'inférence démographique basée sur le SFS



- Modèle neutre standard, ie Wright Fisher
- Modèle avec croissance de population
 - Longues branches externes (N grand)
 - Branches internes plus courtes (N petit)
 - Plus de singletons -> biais SFS

Méthodes d'inférence démographique basée sur le SFS



- Modèle neutre standard, ie Wright Fisher

- Modèle avec croissance de population

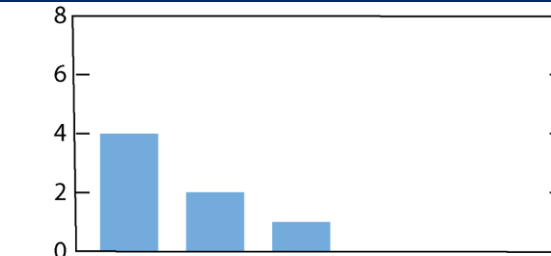
- Longues branches externes (N grand)
- Branches internes plus courtes (N petit)
- Plus de singletons -> biais SFS

- Modèle avec goulet d'étranglement modéré

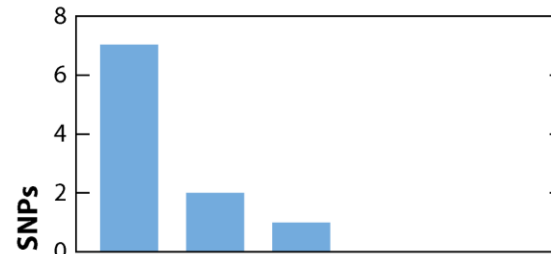
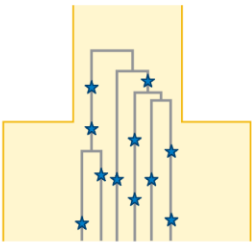
- Pas de coalescence de toutes les lignées
- Après BN, N grand, branche internes longues
- Moins de singletons

Méthodes d'inférence démographique basée sur le SFS

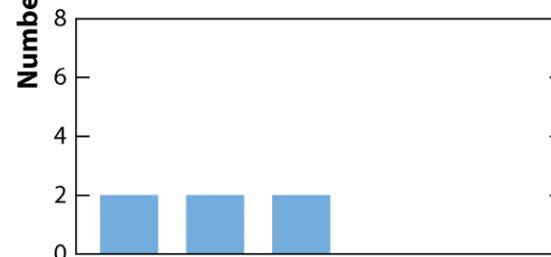
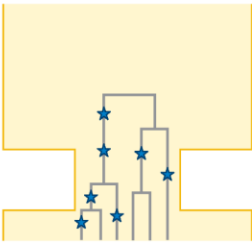
Standard neutral model



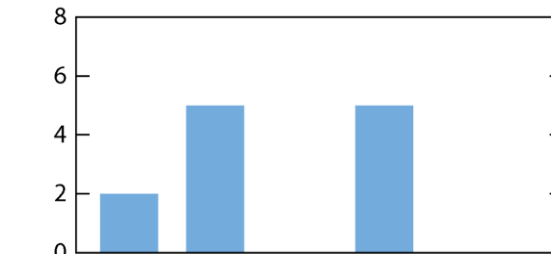
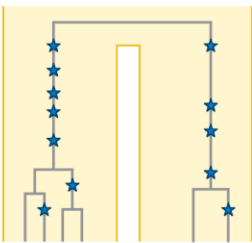
Population growth model



Bottleneck model



Population structure model



Frequency

- Modèle neutre standard, ie Wright Fisher

- Modèle avec croissance de population

- Longues branches externes (N grand)
- Branches internes plus courtes (N petit)
- Plus de singletons -> biais SFS

- Modèle avec goulet d'étranglement modéré

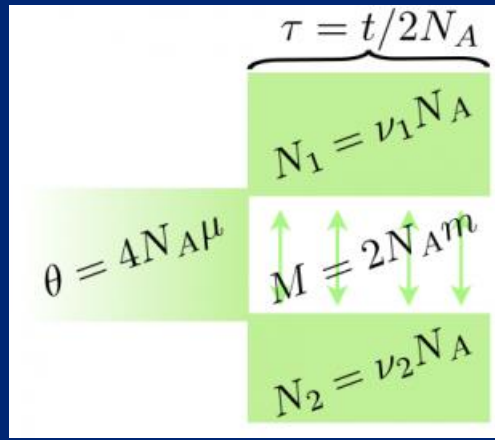
- Pas de coalescence de toutes les lignées
- Après BN, N grand, branche internes longues
- Moins de singletons

- Modèle avec structure de population

- Coalescence plus probable dans chaque sous-population
- Branches internes plus longues
- Si mélange, excès de variant à forte fréquence

Principe général

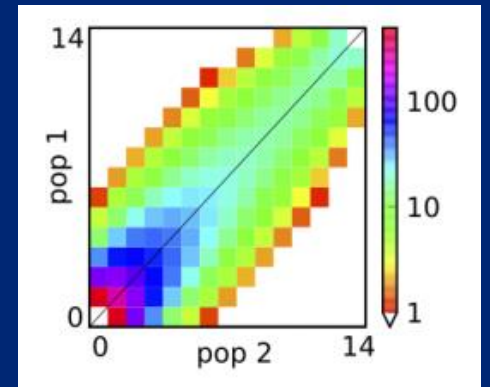
Modèle(s)
démographique (s)



Tirage des valeurs
de paramètres
dans leur
distribution

Simulations
(coalescent):
 $\delta a \delta i$,
fastsimcoal...

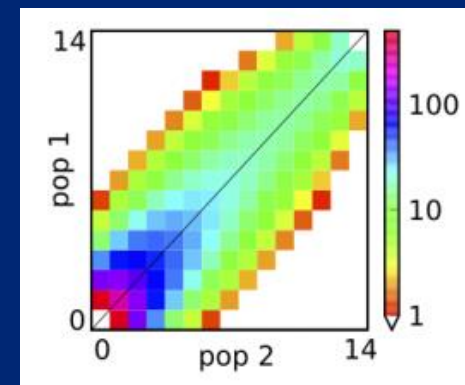
SFS attendu



Données génomiques

	SNPs							
Samples	T	T	G	A	T	A	G	T
	T	T	G	A	T	C	G	T
	T	T	C	A	C	C	A	T
	A	G	G	C	T	C	G	T
	A	G	G	C	T	C	G	A
		3	2	1	2	1	1	1


SFS observé



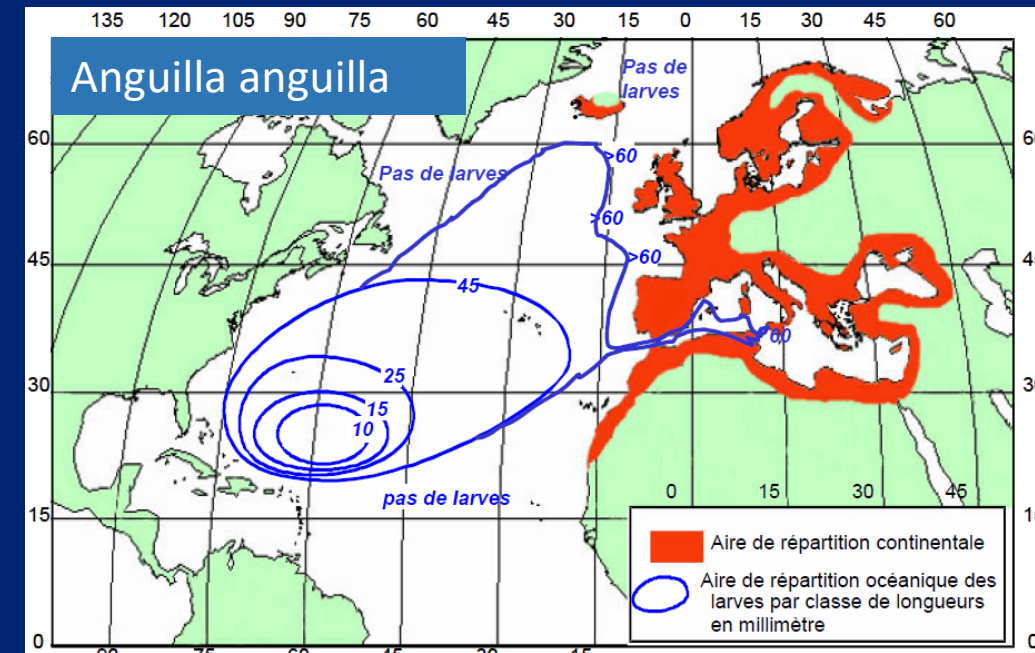
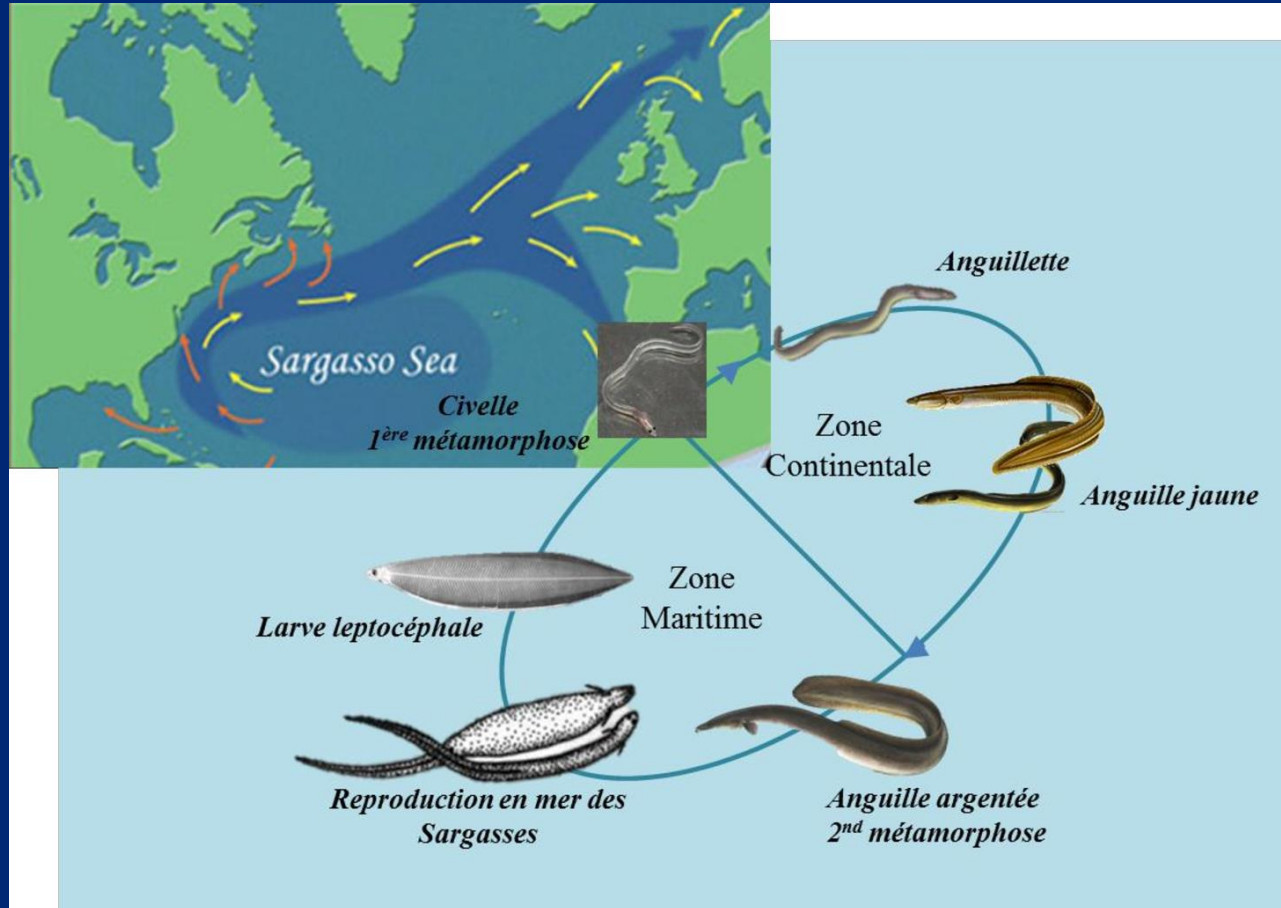
Fonction de
vraisemblance
poissonienne
 $L(\text{données} | \text{modèle})$

Exemple: estimation de l'histoire de la spéciation entre anguille européenne (*Anguilla anguilla*) et américaine (*A. rostrata*)

Speciation history of European (*Anguilla anguilla*) and American eel (*A. rostrata*), analysed using genomic data

Natacha Nikolic¹ | Shenglin Liu² | Magnus W. Jacobsen² | Bjarni Jónsson³ |
Louis Bernatchez⁴ | Pierre-Alexandre Gagnaire⁵ | Michael M. Hansen² 

Anguilla rostrata



Exemple: estimation de l'histoire de la spéciation entre anguille européenne (*Anguilla anguilla*) et américaine (*A. rostrata*)

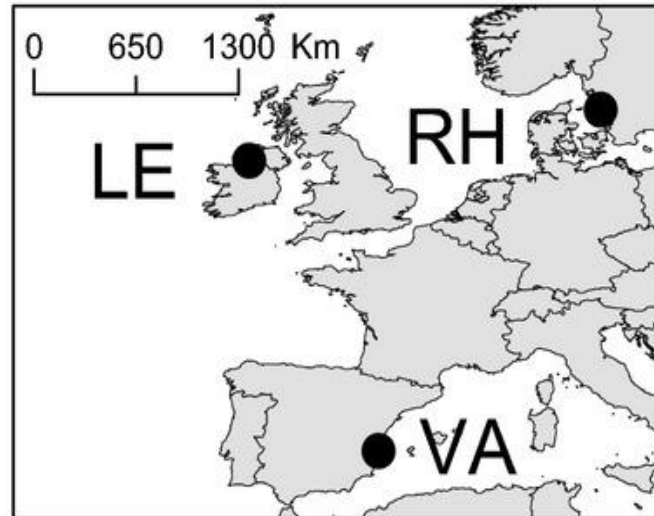
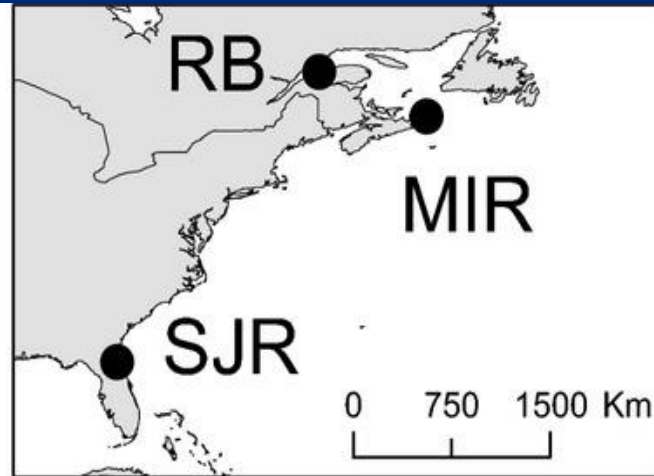
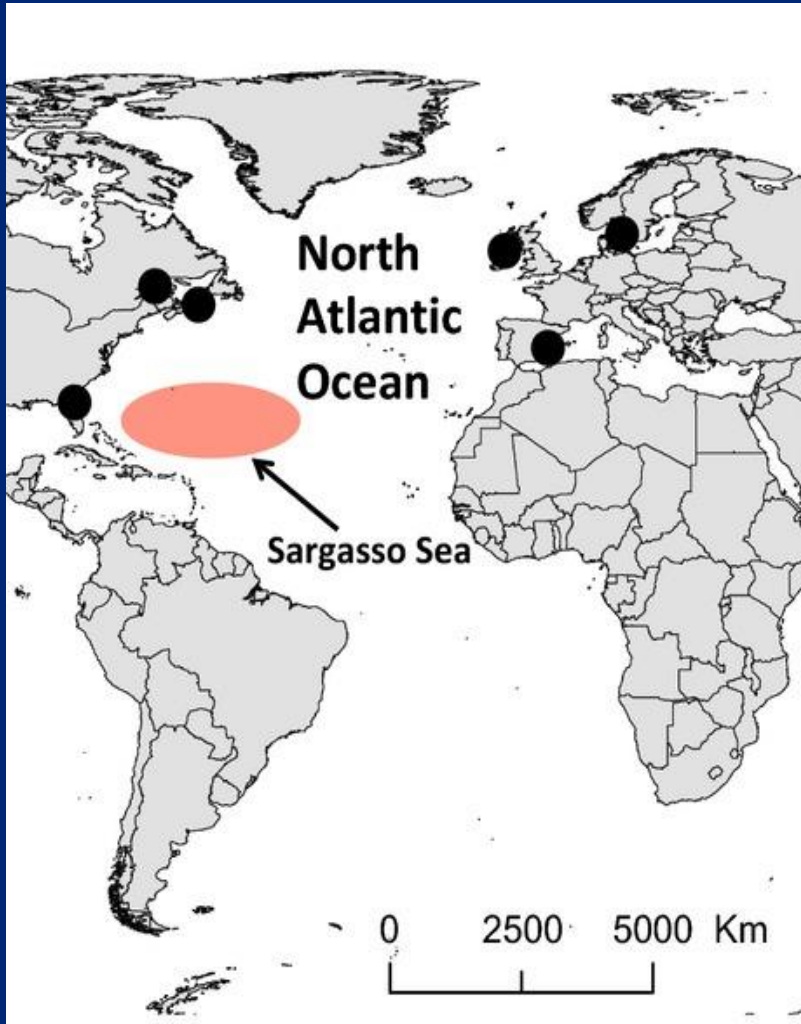
Connaissances préliminaires:

- Hybridation possible mais peu fréquente, suggérant isolement reproducteur entre les deux espèces
- Deux scénarios de spéciation ont été proposés:
 - Allopatrique: disjonction des aires de reproduction pendant les périodes glaciaires
 - Sympatrique: la colonisation de nouveaux habitats sur l'autre continent aurait promu la divergence

Questions spécifiques à l'étude

- a) Quelle est l'histoire de la spéciation des anguilles européennes et américaines, en particulier quelle dynamique temporelle de flux de gènes au cours de la divergence des espèces ?
- b) Toutes les régions génomiques sont-elles également affectées par les flux de gènes interspécifiques ?
- c) Comment les anguilles de l'Atlantique en général peuvent-elles nous informer sur les processus de spéciation dans les environnements océaniques ?

Echantillonnage et données génomiques



30 American and 30 European eels
6 populations
328 300 SNPs (RAD-seq)

Scénarios démographiques étudiés

Basic models	SI	Strict isolation
	IM	Isolation with migration
	AM	Ancient migration
	SC	Secondary contact
	AMSC	Ancient migration followed by strict isolation and secondary contact
Derived models	IM2M	IM with heterogeneous migration rates across genome
	AM2M	AM with heterogeneous migration rates across genome
	SC2M	SC with heterogeneous migration rates across genome
	AMSC2M	AM AMSC with heterogeneous migration rates across genome
	SCG	SC with exponential growth
	SC2MG	SC2M with exponential growth
	AMSC2	MG AMSC2M with exponential growth

Résultats

Observed
JAFS

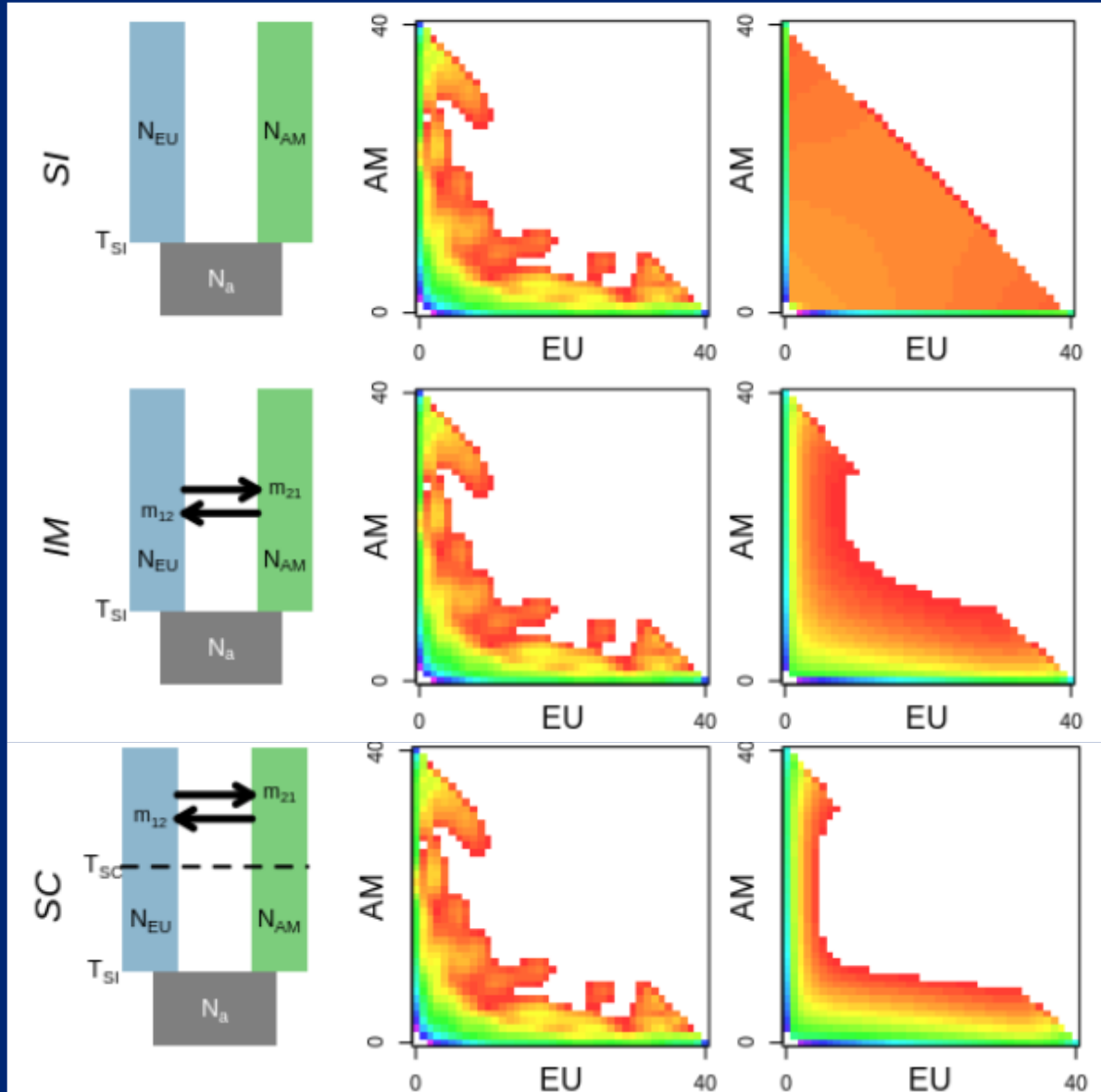
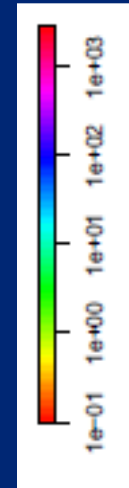
Expected
JAFS

Nombre de SNPs

Strict isolation

Isolation with migration

Secondary contact

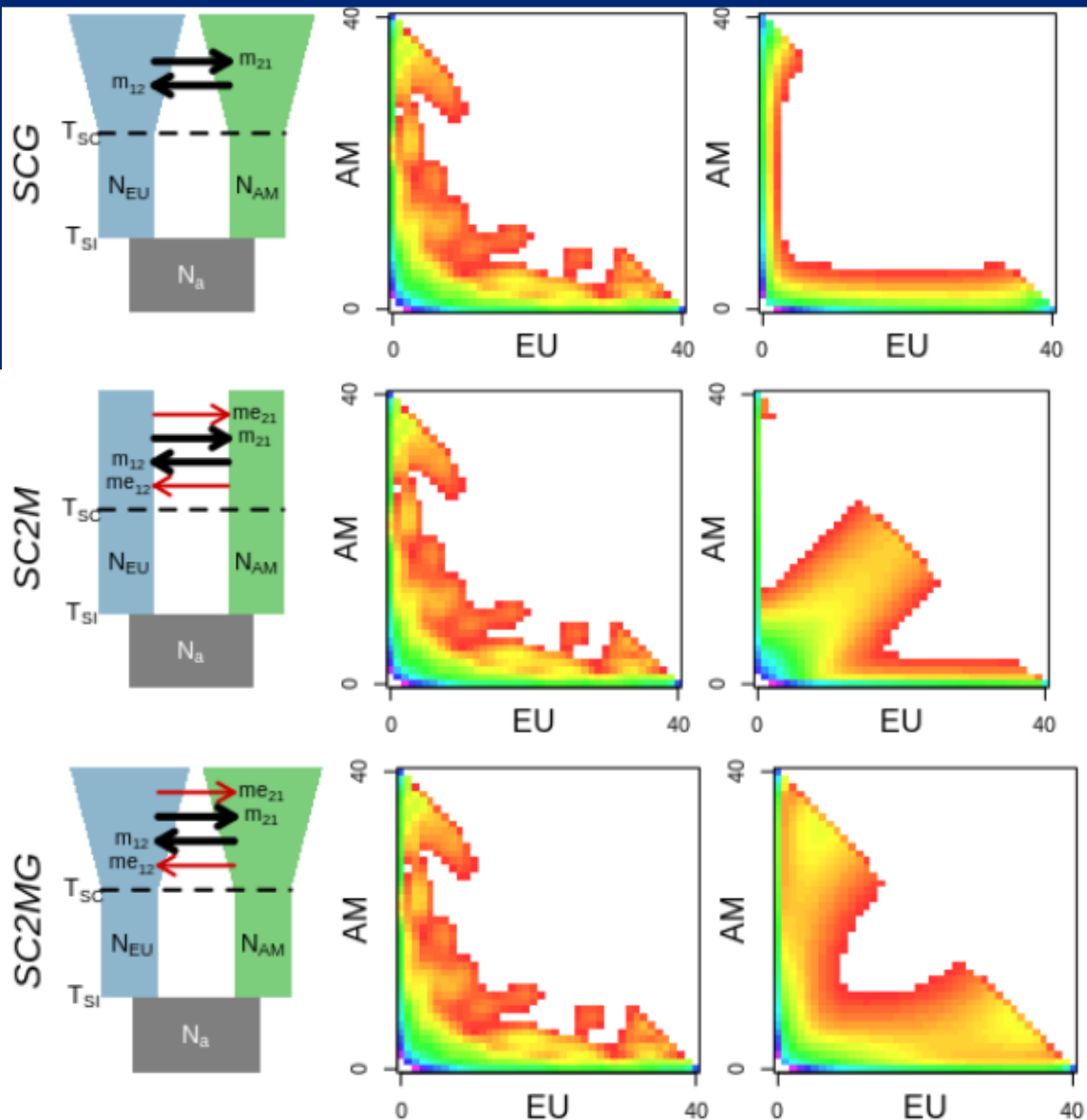
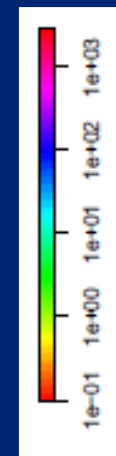


Résultats

Observed
JSFS

Expected
JSFS

Nombre de SNPs



SC with exponential growth

SC with heterogeneous migration rates
across genome

SC2M with exponential growth

Conclusions

a) Quelle est l'histoire de la spéciation des anguilles européennes et américaines ?

➤ Le meilleur scénario, qui est celui d'un contact secondaire exclue la spéciation sympatrique, et pointe vers une disjonction des sites de reproduction pendant le pléistocène (Heinrich)

b) Toutes les régions génomiques sont-elles également affectées par les flux de gènes interspécifiques ?

➤ 64% du génome montre une différenciation forte entre les deux espèces, mais d'autres zones plus sujettes à l'admixture

c) Comment les anguilles de l'Atlantique en général peuvent-elles nous informer sur les processus de spéciation dans les environnements océaniques ?

➤ La spéciation allopatrique paraît bien plus probable en générale que la spéciation sympatrique

➤ Mais chez les espèces océaniques, les grandes tailles de population et les stades larvaires à forte capacité de migration sont supposés favoriser la spéciation sympatrique

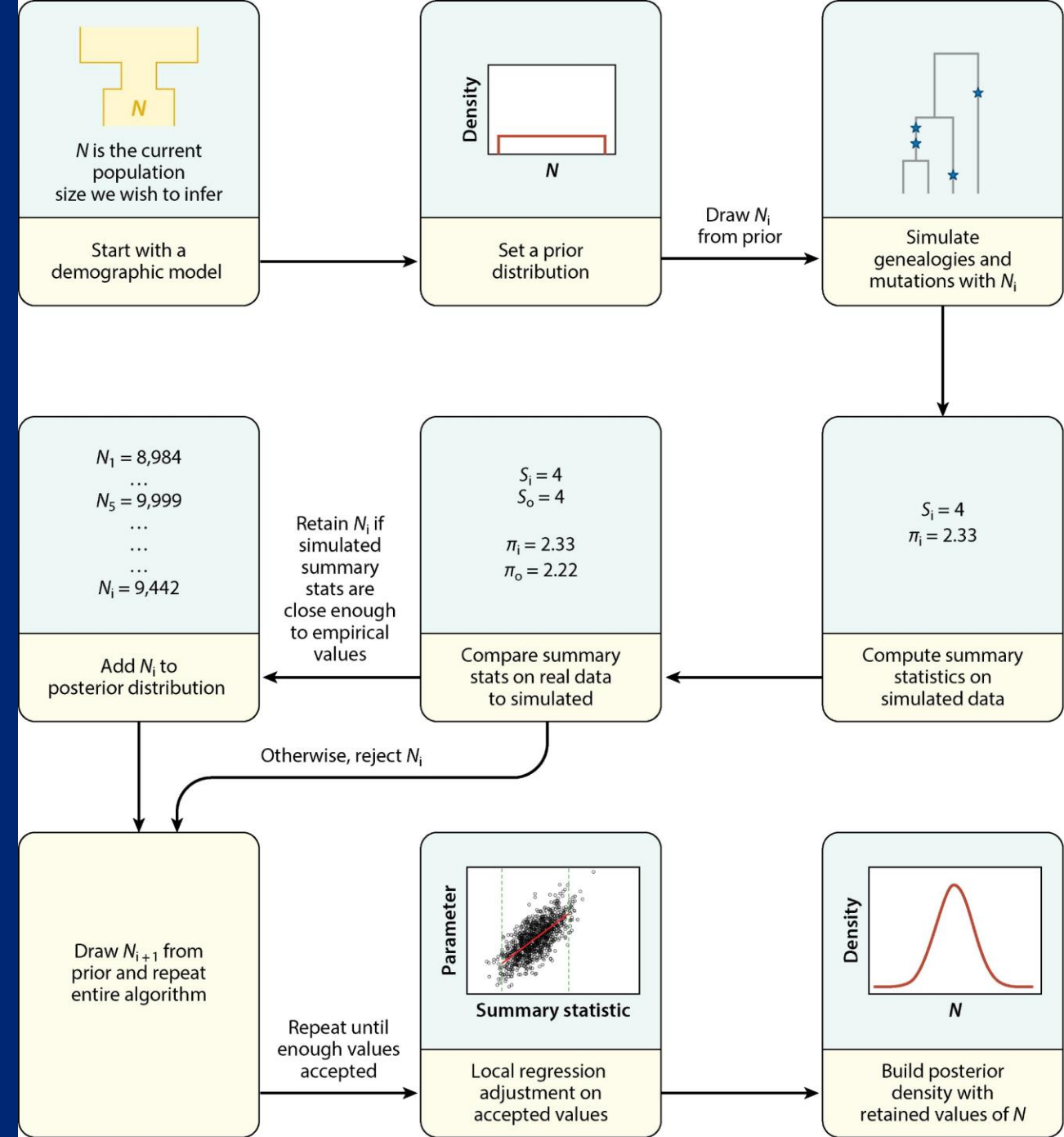
➤ Pas trop supporté par les études génomiques récentes

Méthodes courantes d'inférence démographique

- Méthodes basées sur le SFS (Site Frequency Spectrum) et la théorie du coalescent
- (Méthodes basées sur le coalescent séquentiel markovien par paire)
- Méthodes basées sur l'**ABC** (Approximate Bayesian Computation) avec des statistiques résumées

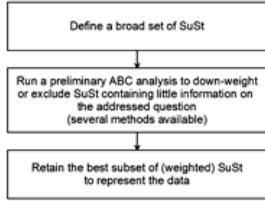
Méthodes basées sur l'ABC

- Objectif = Inférer les paramètres démographiques à partir de plusieurs statistiques résumées, telles que le nombre moyen de différences par paire entre séquences (π), le F_{ST} ou la décroissance du LD
- Stat résumées calculées (1) sur données réelles et (2) sur données simulées
- Contourne la formulation explicite de la vraisemblance \rightarrow permet d'explorer des modèles démographiques complexe

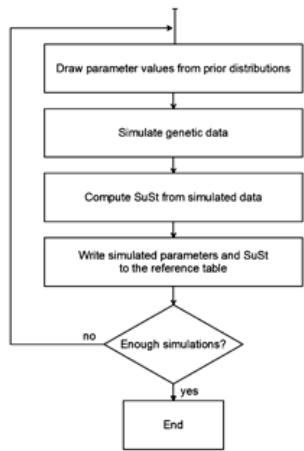


ABC in 9 steps

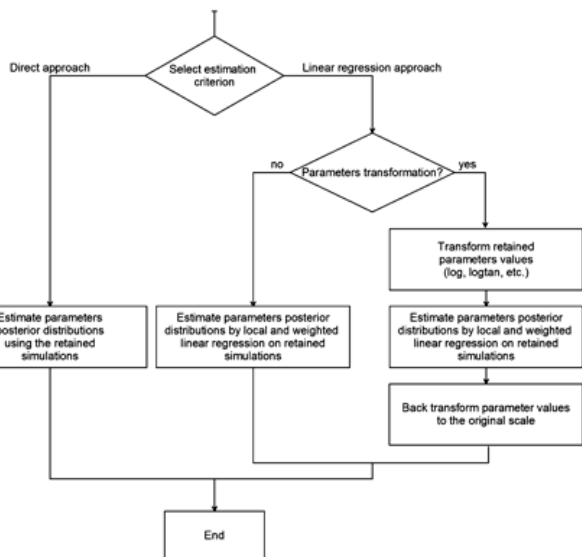
STEP 3 - Choosing the summary statistics (SuSt)



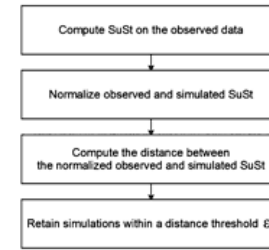
STEP 4 - Simulating the model(s)



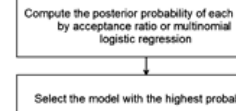
STEP 8 - Parameters estimation



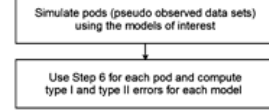
STEP 5 - Filtering the simulations



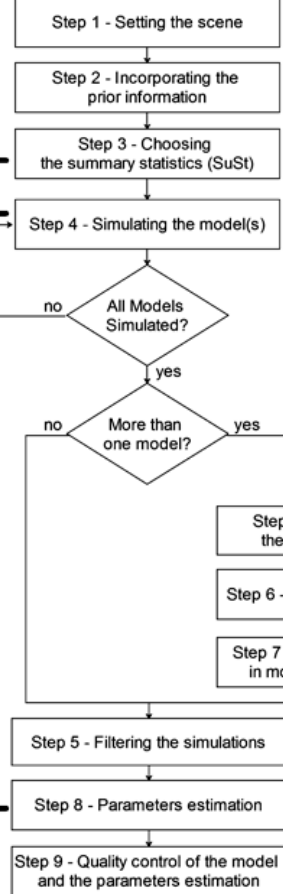
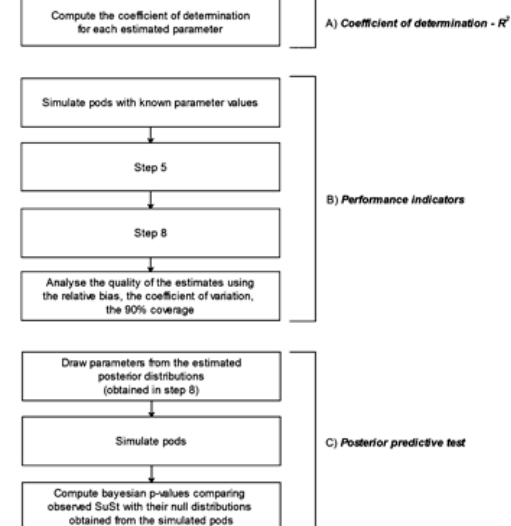
STEP 6 - Model selection



STEP 7 - Quality control in model selection



STEP 9 - Quality control of the model and the parameters estimation



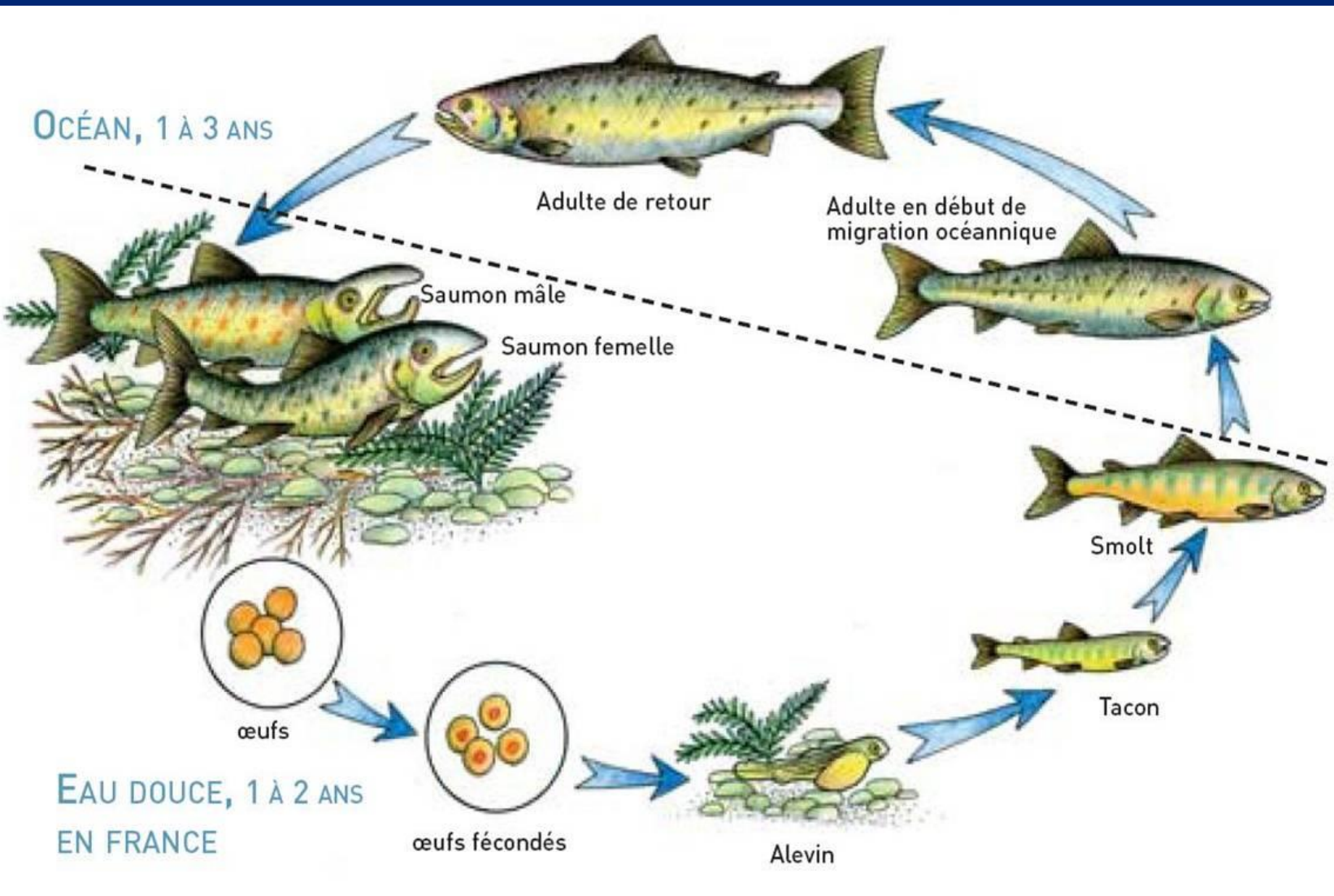
Exemple: estimation de l'histoire démographique du saumon Atlantique (Rougemont & Bernatchez, Evolution 2018)

The demographic history of Atlantic salmon (*Salmo salar*) across its distribution range reconstructed from approximate Bayesian computations*

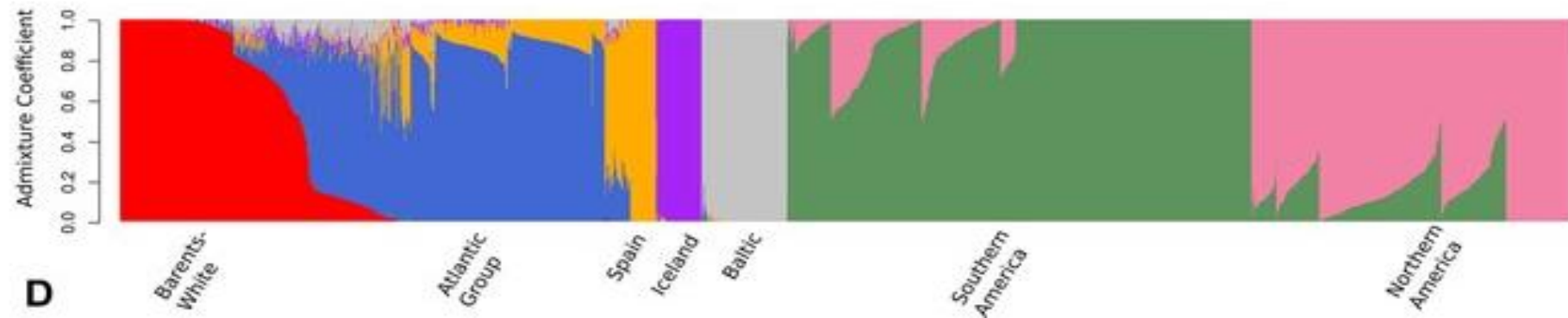
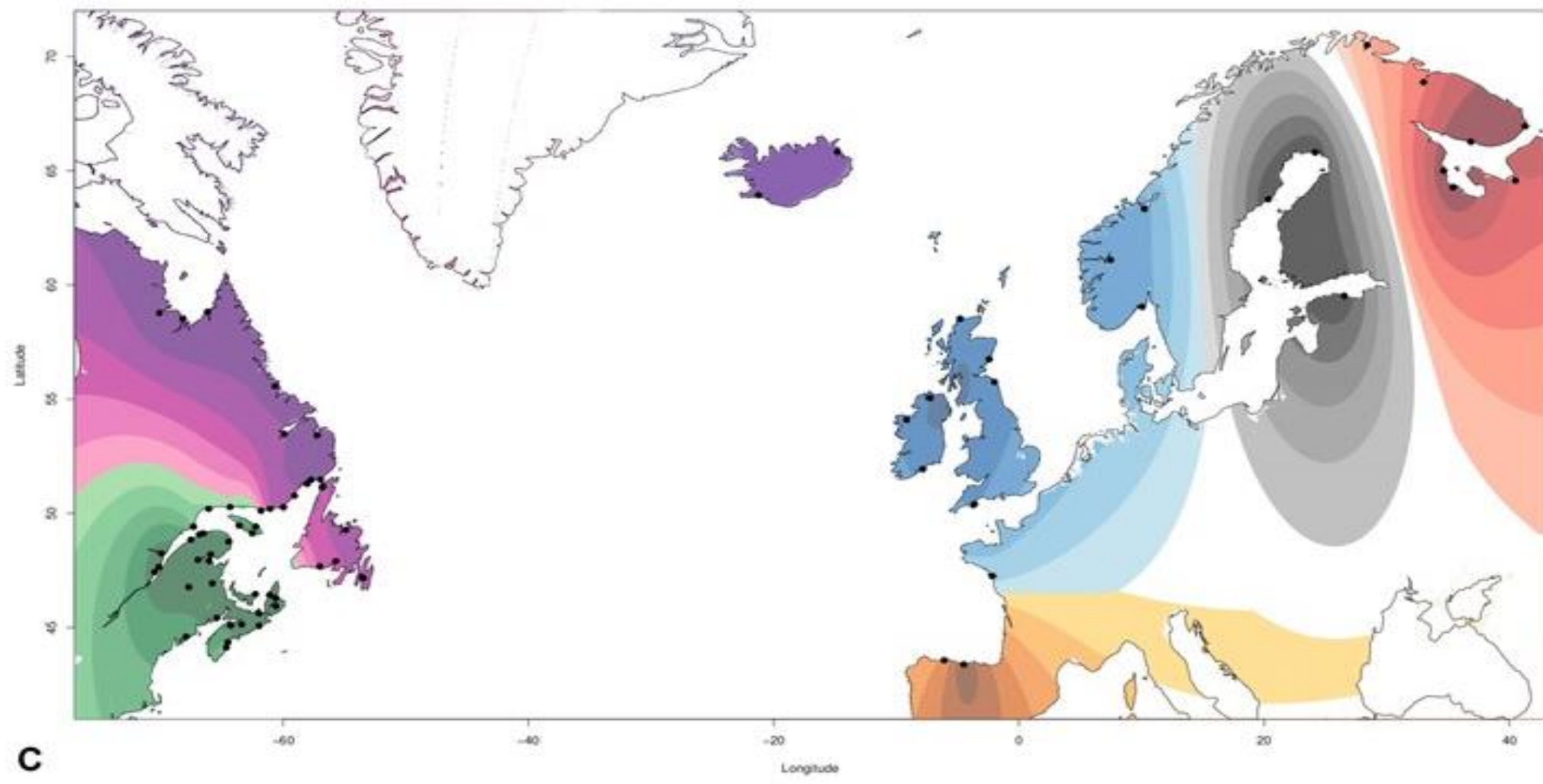
Quentin Rougemont^{1,2} and Louis Bernatchez¹



Cycle de vie: anadromie



- Homing favoriserait adaptation locale et forte structure de populations
- Structure plus marquée pour les populations européennes, que pour les pop américaines
- Forte divergence entre les populations européennes et américaines, avec cependant possibles contacts secondaires



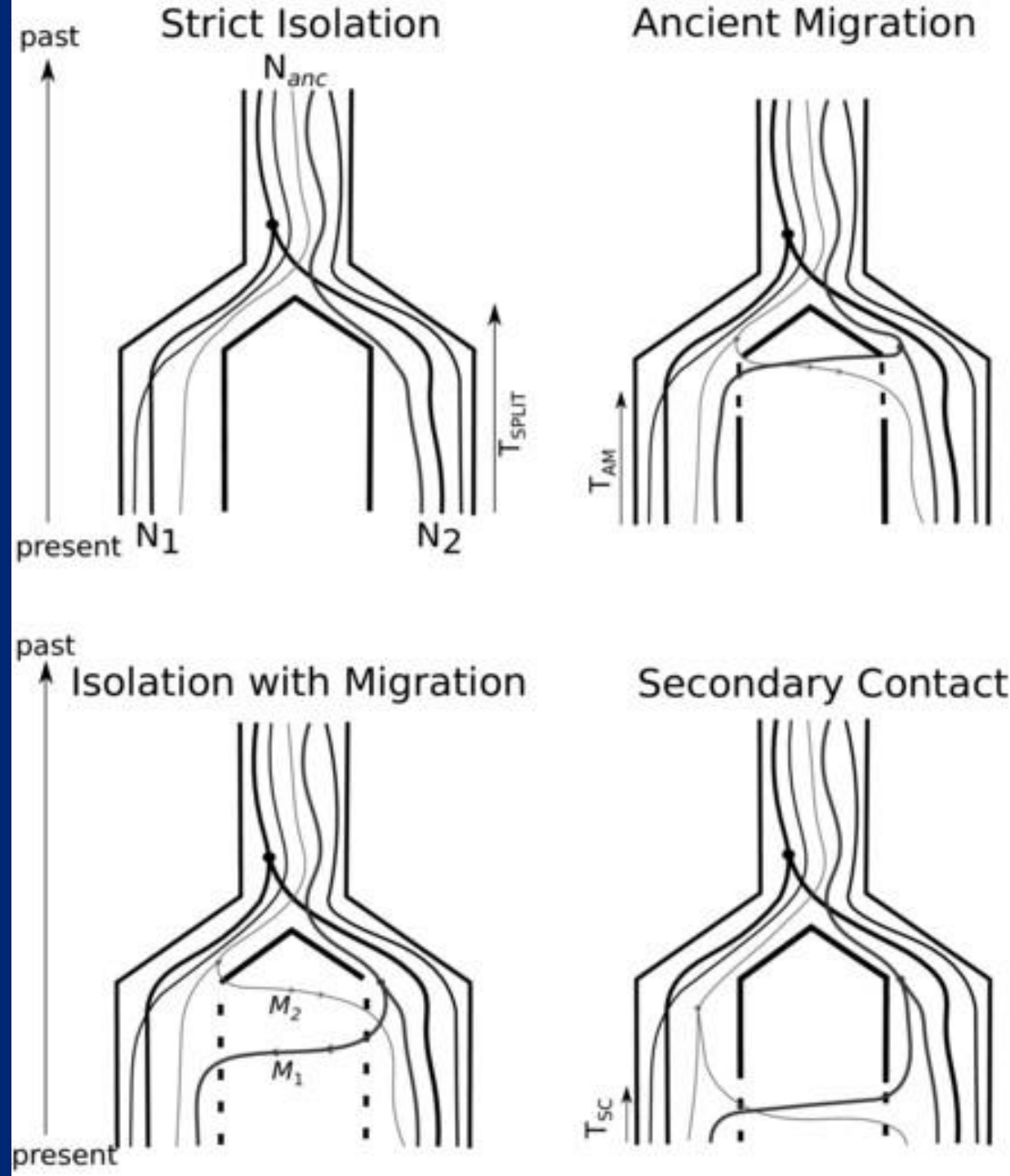
Scénarios démographiques étudiés

T_{split} : # de générations de divergence
 N_{anc} , N_1 , N_2 : taille efficace de la population ancestrale, et des deux pop filles.

M_1 , M_2 : taux de migration efficace par génération.

T_{am} : # de générations depuis que les deux populations ont divergé sans flux génétique.

T_{sc} : # de générations depuis que les populations ont commencé à échanger des allèles (contact secondaire) après une période d'isolement.



Prise en compte explicite de l'hétérogénéité de la taille efficace N_e et des flux de gènes m le long du génome

Modèle de divergence adaptative avec flux de gènes: les zones génomiques soumises à sélection (sélection positive ou purifiante) vont se différencier génétiquement rapidement entre lignées, alors que les flux de gènes érodent la différenciation dans les autres zones du génome

(La sélection contre l'introgession réduit localement les flux de gènes entre populations, tandis que la sélection purifiante renforce l'effet de la dérive en réduisant localement N_e)

Problème: la divergence due à l'isolement (allopatrie) suivi d'un contact secondaire peut produire les mêmes signatures génomiques (hétérogénéité de N_e et m le long du génome)

Solution : tenir compte explicitement de cette hétérogénéité

<i>NhomoMhomo</i> : N_e et m homogènes	<i>NhomoMhetero</i>: N_e homogène, m hétérogène
<i>NheteroMhomo</i> : N_e hétérogène, m homogène	<i>NheteroMhetero</i> : N_e et m hétérogènes

Résultats

	All models				Best model	
	P(SI)	P(AM)	P(IM)	P(SC)	P(SC homo)	P(SC hetero)
Within America	0.000	0.000	0.013	0.984	0.0040	0.9500
Between Continent	0.000	0.000	0.005	0.993	0.0115	0.9760
Within Europe	0.000	0.003	0.024	0.967	0.0385	0.8615

- Scénario de contacts secondaires privilégié à la fois entre et au sein des populations américaines et européennes de saumon atlantique → histoire démographique réticulée.
- Bien que la sélection n'ait pas été explicitement simulée dans cette étude, les meilleures performances des modèles avec hétérogénéité de N_e ou des flux de gènes suggèrent l'action de la sélection (positive ou purifiante)

L'essentiel à retenir

- Les données génomiques contiennent **une mine d'informations** sur l'histoire démographique des populations dont elles sont issues.
- De nombreuses **procédures d'inférence statistique sophistiquées** ont été mises au point pour déduire l'histoire démographique des populations à partir de ce type de données génomiques
- Des progrès sont encore à faire pour estimer les variations dans le temps de paramètres comme la taille efficace à une échelle contemporaine, et utiliser ces méthodes pour guider la gestion ou la conservation

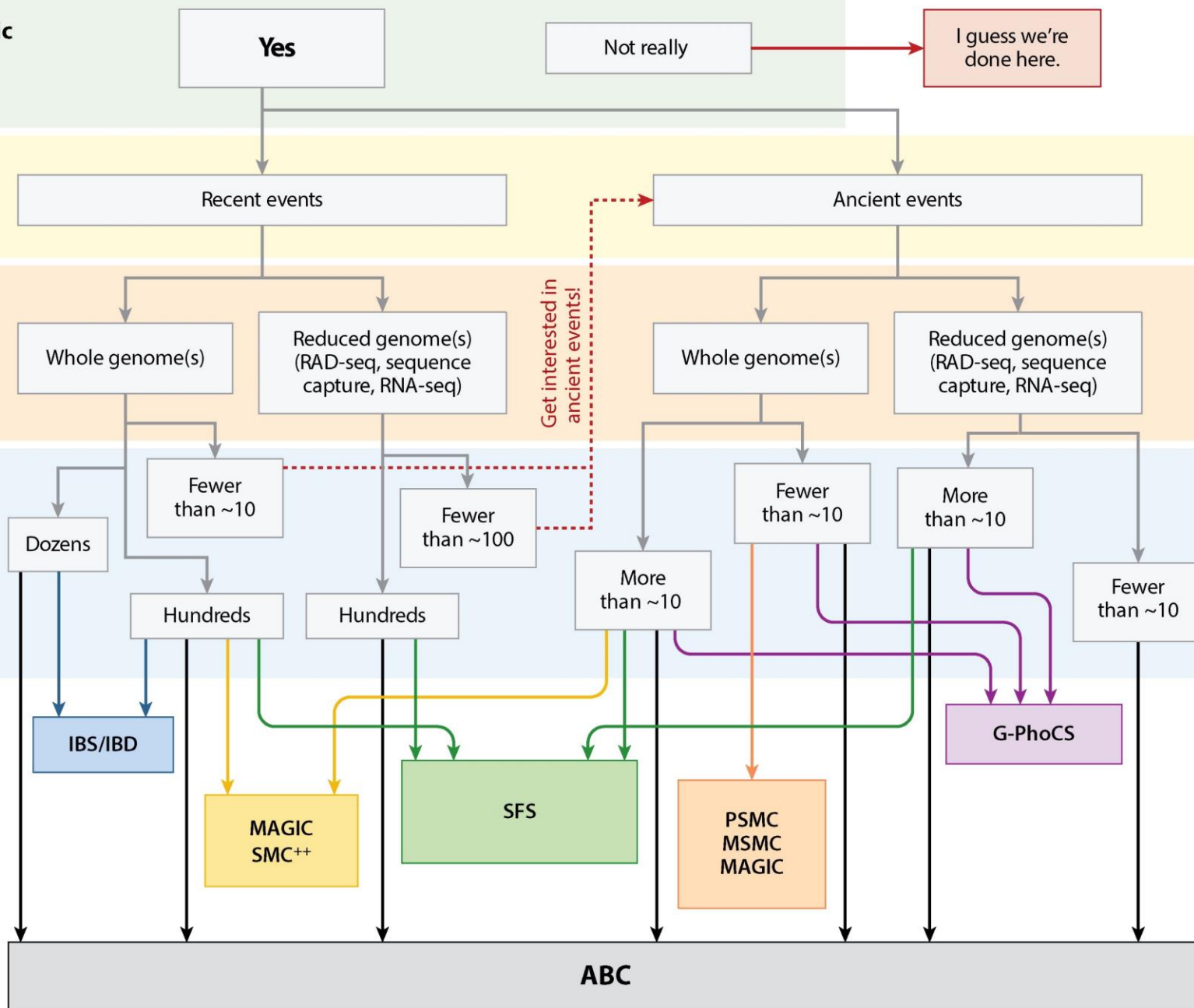
So you want to do demographic inference from genomic data?

What time period are you interested in?

What kind of data do you have?

How many individuals have been sequenced?

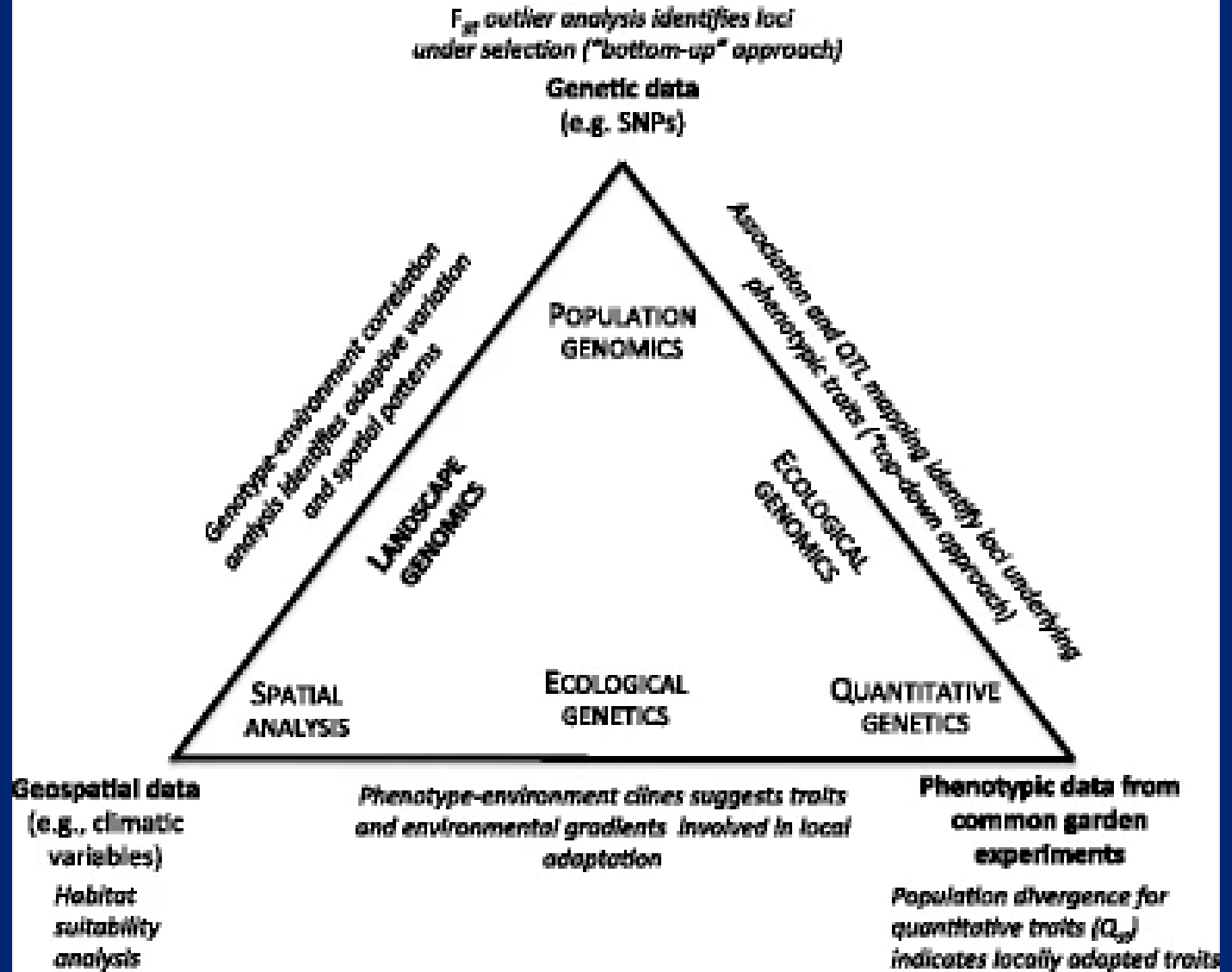
Consider these methods:



Plan du cours

- I. Introduction
- II. Rappels de base en génétique des populations
- III. Estimation de la démographie historique : principes fondamentaux, méthodes courantes, étude de cas
- IV. Estimation de l'histoire sélective et de l'adaptation génétique: principes fondamentaux, méthodes courantes, étude de cas
- V. Vers l'inférence conjoint de processus démographiques et sélectifs

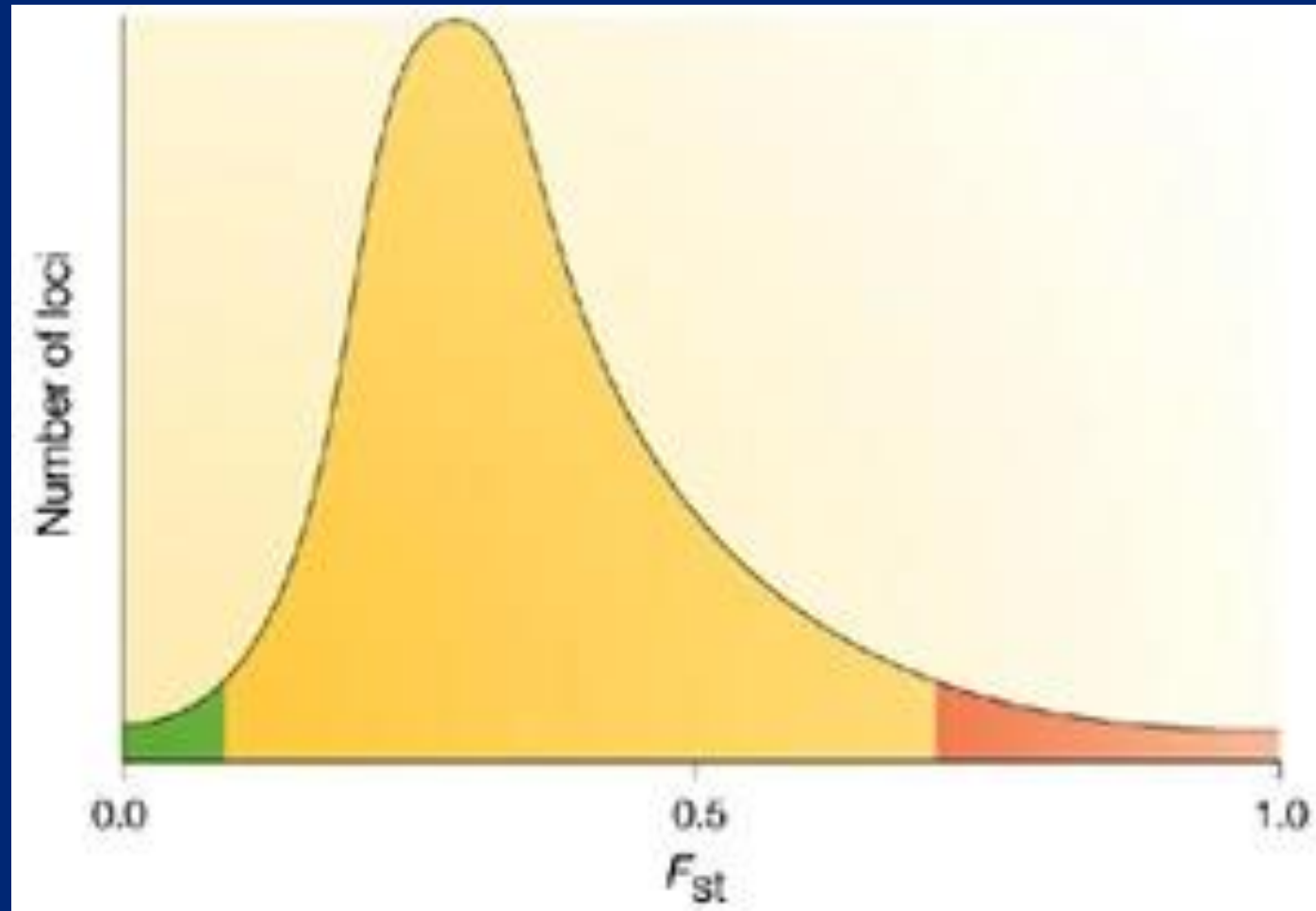
Les différentes approches de la variation génétique adaptative



Méthodes courantes d'inférence de l'histoire sélective et de l'adaptation locale

- Méthodes basées sur la différenciation génétique entre populations (F_{ST}): « genome scan » ou « F_{ST} outliers tests »
- Méthodes basées sur les corrélations génotype-environnement
- Méthodes basées sur les associations génotype-phénotype

Méthodes « genome scan » ou « F_{ST} outliers tests »



Les loci sous sélection divergente/balancée montrent une différenciation plus forte/faible qu'attendue

Exemple: estimation des bases génomiques de l'adaptation à un parasite chez le saumon atlantique (Zueva et al. Marine Genomics 2018)

Genomic signatures of parasite-driven natural selection in north European Atlantic salmon (*Salmo salar*)

Ksenia J. Zueva^a, Jaakko Lumme^b, Alexey E. Veselov^c, Matthew P. Kent^d, Craig R. Primmer^{a,e,f,*}



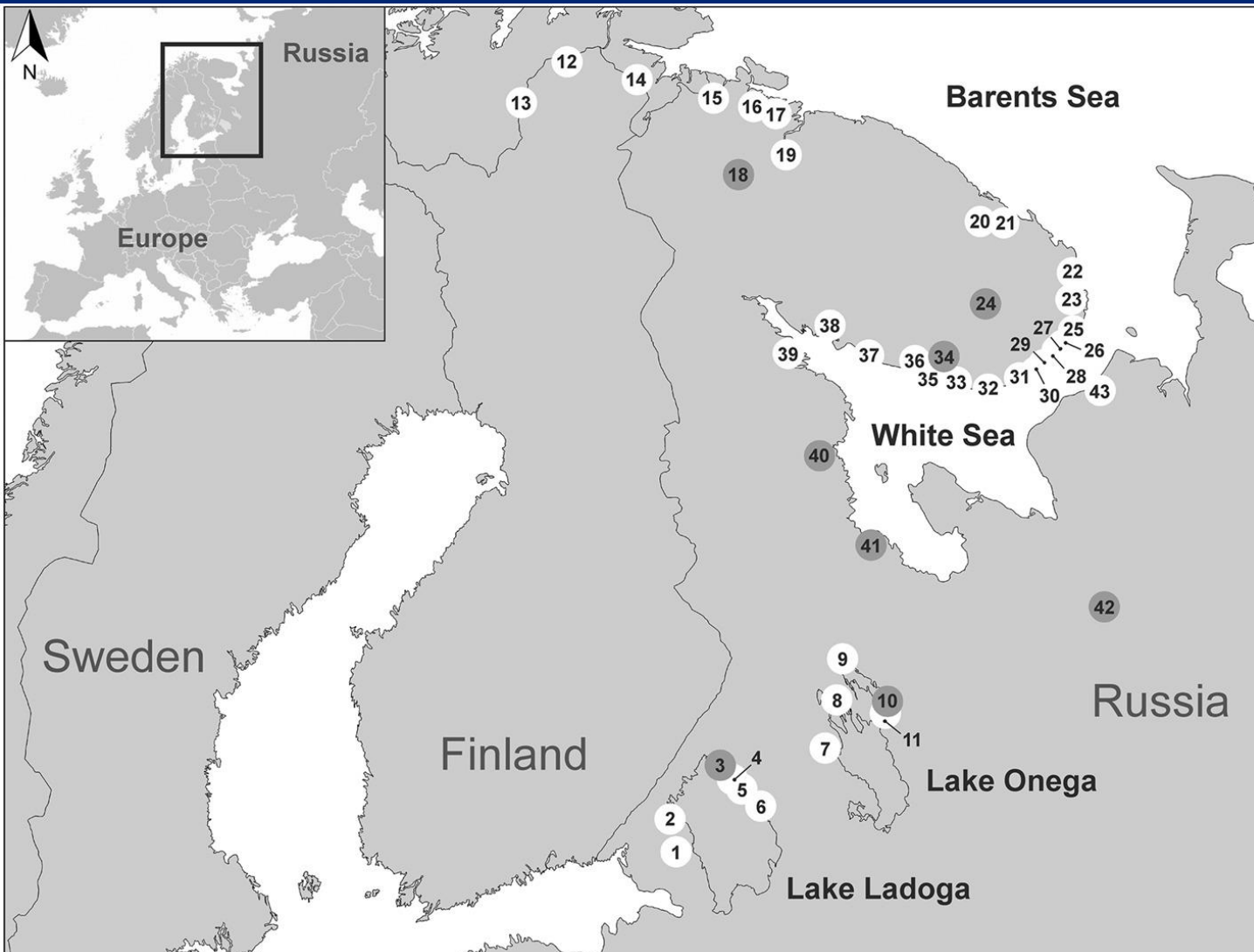
Gyrodactylus salaris, la Gyrodactylose



- Impliqué dans la diminution des populations de saumons atlantique dans les fjords norvégiens
- Parasite invisible à l'œil nu, vivipare
- Blessures épidermiques favorisant les infections secondaires

Echantillonnage et génotypage

43 populations, 220,000 SNPs

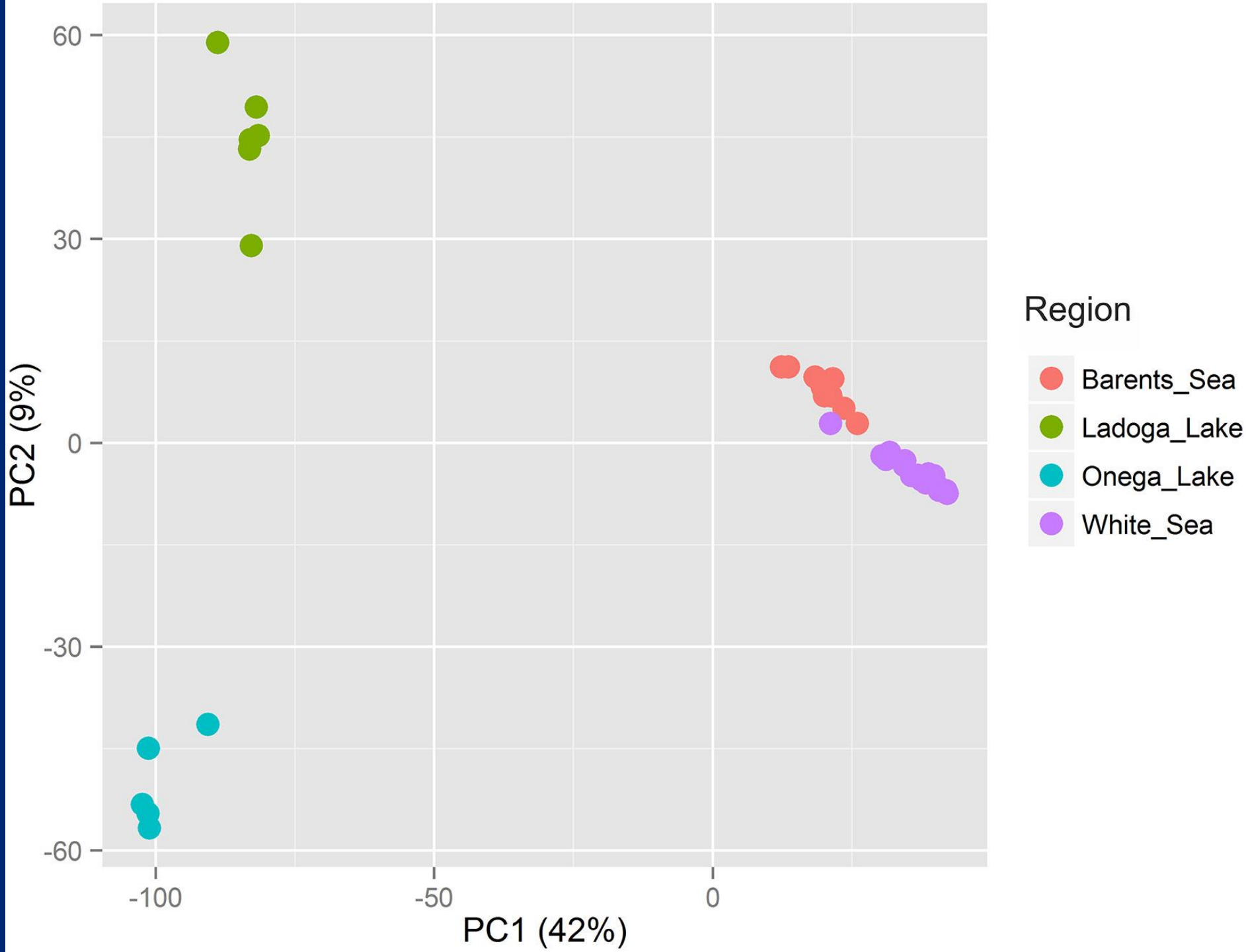


Mer de Barents, mer blanche:
populations sensibles (95%
mortalité)

Lacs Onega et Ladoga:
populations résistantes

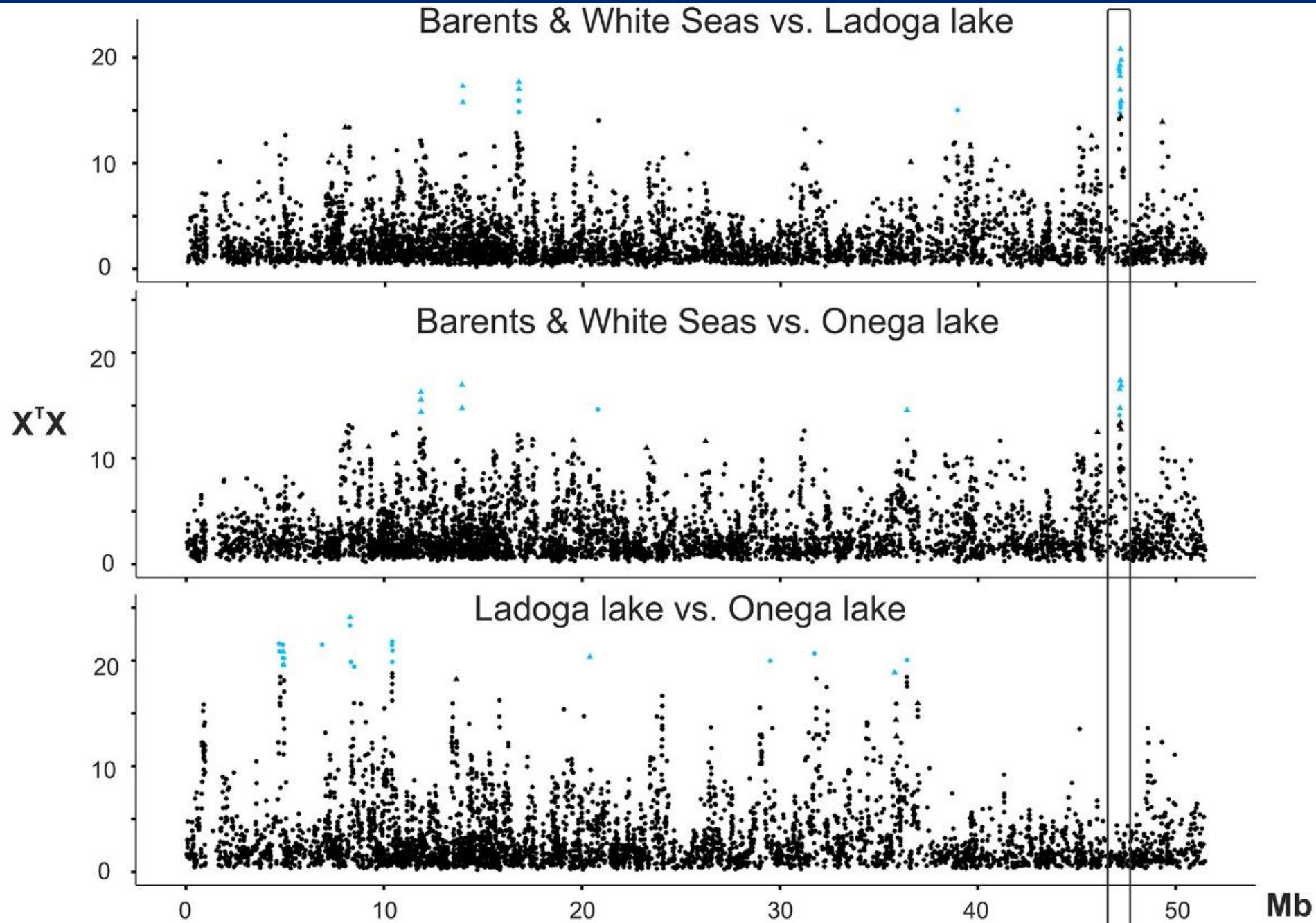
Résultats

Forte structure,
source potentielle
de faux positifs



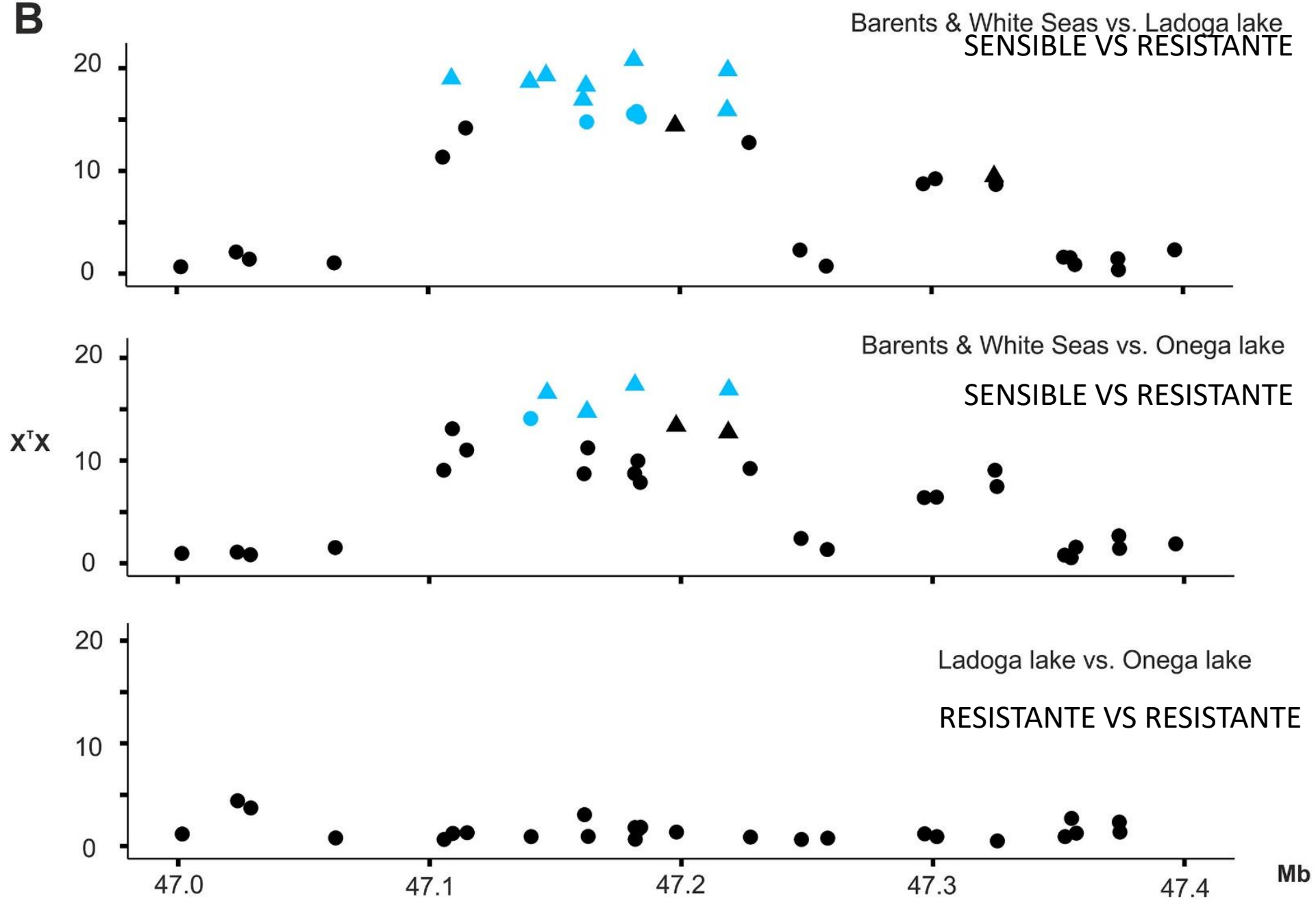
Résultats

Statistique de différenciation prenant en compte la structure

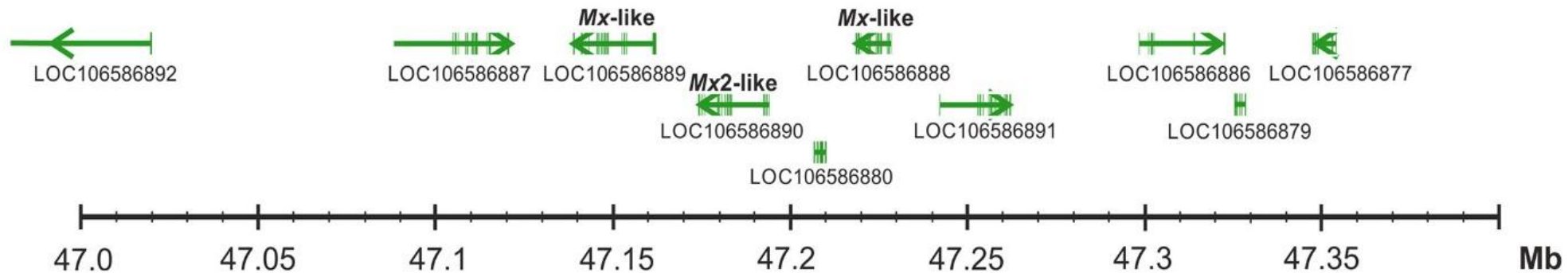


Résultats

B



Résultats et conclusions



- 57 gènes candidats potentiellement soumis à une sélection positive ont été identifiés, dont 50 ont fait l'objet d'une annotation fonctionnelle.
- Plusieurs termes GO liés à l'immunité innée et acquise ont été enrichis parmi les 57 gènes candidats.
- Source de résistance potentielle aux parasites, évitant les traitements chimique pour leur éradication

Méthodes courantes d'inférence de l'histoire sélective et de l'adaptation locale

- Méthodes basées sur la différenciation génétique entre populations (F_{ST}): « genome scan » ou « F_{ST} outliers tests »
- Méthodes basées sur les corrélations génotype-environnement (GEA)
- Méthodes basées sur les associations génotype-phénotype

GEA

Les loci sous
sélection
divergente
montrent une
corrélations avec
les variables
environnementales
pilotant
l'adaptation

Molecular Ecology (2015) doi: 10.1111/mec.13322

INVITED REVIEWS AND SYNTHESSES

A practical guide to environmental association analysis in landscape genomics

CHRISTIAN RELLSTAB,* FELIX GUGERLI,* ANDREW J. ECKERT,† ANGELA M. HANCOCK‡ and ROLF HOLDEREGGER*§

Landscape Genomics: Understanding Relationships Between Environmental Heterogeneity and Genomic Characteristics of Populations

Niko Balkenhol, Rachael Y. Dudaniec, Konstantin V. Krutovsky, Jeremy S. Johnson, David M. Cairns, Gernot Segelbacher, Kimberly A. Selkoe, Sophie von der Heyden, Ian J. Wang, Oliver Selman and Stéphane Joost

MOLECULAR ECOLOGY

ORIGINAL ARTICLE

Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations

Brenna R. Forester✉, Jesse R. Lasky, Helene H. Wagner, Dean L. Urban

INVITED REVIEWS AND SYNTHESSES WILEY MOLECULAR ECOLOGY

The search for loci under selection: trends, biases and progress

Collin W. Ahrens¹ | Paul D. Rymer¹ | Adam Stow² | Jason Bra
Kate D. L. Umbers^{1,5} | Rachael Y. Dudaniec²

¹Max Planck Institute for the Environment

VOL. 188, NO. 4 THE AMERICAN NATURALIST OCTOBER 2016

SYNTHESIS

Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions

Sean Hoban,^{1,†} Joanna L. Kelley,^{2,*,†} Katie E. Lotterhos,^{3,†} Michael F. Antolin,⁴ Gideon Bradburd,⁵ David B. Lowry,⁶ Mary L. Poss,⁷ Laura K. Reed,⁸ Andrew Storfer,² and Michael C. Whitlock⁹

Exple: bases génomiques de la divergence entre écotypes d'eau salée et d'eau douce chez l'anguille américaine (Pavey et al. Current biol. 2015)

RAD Sequencing Highlights Polygenic Discrimination of Habitat Ecotypes in the Panmictic American Eel

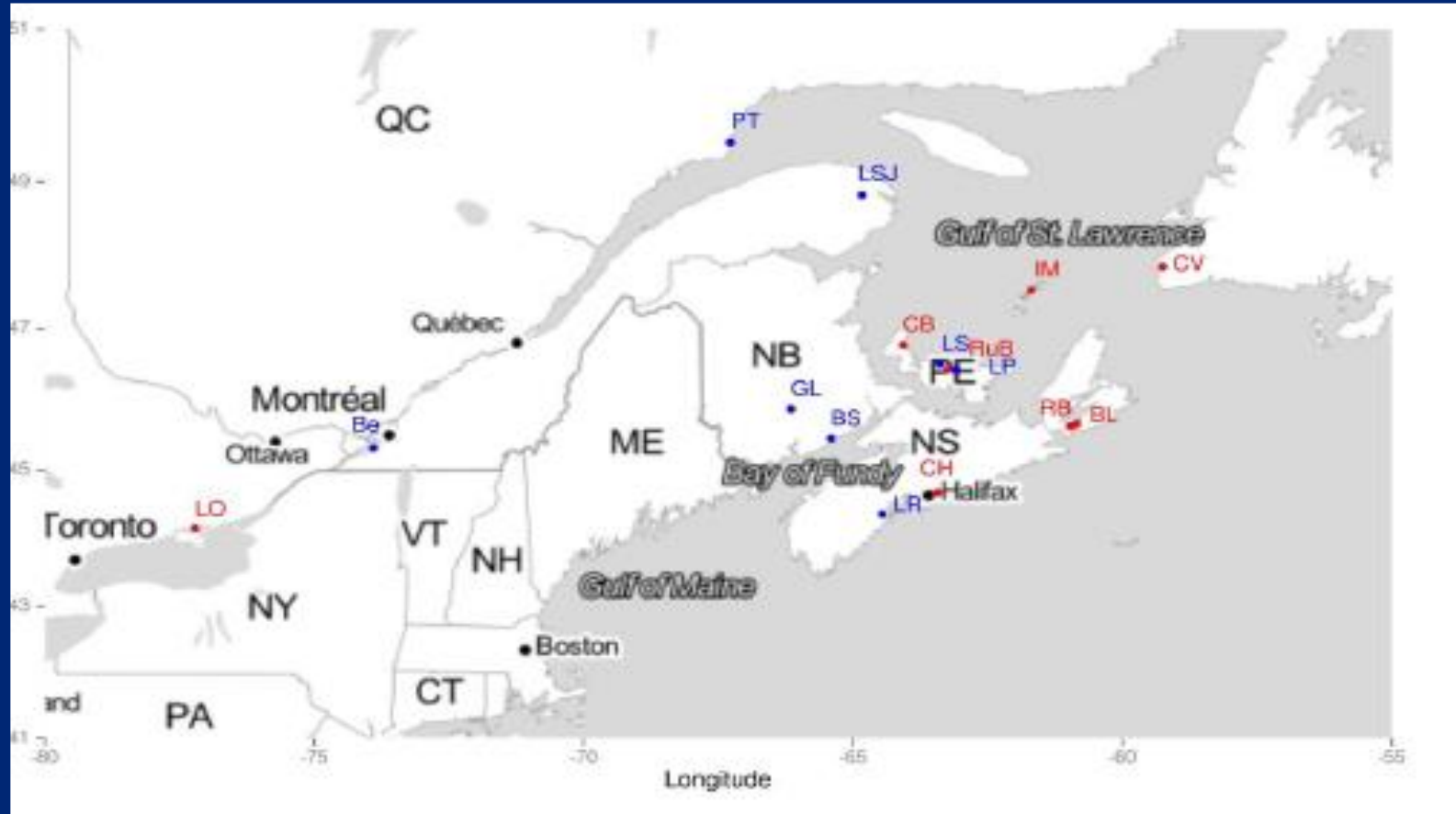
Scott A. Pavey,^{1,*} Jérémy Gaudin,¹ Eric Normandeau,¹ Mélanie Dionne,² Martin Castonguay,³ Céline Audet,⁴ and Louis Bernatchez¹



écotype à croissance lente et à maturation tardive (>20 ans); caractérise le lac Ontario et le haut Saint-Laurent.

écotype des eaux saumâtres/salées, à croissance rapide et maturation précoce (environ 5 ans).

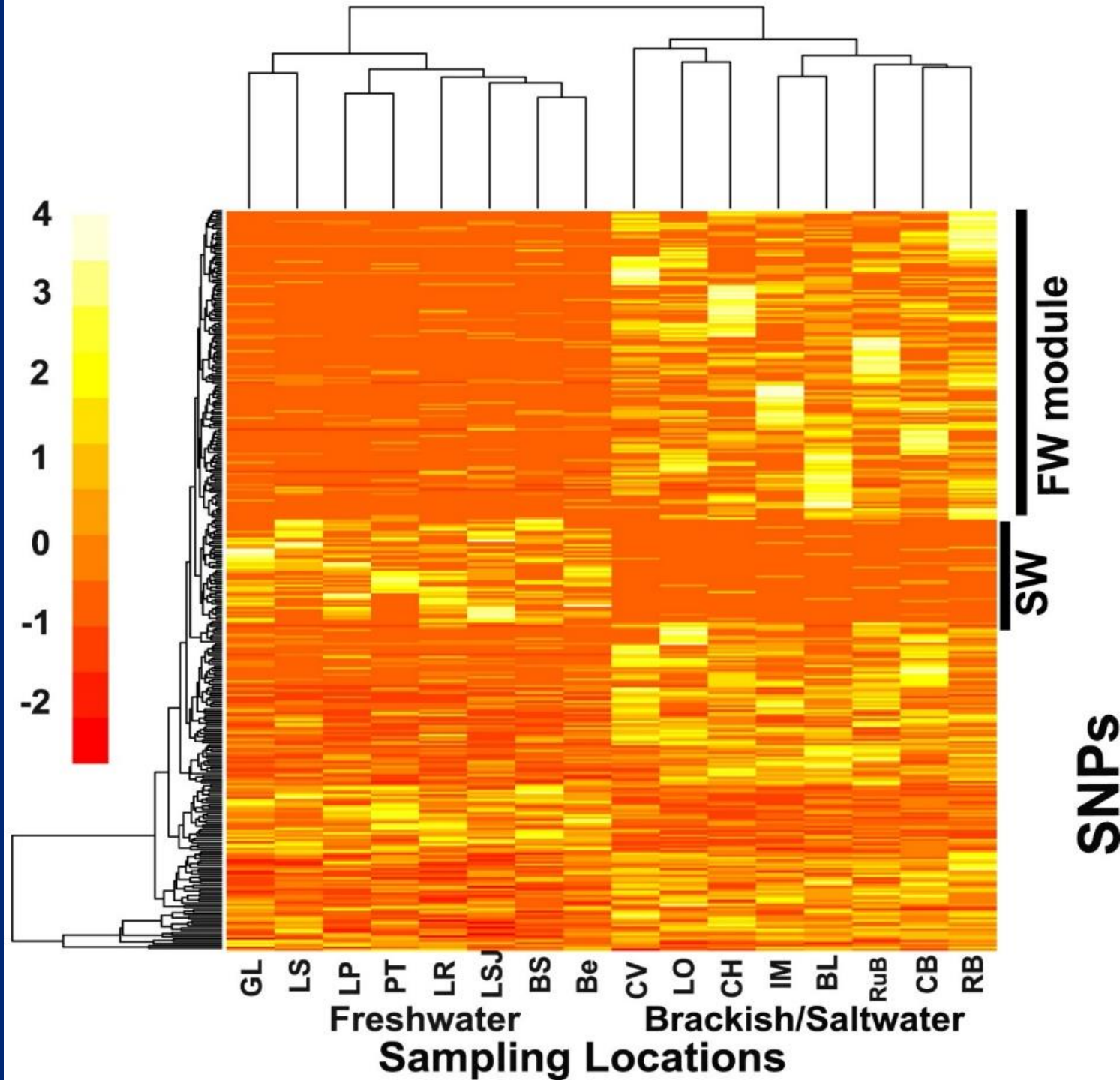
Echantillonnage/génotypage



- 379 anguilles américaines (sarde jaune/argenté)
- 16 populations (8 eaux douces / 8 eaux salées ou saumâtres, 21-24 individus/pop)
- Depuis St. Lawrence River jusqu'à la côte Atlantique
- 42,424 SNPs

Résultats

- 331 locus/ 40k associés à la divergence entre écotypes
- La moitié des marqueurs sont presque fixes dans un écotype et comparativement variables dans l'autre.
- Ces 331 SNP sont associés à 101 gènes impliqué dans le développement vasculaire et morphologique, la régulation du calcium, les facteurs de croissance et de transcription et les récepteurs olfactifs.

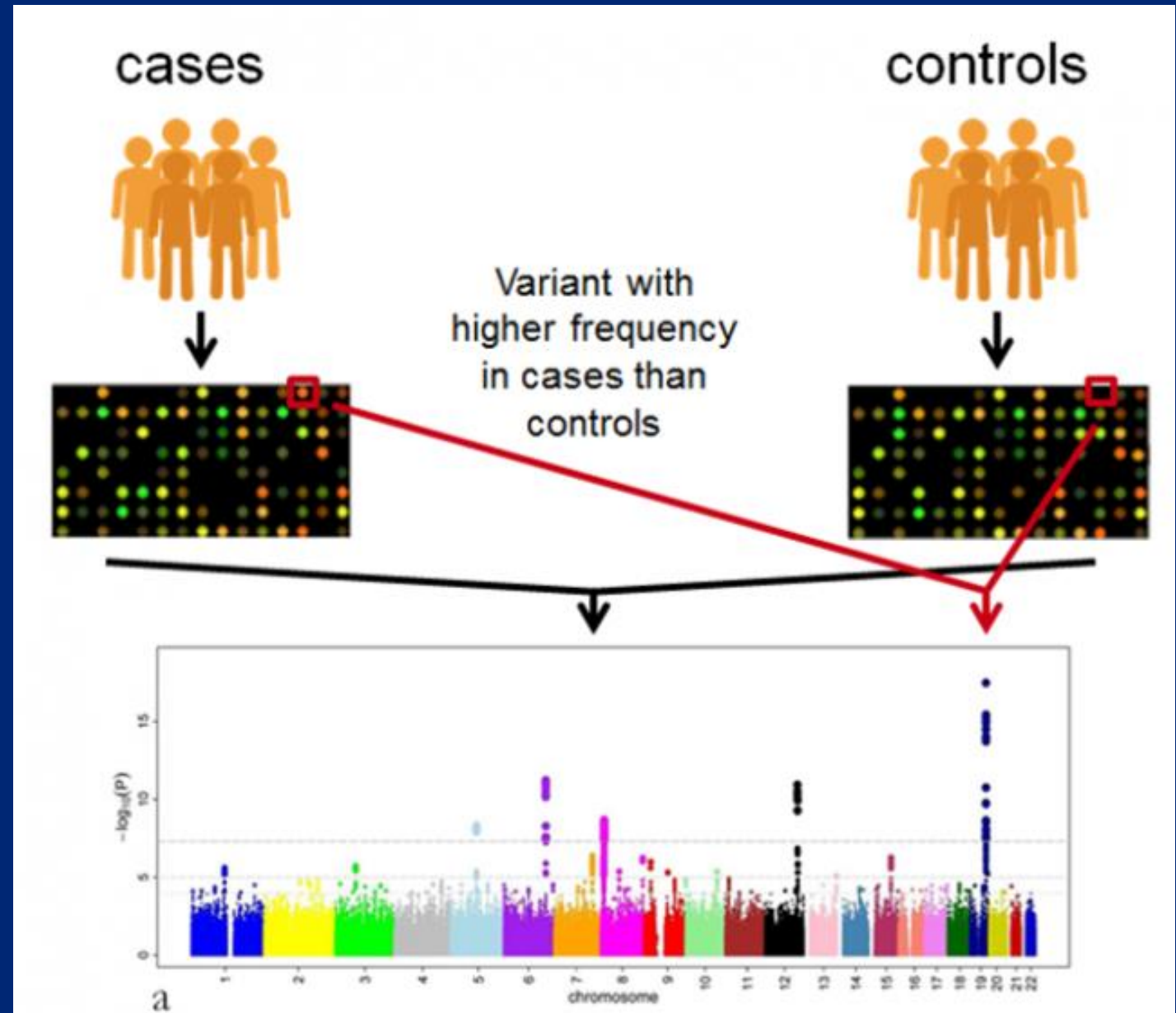


Méthodes courantes d'inférence de l'histoire sélective et de l'adaptation locale

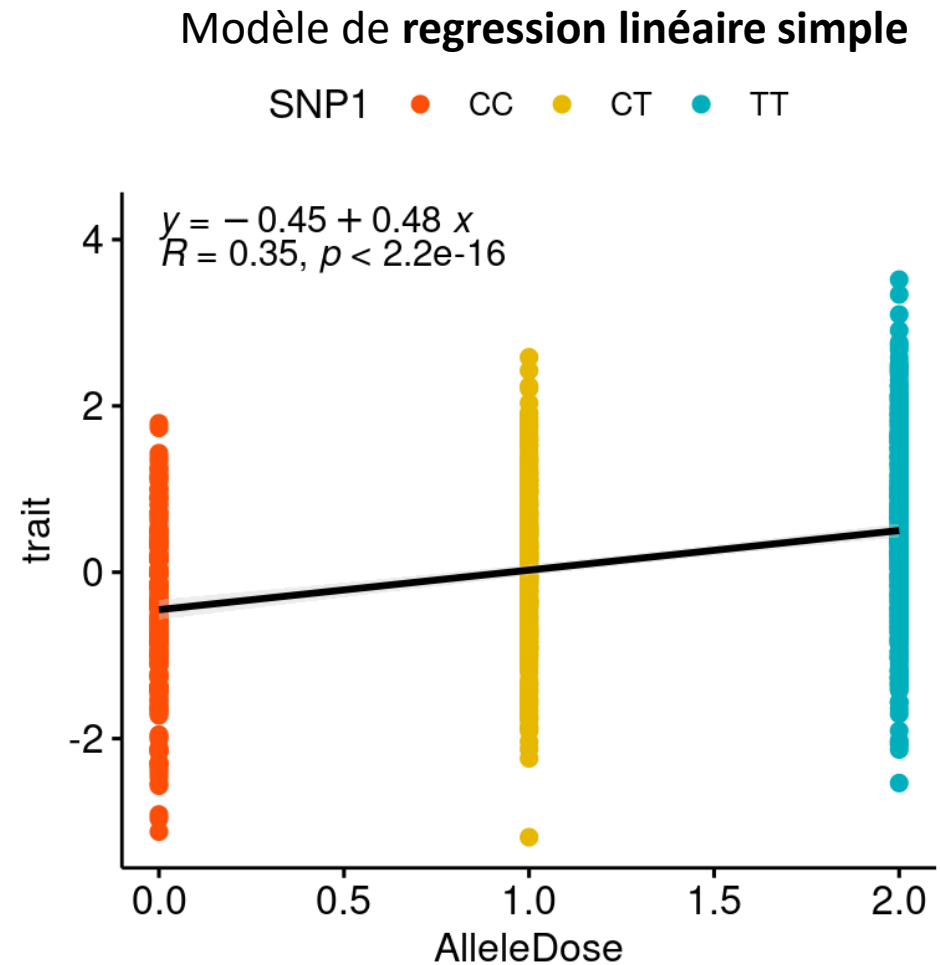
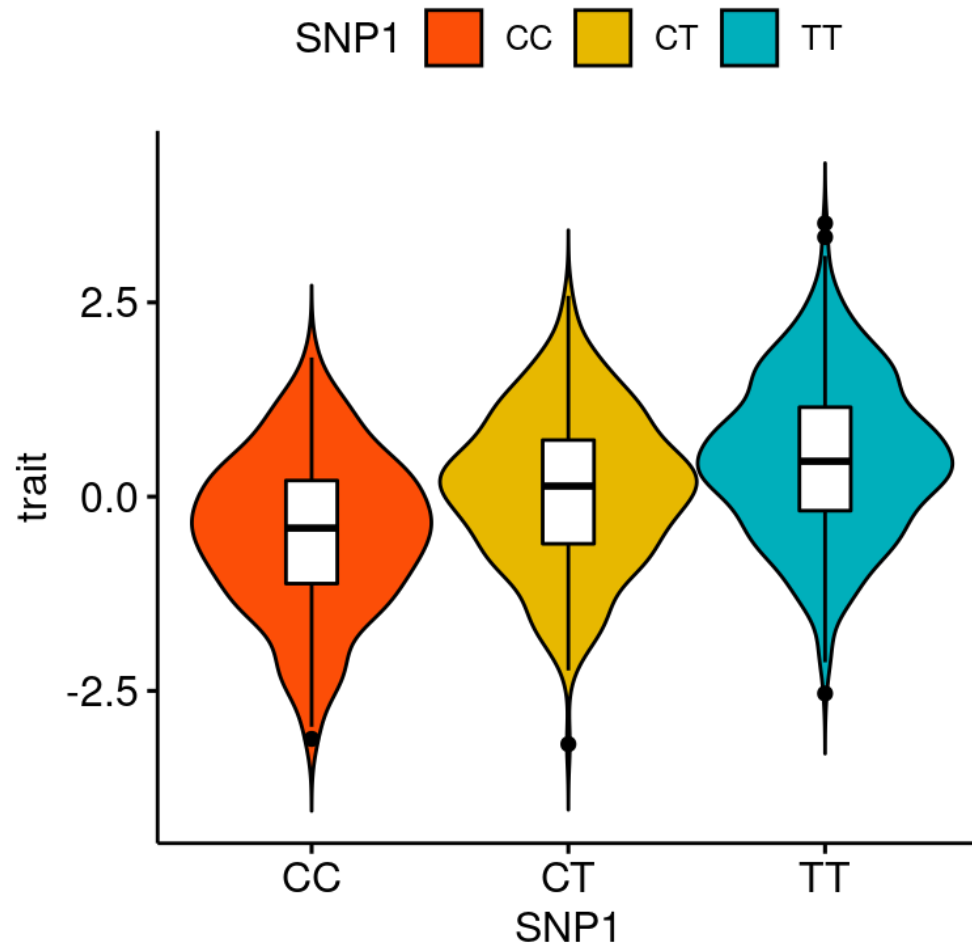
- Méthodes basées sur la différenciation génétique entre populations (F_{ST}): « genome scan » ou « F_{ST} outliers tests »
- Méthodes basées sur les corrélations génotype-environnement (GEA)
- Méthodes basées sur les associations génotype-phénotype (GWAS)

Etude d'association pangénomique (genome-wide association study, GWAS)

Analyse de nombreuses variations génétiques chez de nombreux individus, afin d'étudier leurs corrélations avec des traits phénotypiques.



Génétique d'association : Association statistique entre phénotypes (caractère d'intérêt) et génotypes à des marqueurs moléculaires (SNP) au sein d'une population



Exple: Age à la maturité, VGIL3, et dominance liée au sexe chez le saumon atlantique (Barson et al. Nature 2015)

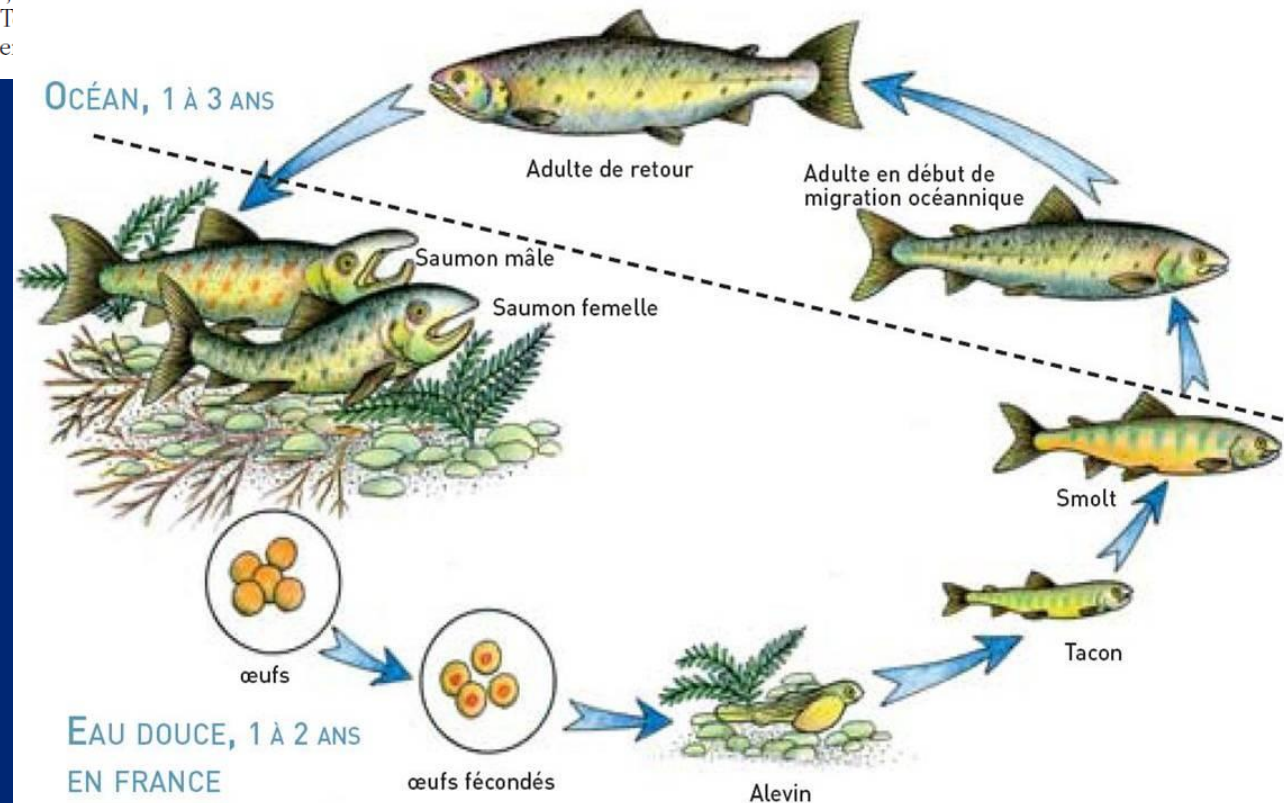
LETTER

doi:10.1038/nature16062

Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon

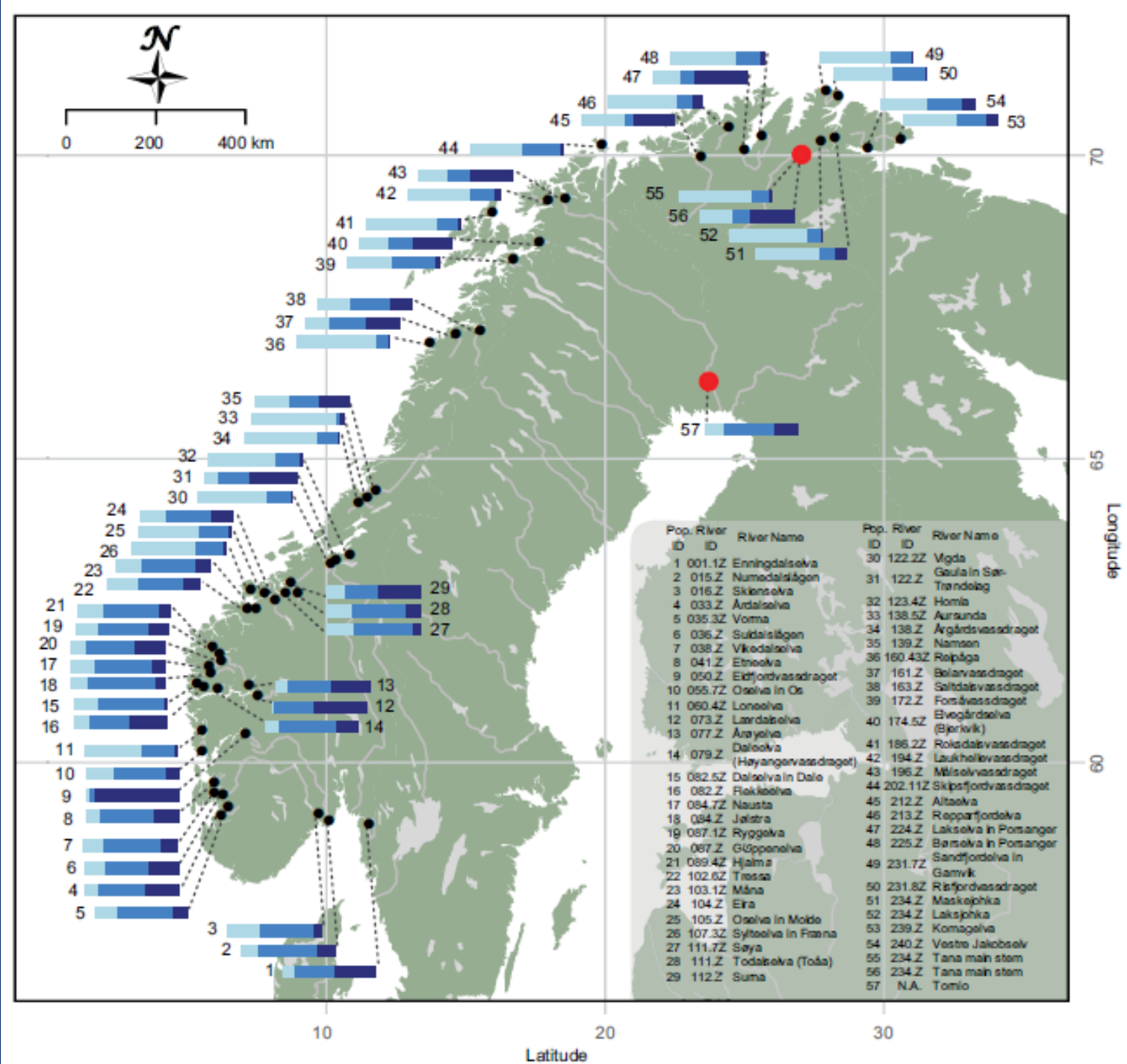
Nicola J. Barson^{1*}, Tutku Aykanat^{2*}, Kjetil Hindar³, Matthew Baranski⁴, Geir H. Bolstad³, Peder Fiske³, Céleste Jacq⁴, Arne J. Jensen³, Susan E. Johnston³, Sten Karlsson³, Matthew Kent¹, Thomas Moen⁶, Eero Niemelä⁷, Torfinn Nome¹, T Panu Orell⁷, Atso Romakkaniemi⁷, Harald Sægrov⁸, Kurt Urdal⁸, Jaakko Erkinaro⁷, Sigbjørn Lien¹ & Craig R. Primme

Age à la maturité = trait morphologique important pour lequel la sélection favorise une maturation plus précoce chez les mâles que chez les femelles chez le saumon atlantique.

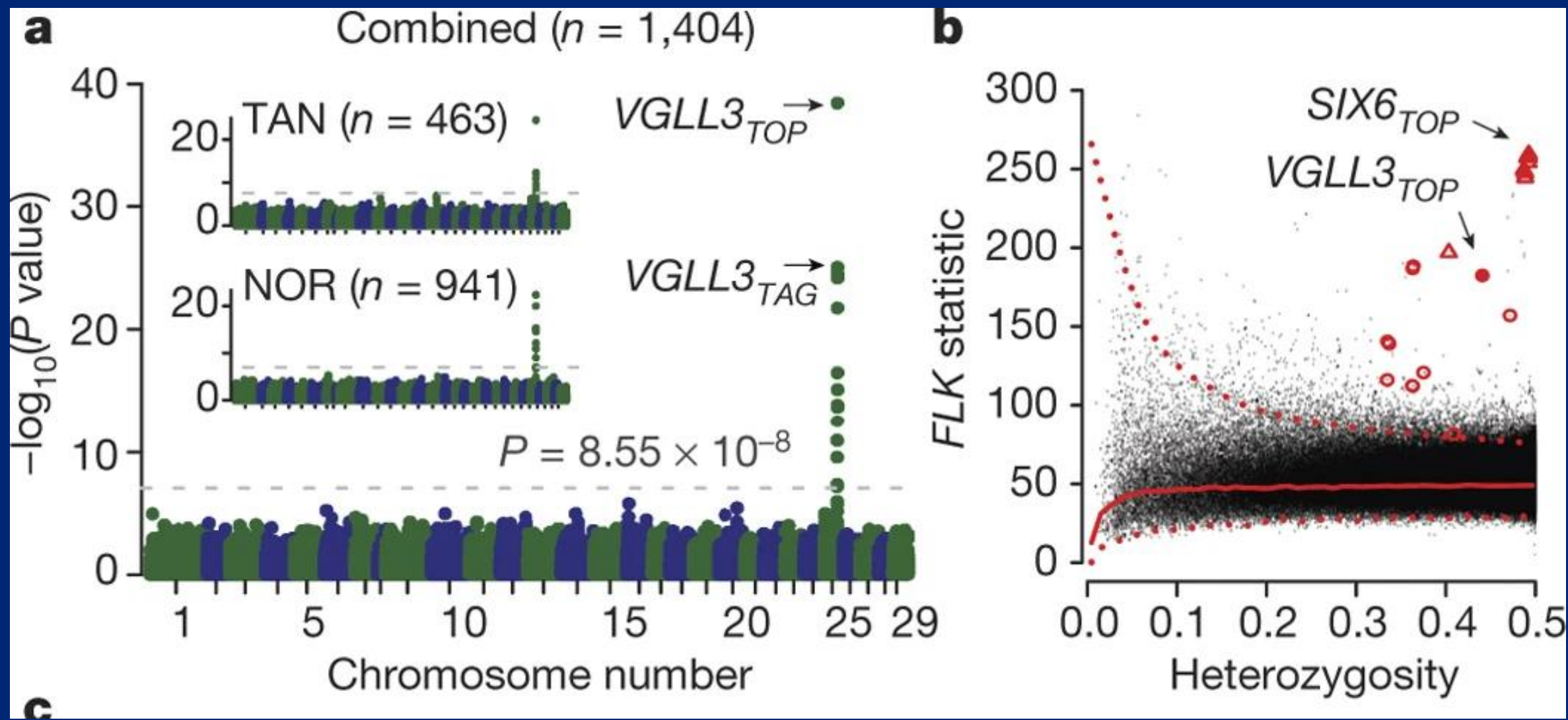


Echantillonnage/ génotypage

- 57 populations sauvages, 1518 saumons
- Age à la maturité lue dans les écailles
- 220,000 SNPs

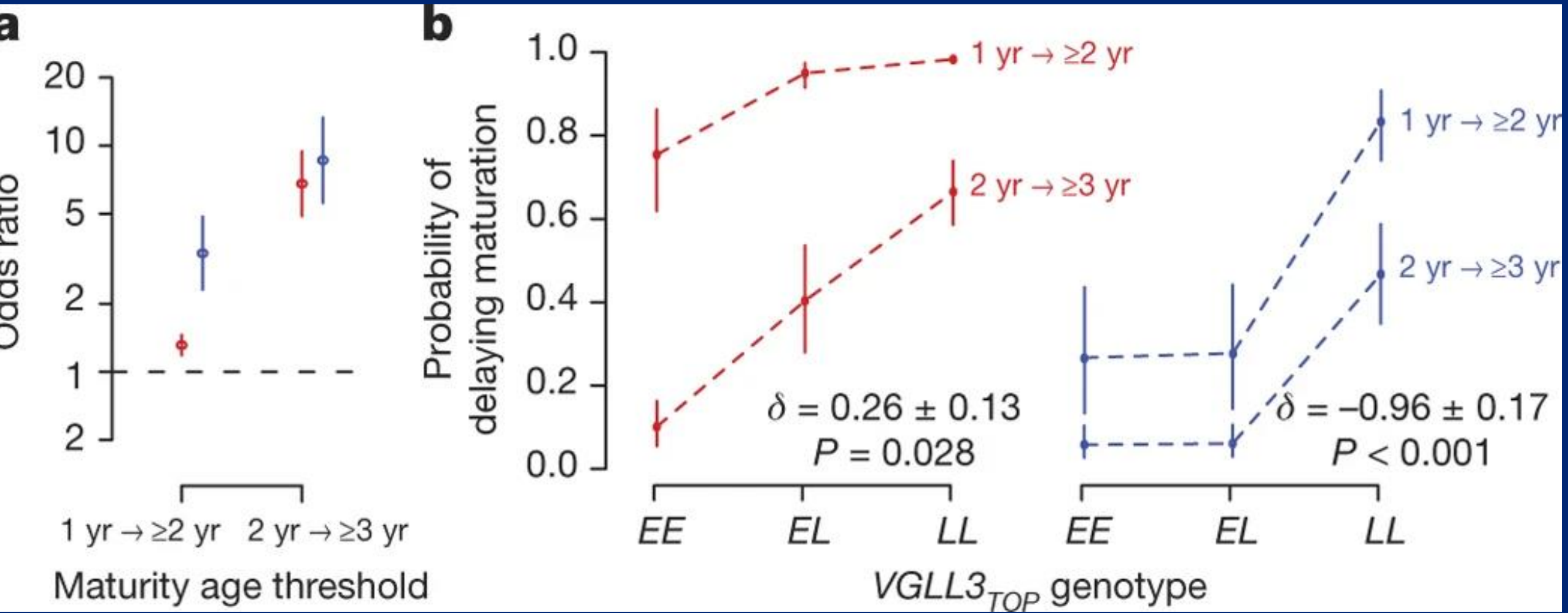


Résultats: identification d'un locus à effet majeur, VGLL3



VGLL3, un régulateur de l'adiposité associé à la taille et à l'âge de maturité chez l'homme, explique 39% de la variation phénotypique

Résultats: effet contrasté du génotype sur l'âge à la maturité chez les males (bleu) et les femelles (rouge)



Femelles avec le génotype EL ou LL ont une maturité tardive, alors que seuls les males avec génotype LL ont une maturité tardive

L'essentiel à retenir

- Les données génomiques contiennent une mine d'informations sur l'adaptation des populations à leur environnement
- De nombreuses procédures d'inférence statistique sophistiquées ont été mises au point pour inférer les bases génomiques de l'adaptation des populations à partir de ce type de données génomiques
- Il est toujours intéressant d'utiliser plusieurs approches (genome scan, GEA, GWAS) pour une même question

L'essentiel à retenir

- Les méthodes d'inférence de la démographie et de la sélection prennent rarement en compte les interactions réciproques entre ces processus.
 - les inférences démographiques se basent sur un grand nombre de loci supposés non affectés par la sélection
 - les variations démographiques marquées (e.g. expansion, goulets d'étranglement) biaisent encore souvent les estimations de la sélection, qui reposent sur des hypothèses d'équilibre
- Une stratégie pour inférer conjointement les histoires démographiques et sélective = exploiter l'information contenue dans les observations répétées dans le temps → déconvoluer les effets des processus démographiques et sélectifs sur les variations de fréquences alléliques entre pas de temps

Références

- Précis de génétique des population. JP Henry, PH Gouyon
- Principles of Population Genetics, 4th Edition. Hartl DL, Clark AG
- Cours de génétique des populations de Maxime Bonhomme:
http://www.m2p-bioinfo.ups-tlse.fr/site/images/1/11/GEQ_pgen.pdf
- Cours de Laurent Excoffier:
http://genet.univ-tours.fr/EXCOFFIER/Laurent/GMDP_intro.htm.
- Cours sur la coalescence de Mathieu Foll
<https://slideplayer.fr/amp/1722390/>