Genome scans (Cours)

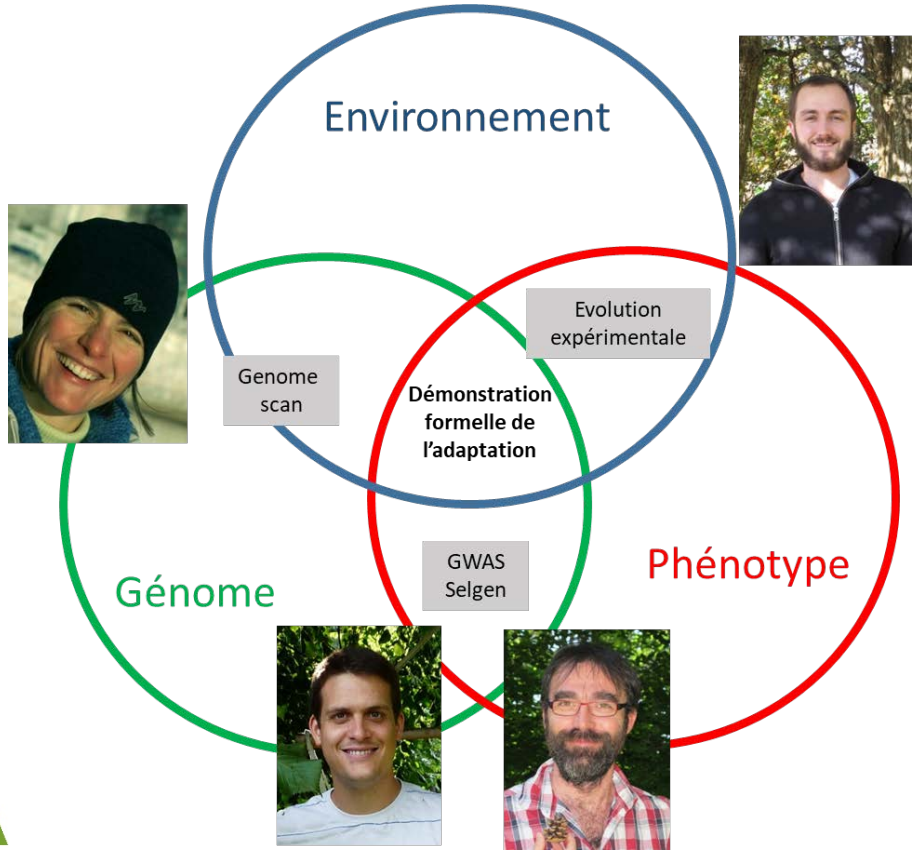
Véronique Jorge

GENOME SCANS

Véronique Jorge

UE génomique des populations

UE GÉNOMIQUE DES POPULATIONS
DÉTERMINISME DE L'ADAPTATION DES ORGANISMES À LEUR ENVIRONNEMENT



SOMMAIRE

1. **Introduction : définitions et objectifs des scans génomiques**
2. Rappels de génétique des populations : statistiques de diversité appliquées aux séquences
3. La sélection et ses effets sur les séquences ADN
4. Outils pour la détection de traces de sélection
5. La génomique du paysage
6. La validation des résultats de scans génomiques

INTRODUCTION

- Génétique évolutive => identification des caractères adaptatifs i.e. soumis à sélection.
- Génomique des populations* :

  « Le processus d'échantillonnage simultané de nombreux loci le long du génome et l'inférence d'effets spécifiques à certains loci à partir des distributions de ces effets. »
- Hypothèse sous-jacente au scans génomiques :
  - La sélection naturelle agit sur les phénotypes;
  - Et indirectement sur la séquence;
  - En analysant les séquences, on peut détecter des traces de sélection.

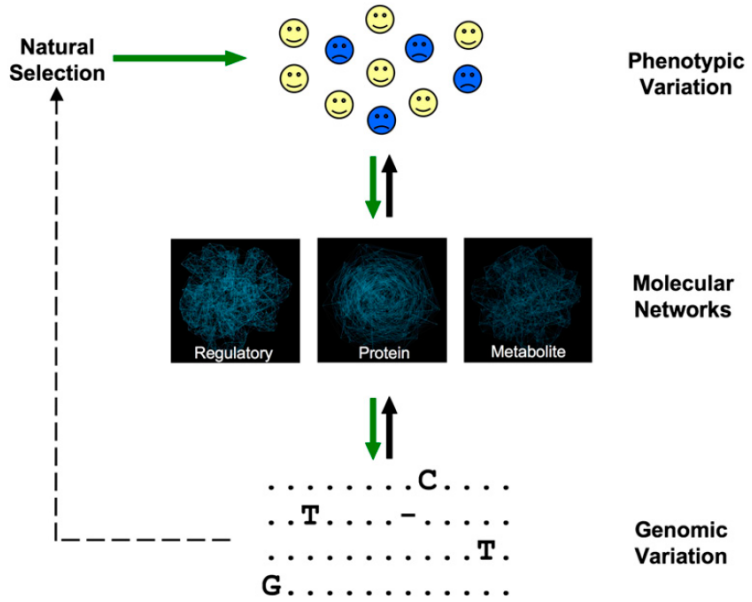


Figure 3. Bottom-up population genomics. Genome-wide scans of positive selection are agnostic to phenotypic data and make inferences of selection directly from patterns of genetic variation (dashed black arrow). However, selection acts directly on phenotypic variation and only indirectly on DNA sequence variation (dark green arrows). Solid black arrows show that the path from genetic to phenotypic variation runs through dynamic molecular networks (such as regulatory, protein, and metabolite). Scale-free molecular networks were simulated with the R package igraph and visualized in CytoScape (Cline et al. 2007).

Figure : Akey *Genome Res.* **2009** 19: 711-722

* Black IV et al. Annu. Rev. Entomol. **2001** 46:441-469

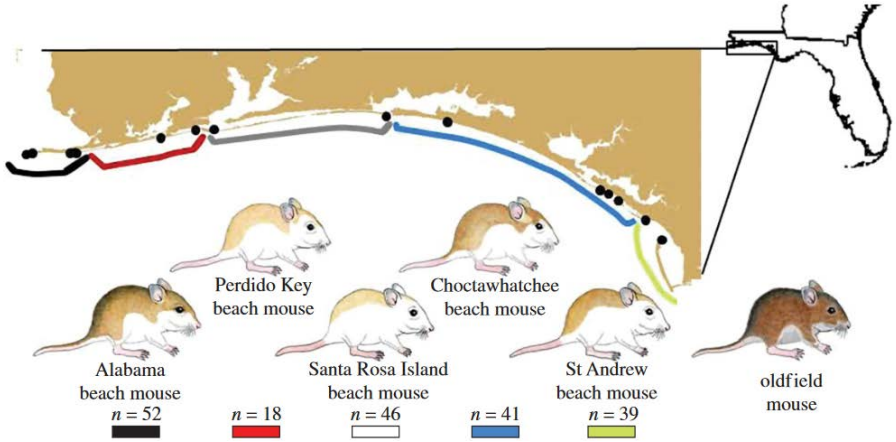
L'APPROCHE « GÈNE CANDIDAT »

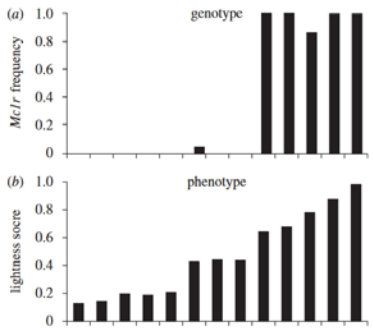
6 gènes contrôlant la pigmentation de la peau chez l'homme

Fréquences alléliques ASIP A8818G

Norton et al. *Mol. Biol. Evol* 2007. 24(3):710–722.

1 gène contrôlant la pigmentation du pelage chez la souris



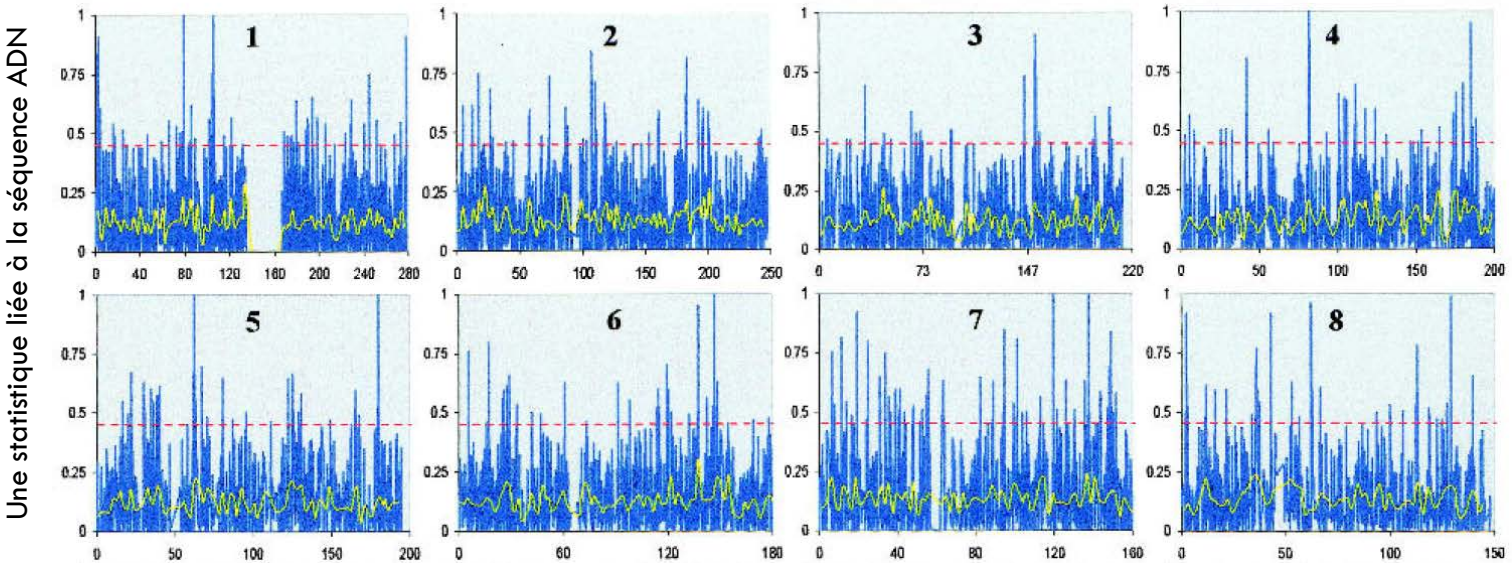


Mullen et al. *Proc. R. Soc. B* 2009. 276, 3809–3818

- Fort a priori sur la fonction des gènes et leur lien avec l'adaptation
- Effet confondant entre sélection et démographie

« GENOME SCAN »

Détection de **signatures de sélection** à l'échelle de génomes entiers, grâce à la disponibilité croissante de données à l'échelle des génomes



Chez l'homme :

- 3 populations analysées (42 East Asian, 42 African-American, and 42 European-American)
- 26,530 SNP
- 174 gènes avec une signature de sélection

Akey et al. 2002. *Genome Res.* 12: 1805–1814.

SOMMAIRE

1. Introduction : définitions et objectifs des scans génomiques
2. **Rappels de génétique des populations : statistiques de diversité appliquées aux séquences**
3. La sélection et ses effets sur les séquences ADN
4. Outils pour la détection de traces de sélection
5. La génomique du paysage
6. La validation des résultats de scans génomiques

RAPPELS DE GÉNÉTIQUE DES POPULATIONS

Indices de diversité génétique

Spectres de fréquences allélique

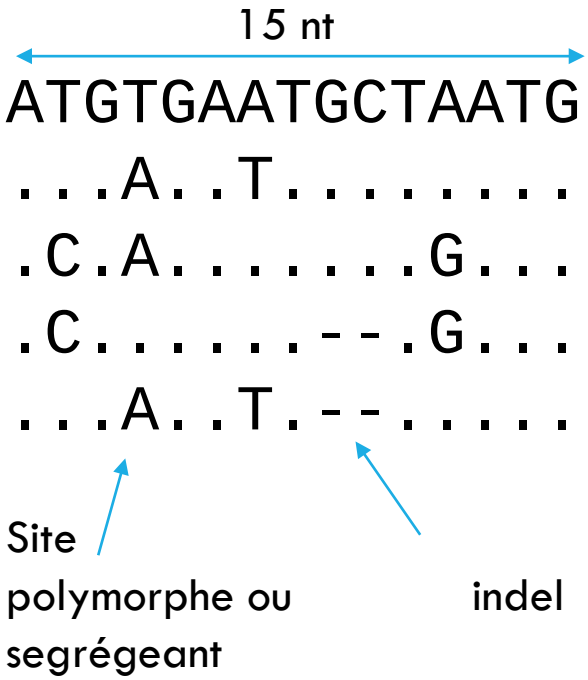
Différenciation entre populations

Déséquilibre de liaison

Modèle neutre

Théorie de la coalescence

LES STATISTIQUES QUI MESURENT LE POLYMORPHISME NUCLÉOTIDIQUE



Nb de site en ségrégation (**S**) = 4

Nb moyen de différence entre paires (**π**) = 2.4

par site = 0.16

	2	**3**	**4**	**5**
1	2	3	2	2
2		3	4	0
3			1	3
4				4

Nb d'haplotypes ..... = 4

(D'après G. McVean 2001)

LES STATISTIQUES QUI MESURENT LE POLYMORPHISME NUCLÉOTIDIQUE

Mutations synonymes & non synonymes

```
Arg Gln Val
AGA CAA GTA
```

```
CAG CGA GTA
Arg Arg Val
```

```
Arg Gln Val
AGA CAA GTA
```

```
AGA CAG GTA
Arg Gln Val
```

Dégénérescence du code génétique

D. simulans $\pi_{total} = 0.010$ per site

$\pi_{silent} = 0.038$

$\pi_{noncoding} = 0.023$

(D'après G. McVean 2001)

HÉTÉROZYGOTIES ATTENDUE ET OBSERVÉE (À L'ÉCHELLE DE POPULATIONS)

- H_o (observed Heterozygosity) = moyenne de l'hétérozygotie observée dans l'échantillon.

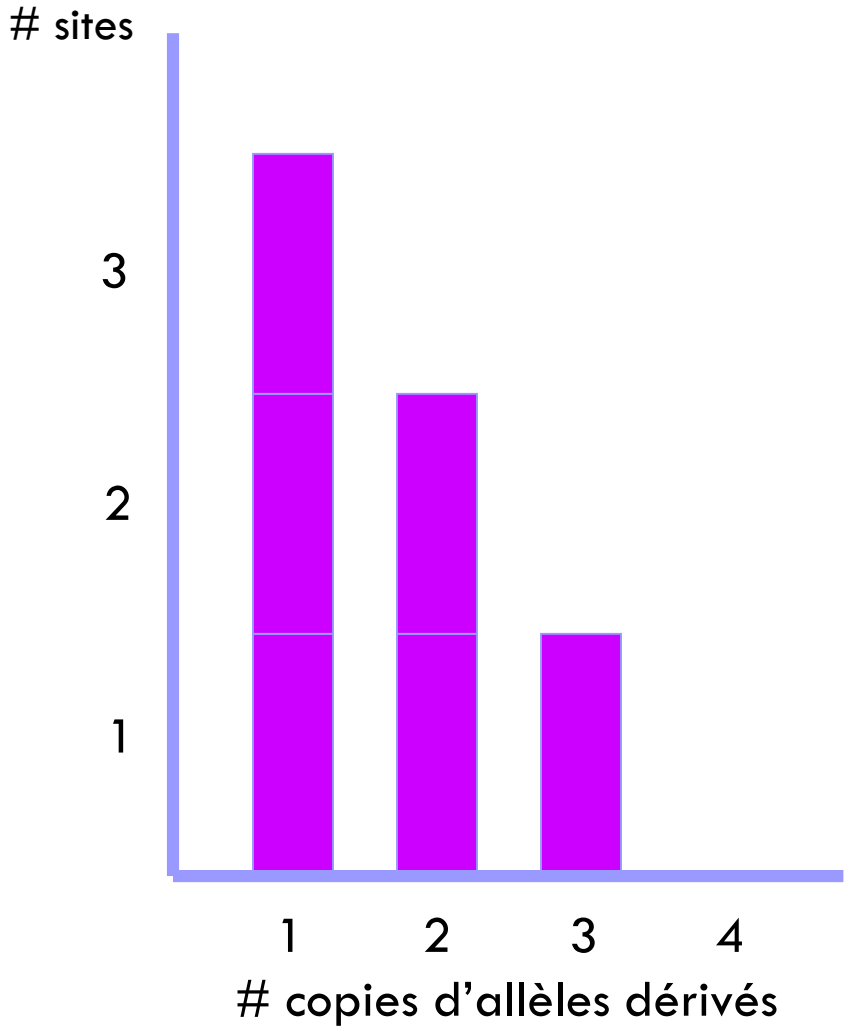
- H_e (expected Heterozygosity) = probabilité que 2 alleles tires au hazard soient différents

$$H_e = 1 - \Sigma\, p_i^2 = 1 - (p_1^2 + p_2^2 + \ldots + p_n^2) = \Sigma\, p_i p_k \quad (\text{for } i < k)$$

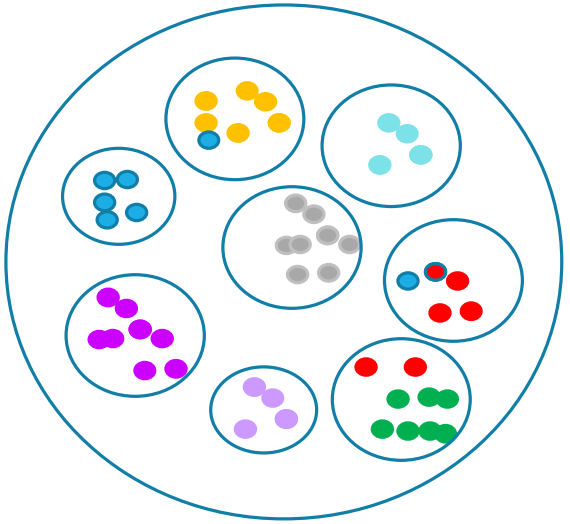
p_i frequence de l'allèle *i*.

LE SPECTRE DE FRÉQUENCES ALLÉLIQUES (SITE FREQUENCY SPECTRUM – SFS)

Orang-Outan	*A T C A G T*
Chimpanzé	*A T C A G T*
Homme	A T **G** A G T
	A **A** C A G T
	C T C A G T
	A T **G** A G T
	A T C A G T
	A T C A G **G**
	A T C **C** **T** T
	A T **G** A **T** T
	A T C **C** G T
dérivés	1 1 3 2 2 1



STRUCTURE DES POPULATIONS



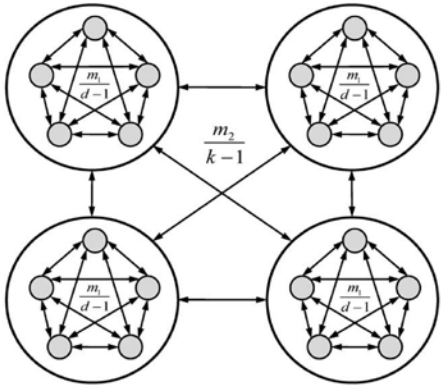
Les population naturelles sont subdivisées à cause de :

- Habitats discontinus
  - Montagnes, lacs, océans
  - Ressources
  - Système hôte parasites
  - Saisonnalité

- Comportement

Le patron de migration et la date de séparation a un effet fort sur le niveau de structuration des populations.

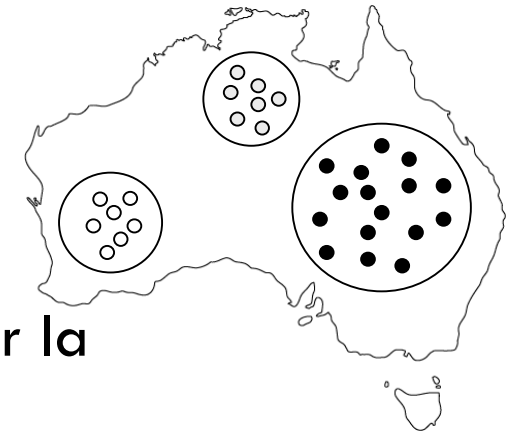
Les subdivisions peuvent être hiérarchisées.



Excoffier et al. Heredity (2009) 103, 285–298

Slide by Excoffier

MESURER LA DIFFERENTIATION GÉNÉTIQUE



Le F_{ST} *de* Wright : Part de la diversité expliquée par la subdivision en population.

Hétérosigotie moyenne à l'intérieur des populations

Heterozygotie entre toutes les populations

$$= \frac{H_T - \overline{H}_S}{H_T}$$

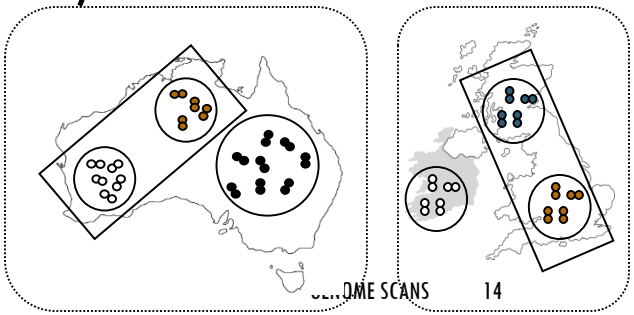
(Nei's Gst)

Detect significant values by permutation

Hierarchical nature of *F* statistics (fixation indices)

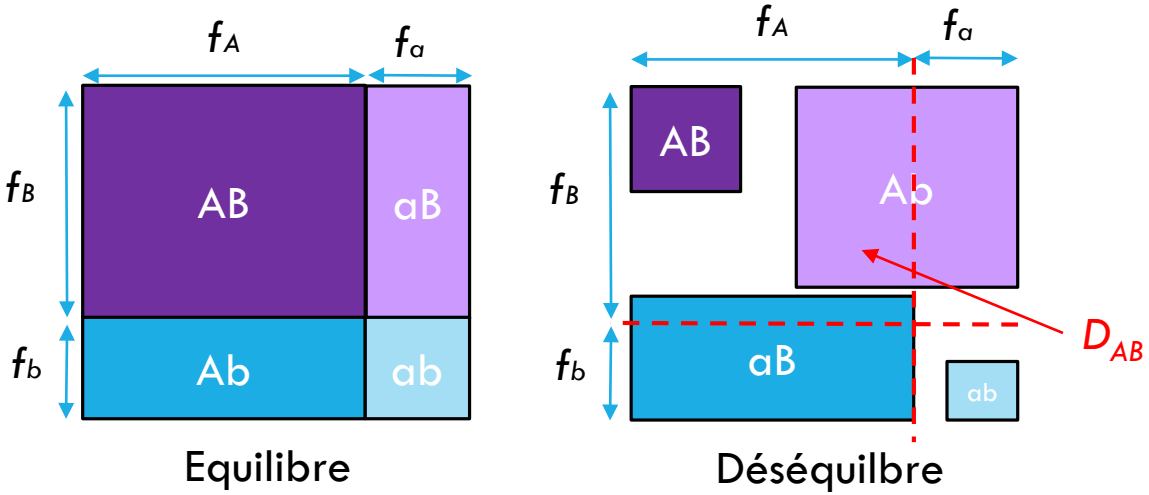
$$H_{Individual} < H_{Subpopulation} < H_{Population} < H_{Region} < H_{Total}$$

Nst for a fragment/gene : $Nst = \pi t - \pi s / \pi s$,



DÉSÉQUILIBRE GAMÉTIQUE OU DE LIAISON

Définition : association préférentielle entre allèles à 2 locus



Déséquilibre maximal

$D_{max} = \min(f_A.f_b ; f_B.f_a)$

(si $f_A > f_a$ et $f_B > f_b$)

Déséquilibre normalisé

$D' = D/D_{max}$

Lewontin, 1964

Equilibre : $f_{AB} = f_A.f_B$

Déséquilibre : $D_{AB} = f_{AB} - f_A.f_B$

$= f_{AB}.f_{ab} - f_{aB}.f_{Ab}$

$= D_{ab} = -D_{Ab} = -D_{aB}$

Corrélation entre sites

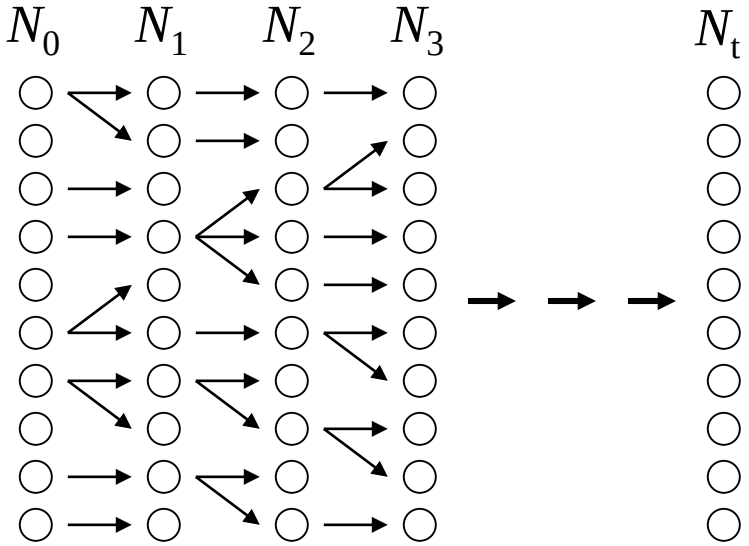
$r^2 = D^2 / (fA.fa.fB.fb) = \rho^2$

Hill et Robertson, 1968

LE MODÈLE NEUTRE EN GÉNÉTIQUE DES POPULATIONS

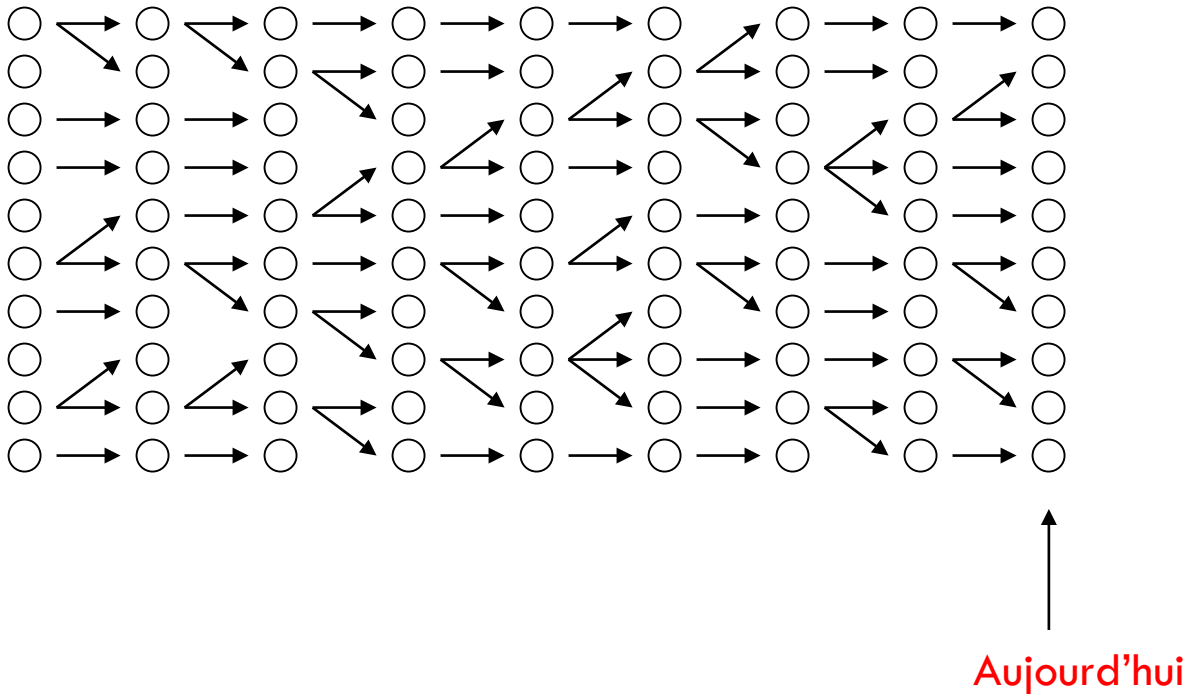
La grande majorité des mutations sont neutres (nothing happens!)

Le modèle neutre de Wright-Fisher

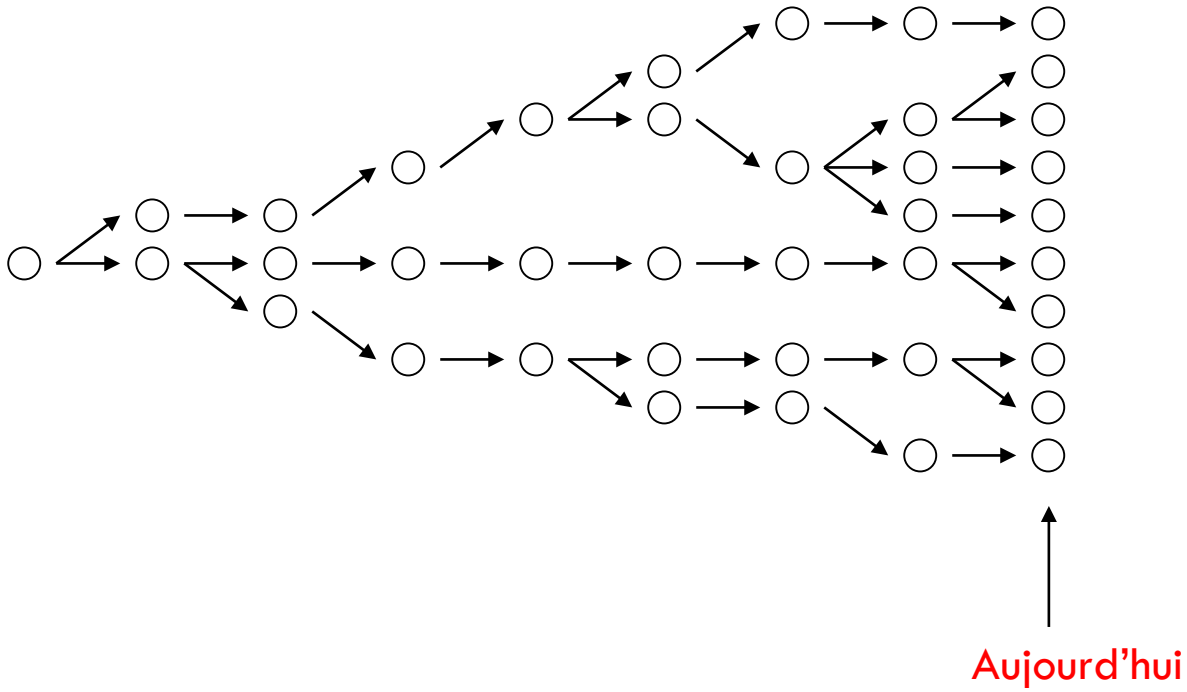


- Des individus diploïdes se reproduisent sexuellement avec possibilité d'autofécondation
- Les croisements sont aléatoires (panmixie)
- Les générations de sont pas recouvrantes
- La taille de la population est constante de taille N (2N allèles)
- Pas de migration, pas de sélection

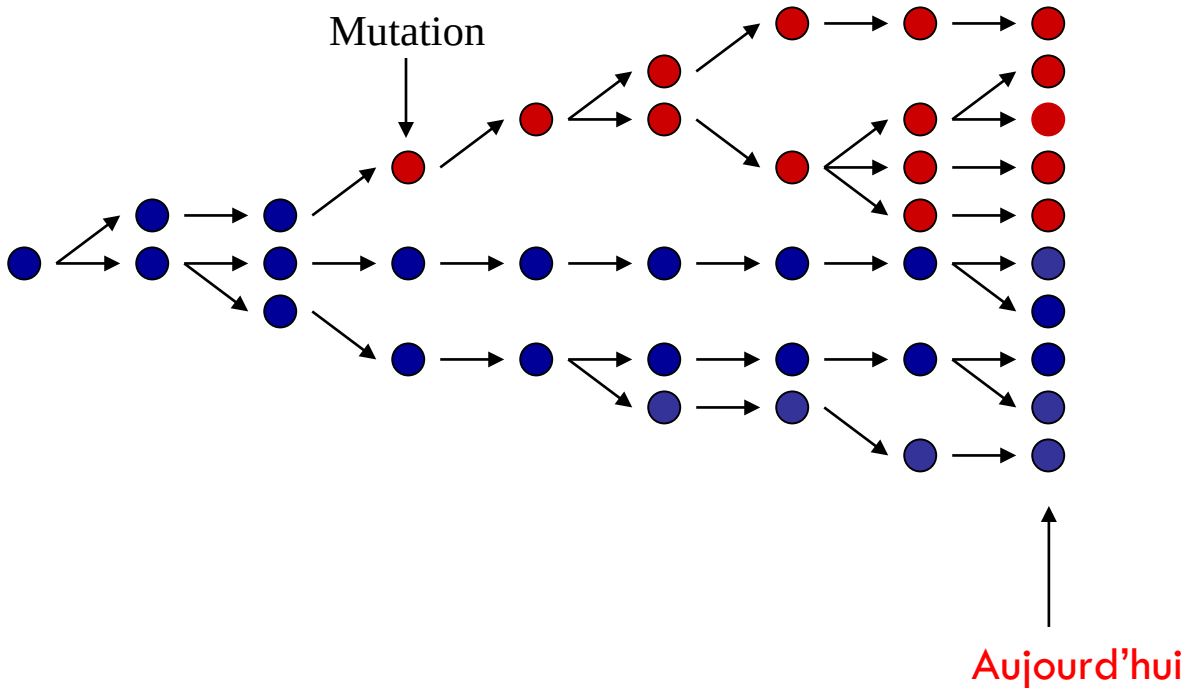
RAPPELS DE GÉNÉTIQUE DES POPULATIONS



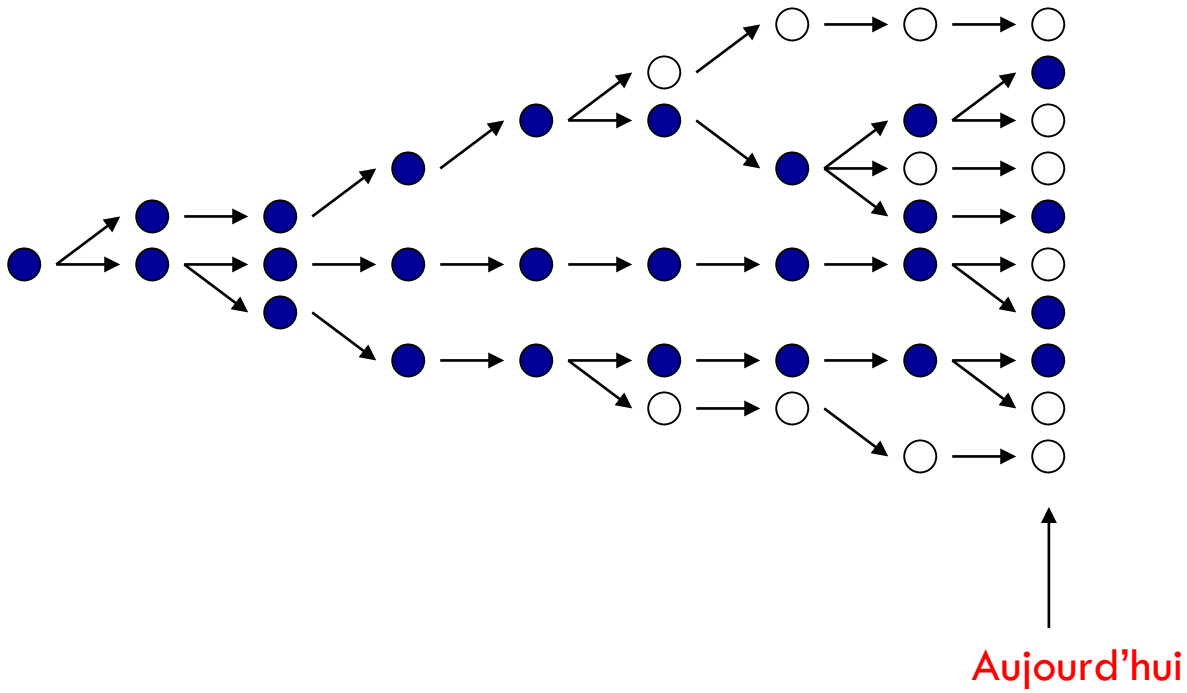
ASCENDANCE DE LA POPULATION ACTUELLE



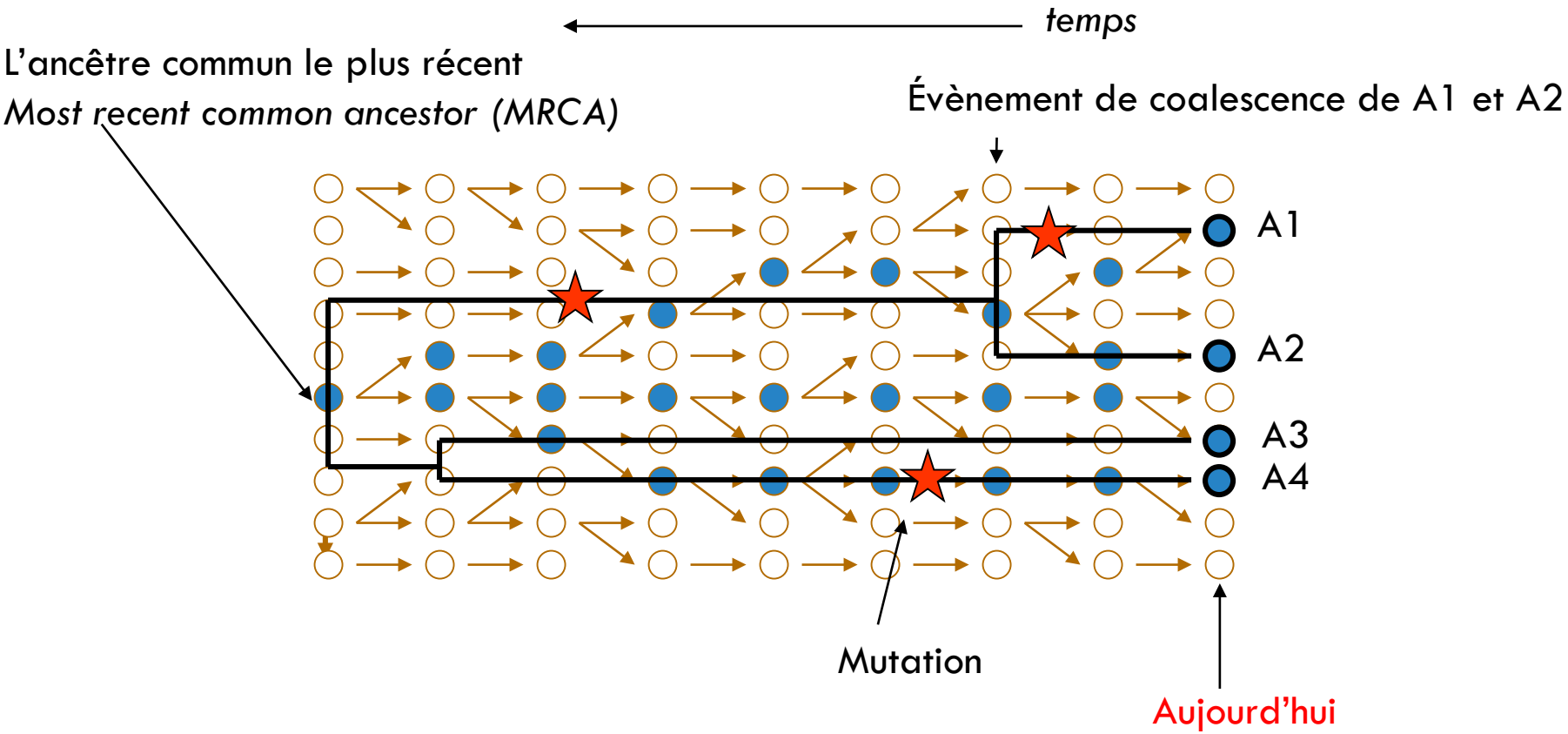
AJOUT D'UNE MUTATION



ASCENDANCE D'UN ÉCHANTILLON



LA COALESCENCE: DES ÉCHANTILLONS DANS DES POPULATIONS



Le nombre moyen de différence attendues entre 2 séquences (A1 et A4) :

$$E[\pi] = 2 \times \mu \times E[T_{MRCA}] = 4\,N\,\mu$$

$$\theta = 4\,N\,\mu$$

CONCEPT D'EFFECTIF EFFICACE (EFFECTIVE POPULATION SIZE) NE

Dans les populations naturelles, tous les individus ne participent pas forcément au processus reproductif

→*en général l'effectif de la population N (qui détermine le rythme de la dérive génétique) n'est pas égale à l'effectif de recensement de la population (censu size).*

→ *On défini donc l'effectif efficace de la population (ou taille efficace) comme l'effectif d'une population idéale (suivant un modèle de Wright-Fisher) pour laquelle on aurait une fluctuation du polymorphisme équivalente à celle de la population naturelle.*

→ *C'est donc le nombre d'individus (d'une population idéale) pour lequel on aurait un degré de dérive génétique équivalent à celui de la population réelle.*

→*Plusieurs estimateurs (types) d'effectif efficace, selon à quel effet de la dérive génétique on s'intéresse: inbreeding, variance of allelic frequencies, hétérozygotie*

ESPRIT DE L'APPROCHE "COALESCENCE"

- « *Coalescent theory is a probabilistic description of the genealogical process for samples of chromosomes in large populations.* » (*Kingman 1982*)

  → description mathématique du processus généalogique dans une population idéalisée

- Coalescence versus phylogénie moléculaire

	Coalescence	Phylogénie
Généalogie des séquences	Au sein d'une espèce	Entre espèces
L'arbre	Quelles sont les forces évolutives compatibles ?	Reconstruire le « vrai »

- Déterminer l'ensemble des arbres généalogiques qu'il est possible d'observer pour un modèle de génétique des populations fixe.

- Si on connaît tous les arbres possibles pour un modèle donné, trouver l'ensemble des **paramètres** de ce modèle qui sont le plus compatible avec l'arbre observé.

Les paramètres : taille de la population, taux de migration, taux de croissance de la population, force de la sélection etc.).

=> inférer l'importance des différentes forces évolutives qui sont "responsables" de l'arbre observé.

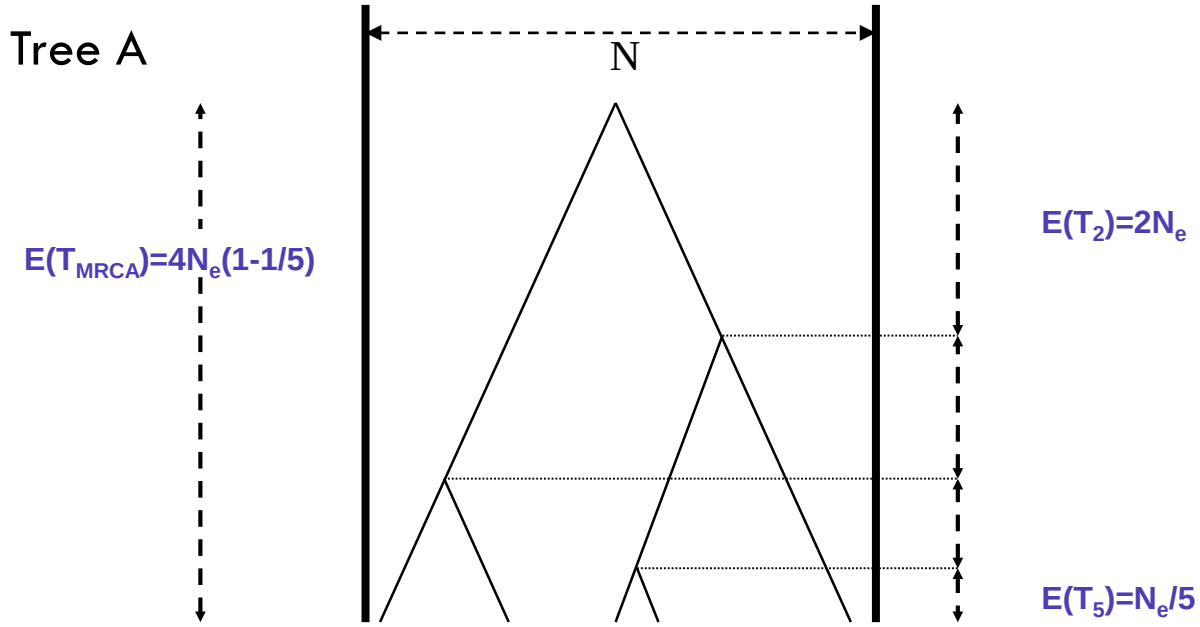
HISTOIRE DÉMOGRAPHIQUE

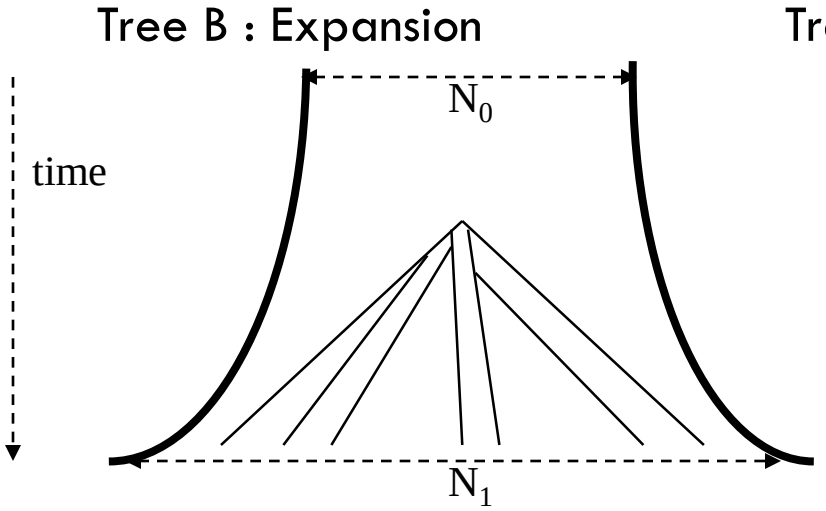
"The reproductive history of a population or group of populations. This can include population sizes, sex ratios, migration rates, population splitting events, variation in reproductive rates and times among organisms, as well as variation over time in all of these quantities."

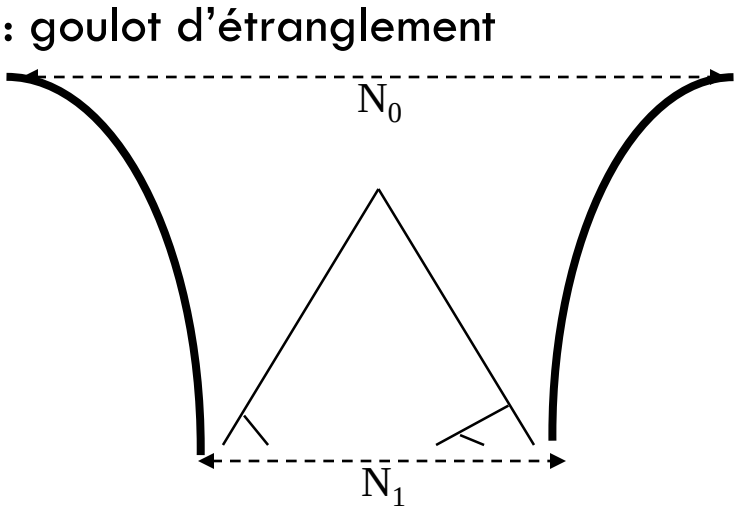
Hey and Machado, NRG, 2003

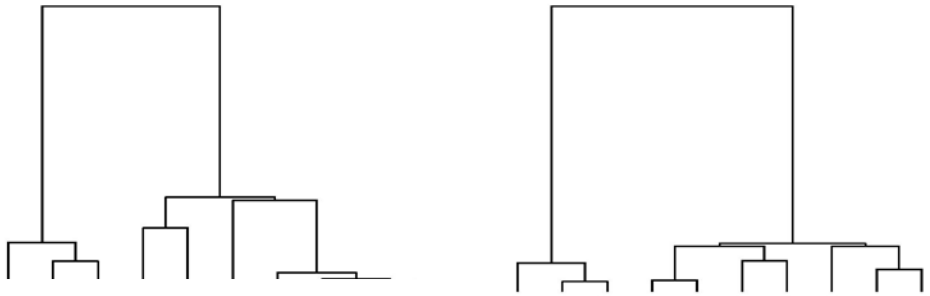
→L'histoire démographique d'une population influence la distribution des temps de coalescence (longueur des branches)

- Donc, elle influence les autres paramètres de diversité.
- On peut donc utiliser ces paramètres aussi pour en déduire l'histoire démographique.

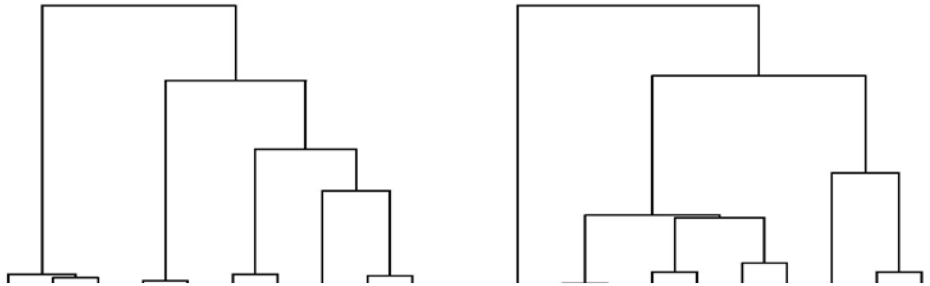

Tree A
N
$E(T_{MRCA})=4N_e(1-1/5)$
$E(T_2)=2N_e$
$E(T_5)=N_e/5$


Tree B : Expansion
time
N_0
N_1

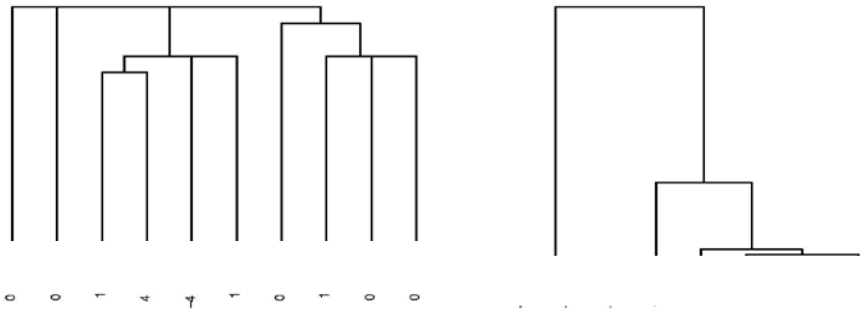

Tree C : goulot d'étranglement
N_0
N_1



Stable



Growing



Contracting

SOMMAIRE

1. Introduction : définitions et objectifs des scans génomiques
2. Rappels de génétique des populations : statistiques de diversité appliquées aux séquences
3. **La sélection et ses effets sur les séquences ADN**
4. Outils pour la détection de traces de sélection
5. La génomique du paysage
6. La validation des résultats de scans génomiques

MODÈLE NEUTRE

Caractéristiques de la théorie neutre

• La majorité des changements dans les séquences protéiques et ADN qui sont fixée entre espèce et polymorphes au sein des espèces n'ont pas d'impact sélectif.

• Le taux de substitution est égal au taux de mutation neutre.

• Le niveau de polymorphisme dans une population est fonction de la taille effective* de la population et du taux mutation neutre.

• Les polymorphismes sont plutôt transitoire que « balancés ».

=> L'HYPOTHÈSE NULLE

* Donc de la dérive

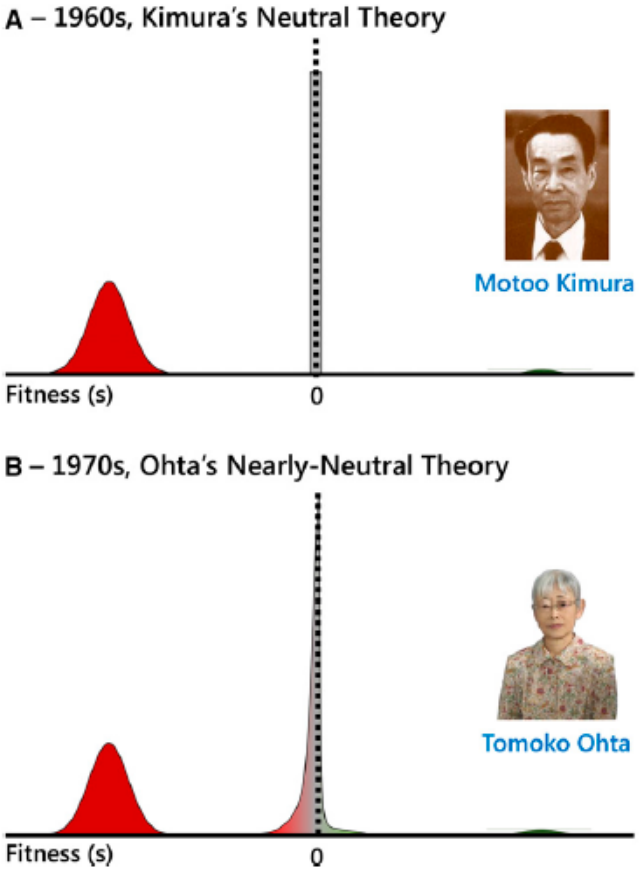


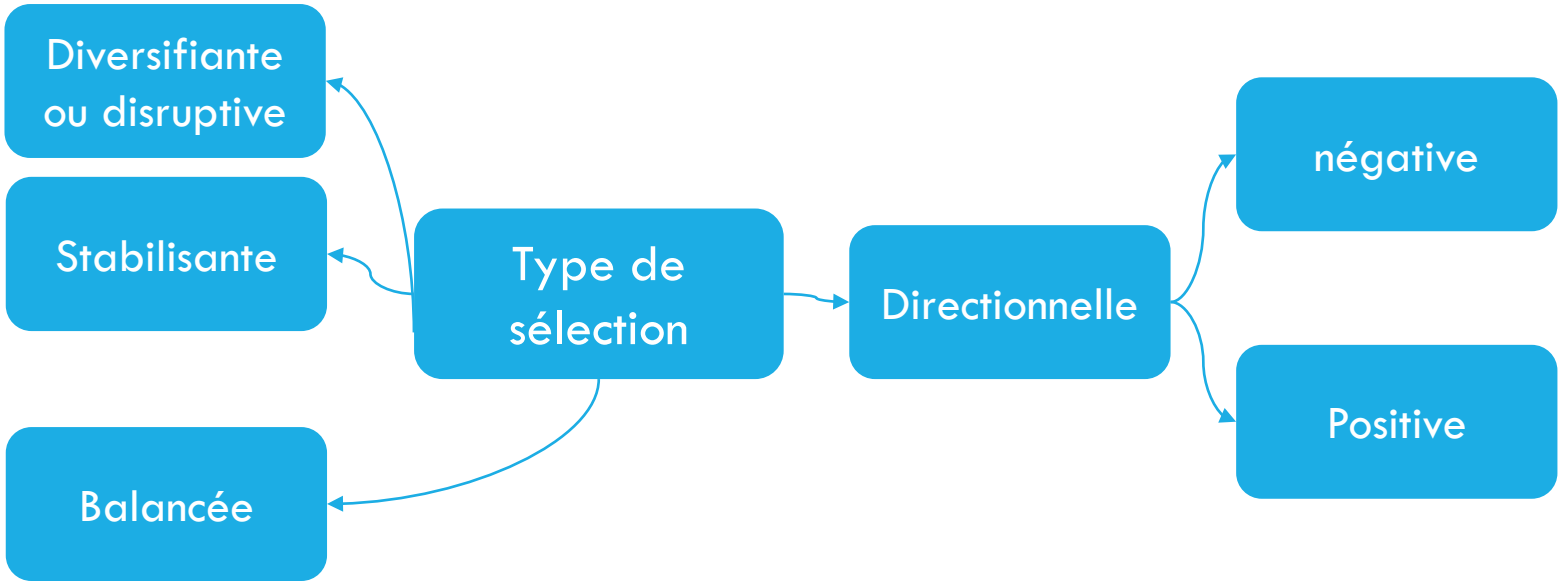
Figure 2 DFE according to the (nearly) neutral theory of molecular evolution. (A) In the 1960s, according to the Kimura's neutral theory. (B) In the 1970s, after the extension of the neutral theory by Ohta. Different selection coefficients of mutations are colored in a gradient from maroon (strongly deleterious), red (slightly deleterious), gray (neutral), light green (slightly advantageous), and dark green (advantageous).

Casillas & Barbadillas
2017 Genetics, Vol. 205, 1003–1035

LES DIFFÉRENTS TYPES DE SÉLECTION

Définition générale

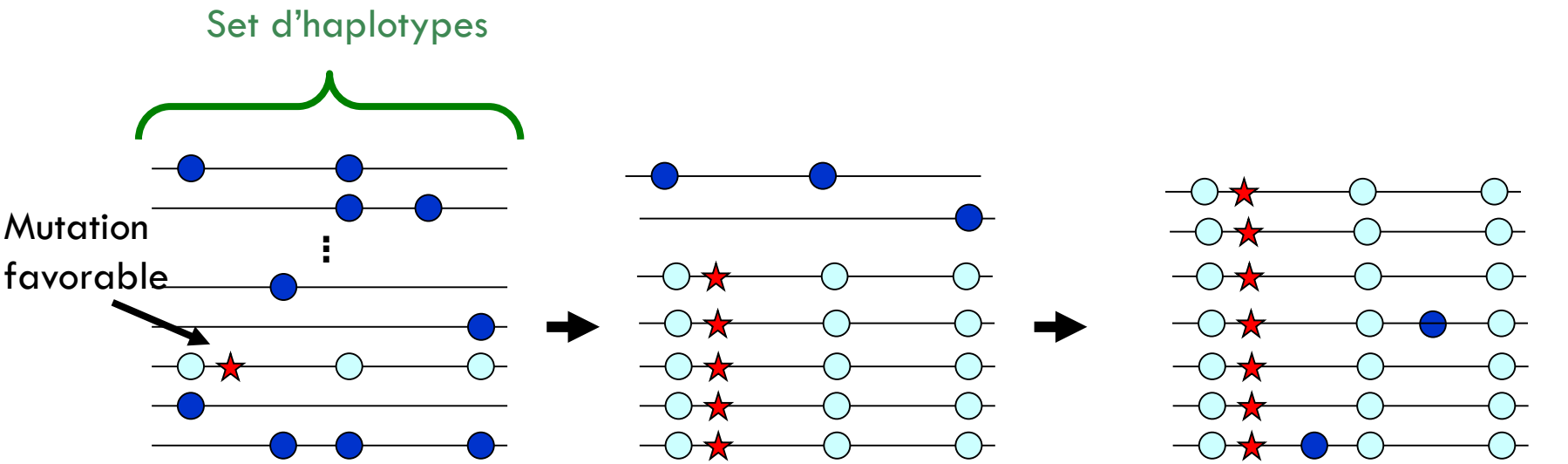
Propagation différentielle non aléatoire d'un allèle.



Organismes diploïdes

Nielsen 2005 Annu Rev. Genet. 39:197-218
Vitti et al 2013 Annu. Rev. Genet. 47:97-120

SÉLECTION DIRECTIONNELLE POSITIVE



Apparition d'une mutation favorable

La fréquence de la mutation augmente sous l'effet de la sélection et entraîne par auto-stop* les sites voisins => baisse de la diversité dans la région

Après fixation de la mutation, la diversité se régénère au rythme des mutations, en faible fréquences → excès de mutations rares.

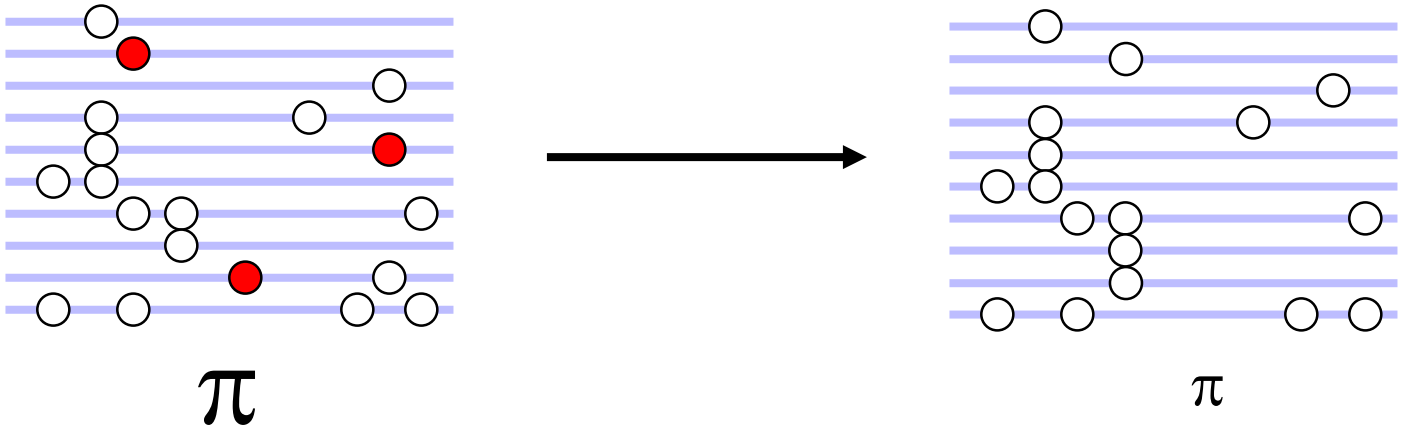
*Hitch-hiking

\+ facile à détecter

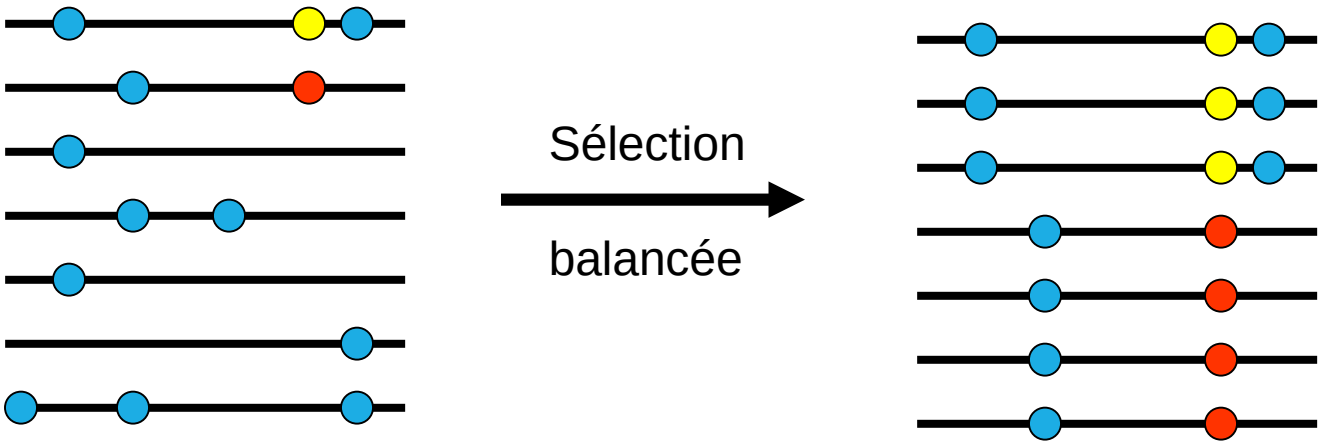
Ex : Adaptation à l'altitude chez les tibétains

SÉLECTION DIRECTIONNELLE NÉGATIVE (PURIFICATRICE)

- Certains gènes très conservés
- Elimination de la moindre mutation.



SÉLECTION BALANCÉE

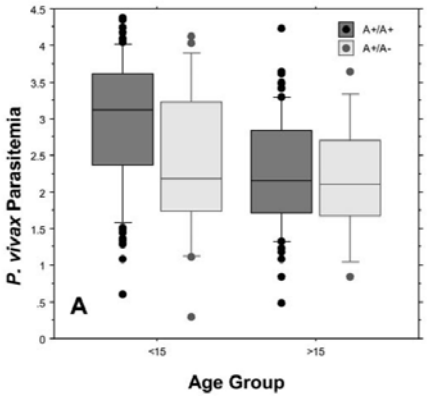


Selected Mutations

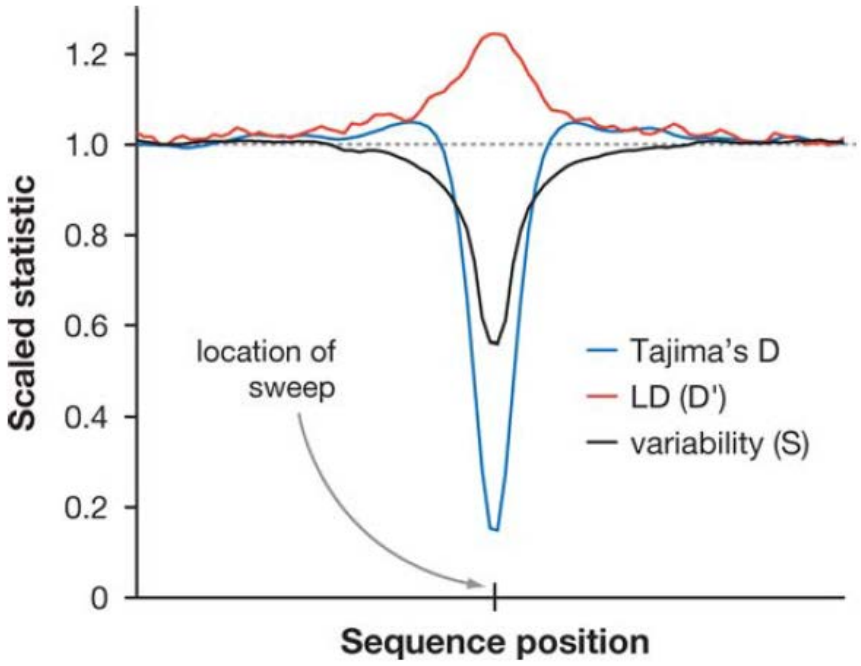
Neutral Mutations

- Plusieurs allèles ont maintenues à des fréquences intermédiaires.
- La diversité augmente au locus et autour
- Mais la structure en haplotype est forte

Exemple :
Mutations causant une résistance à la malaria (haemoglobinopathies)
Kasehagen et al 2007



EFFET DE LA SÉLECTION SUR LES PARAMÈTRES DE DIVERSITÉ

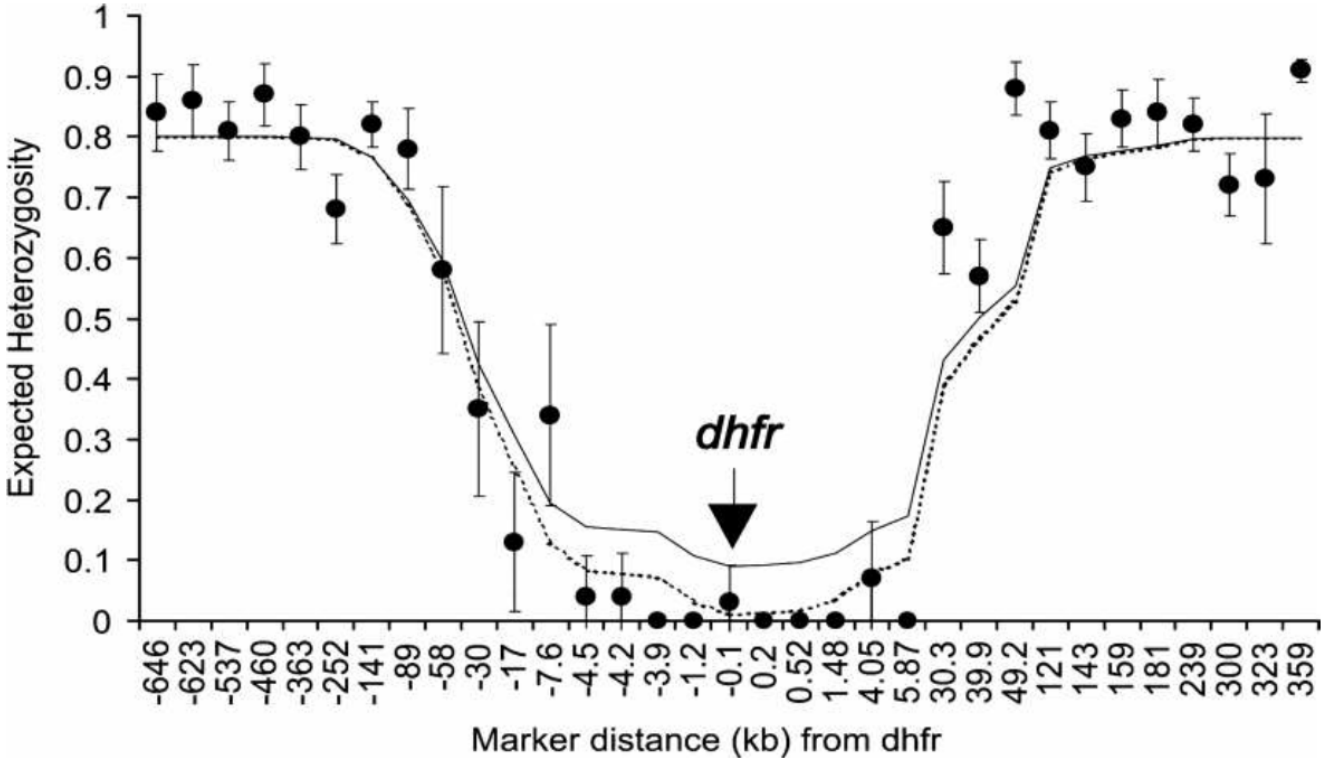


Selective sweep / balayage sélectif

Nielsen Annu. Rev. Genet. 2005. 39:197–218

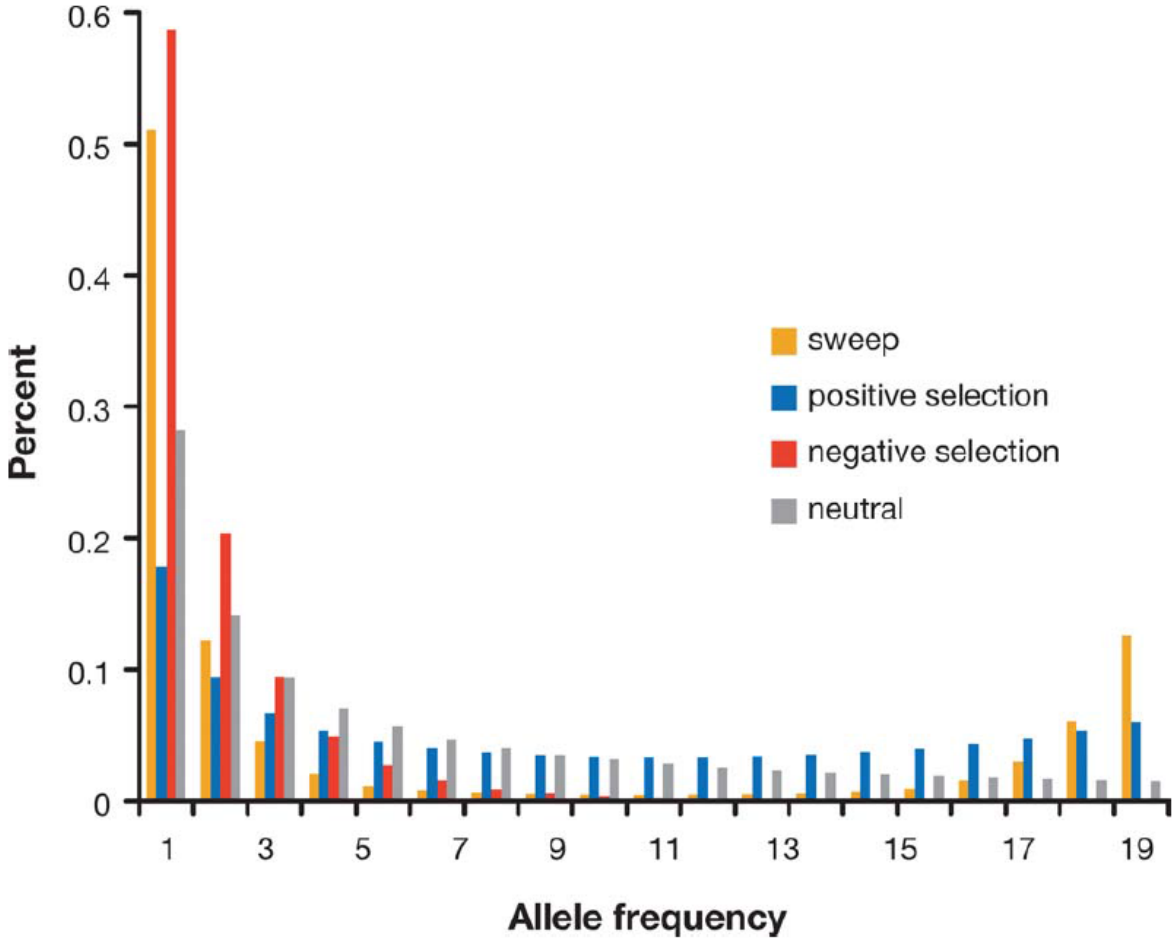
EFFET DE LA SÉLECTION SUR LES PARAMÈTRES DE DIVERSITÉ

Balayage sélectif associé à la fixation d'un allèle de résistance aux médicaments antimalarique au gène *dhfr* chez le parasite *Plasmodium falciparum*.



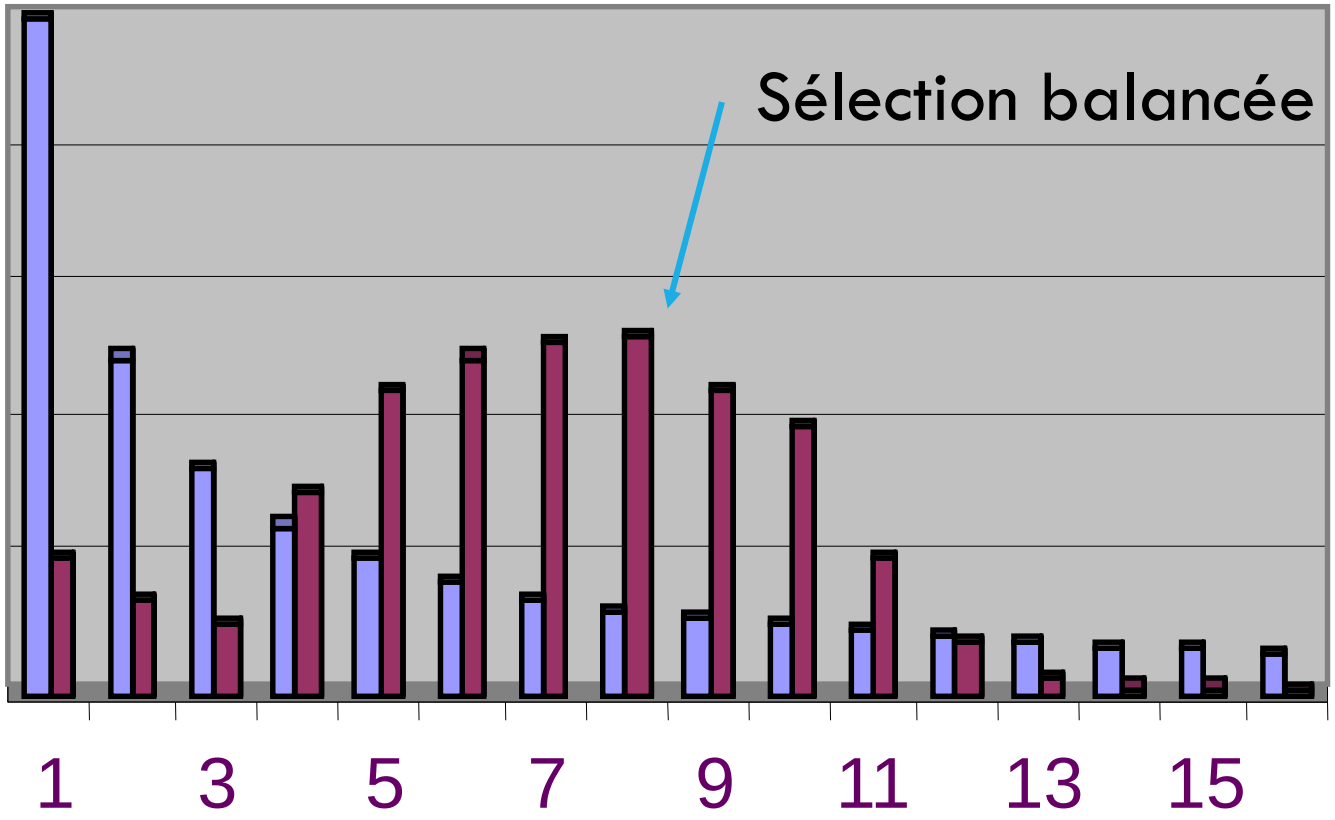
La variabilité (marqueurs microsatellites) est réduite dans un zone de100 kb autour du locus dhfr (chromosome 4 ; populations de la frontière Thailand-Myanmar).
Nair S et al. Mol Biol Evol 2003;20:1526-1536

EFFET DE LA SÉLECTION SUR LE SFS

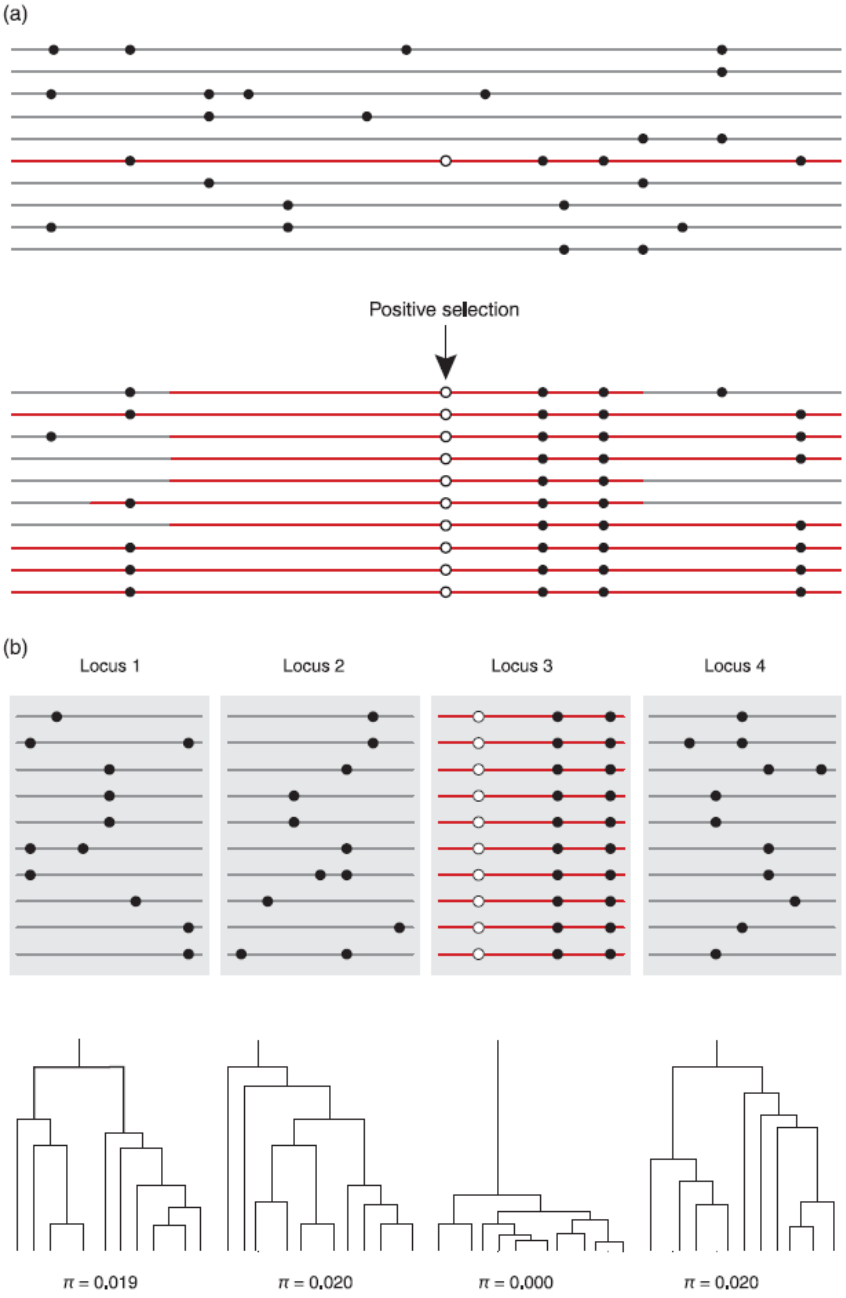


Nielsen Annu. Rev. Genet. 2005. 39:197–218

EFFET DE LA SÉLECTION SUR LE SFS



EFFET DE LA SÉLECTION SUR L'ARBRE DE COALESCENCE

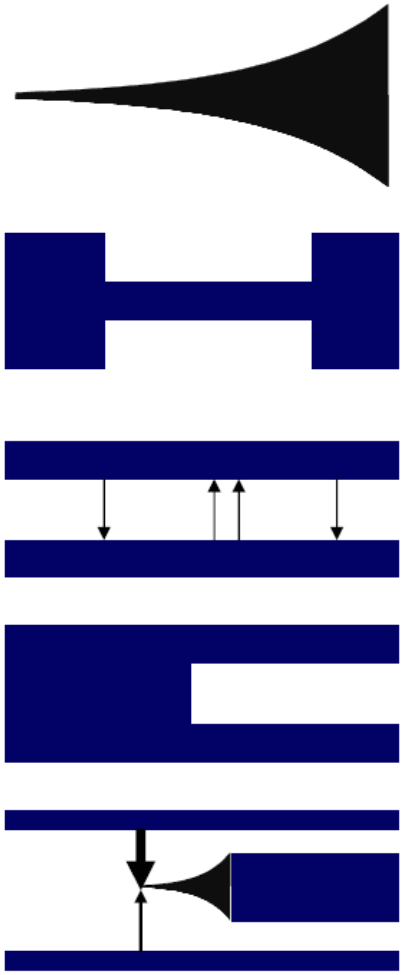


Storz. 2005 Molecular Ecology **14** , 671–688

EFFETS DÉMOGRAPHIQUES ET SÉLECTION

Demographic models

- Population growth
- Population bottlenecks
- Subdivided populations
- Population splits
- Admixture



PAUSE

SOMMAIRE

1. Introduction : définitions et objectifs des scans génomiques
2. Rappels de génétique des populations : statistiques de diversité appliquées aux séquences
3. La sélection et ses effets sur les séquences ADN
4. **Outils pour la détection de traces de sélection**
5. La génomique du paysage
6. La validation des résultats de scans génomiques

OUTILS POUR LA DÉTECTION DE TRACES DE SÉLECTION

Tests basés sur la macro-évolution

Tests basés sur la micro-évolution

- Indices de diversité et SFS
- Déséquilibre de liaison
- Différentiation entre populations

Vitti et al 2013 Annu. Rev. Genet. 47:97-120

TESTS BASÉS SUR LA MACRO-ÉVOLUTION

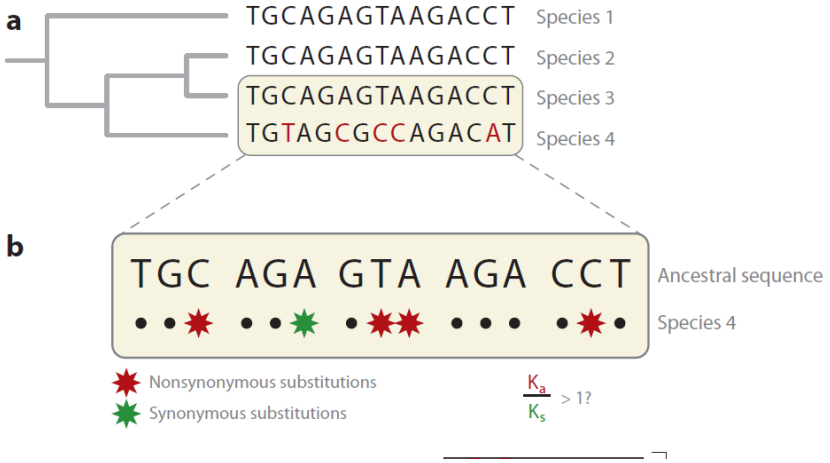
- Détection d'évènement de sélection passée
- Basés sur la comparaison de séquences (gènes) entre différentes espèces (voisines)

Ka/Ks (dN-dS) Rapport ou soustraction

$$\frac{\text{Nombre de substitutions* non synonymes par site non-synonyme } (Ka)}{\text{Nombre de substitution synonyme par site synonyme } (Ks)}$$

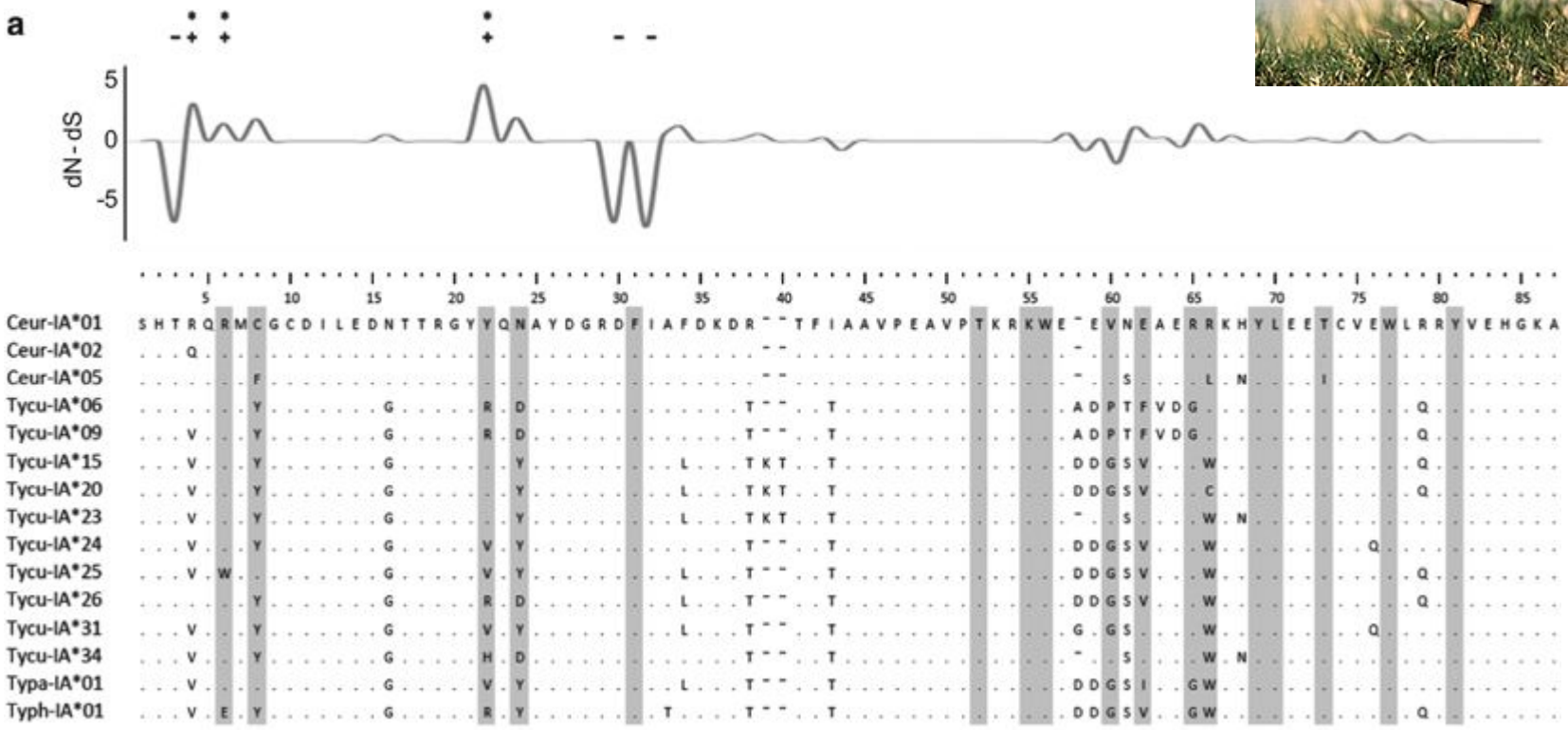
Ka/Ks = 1 pas de sélection
>1 Sélection positive
<1 Sélection négative

Définition substitution ?



Hughes AL, Nei M. (1988). *Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature* **335**: 167–170.
Hurst 2002 TIG 18(9): 486 ;
Vitti et al 2013 Annu. Rev. Genet. 47:97-120

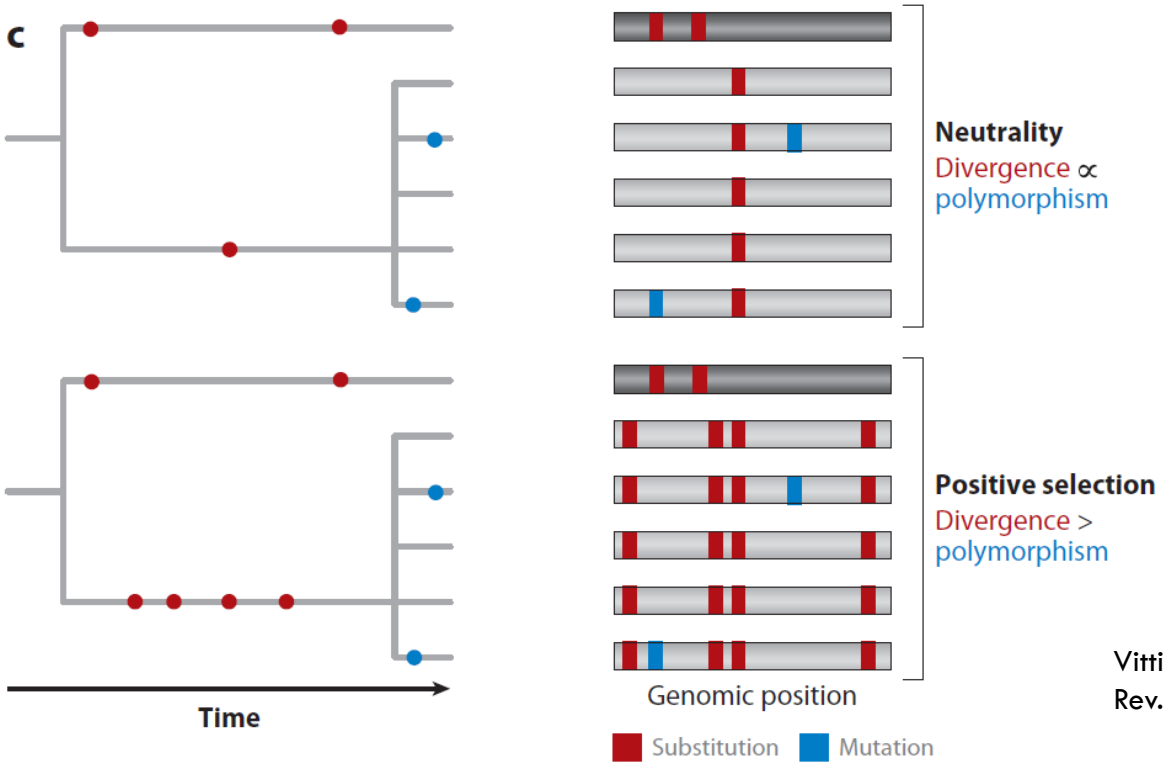
TESTS BASÉS SUR LA MACRO-ÉVOLUTION



Minias et al Heredity **volume 116,** pages 466–476 (2016)

TESTS BASÉS SUR LA MACRO-ÉVOLUTION

Tests de Hudson-Krietman-Aguadé (HKA)
Tests de MacDonald Krietman (MKT)
Nombre de substitutions interspécifiques et nb de polymorphismes intraspécifiques corrélés



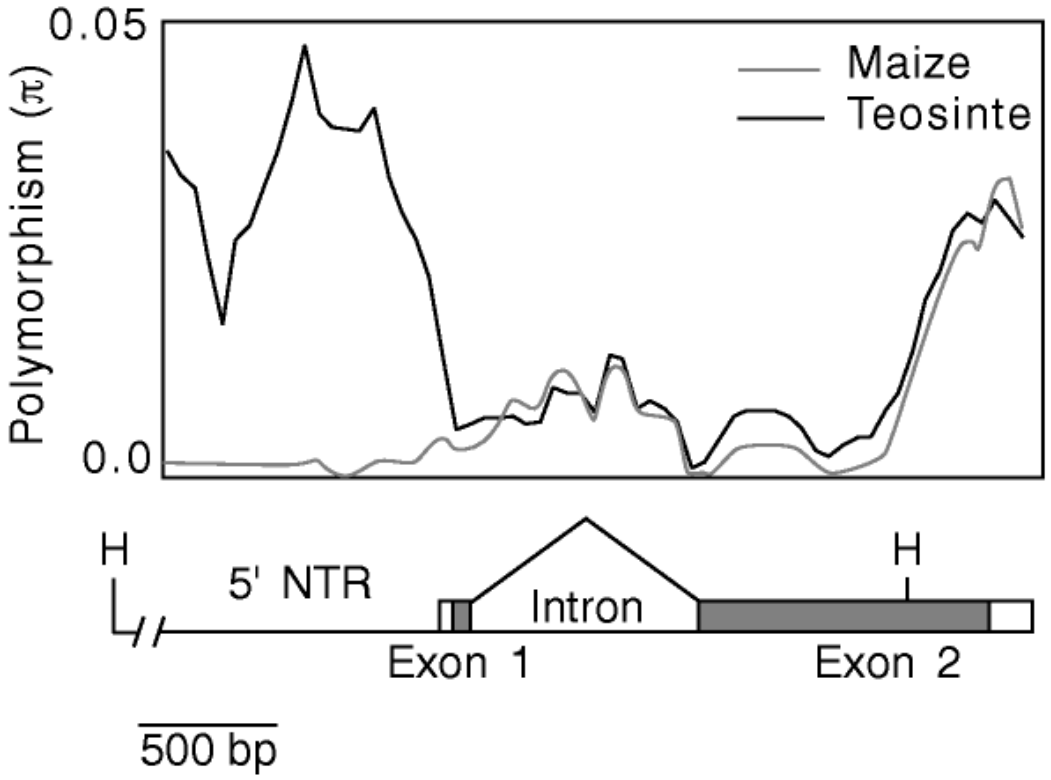
Vitti et al 2013 Annu. Rev. Genet. 47:97-120

Refs :
Hudson, R. R., M. Kreitman and M. Aguadé, 1987. A test of neutral molecular evolution based on nucleotide data. Genetics 116: 153–159
McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. Nature. 1991;351:652–654.

TESTS BASÉS SUR LA MACRO-ÉVOLUTION

Exemple: Sélection sur le gène *branched1* de la teosinte durant la domestication du maïs

- Dans la zone non traduite, π est plus faible chez le maïs que chez la teosinte
- À la limite NTR & TU: π chute pour la teosinte
- Après le codon stop, π remonte pour la teosinte et le maïs.



→**Test HKA**: Comparer le polymorphisme et la divergence des différentes partie du gène avec d'autre gènes « neutres ».

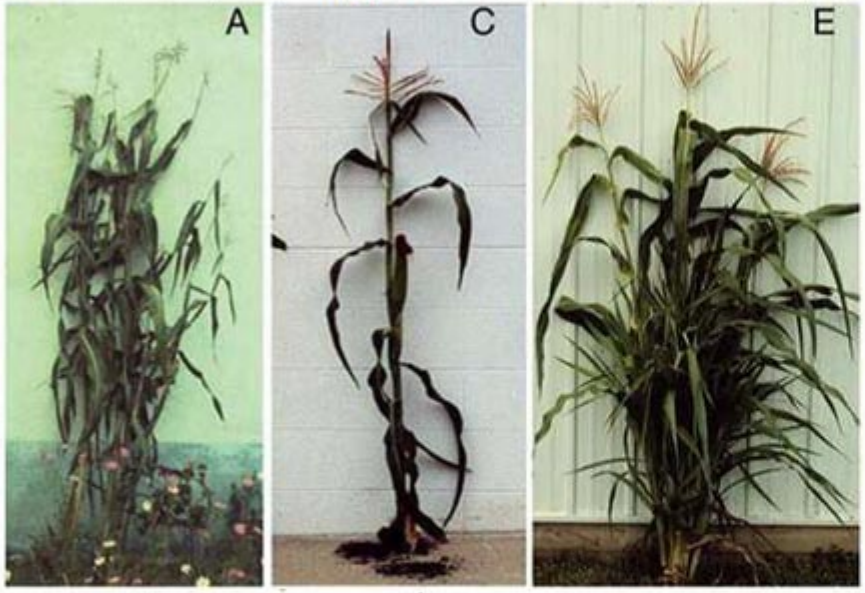
Test loci	*tb1-NTR*	*tb1-TU*	*tb1-NTR*
Control	*adh1,2*	*adh1,2*	*tb1-TU*
X^2	**13.58**	2.7	**8.24**
Proba	0.001	0.26	0.004

(Wang et al. Nature 1999. 398:236-239)

TESTS BASÉS SUR LA MACRO-ÉVOLUTION

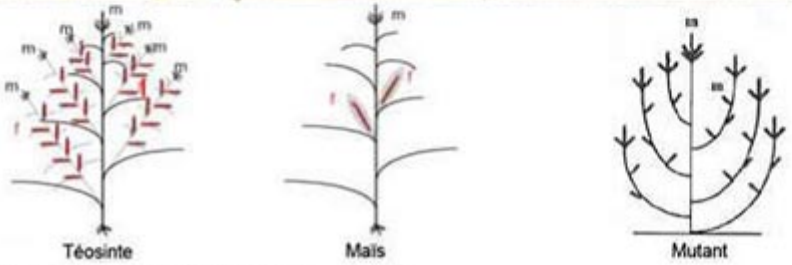
Un mutant de maïs : le mutant tb1 (teosinte branched 1)

Les chercheurs ont isolé un mutant du maïs présentant une architecture étrange. Les images en montrent les caractéristiques par rapport à un pied de maïs normal et par rapport à la téosinte.



A : plant de téosinte
C : Pied de maïs normal
E : Pied de maïs mutant (tb1)

REMARQUE : Le croisement entre maïs normal et mutant (jouant le rôle de plante mâle) engendre une population d'hybrides F1 ayant tous le phénotype normal. En F2 (F1 x F1) sur 99 plantes, 72 ont le phénotype normal et 27 le phénotype mutant.

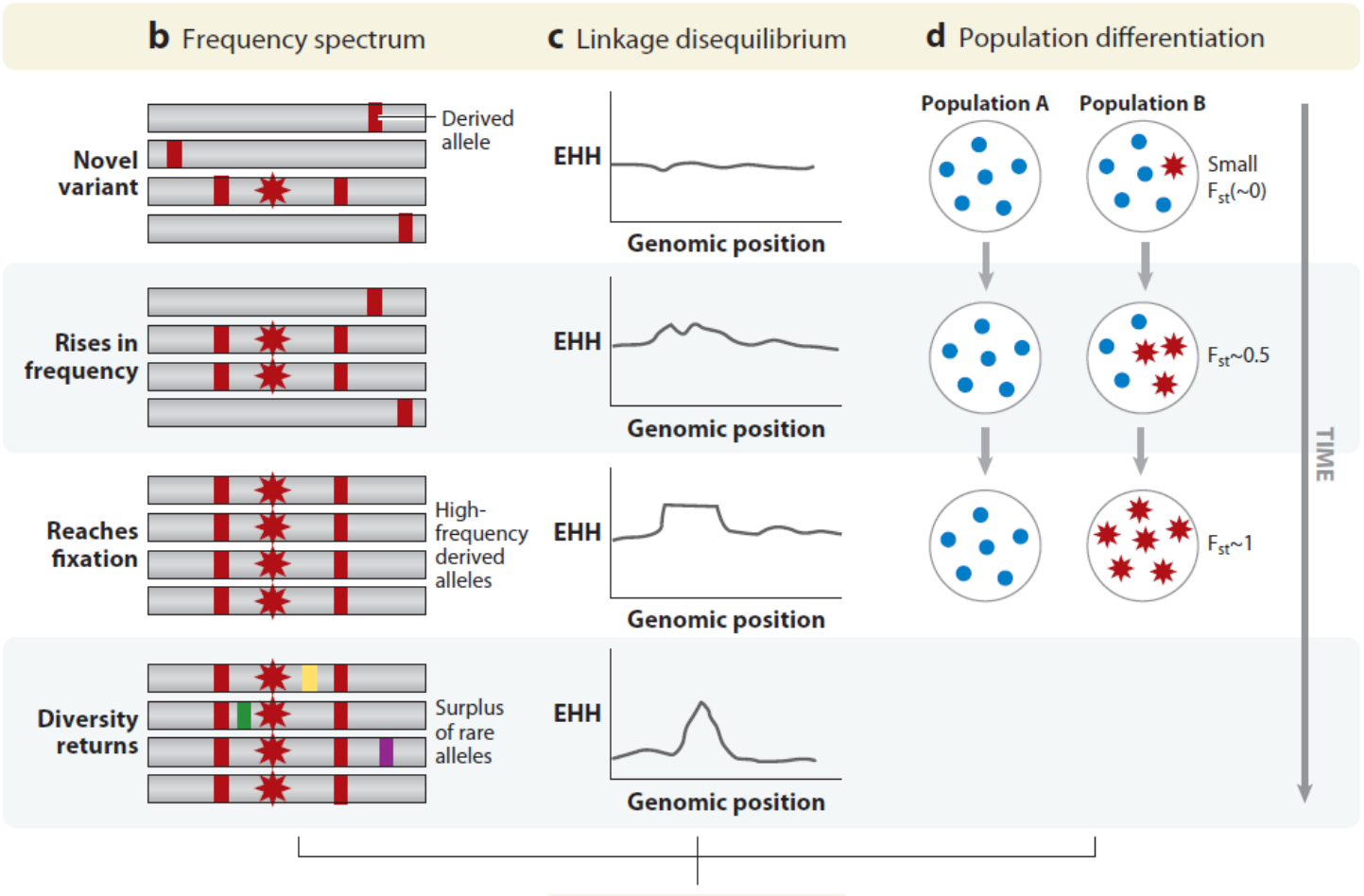


NB : m= inflorescence mâle, f= inflorescence femelle

Les différences principales entre les 3 types de plantes sont schématisées ci-contre.

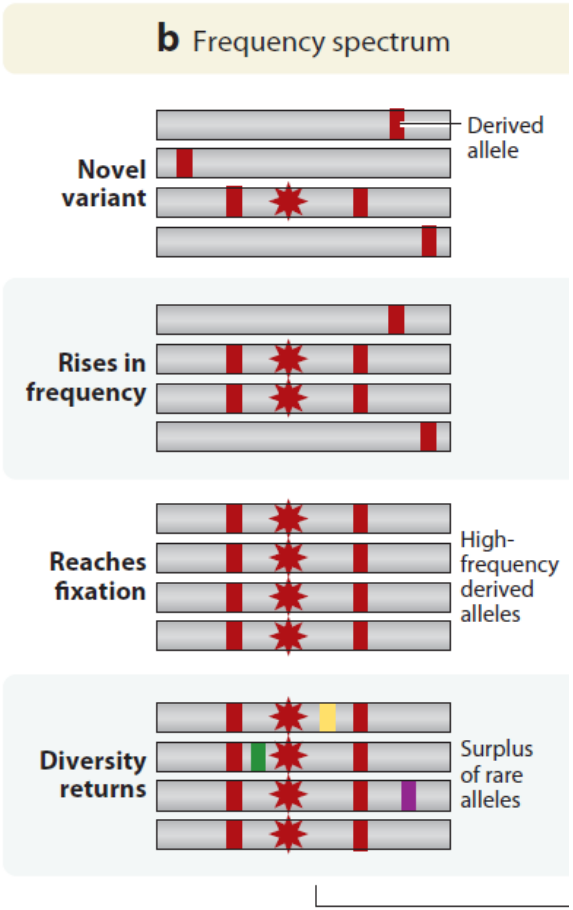
Source : The genetics of Maize evolution .John Doebley Annual review of genetics 2004 38.37-59

TESTS BASÉS SUR LA MICROÉVOLUTION



Vitti et al 2013 Annu. Rev. Genet. 47:97-120

TESTS BASÉS SUR LA MICROÉVOLUTION



Vitti et al 2013 Annu. Rev. Genet. 47:97-120

- ✓ Un allèle sélectionné
- ✓ Entrainement par balayage sélectif des allèles voisins
- ✓ Réduction de la diversité autour du locus
- ✓ De nouvelles mutations apparaissent, créant un surplus d'allèles rares

Une statistique capte ce signal:
Le D de Tajima
Tajima 1989 Genetics 123: 585-595

TESTS BASÉS SUR LA MICROÉVOLUTION

Test du D de Tajima (1989)

Basé sur la comparaison de 2 estimateurs $\vartheta = 4Ne\mu$:

- ϑ_S, basé sur le nb de sites polymorphes **S**
- ϑ_π, basé sur le nb moyen de différences entre paires d'haplotypes (séquences) → hétérozygotie au niveau nucléotidique.

$$D = \frac{\hat{\theta}_\pi - \hat{\theta}_S}{\sqrt{\mathrm{Var}(\hat{\theta}_\pi - \hat{\theta}_S)}}$$

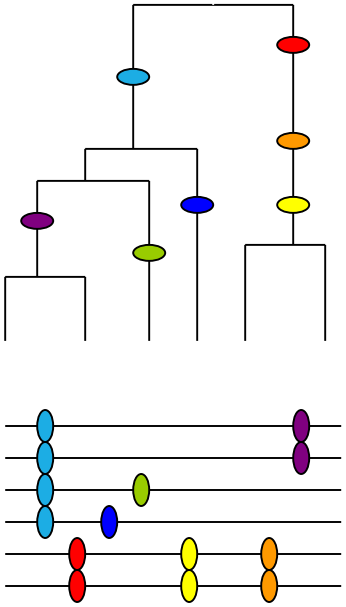
D Standardisé / déviations standard de la différence

- Comme S est plus sensible que π aux allèles rares, s'il y a un excès de mutations rares → D < 0 => sélection directionnelle ou expansion
- Réciproquement, π est affecté par allèles à fréquence intermédiaire. → si D> 0 → excès d'allèle à freq. Intermédiaire => sélection balancée ou bottleneck.
- Des simulations démographiques peuvent résoudre le problème.

TESTS BASÉS SUR LA MICROÉVOLUTION

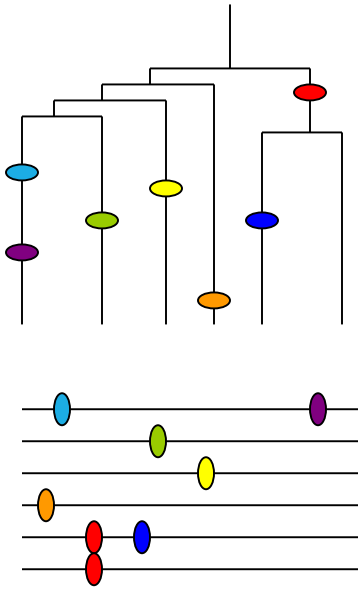
Une population en croissance présentera aussi un excès d'allèles rares

Modèle neutre standard

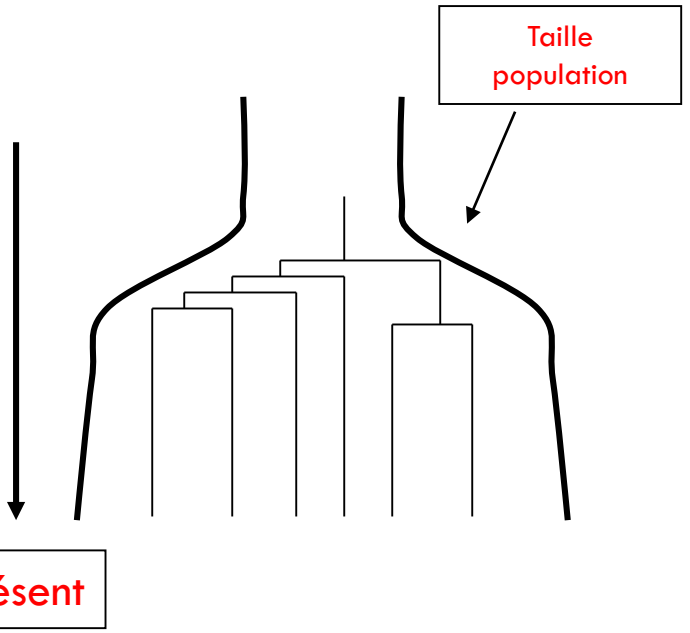


Souvent deux haplotypes fréquents, quelques mutations rares

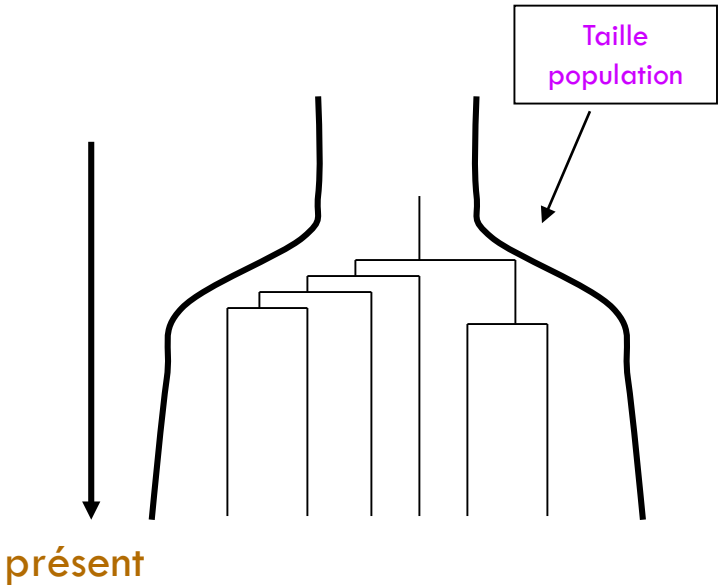
Augmentation de la taille de population



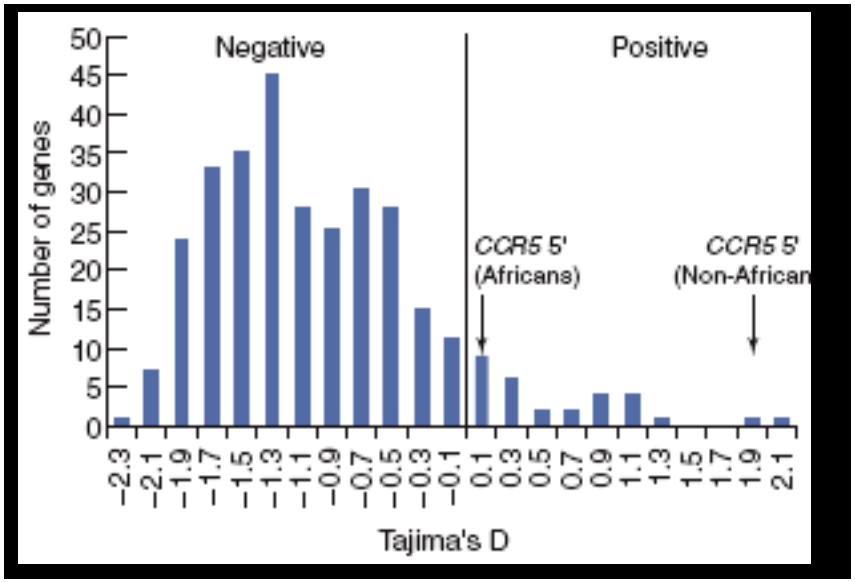
La plupart des allèles sont rares



TESTS BASÉS SUR LA MICROÉVOLUTION



→ D négatif en moyenne dans le génome humain de nombreuses populations



Ref ?

TESTS BASÉS SUR LA MICROÉVOLUTION

Autres tests basés sur les distributions des fréquences alléliques

- **Test basé sur les statistiques *Fs* (Fu 1997)** → $\vartheta\pi$ est utilisé pour calculer la probabilité d'observer k ou plus d'allèles dans l'échantillon

$$S' = \Pr(K \geq k_{obs} \mid \theta = \hat{\theta}_\pi) = 1 - \sum_{k=1}^{k_{obs}-1} \frac{\left|S_n^k\right| \hat{\theta}_\pi^k}{S_n(\hat{\theta}_\pi)}$$

$$F_S = \ln\left(\frac{S'}{1-S'}\right)$$

Interpretation: si K_{obs} *(le nb d'haplotypes observés)* > K attendu, S' sera petit et $F_S < 0$ << excès d'allèles rares.

- Famille de tests basés sur le nombre de mutations dérivées avec une fréquence particulière dans l'échantillon (Fu 1995):

$$E[\xi i] = \theta / i \rightarrow \theta_H$$

→ *H* test of Hitch-hiking from Fay et Wu (2000)

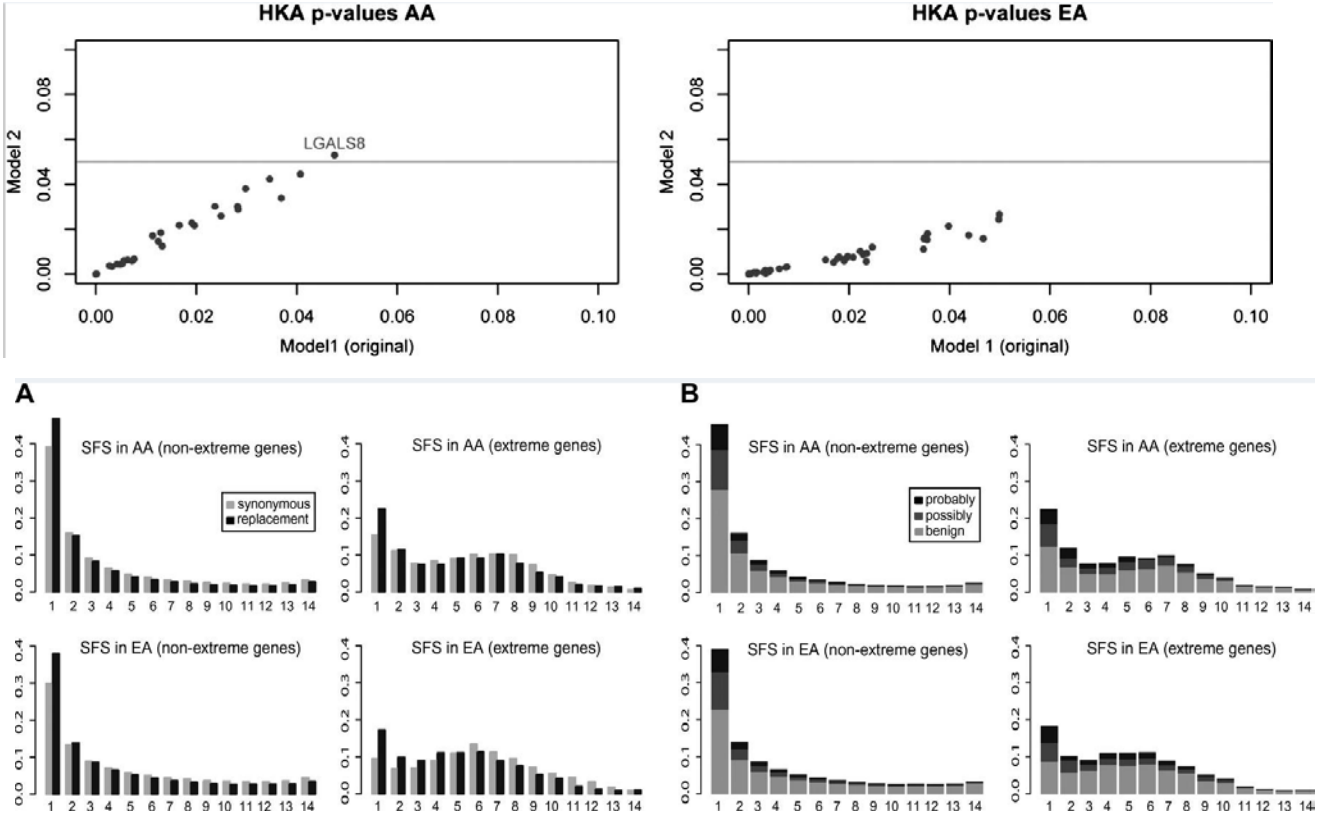
$$H = (\pi - \theta_H) / \text{SQRT}(\text{Var}(\pi - \theta_H))$$

S'il y a un excès de variants derivés, la valeur observée de la $H < 0$

→ hitch-hiking dans un passé récent

TESTS BASÉS SUR LA MICROÉVOLUTION

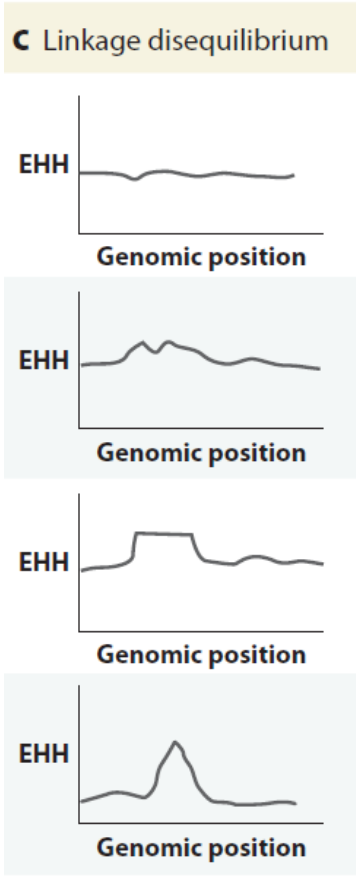
Combinaisons d'approches



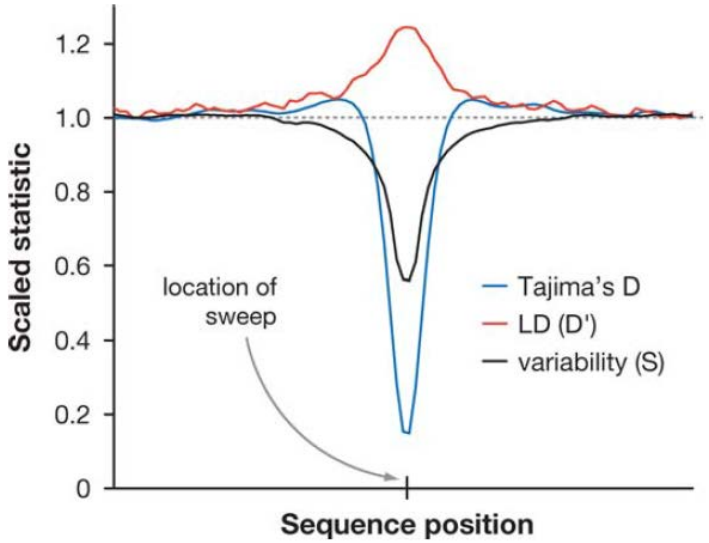
Andrés AM, Hubisz MJ, Indap A, Torgerson DG, Degenhardt JD, et al. 2009. Targets of balancing selection in the human genome. Mol. Biol. Evol. 26(12):2755–64

TESTS BASÉS SUR LA MICROÉVOLUTION

Approches basées sur le LD



Hypothèse :
Le balayage sélectif « entraine » les allèle voisins
Le DL persiste tant que la recombinaison n'intervient pas. (temps, distance)



Nielsen Annu. Rev. Genet. 2005. 39:197–218

Le « plus » de l'approche :
Détection d'évènements de balayage sélectif partiel ou récents.

TESTS BASÉS SUR LA MICROÉVOLUTION

Approches basées sur le LD

EHH: Extended Haplotype Homozygosity
Probabilité que 2 chromosomes tirés au sort dans une population portant la région « cœur » soit identiques par filiation (Identity By Descent – IBS)

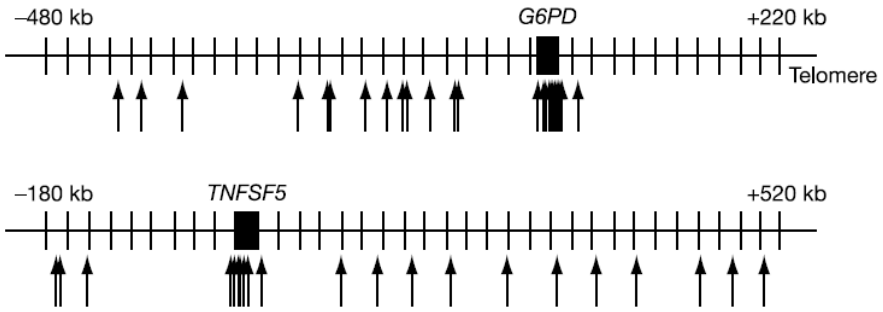
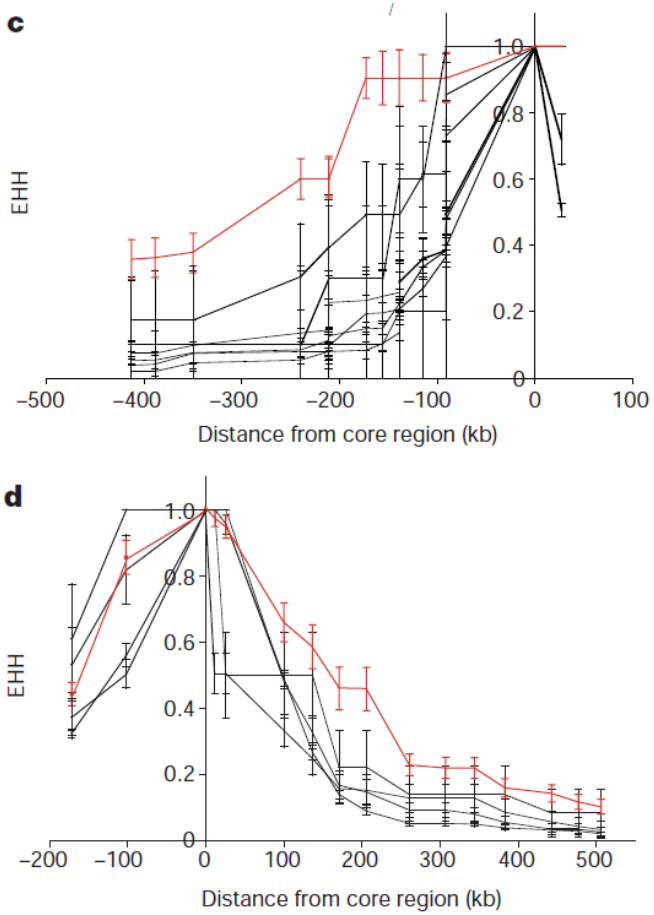


Figure 1 Experimental design of core and long-range SNPs for *G6PD* and *TNFSF5*. The core region is highlighted by a cluster of densely spaced SNPs (arrows) at the gene. Additional, widely separated flanking SNPs, used to examine the decay of LD from each core haplotype, are also shown. Markers distal to *G6PD* were within repetitive subtelomeric sequence and could not be genotyped.



Méthode développée sur des populations humaines / résistance à la malaria.
Sabeti et al 2002 Nature 419: 832-837

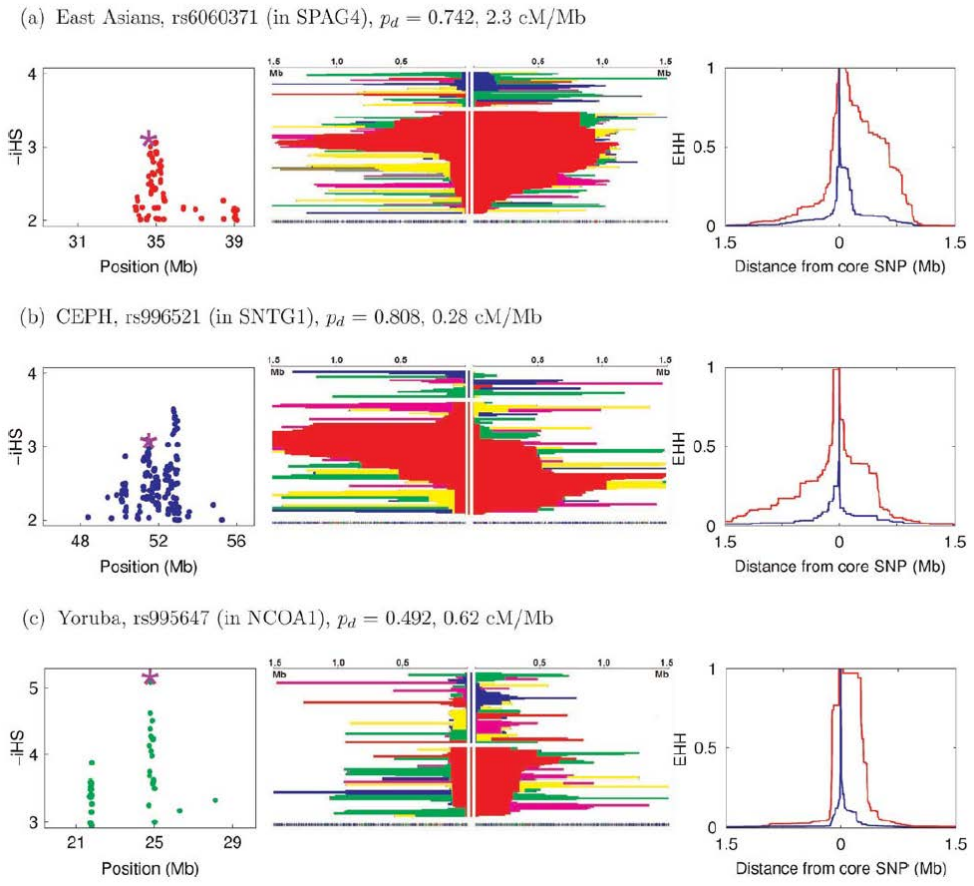
TESTS BASÉS SUR LA MICROÉVOLUTION

Approches basées sur le LD

iHS

Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4(3):e72

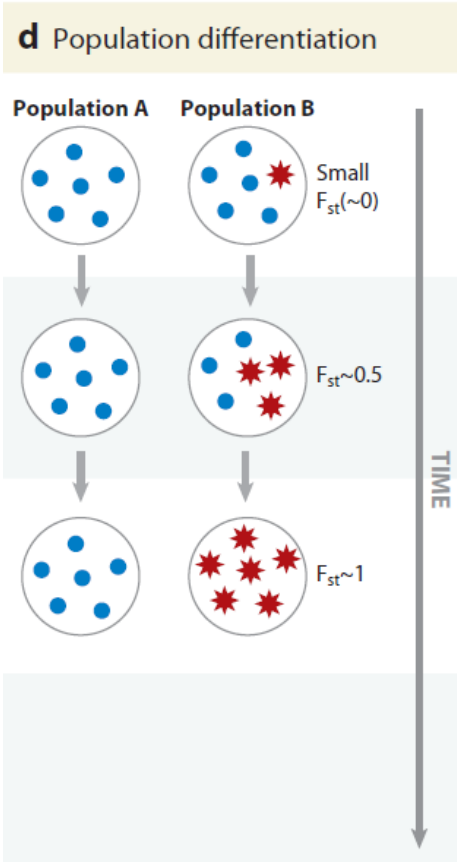
Intégrale de EHH (aire sous la courbe)



Autres méthodes

LRH, XP-EHH, LDD => voir Vitti et al 2013

TESTS BASÉS SUR LA MICROÉVOLUTION



Vitti et al 2013 Annu. Rev. Genet. 47:97-120

Tests basés sur la différenciation entre populations

Hypothèse : les différentes populations sont soumises à des environnements différents

Méthodes avec une longue histoire :
Lewontin et Krakauer 1973. Genetics 74: 175-195

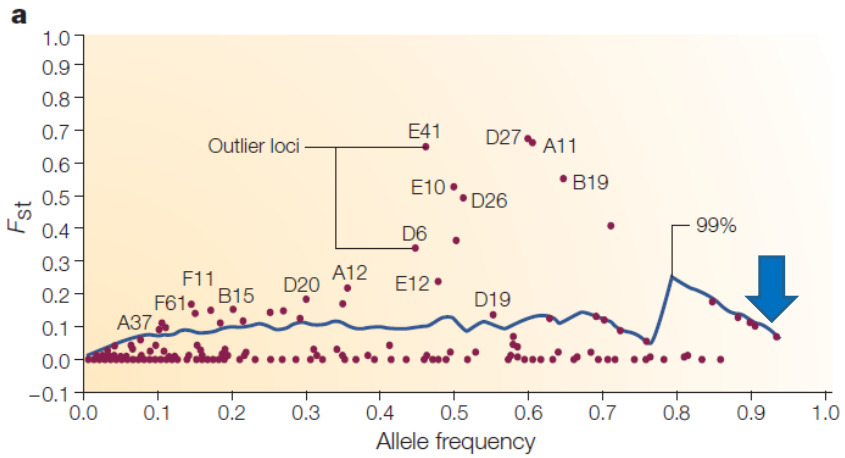
Comparer les valeurs empiriques de **Fst** le long du génome à une valeur théorique (sans sélection) obtenue par simulation.

TESTS BASÉS SUR LA MICROÉVOLUTION

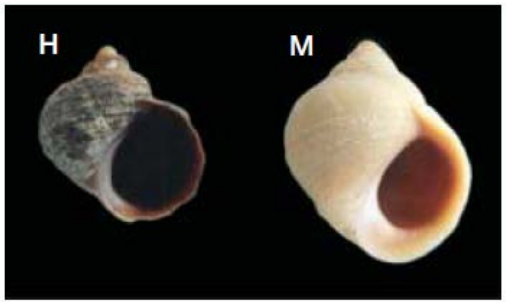
Tests basés sur la différenciation entre populations

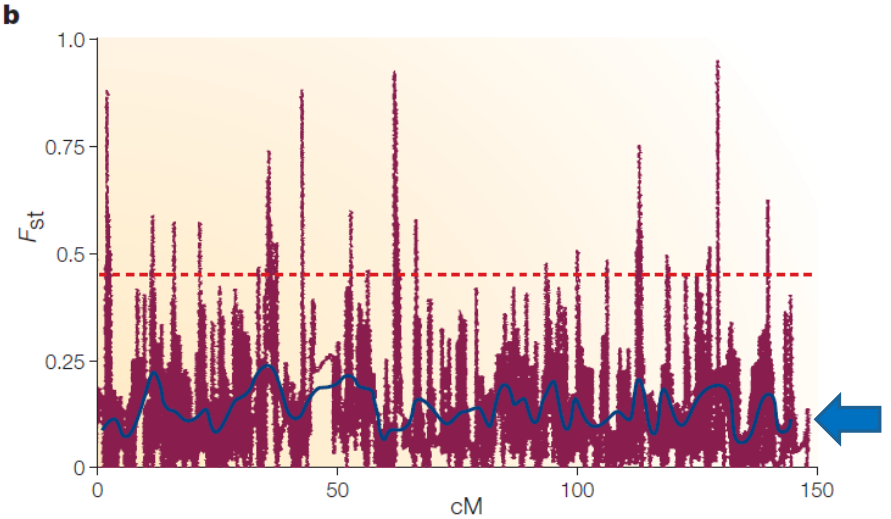
=> Détection de « **outliers** »

(d'après Luikart et al. 2003. Nature reviews Genetics)



Genome scan of Fst in *Littorina saxatilis* (Wilding et al. 2001. J Evol Biol)

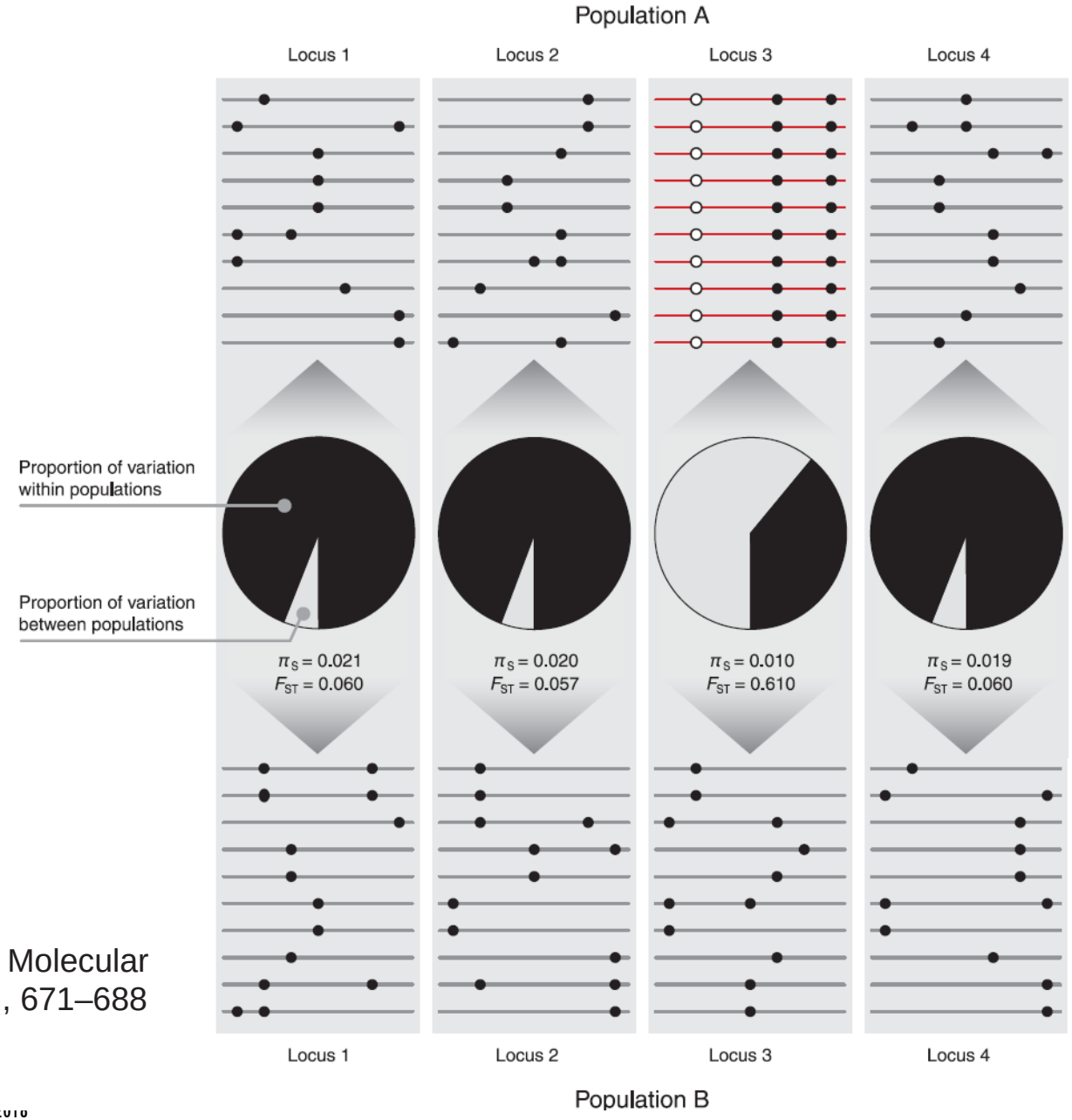




Scan of Fst at SNPs on human Chr. 8 (Akey et al. 2002 Genome Res. 12, 1805–1814)

Valeur théorique

OUTILS POUR LA DÉTECTION DE TRACES DE SÉLECTION



Storz. 2005 Molecular Ecology **14** , 671–688

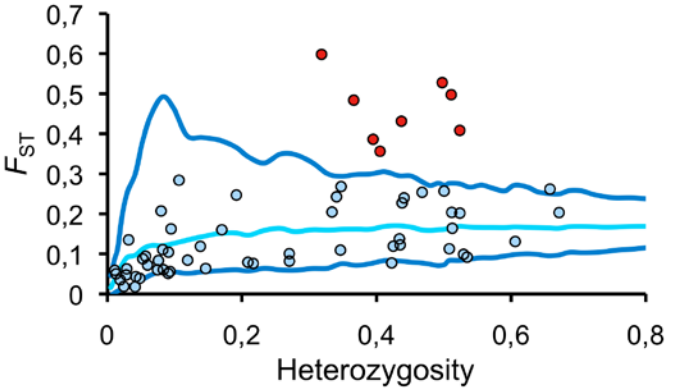
TESTS BASÉS SUR LA MICROÉVOLUTION

Tests basés sur la différenciation entre populations :

Faux positifs, effets de la démographie, de la migration, de la mutation …

- Méthodes basées sur un modèle : **Beaumont and Nichols (1996)** → program Fdist2:
  - → la distribution des F_{ST} est representée comme une fonction de l'hétérozygotie
  - modèle en nombre infini de dèmes (iles)
  - À partir d'un Fst moyen, simulations de l'enveloppe neutre et détection d' « outliers »
  - Robustesse de la méthode affectée par tx de mutation , taille échantillon, Non-équilibre, certains modèles démographiques (Stepping-Stone sauf si pop. trop proches,…)

- **Vitalis, Dawson et Boursot (2001)** → logiciel Detsel
- **Beaumont and Balding (2004)** → Extension Bayésienne de la méthode de Beaumont et Nichols mais + de flexibilité pour le modèle et les tx de migration entre populations et possibilité de distinguer effets locus et populations dans un patron atypique

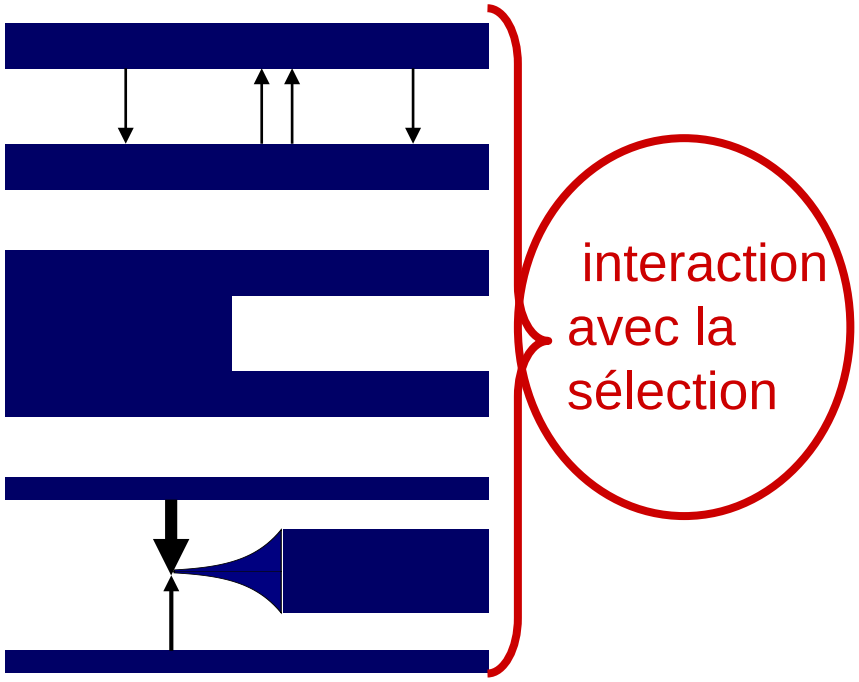


MODELS OF POPULATION STRUCTURE

Origins of genetic differenciation among populations?

There are several different ways in which populations can become genetically distinct

- Restricted gene flow
  + genetic drift
- Population splits
  +isolation
- Admixture



→ Séparer les effets de la sélection des autres processus?

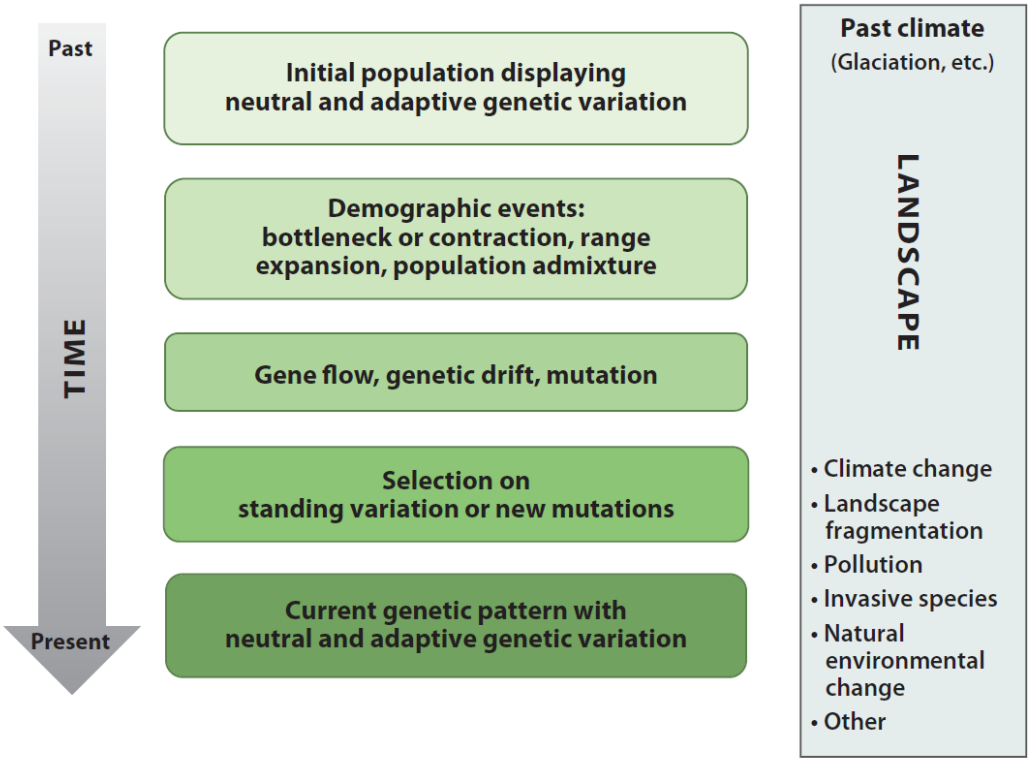
Slide modified < Gil Mc Vean

SOMMAIRE

1. Introduction : définitions et objectifs des scans génomiques
2. Rappels de génétique des populations : statistiques de diversité appliquées aux séquences
3. La sélection et ses effets sur les séquences ADN
4. Outils pour la détection de traces de sélection
5. **La génomique du paysage**
6. La validation des résultats de scans génomiques

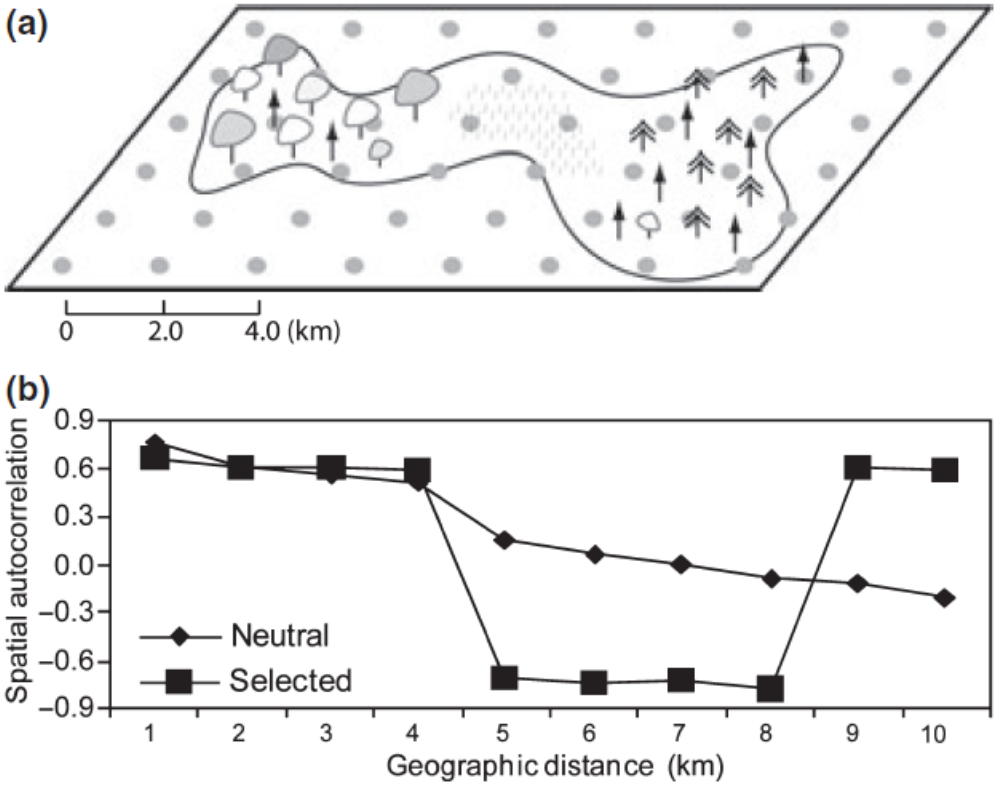
LA GÉNOMIQUE DU PAYSAGE

Ou comment le paysage et ses propriétés interagissent avec les processus de microévolution (flux de gène, dérive, sélection …) à l'échelle génomique



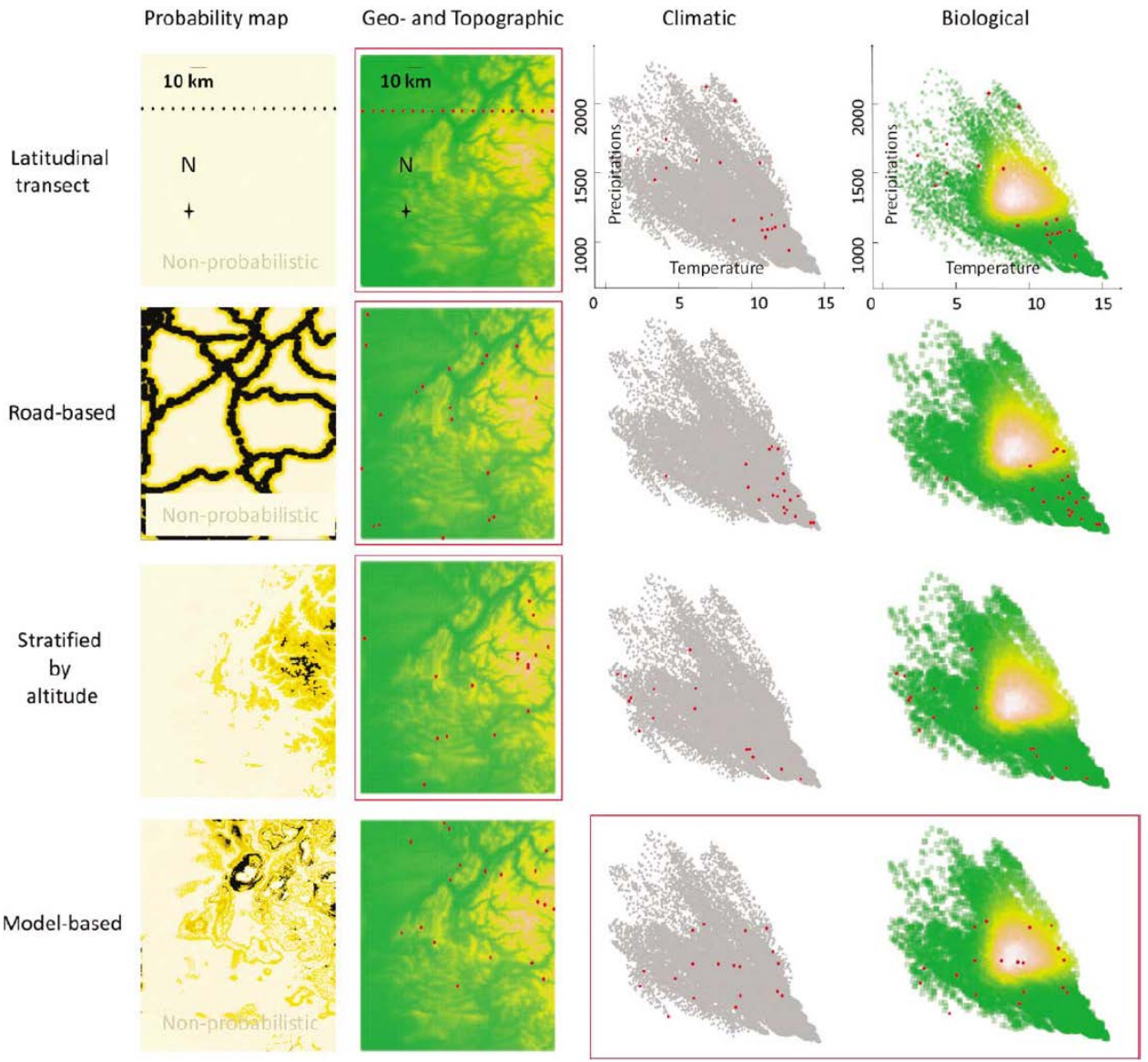
Schoville et al 2012 Annu. Rev. Ecol. Evol. Syst. 2012.43:23-43

DONNÉES SPATIALISÉES



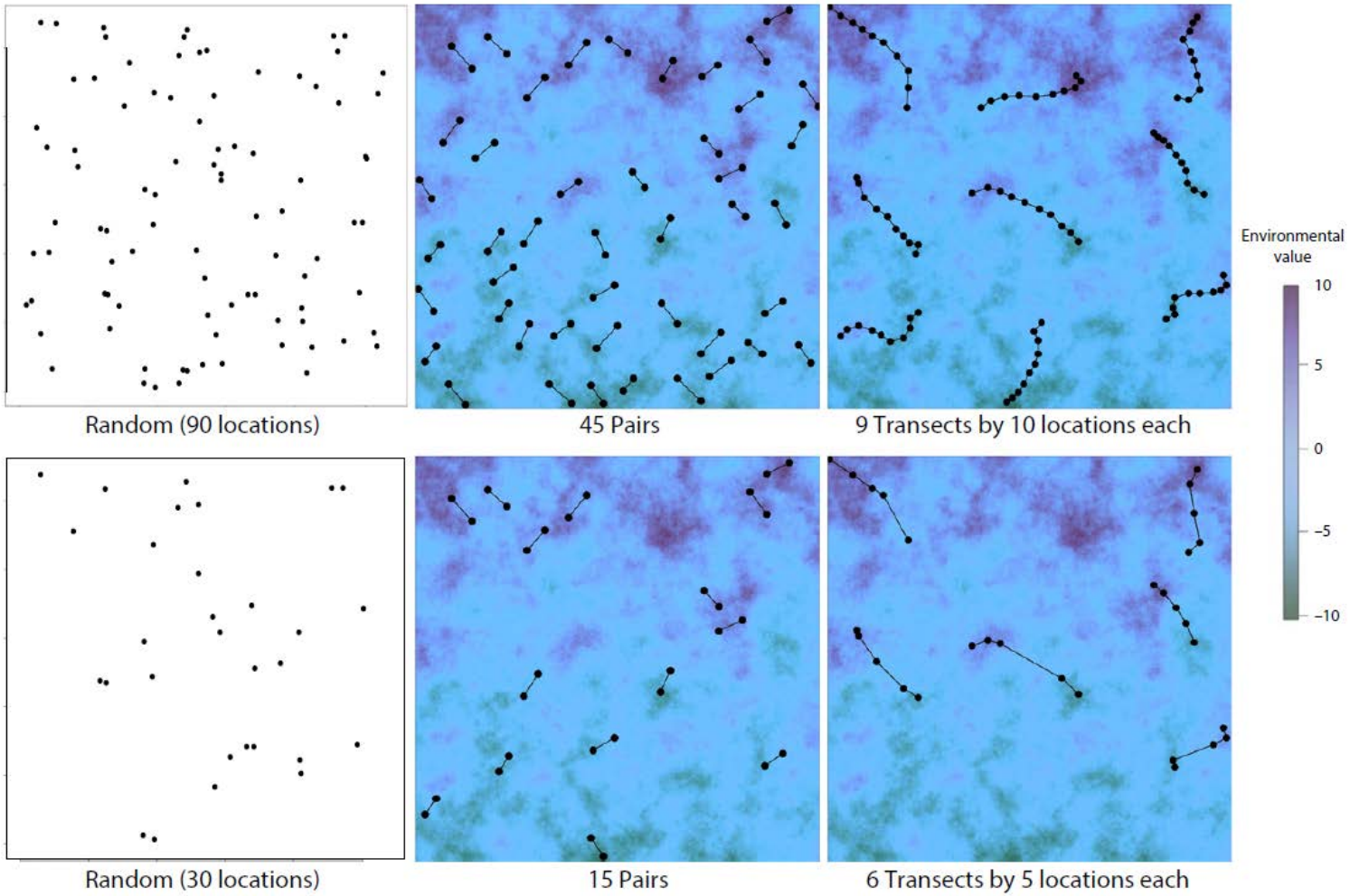
Manel et al. 2010 Molecular Ecology 19, 3760–3772

L'ÉCHANTILLONNAGE



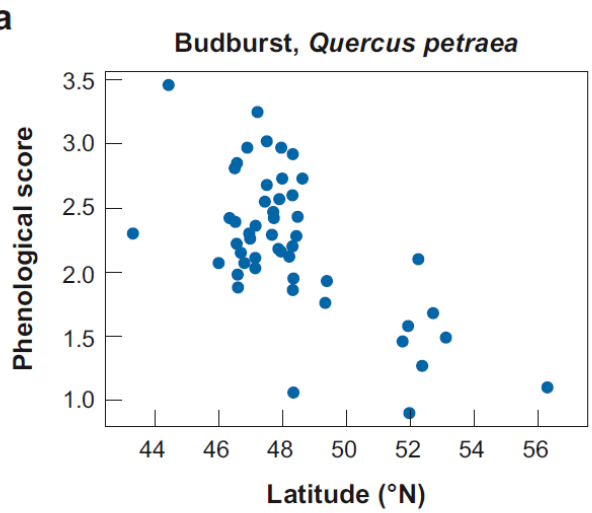
Albert et al 2010
Ecography 33: 10281037

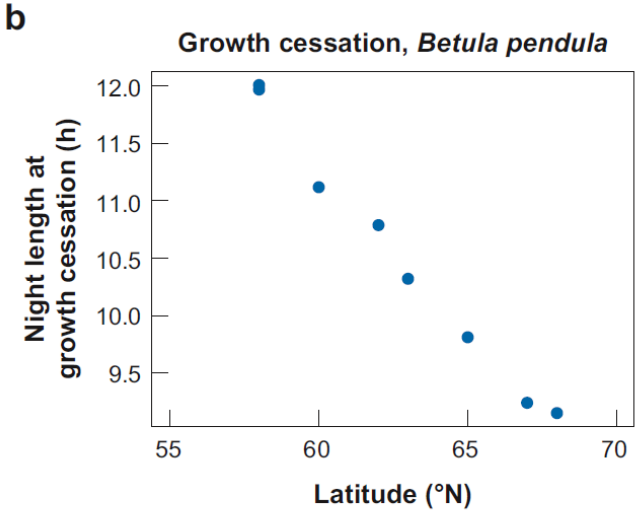
L'ÉCHANTILLONNAGE

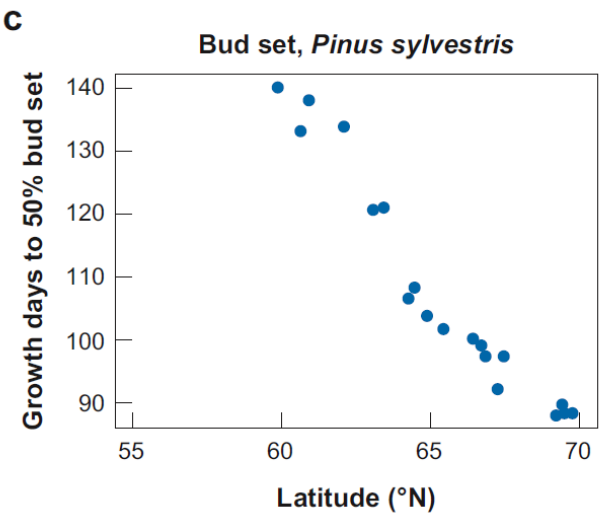


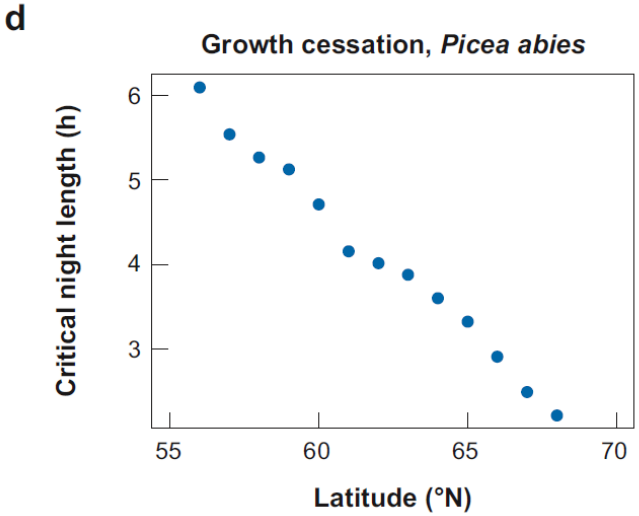
Lotterhos & Whitlock, Molecular Ecology (2015) 24, 1031–1046

Variation de caractères adaptatifs chez les arbres forestiers en fonction de la latitude









Savolainen et al 2007 Annu. Rev. Ecol. Evol. Syst. 2007. 38:595–619

LES DONNÉES ENVIRONNEMENTALES

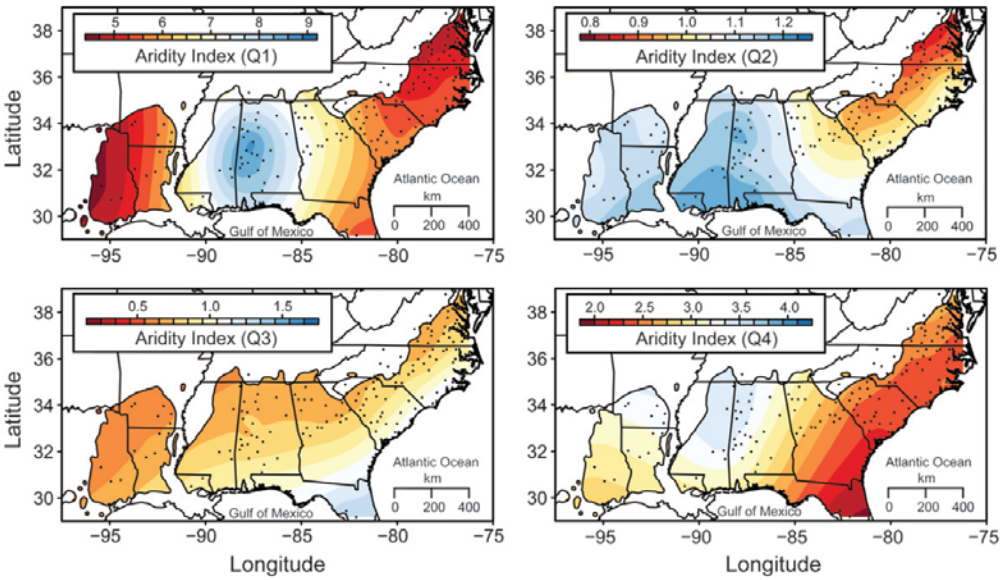
- Bases de données publiques – SIG

The Global Map Project (http://www.globalmap.org/)
WorldClim (http://www.worldclim.org)

- Mesures in situ

Quelles sont les variables pertinentes ?

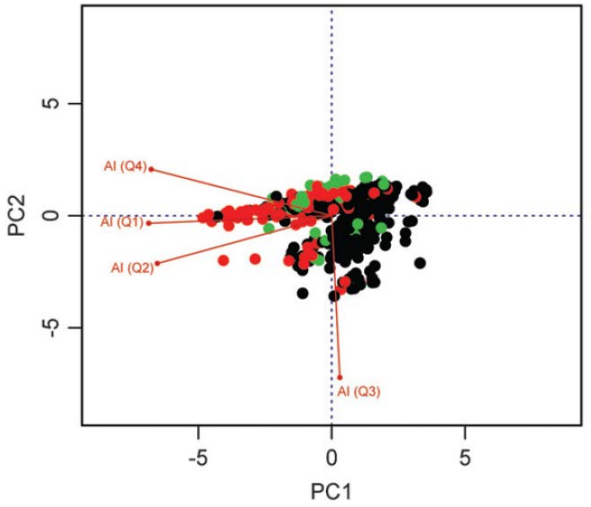
- latitude/longitude/altitude
- températures : quels modèles ?
- période ?



Variables dérivées :

- degrés.jours cumulés
- Axes ACP de toutes les variables

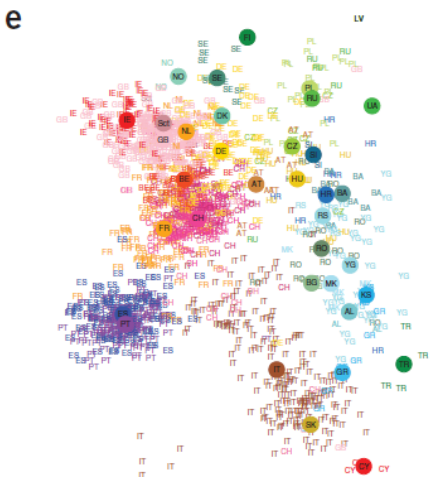
(ex: Eckert et al 2010 Genetics 185: 969–982; populations de pin taeda et index d'aridité)

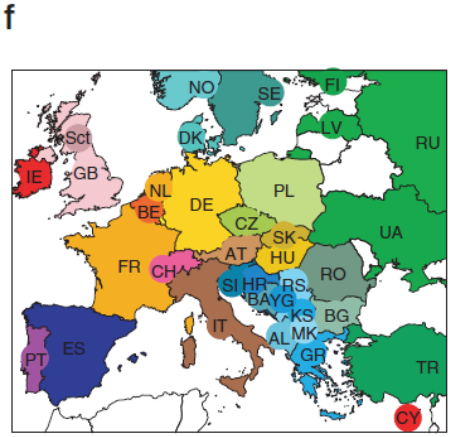


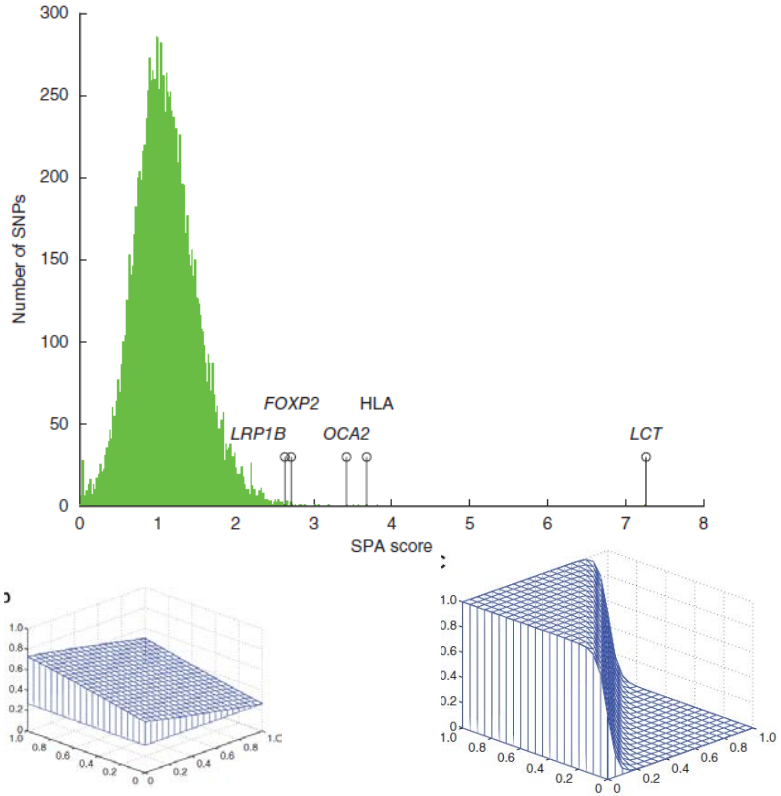
MÉTHODES

SPA –Spatial Ancestry analysis (Yang et al 2012 Nat genet)

- Modèle : la fréquence de chaque allèle est fonction de l'origine géographique (continue)
- => On peut prédire l'origine géographique des ind. avec les données génétiques
- => Les SNP avec de forts gradients de freq. allélique => candidats soumis à sélection

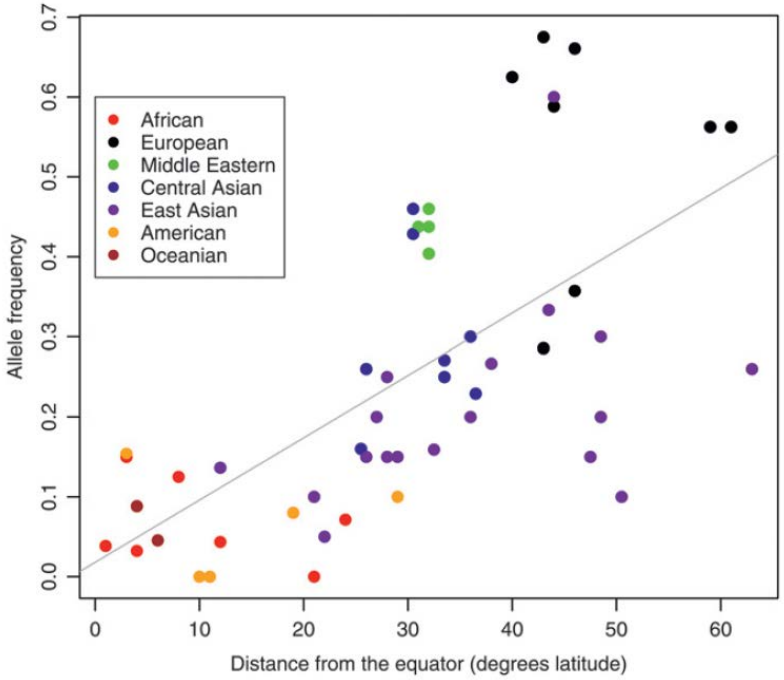






MÉTHODES

Bayenv 1.0 (*Hancock et al. 2008, PlosONE; Coop et al 2010, Genetics*)
Bayenv 2.0 (Gunther & Coop, 2013)



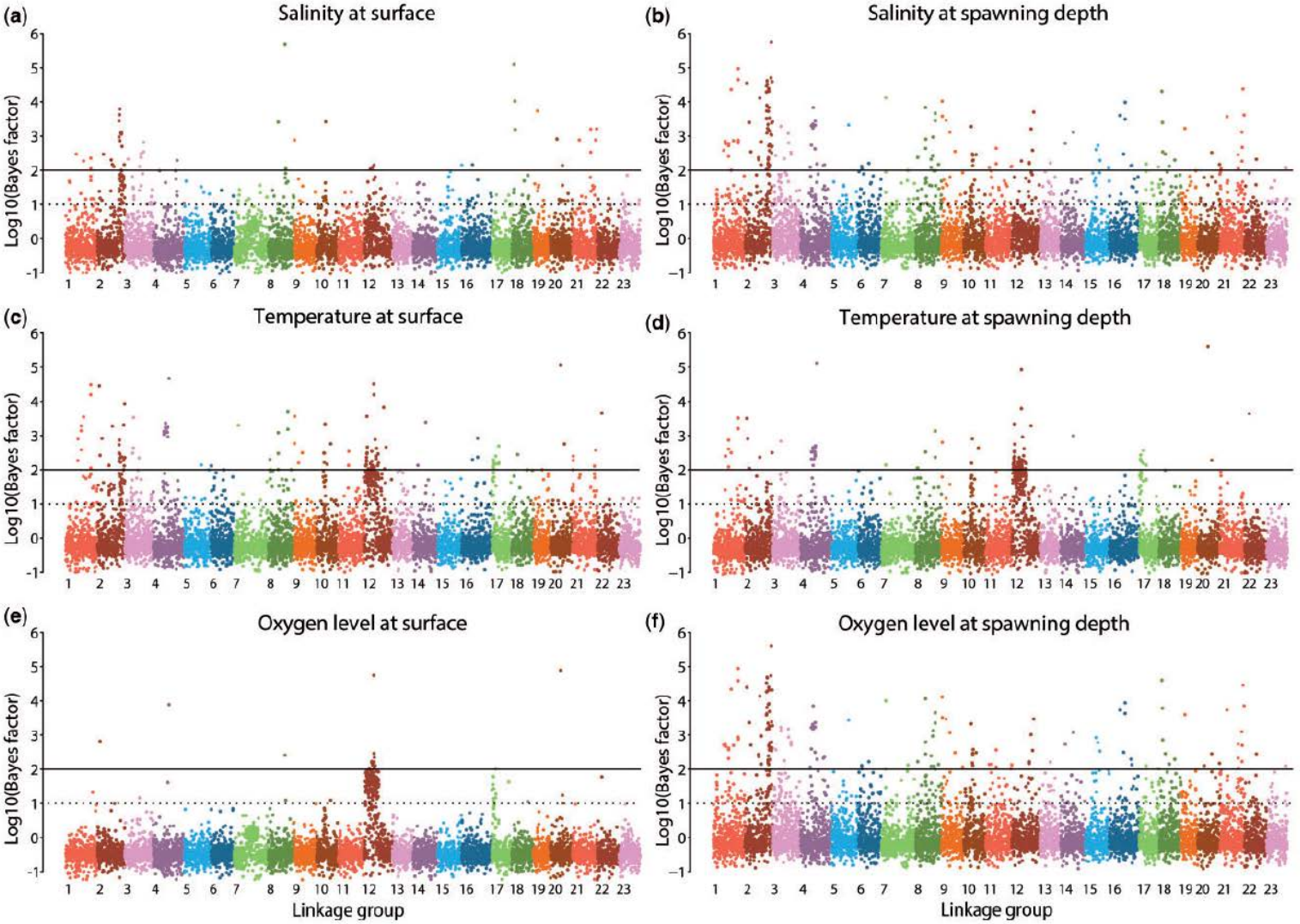
Corrélation entre fréquence allélique et variable environnementale
⇒ Effet de l'échantillonnage / ≠ tailles populations
⇒ Non indépendance entre les pop.

- Un set de SNP contrôles => modèle nul de la covariation des freq. entre pop.
- modèle bayésien
- BF = rapport de proba postérieures, [P(alt)/P(nul)].

MÉTHODES

Berg et al 2015 Genome Biol. Evol. 7(6):1644–1663

Bayenv



MÉTHODES

SAM (Spatial Analysis Method; Joost et al 2007)
Samβada (Stucki et al al 2014)

- Régression logistique (SAM, => P(Y/X))
- Plus rapide que SAM
- Analyse multivariée (incl. Structure, Q)
- Mesures d'auto-corrélation spatiale

Temps de calculs (h)

	41,215 SNPs 804 samples	634,849 SNPs 102 samples
Samβada	1.2	2.9
Samβada biv.	8.7	18.4
BayEnv	41.3	62.,2
LFMM	3.2	16.0
LFMM (mono)	6.1	58.1

LFMM (Frichot et al 2013)
SGLMM (Guillot et al 2014) …

25 loci les plus significatifs

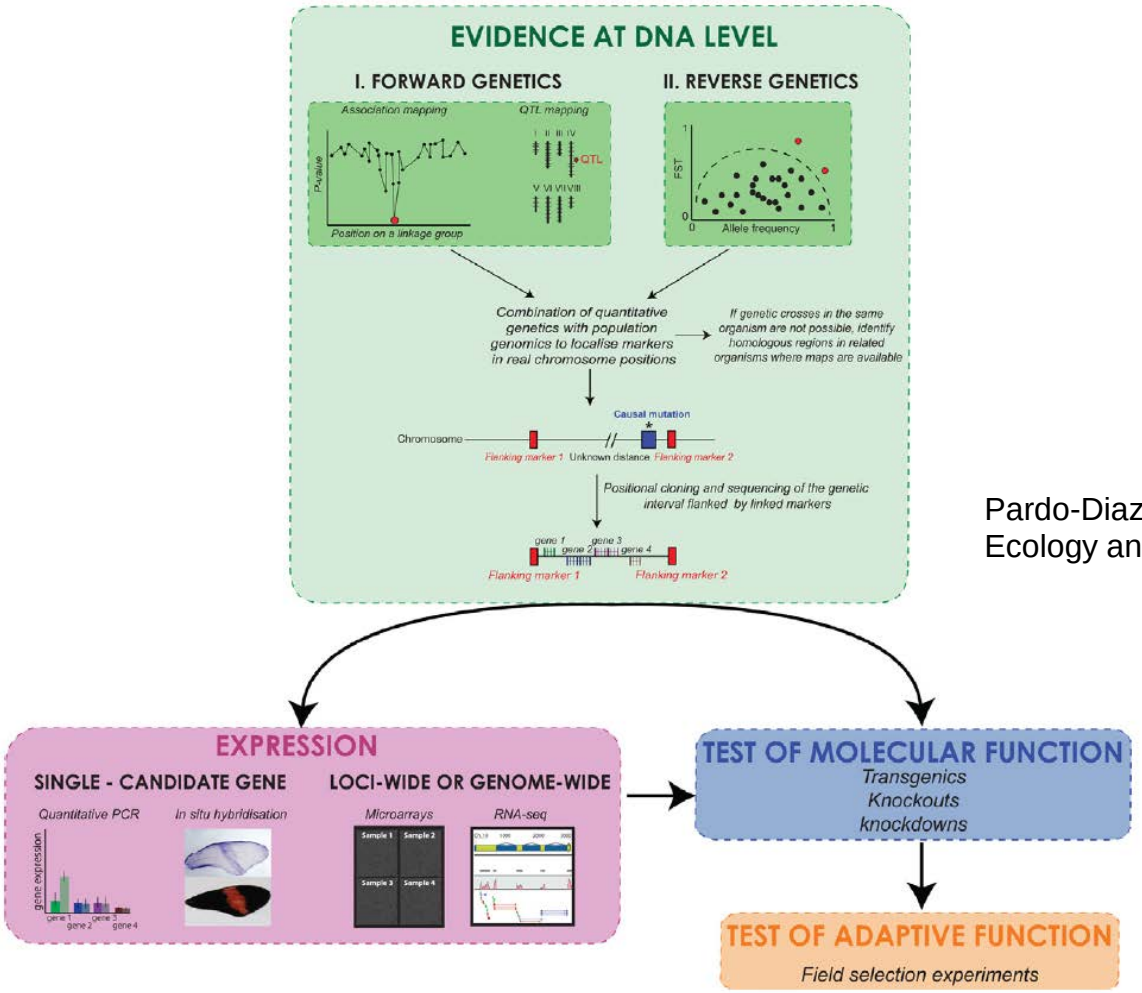
	Loci	Chr.	Pos. [Mbp]	Samβada	BayEnv	LFMM	Arlequin	Detections
1	ARS-BFGL-NGS-113888	5	48.32	1	1	0	0	2
2	Hapmap41074-BTA-73520	5	48.35	1	1	0	0	2
3	Hapmap41762-BTA-117570	5	18.94	1	1	0	0	2
4	ARS-BFGL-NGS-46098	20	2.95	1	1	0	0	2
5	Hapmap41813-BTA-27442	5	49.04	1	1	0	0	2
6	BTA-73516-no-rs	5	48.75	1	1	0	0	2
7	**Hapmap28985-BTA-73836**	**5**	**70.34**	**1**	**1**	**1**	**0**	**3**
8	Hapmap31863-BTA-27454	5	48.99	1	1	0	0	2
9	**ARS-BFGL-NGS-106520**	**5**	**70.20**	**1**	**1**	**1**	**0**	**3**
10	**BTA-73842-no-rs**	**5**	**70.18**	**1**	**1**	**1**	**0**	**3**
11	Hapmap50523-BTA-98407	5	46.74	1	1	0	0	2
12	BTB-01400776	20	2.70	1	1	0	0	2
13	Hapmap23956-BTA-36867	15	47.20	1	1	0	0	2
14	ARS-BFGL-NGS-10586	2	128.64	1	1	0	0	2
15	ARS-BFGL-NGS-43694	5	49.65	1	1	0	0	2
16	BTA-122374-no-rs	14	16.44	1	1	0	0	2
17	BTB-01356178	20	2.49	1	1	0	0	2
18	**ARS-BFGL-NGS-94862**	**11**	**103.53**	**1**	**1**	**1**	**0**	**3**
19	BTA-108359-no-rs	14	16.31	1	0	0	0	1
20	ARS-BFGL-NGS-15960	5	28.02	1	1	0	0	2
21	ARS-BFGL-NGS-116294	2	128.58	1	1	0	0	2
22	INRA-566	13	57.94	1	0	1	0	2
23	**BTA-49720-no-rs**	**5**	**69.66**	**1**	**1**	**1**	**0**	**3**
24	ARS-BFGL-NGS-56387	13	24.36	1	1	0	0	2
25	BTA-28185-no-rs	26	22.78	1	0	0	0	1

SOMMAIRE

1. Introduction : définitions et objectifs des scans génomiques
2. Rappels de génétique des populations : statistiques de diversité appliquées aux séquences
3. La sélection et ses effets sur les séquences ADN
4. Outils pour la détection de traces de sélection
5. La génomique du paysage
6. **La validation des résultats de scans génomiques**

CONCLUSION

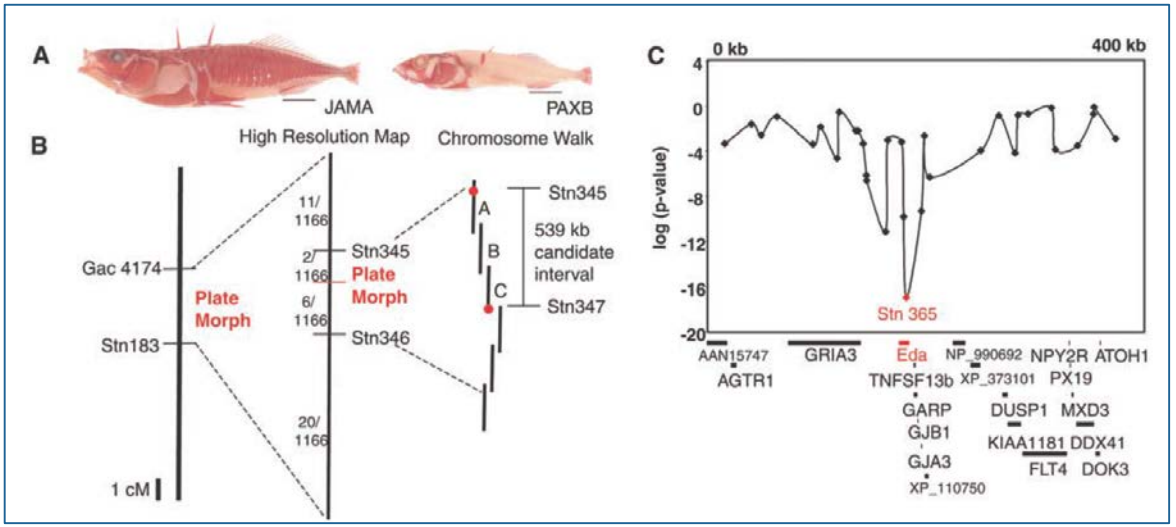
et la validation des résultats de scans génomiques?



Pardo-Diaz et al. 2015 Methods in Ecology and Evolution, 6, 445–464

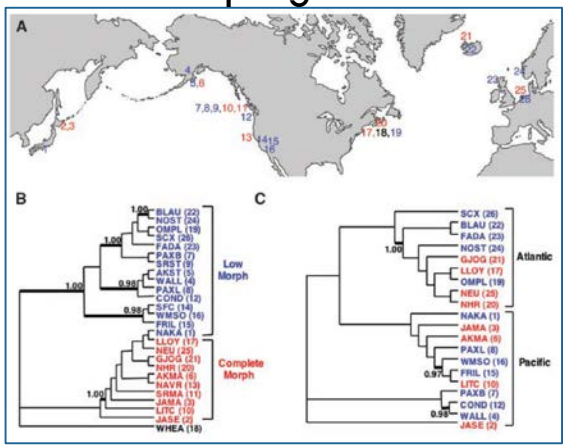
Un exemple : variation du nombre d'écailles de la queue chez l'épinoche à trois épines (*Gasterosteus aculeatus*)

Cartographie génétique et physique ('Foward genetics')

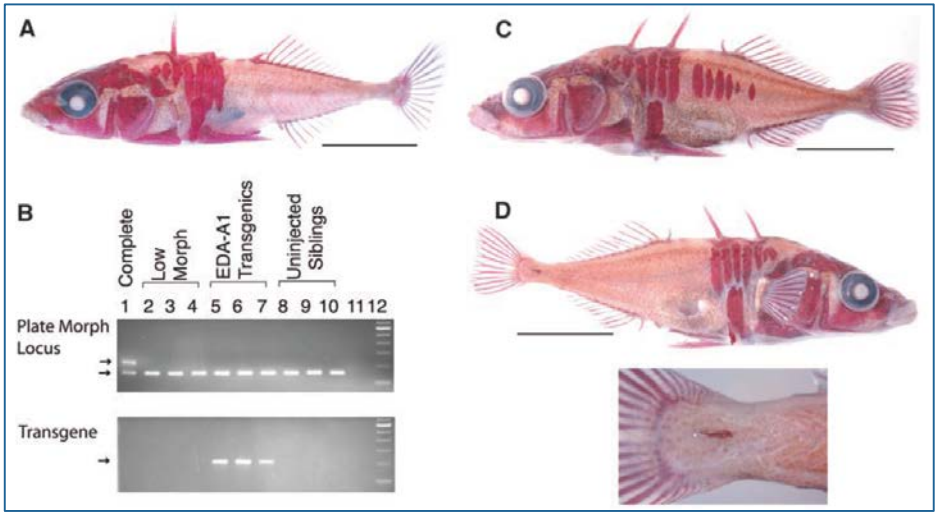


Colosimo et al 2005

Association mapping / landscape genomics



Validation fonctionnelle



BIBLIOGRAPHIE

Akey, J.M. (2009). Constructing genomic maps of positive selection in humans: where do we go from here? Genome Res. *19*, 711–722.

Akey, J.M., Zhang, G., Zhang, K., Jin, L., and Shriver, M.D. (2002). Interrogating a High-Density SNP Map for Signatures of Natural Selection. Genome Res. *12*, 1805–1814.

Albert, C.H., Yoccoz, N.G., Edwards, T.C., Graham, C.H., Zimmermann, N.E., and Thuiller, W. (2010). Sampling in ecology and evolution – bridging the gap between theory and practice. Ecography *33*, 1028–1037.

Andrés, A.M., Hubisz, M.J., Indap, A., Torgerson, D.G., Degenhardt, J.D., Boyko, A.R., Gutenkunst, R.N., White, T.J., Green, E.D., Bustamante, C.D., et al. (2009). Targets of Balancing Selection in the Human Genome. Mol. Biol. Evol. *26*, 2755–2764.

Beaumont, M.A., and Balding, D.J. (2004). Identifying adaptive genetic divergence among populations from genome scans. Mol. Ecol. *13*, 969–980.

Beaumont, M.A., and Nichols, R.A. (1996). Evaluating loci for use in the genetic analysis of population structure. Proc R Soc Lond B *263*, 1619–1626.

Berg, P.R., Jentoft, S., Star, B., Ring, K.H., Knutsen, H., Lien, S., Jakobsen, K.S., and André, C. (2015). Adaptation to Low Salinity Promotes Genomic Divergence in Atlantic Cod (Gadus morhua L.). Genome Biol. Evol. *7*, 1644–1663.

Black IV, W.C., Baer, C.F., Antolin, M.F., and DuTeau, N.M. (2001). POPULATION GENOMICS: Genome-Wide Sampling of Insect Populations. Annu. Rev. Entomol. *46*, 441–469.

Casillas, S., and Barbadilla, A. (2017). Molecular Population Genetics. Genetics *205*, 1003–1035.

Colosimo, P.F., Hosemann, K.E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R.M., Schluter, D., and Kingsley, D.M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science *307*, 1928–1933.

Coop, G., Witonsky, D., Di Rienzo, A., and Pritchard, J.K. (2010). Using environmental correlations to identify loci underlying local adaptation. Genetics *185*, 1411–1423.

Eckert, A.J., van Heerwaarden, J., Wegrzyn, J.L., Nelson, C.D., Ross-Ibarra, J., Gonzalez-Martinez, S.C., and Neale, D.B. (2010). Patterns of Population Structure and Environmental Associations to Aridity Across the Range of Loblolly Pine (Pinus taeda L., Pinaceae). Genetics *185*, 969–982.

Excoffier, L., Hofer, T., and Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. Heredity *103*, 285–298.

Fay, J.C., and Wu, C.I. (2000). Hitchhiking under positive Darwinian selection. Genetics *155*, 1405–1413.

Frichot, E., Schoville, S.D., Bouchard, G., and François, O. (2013). Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. Mol. Biol. Evol. *30*, 1687–1699.

Fu, Y.X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics *147*, 915–925.

Guillot, G., Vitalis, R., Rouzic, A. le, and Gautier, M. (2014). Detecting correlation between allele frequencies and environmental variables as a signature of selection. A fast computational approach for genome-wide studies. Spat. Stat. *8*, 145–155.

Gunther, T., and Coop, G. (2013). Robust identification of local adaptation from allele frequencies. Genetics *195*, 205–220.

BIBLIOGRAPHIE

Hancock, A.M., Witonsky, D.B., Gordon, A.S., Eshel, G., Pritchard, J.K., Coop, G., and Di Rienzo, A. (2008). Adaptations to climate in candidate genes for common metabolic disorders. PLoS Genet *4*, e32.

Hey, J., and Machado, C.A. (2003). The study of structured populations — new hope for a difficult and divided science. Nat. Rev. Genet. *4*, 535–543.

Hudson, R.R., Kreitman, M., and Aguadé, M. (1987). A Test of Neutral Molecular Evolution Based on Nucleotide Data. Genetics *116*, 153–159.

Hughes, A.L., and Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature *335*, 167–170.

Hurst, L.D. (2002). The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet. TIG *18*, 486.

Joost, S., Bonin, A., Bruford, M.W., Despres, L., Conord, C., Erhardt, G., and Taberlet, P. (2007). A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. Mol Ecol *16*, 3955–3969.

Kasehagen, L.J., Mueller, I., Kiniboro, B., Bockarie, M.J., Reeder, J.C., Kazura, J.W., Kastens, W., McNamara, D.T., King, C.H., Whalen, C.C., et al. (2007). Reduced Plasmodium vivax Erythrocyte Infection in PNG Duffy-Negative Heterozygotes. PLOS ONE *2*, e336.

Kingman, J.F.C. (1982). The coalescent. Stoch. Process. Their Appl. *13*, 235–248.

Lewontin, R.C., and Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics *74*, 175–195.

Lotterhos, K.E., and Whitlock, M.C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol. Ecol. *24*, 1031–1046.

Manel, S., Joost, S., Epperson, B.K., Holderegger, R., Storfer, A., Rosenberg, M.S., Scribner, K.T., Bonin, A., and Fortin, M.-J. (2010). Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. Mol. Ecol. *19*, 3760–3772.

McDonald, J.H., and Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. Nature *351*, 652–654.

Minias, P., Bateson, Z.W., Whittingham, L.A., Johnson, J.A., Oyler-McCance, S., and Dunn, P.O. (2016). Contrasting evolutionary histories of MHC class I and class II loci in grouse—effects of selection and gene conversion. Heredity *116*, 466.

Mullen, L.M., Vignieri, S.N., Gore, J.A., and Hoekstra, H.E. (2009). Adaptive basis of geographic variation: genetic, phenotypic and environmental differences among beach mouse populations. Proc. R. Soc. B Biol. Sci. *276*, 3809–3818.

Nair, S., Williams, J.T., Brockman, A., Paiphun, L., Mayxay, M., Newton, P.N., Guthmann, J.-P., Smithuis, F.M., Hien, T.T., White, N.J., et al. (2003). A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. Mol. Biol. Evol. *20*, 1526–1536.

Nielsen, R. (2005). Molecular signatures of natural selection. Annu. Rev. Genet. *39*, 197–218.

Norton, H.L., Kittles, R.A., Parra, E., McKeigue, P., Mao, X., Cheng, K., Canfield, V.A., Bradley, D.G., McEvoy, B., and Shriver, M.D. (2007). Genetic Evidence for the Convergent Evolution of Light Skin in Europeans and East Asians. Mol. Biol. Evol. *24*, 710–722.

BIBLIOGRAPHIE

Pardo-Diaz, C., Salazar, C., and Jiggins, C.D. (2015). Towards the identification of the loci of adaptive evolution. Methods Ecol. Evol. *6*, 445–464.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. Nature *419*, 832–837.

Savolainen, O., Pyhäjärvi, T., and Knürr, T. (2007). Gene Flow and Local Adaptation in Trees. Annu. Rev. Ecol. Evol. Syst. *38*, 595–619.

Schoville, S.D., Bonin, A., François, O., Lobreaux, S., Melodelima, C., and Manel, S. (2012). Adaptive Genetic Variation on the Landscape: Methods and Cases. Annu. Rev. Ecol. Evol. Syst. *43*, 23–43.

Storz, J.F. (2005). INVITED REVIEW: Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol. Ecol. *14*, 671–688.

Stucki, S., Orozco-terWengel, P., Forester, B.R., Duruz, S., Colli, L., Masembe, C., Negrini, R., Landguth, E., Jones, M.R., Bruford, M.W., et al. (2017). High performance computation of landscape genomic models including local indicators of spatial association. Mol. Ecol. Resour. *17*, 1072–1089.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics *123*, 585–595.

Vitalis, R., Dawson, K., and Boursot, P. (2001). Interpretation of variation across marker loci as evidence of selection. Genetics *158*, 1811–1823.

Vitti, J.J., Grossman, S.R., and Sabeti, P.C. (2013). Detecting Natural Selection in Genomic Data. Annu. Rev. Genet. *47*, 97–120.

Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A Map of Recent Positive Selection in the Human Genome. PLOS Biol. *4*, e72.

Wang, R.L., Stec, A., Hey, J., Lukens, L., and Doebley, J. (1999). The limits of selection during maize domestication. Nature *398*, 236–239.

Wilding, C.S., Butlin, R.K., and Grahame, J. (2001). Differential gene exchange between parapatric morphs of Littorina saxatilis detected using AFLP markers. J. Evol. Biol. *14*, 611–619.

Yang, W.-Y., Novembre, J., Eskin, E., and Halperin, E. (2012). A model-based approach for analysis of spatial structure in genetic data. Nat. Genet. *44*, 725–731.