



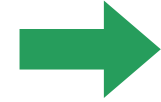
Inférence de la démographie récente d'une plante méditerranéenne endémique en déclin à partir de segments identiques par descendance

Océane EYCHENNE

Génomique de la Conservation & Démographie Récente



*Problématiques de
conservation des espèces*



Démographie des Populations →

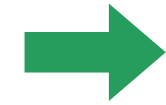


Patterns de **migration**
Taille de population efficace
=> **Connectivité**

Génomique de la Conservation & Démographie Récente



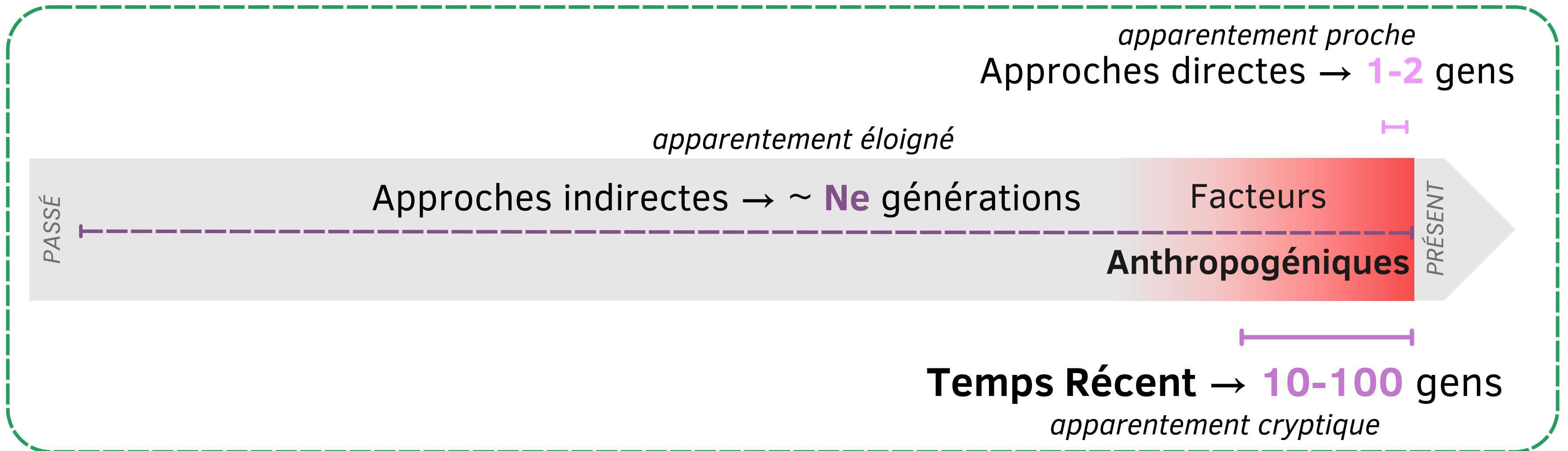
*Problématiques de
conservation des espèces*



Démographie des Populations →

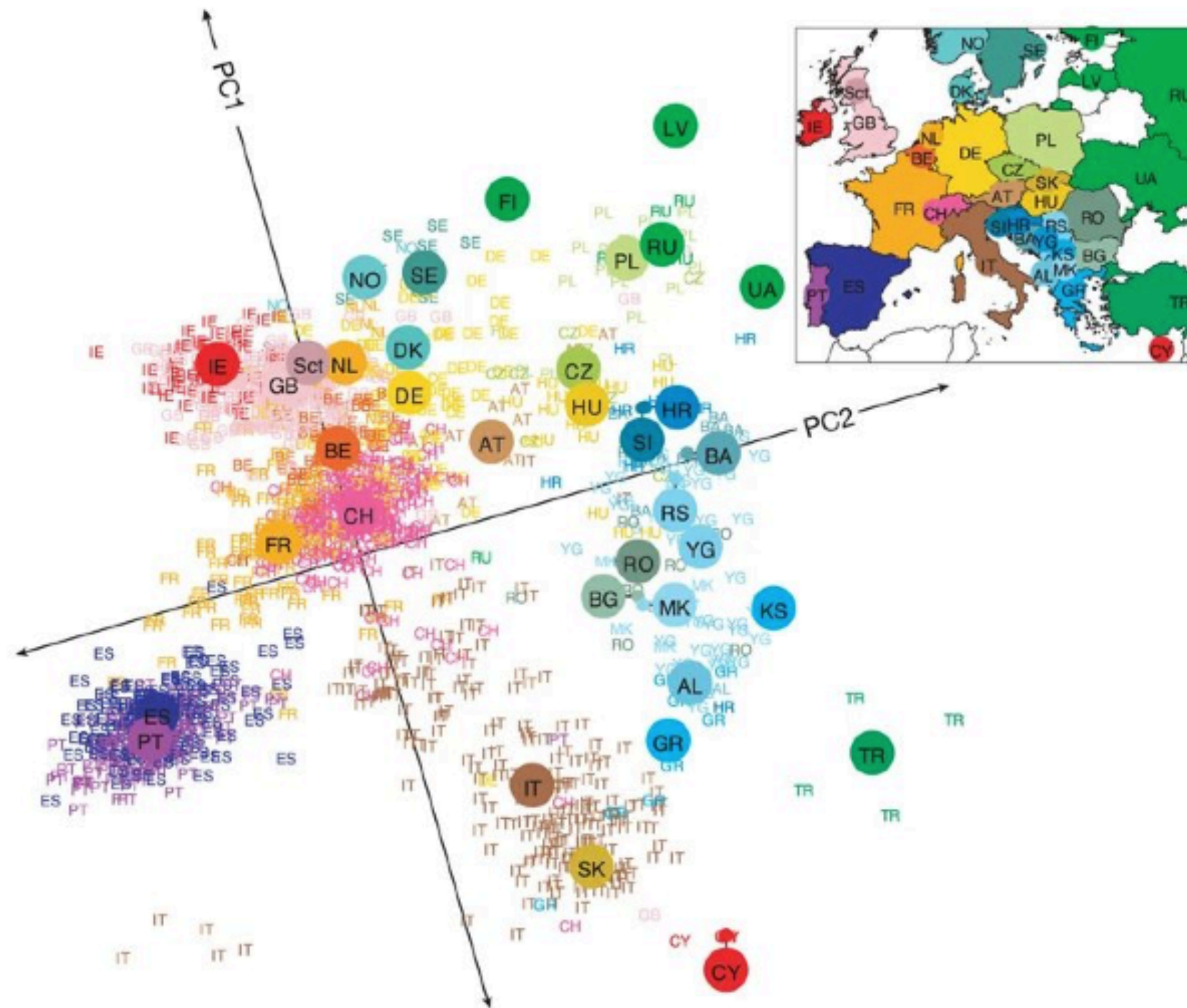
Patterns de **migration**
Taille de population efficace
=> **Connectivité**

ÉCHELLES DE TEMPS

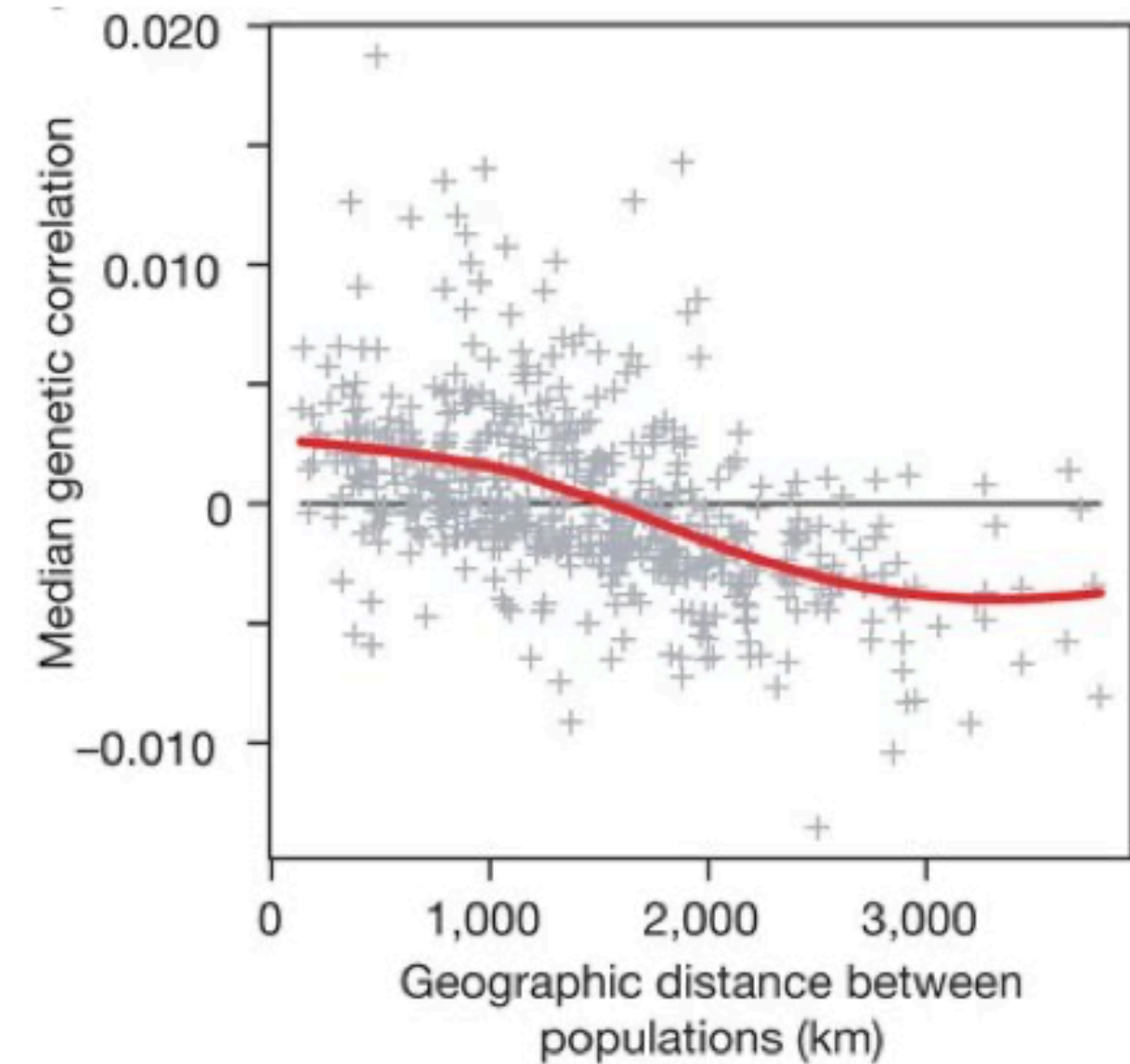


Dimension spatiale de la connectivité

Le degré d'isolement par la distance est directement lié au **flux génétique**.



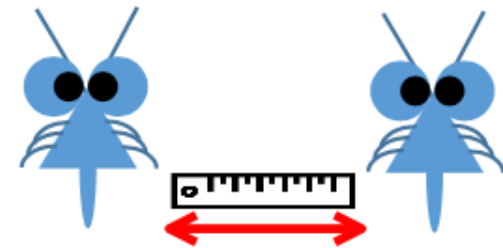
Genetic similarity vs geographic distance



Novembre et al, 2008

Dimension temporelle (généalogique) de la connectivité

Connectivité génétique est le **taux de déclin de la parenté génétique** entre individus en fonction de la **distance géographique** qui les sépare...

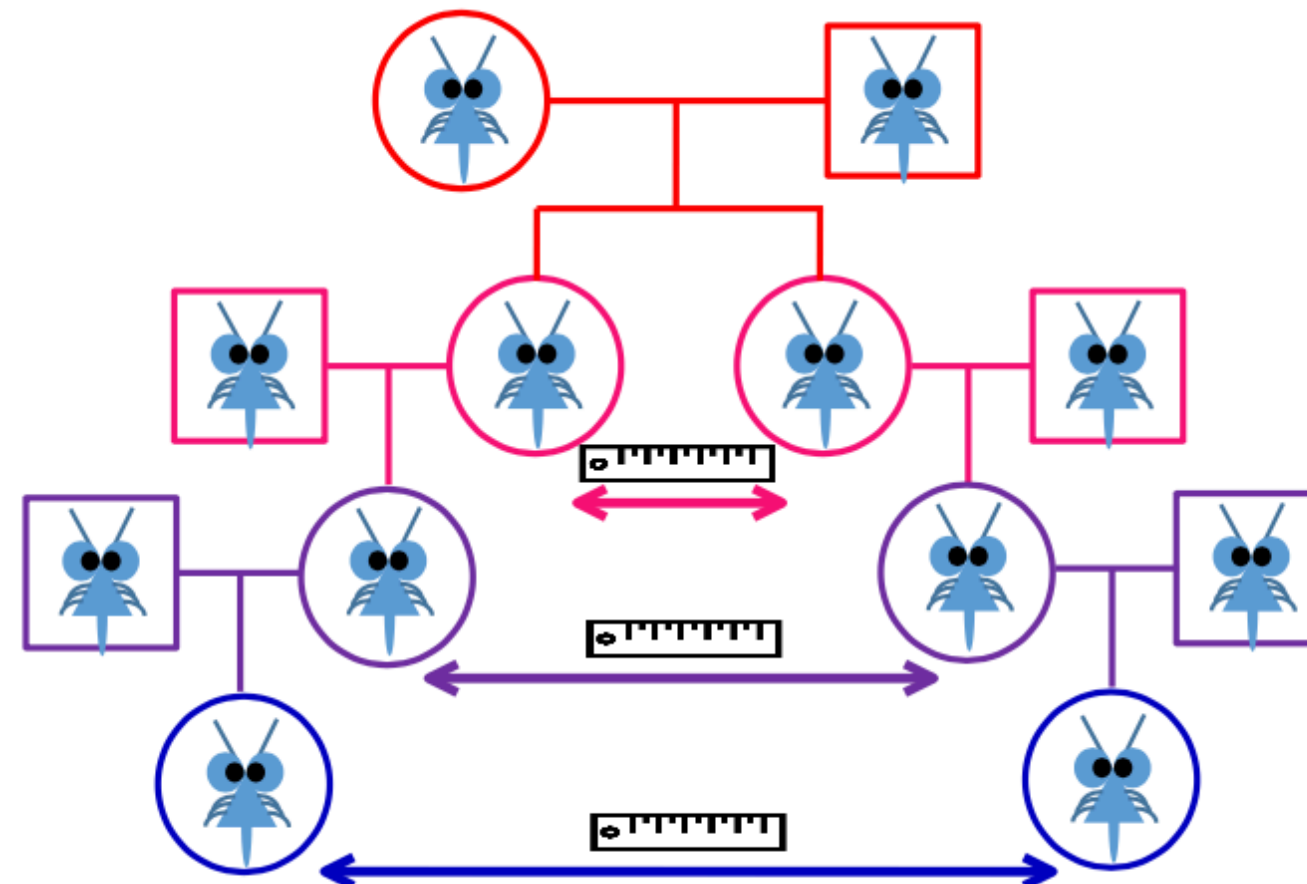


Dimension temporelle (généalogique) de la connectivité

Connectivité génétique est le **taux de déclin de la parenté génétique** entre individus en fonction de la **distance géographique** qui les sépare...

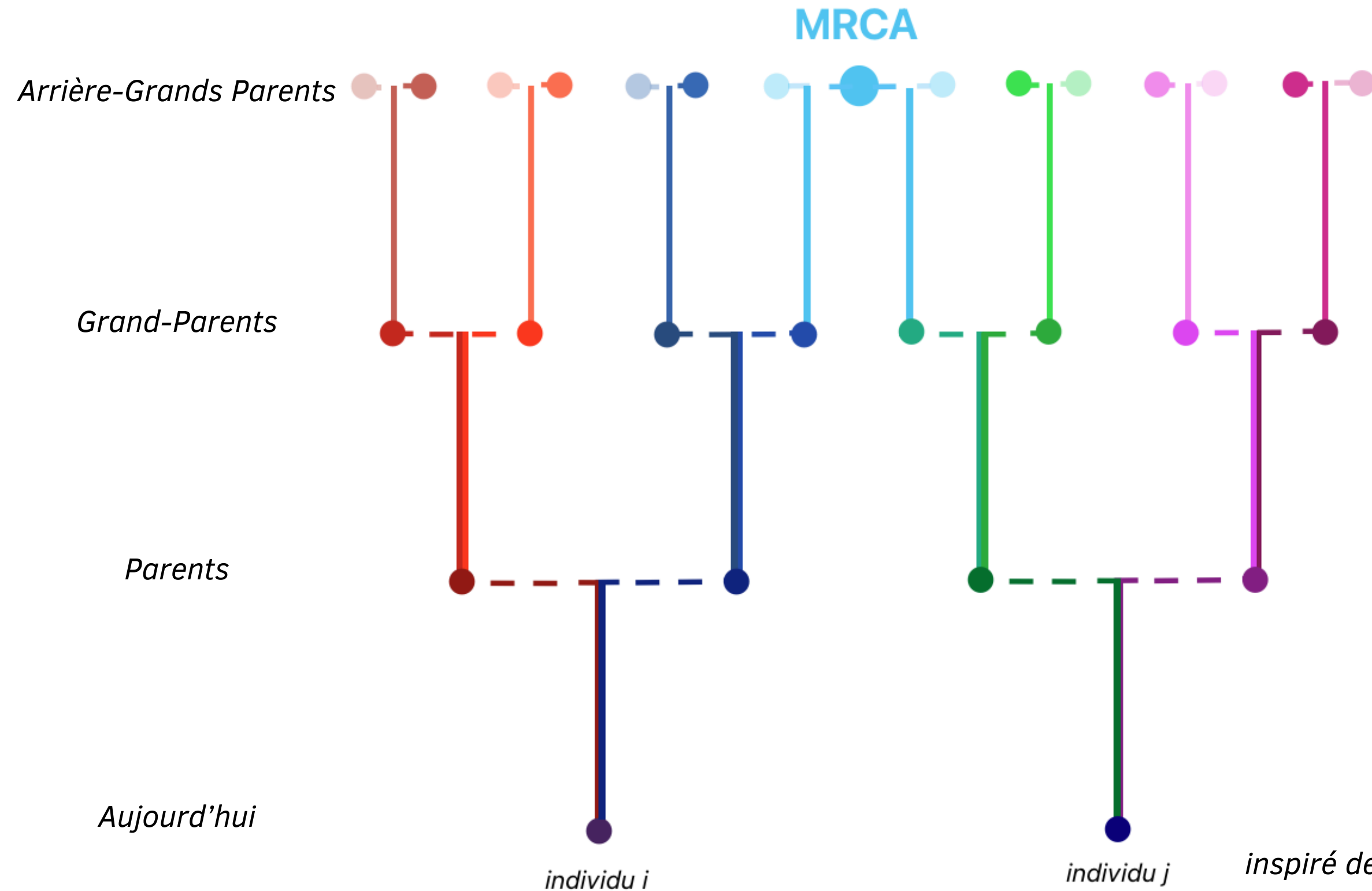


...mais elle est également fonction du **temps qui relie les individus** entre eux par le **biais de leurs ancêtres communs** dans leur pedigree, en raison de la dispersion multi-générationnelle.



Pedigree Populationnel

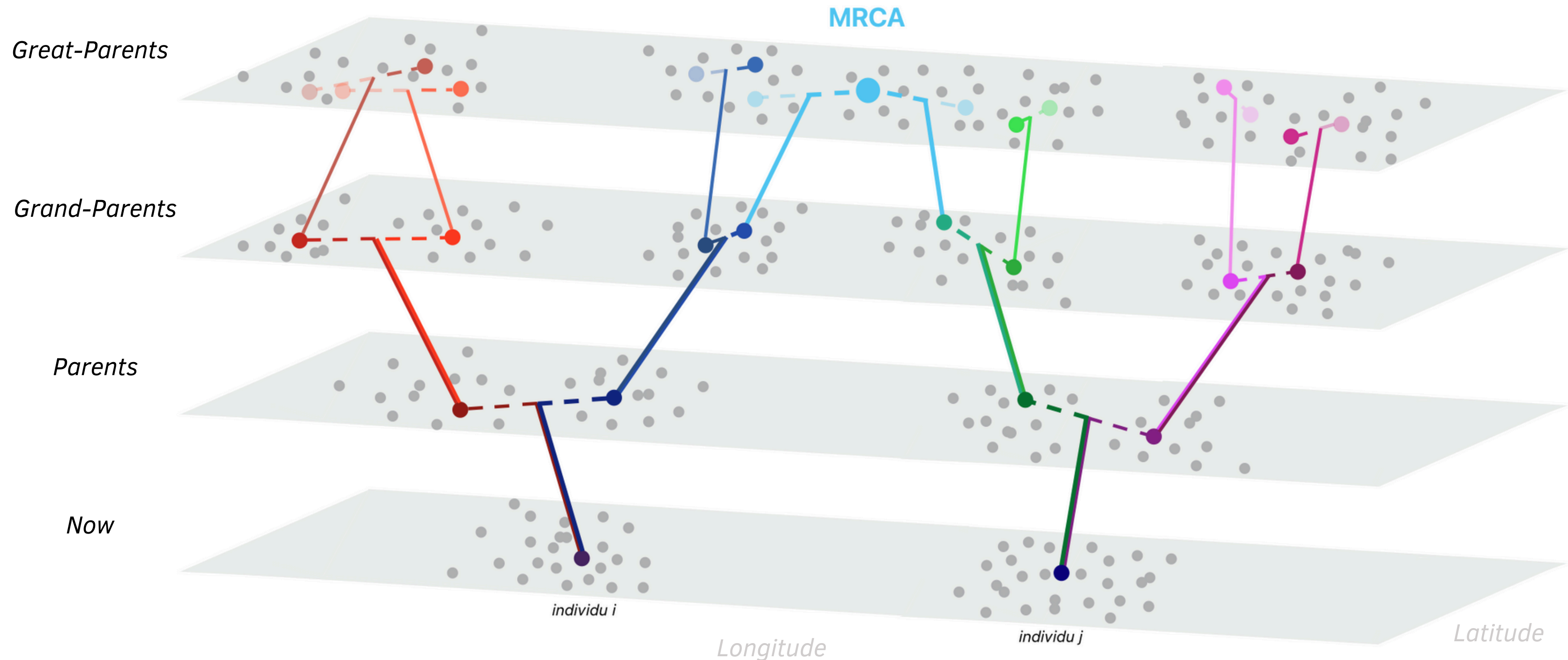
Exploiter l'information d'ascendance **généalogique** dans le pedigree populationnel



inspiré de Bradburd & Ralph, 2019

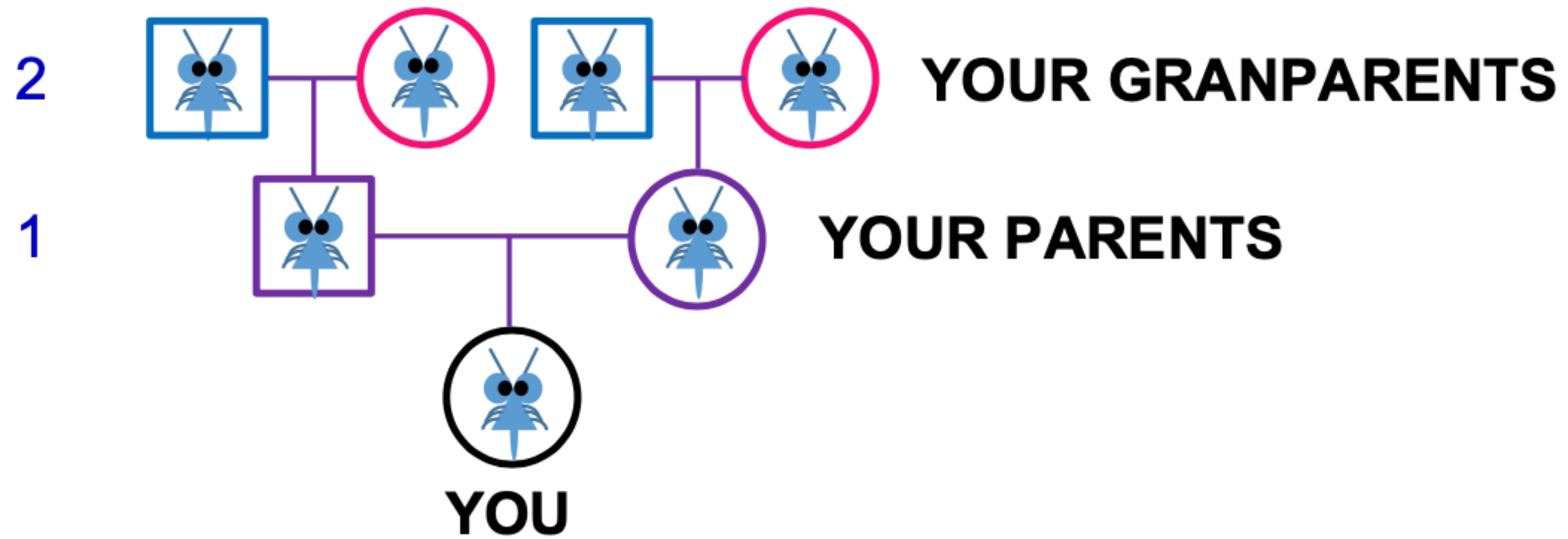
Pedigree Spatial Populationnel

Combiner le pedigree populationnel avec les informations **spatiales**

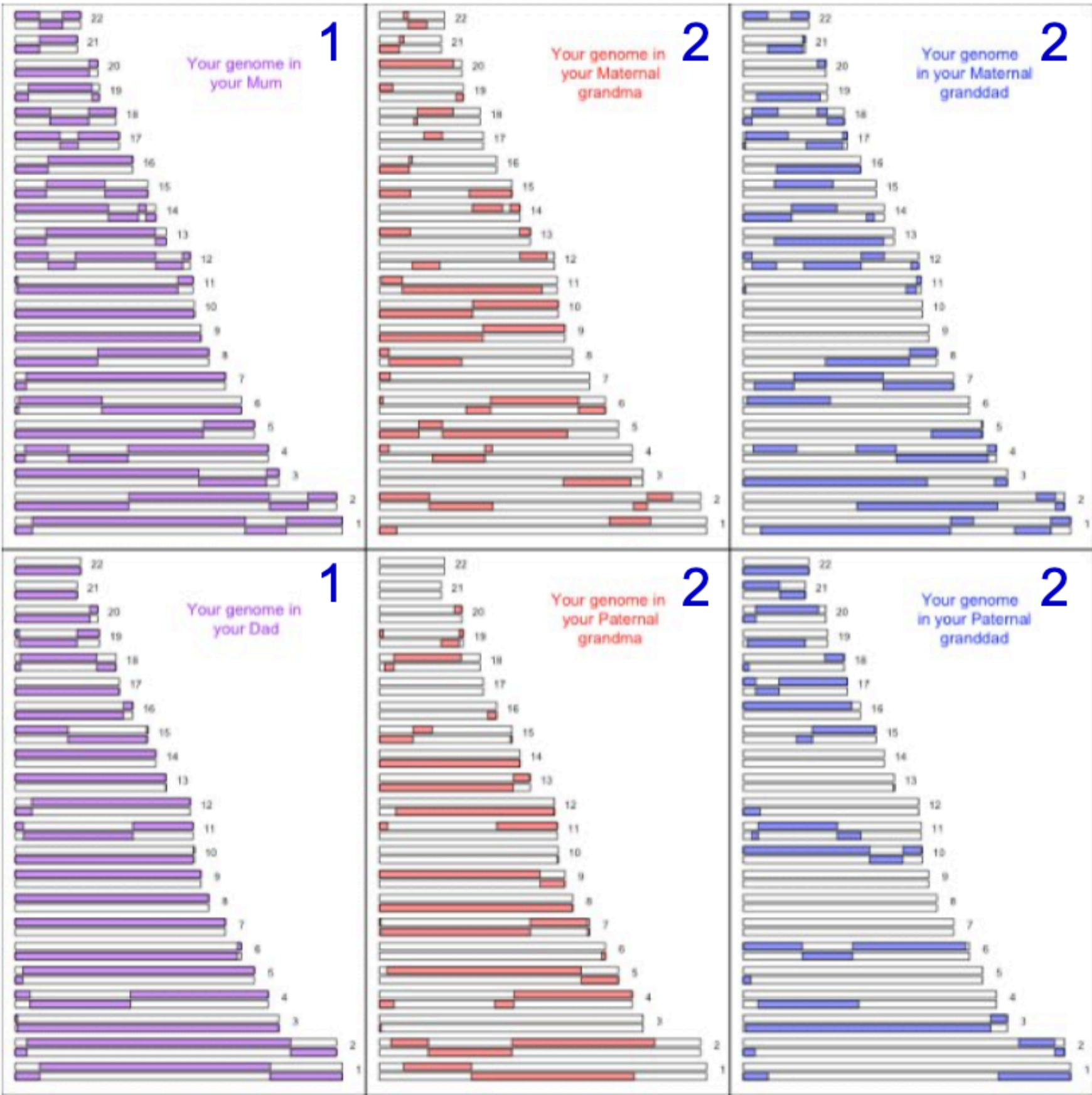
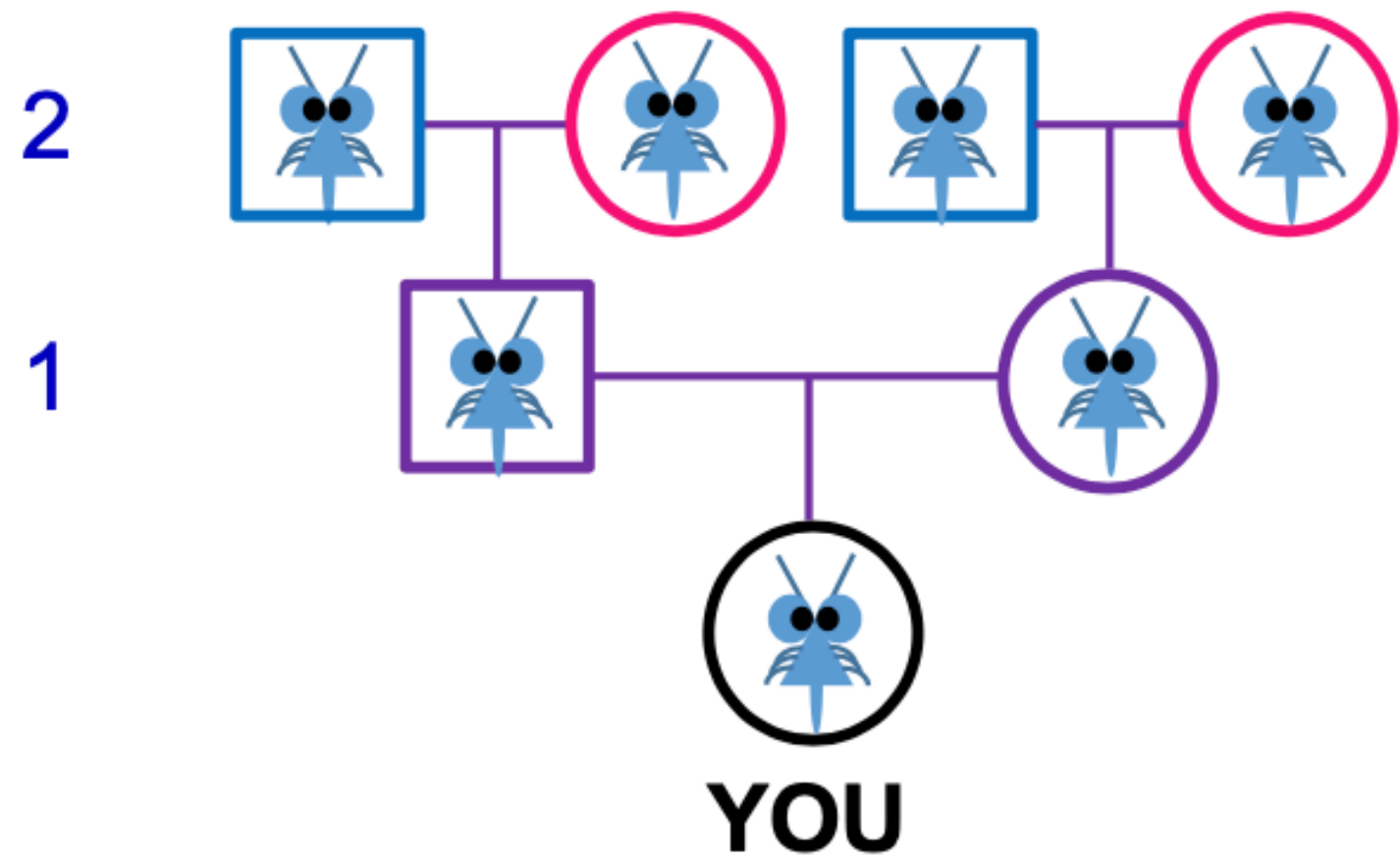


inspiré de Bradburd & Ralph, 2019

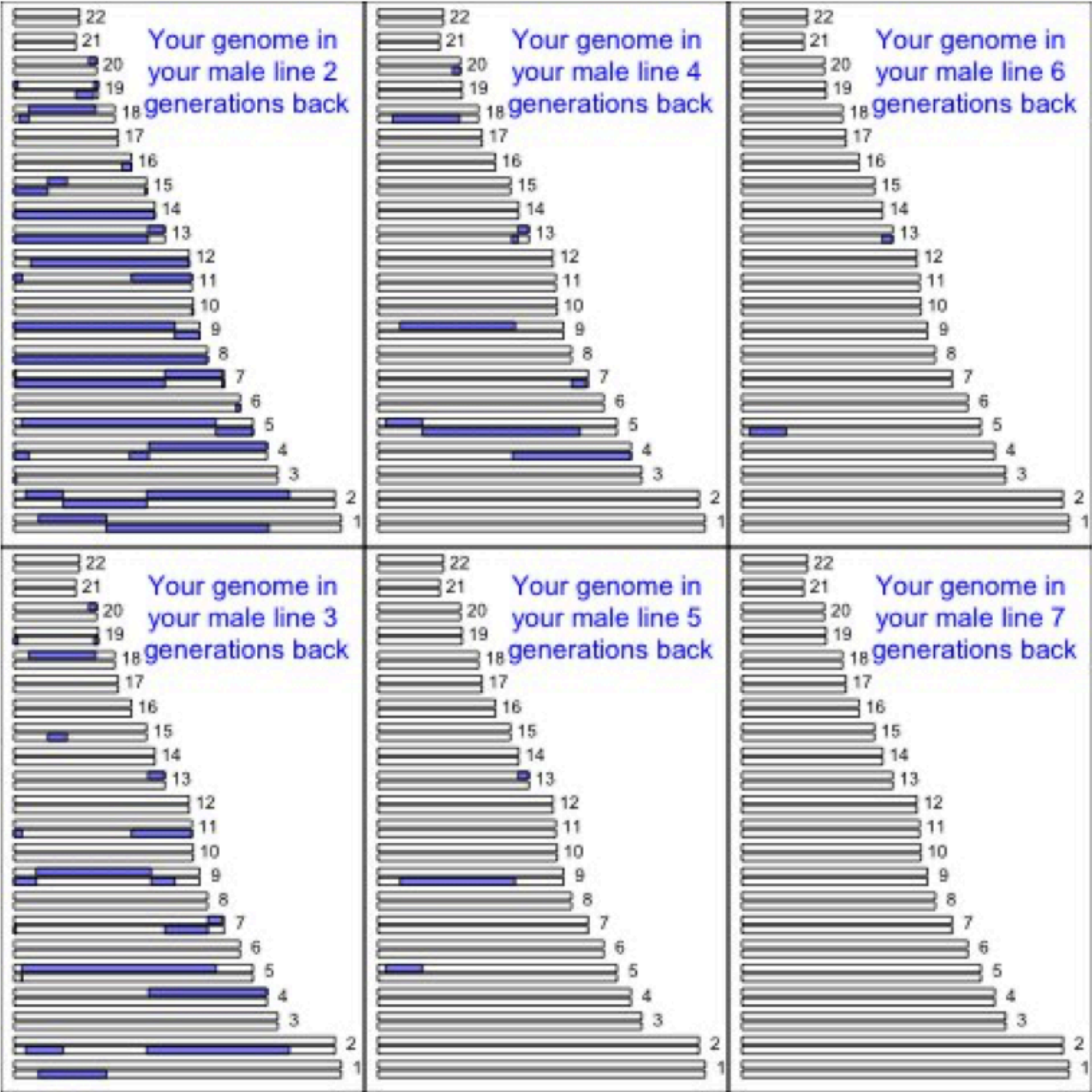
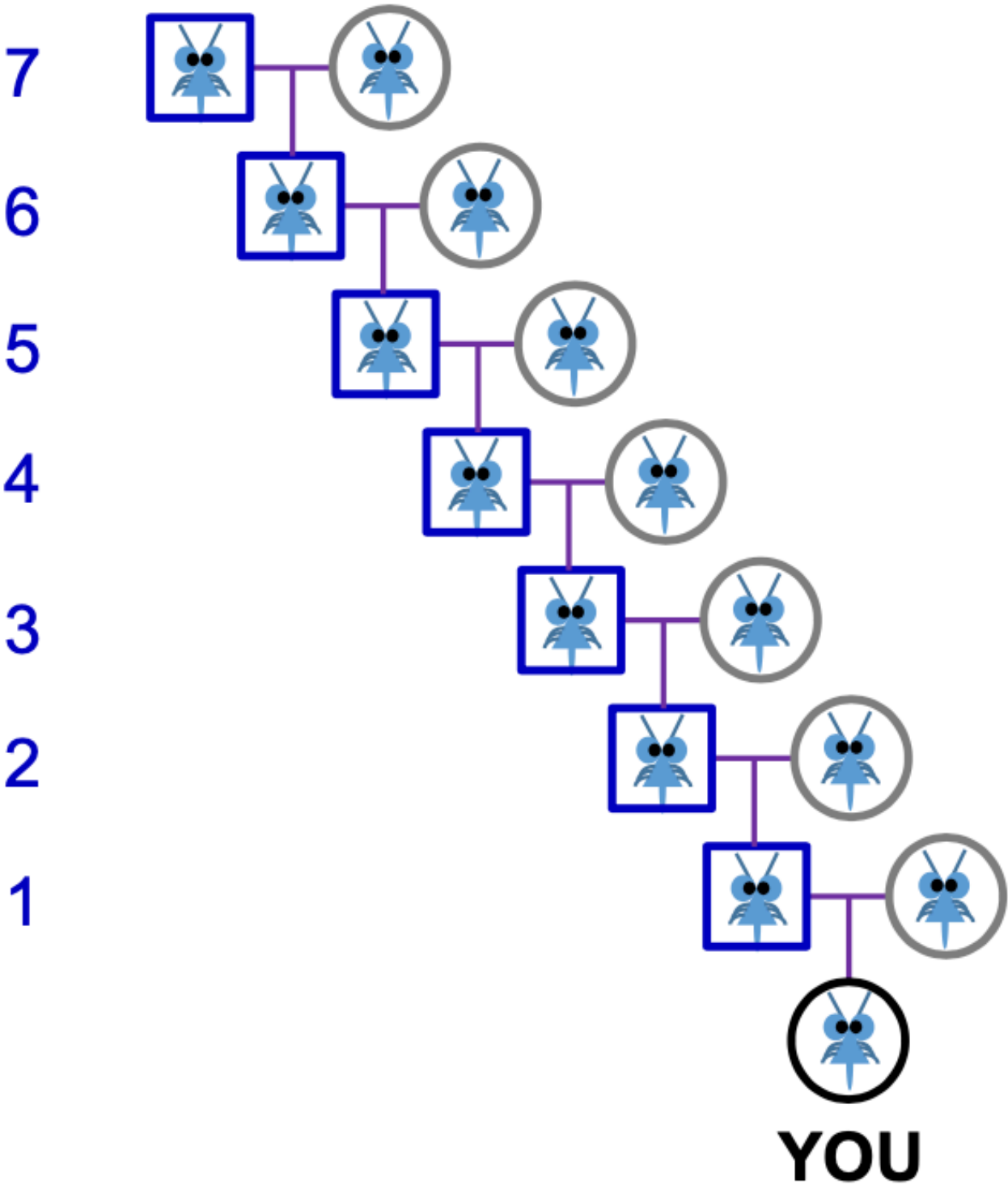
Pourquoi la génétique n'enregistre pas tout le pedigree généalogique ?



Pourquoi la génétique n'enregistre pas tout le pedigree généalogique ?

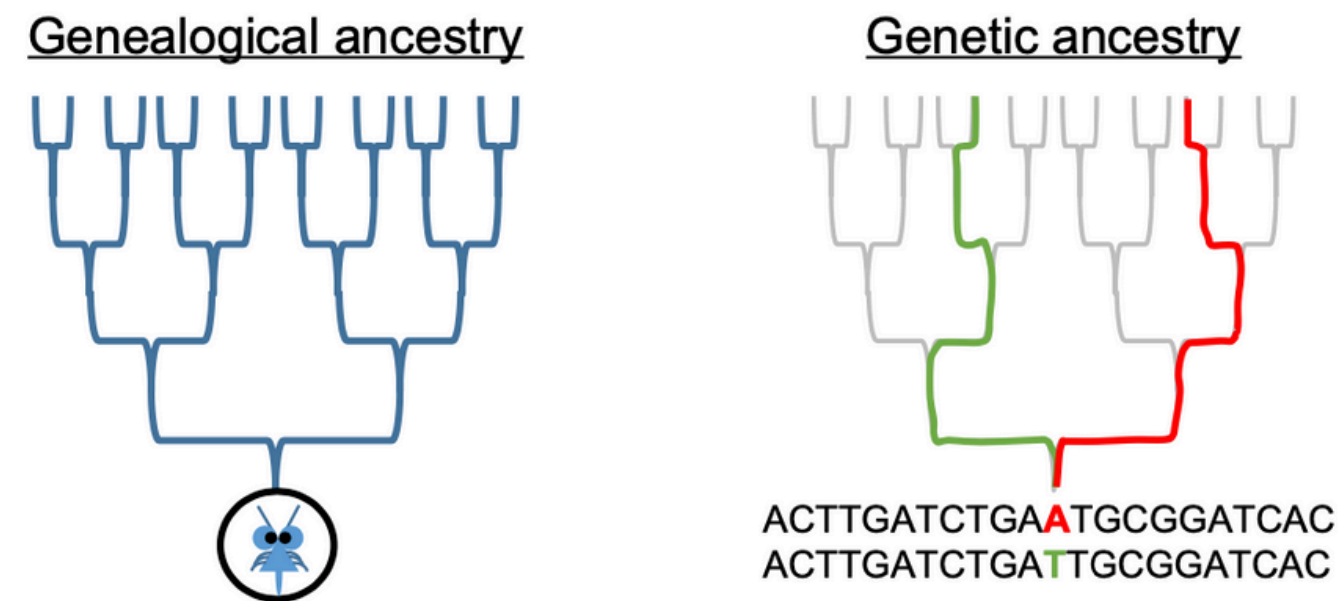


Pourquoi la génétique n'enregistre pas tout le pedigree généalogique ?



Que peut-on apprendre sur nos ancêtres à partir des données génétiques ?

Les ancêtres génétiques représentent un **sous-ensemble aléatoire** de tous les ancêtres généalogiques.



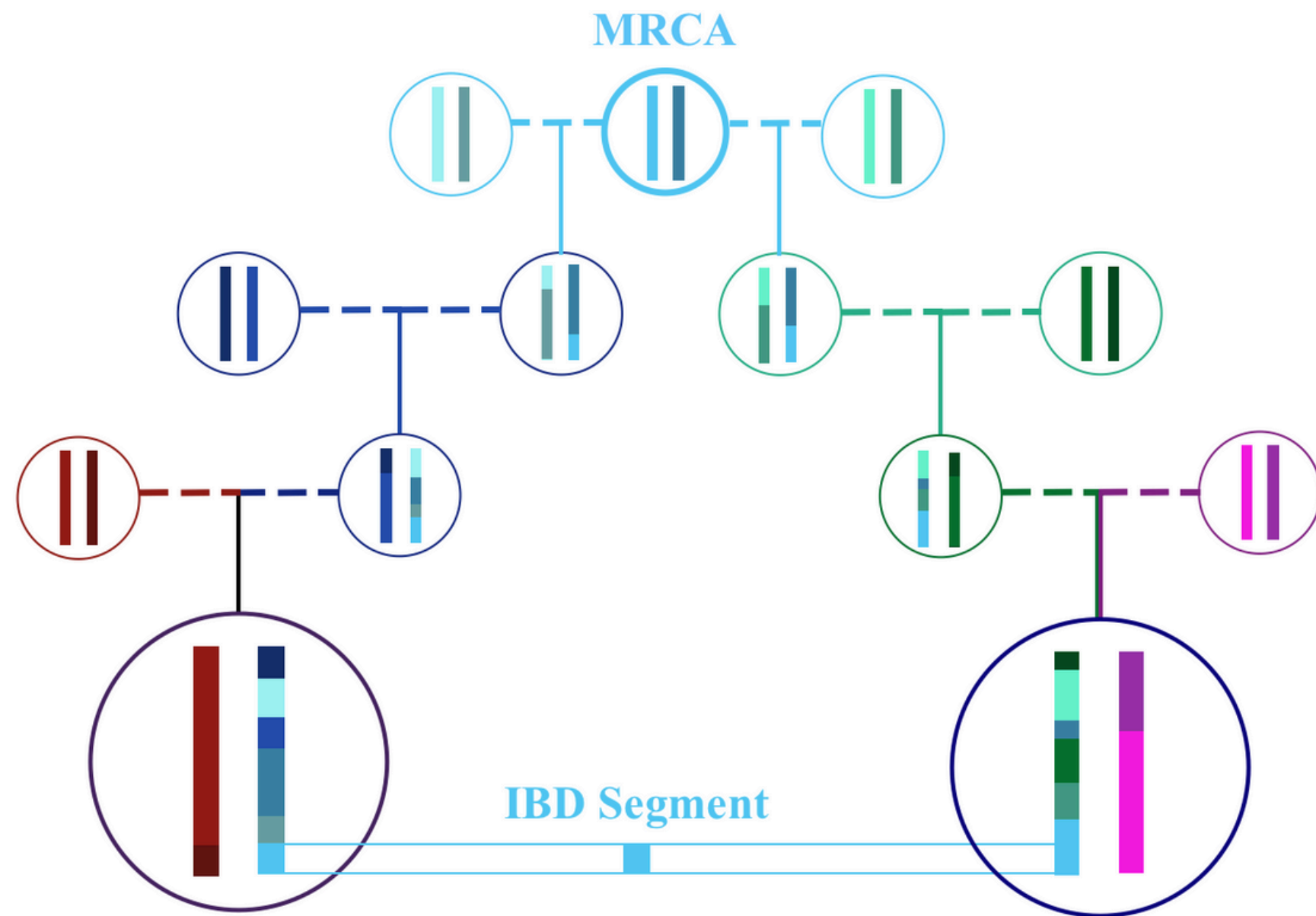
Le pedigree comporte de nombreux **trous génétiques aléatoires**.



Ascendance génétique : informations sur la connectivité démographique récente.

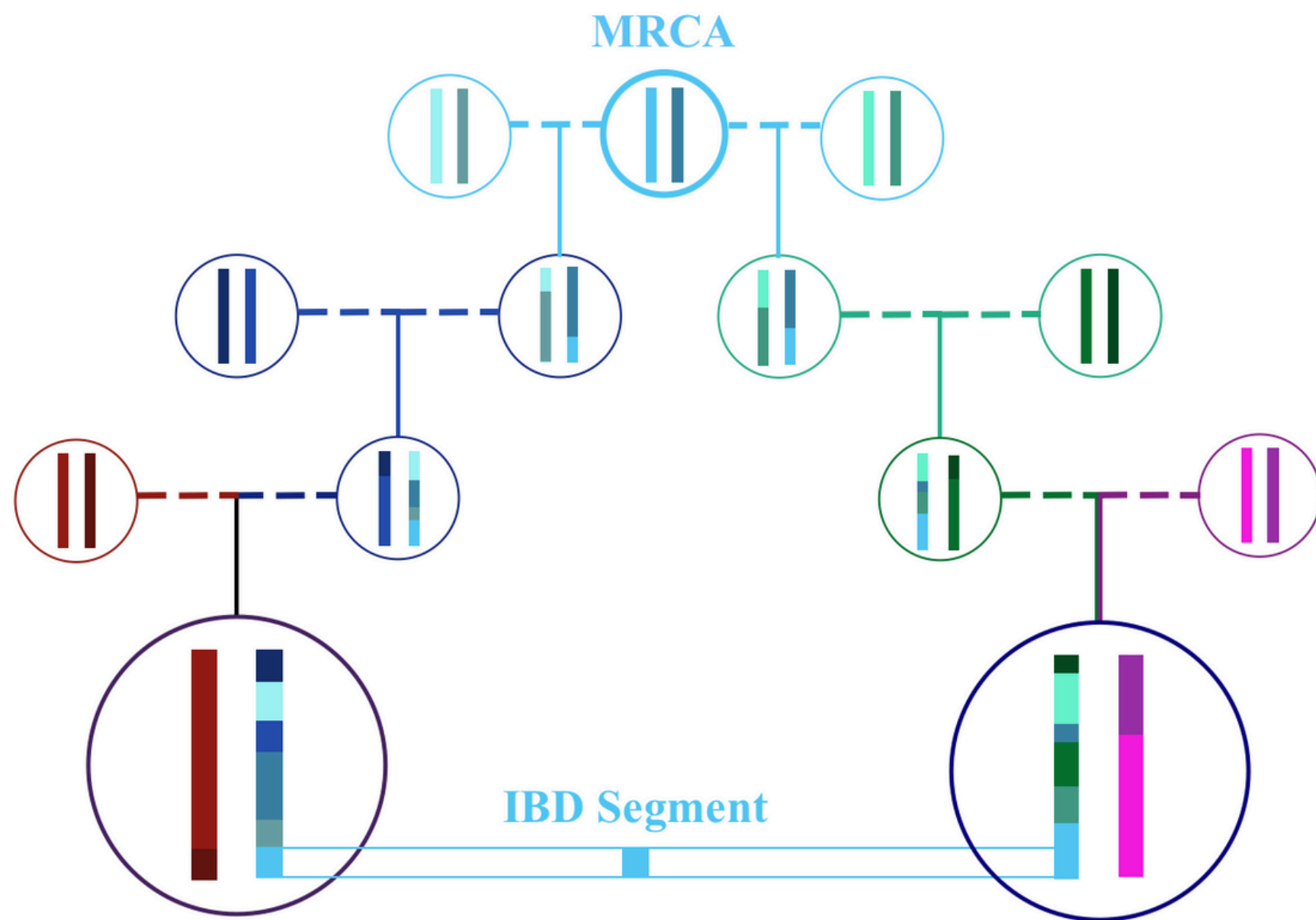
Segments IBD informatifs de la co-ascendance récente

Segments IBD = segments Identique par Descendance

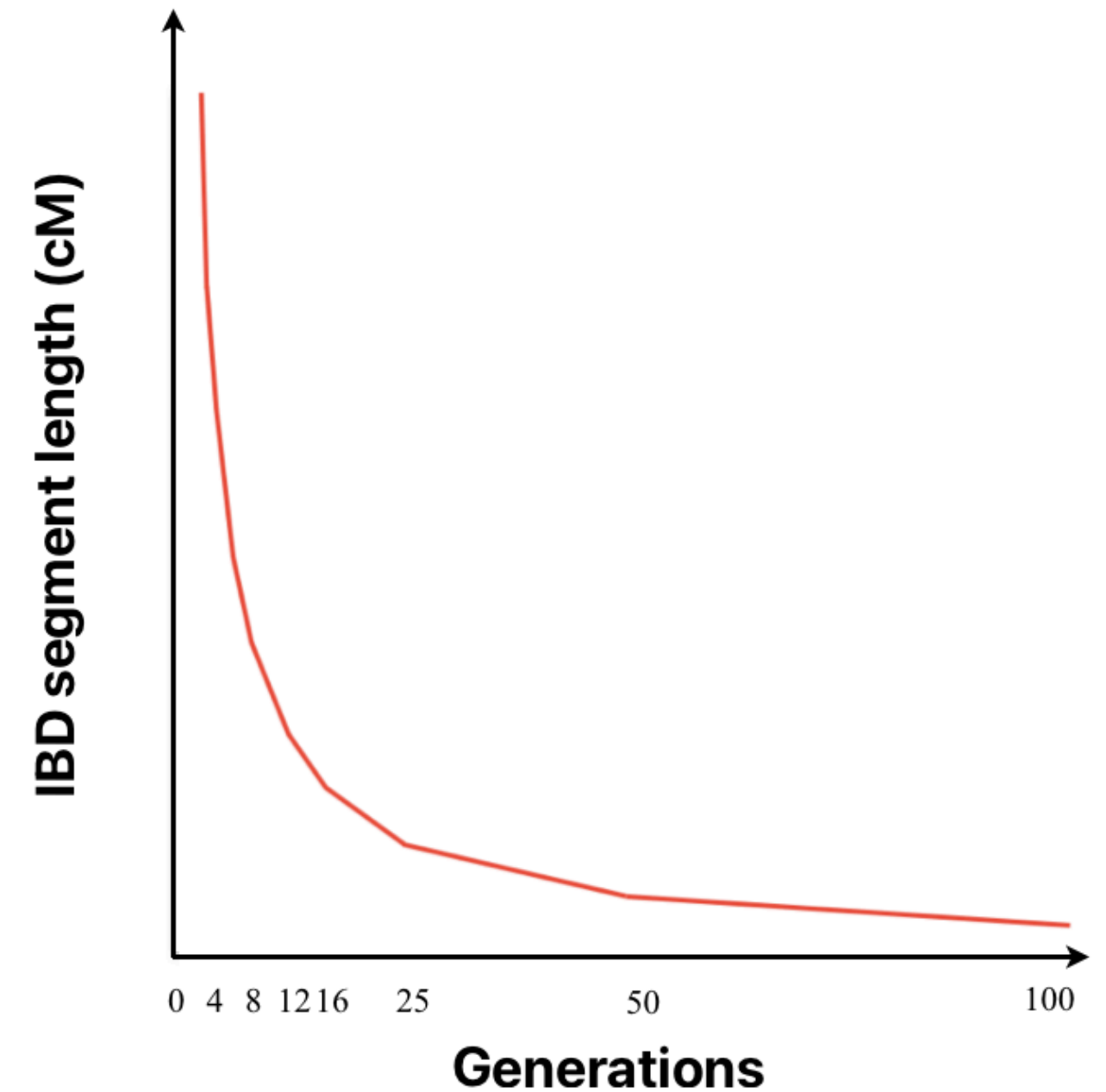


Segments IBD informatifs de la co-ascendance récente

Segments IBD = segments Identique par Descendance



Principe de
l'horloge de
recombinaison

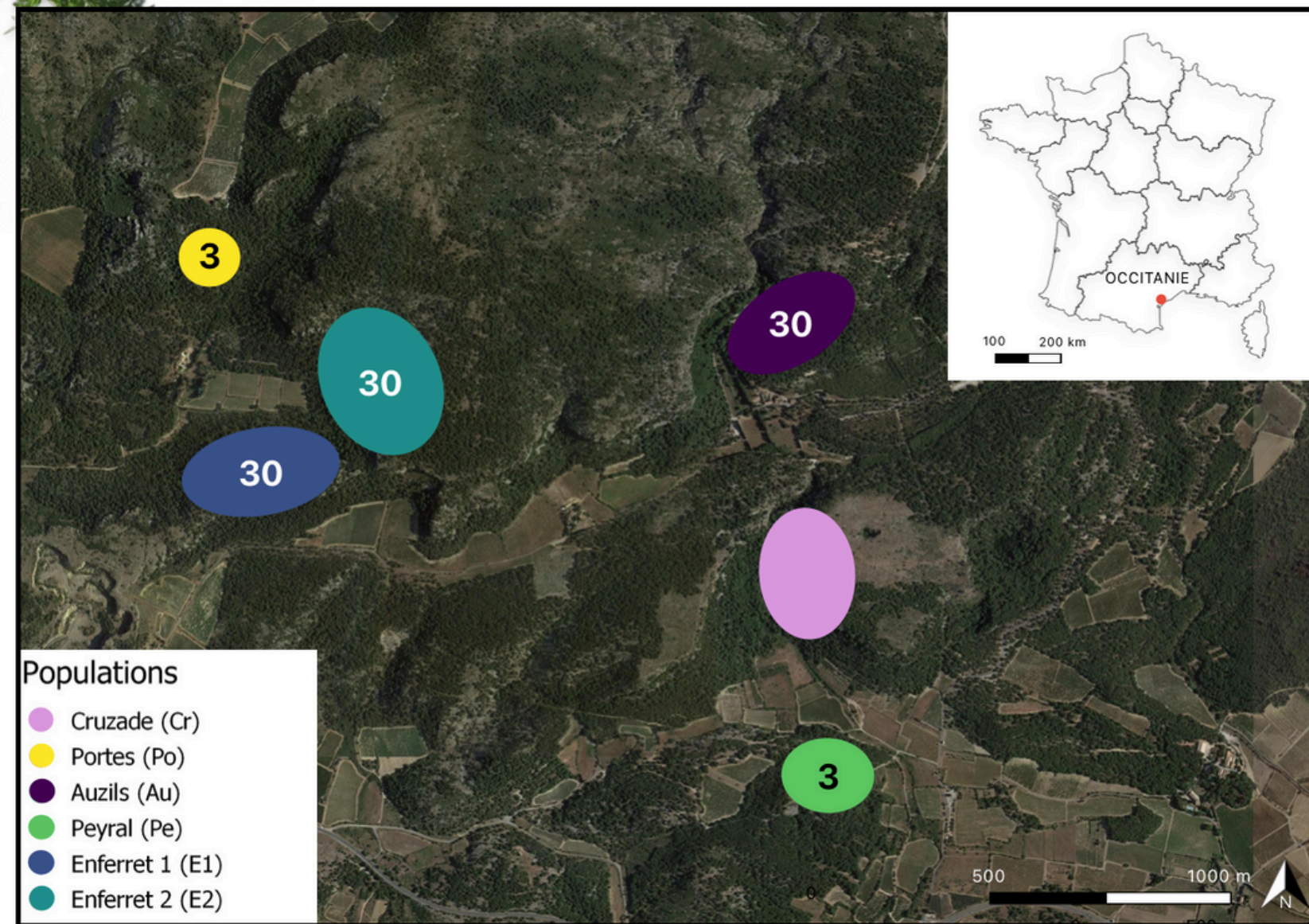


Browning & Browning, 2012

Reconstruction de Génomes Phasés Complets



Centaurea corymbosa (Asteraceae)



Carte d'Échantillonnage (d'après Ducrettet, 2022)

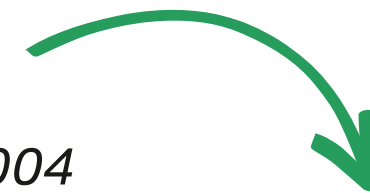
Comprendre la connectivité récente de *C. corymbosa*



Centaurea corymbosa (Asteraceae)

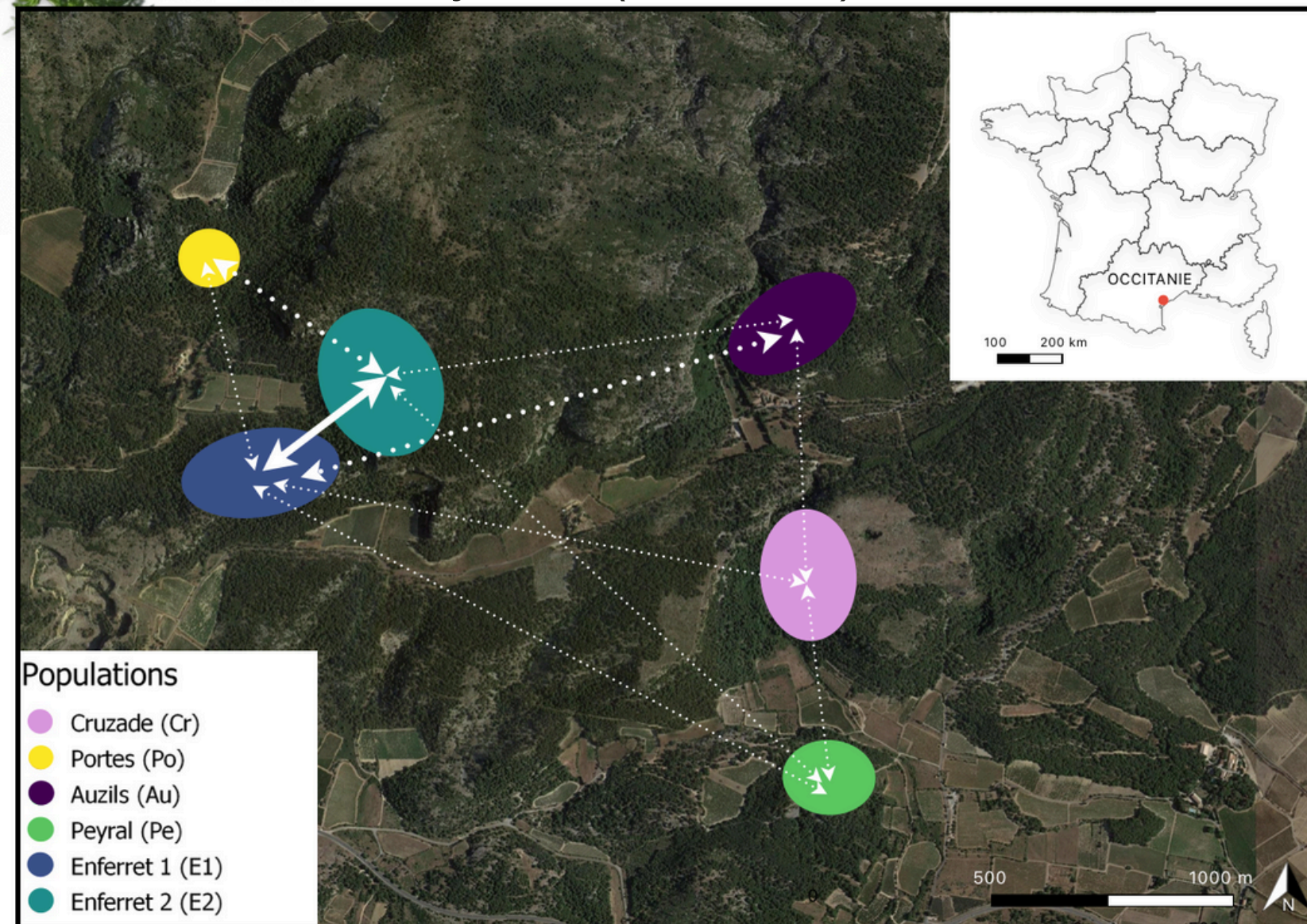
VU

Fréville et al. 2004



Pas de flux de gènes contemporains

Différenciation forte ($F_{ST} \sim 0.25$)



Populations

- Cruzade (Cr)
- Portes (Po)
- Auzils (Au)
- Peyral (Pe)
- Enferret 1 (E1)
- Enferret 2 (E2)

Carte d'Échantillonnage (d'après Ducrettet, 2022)

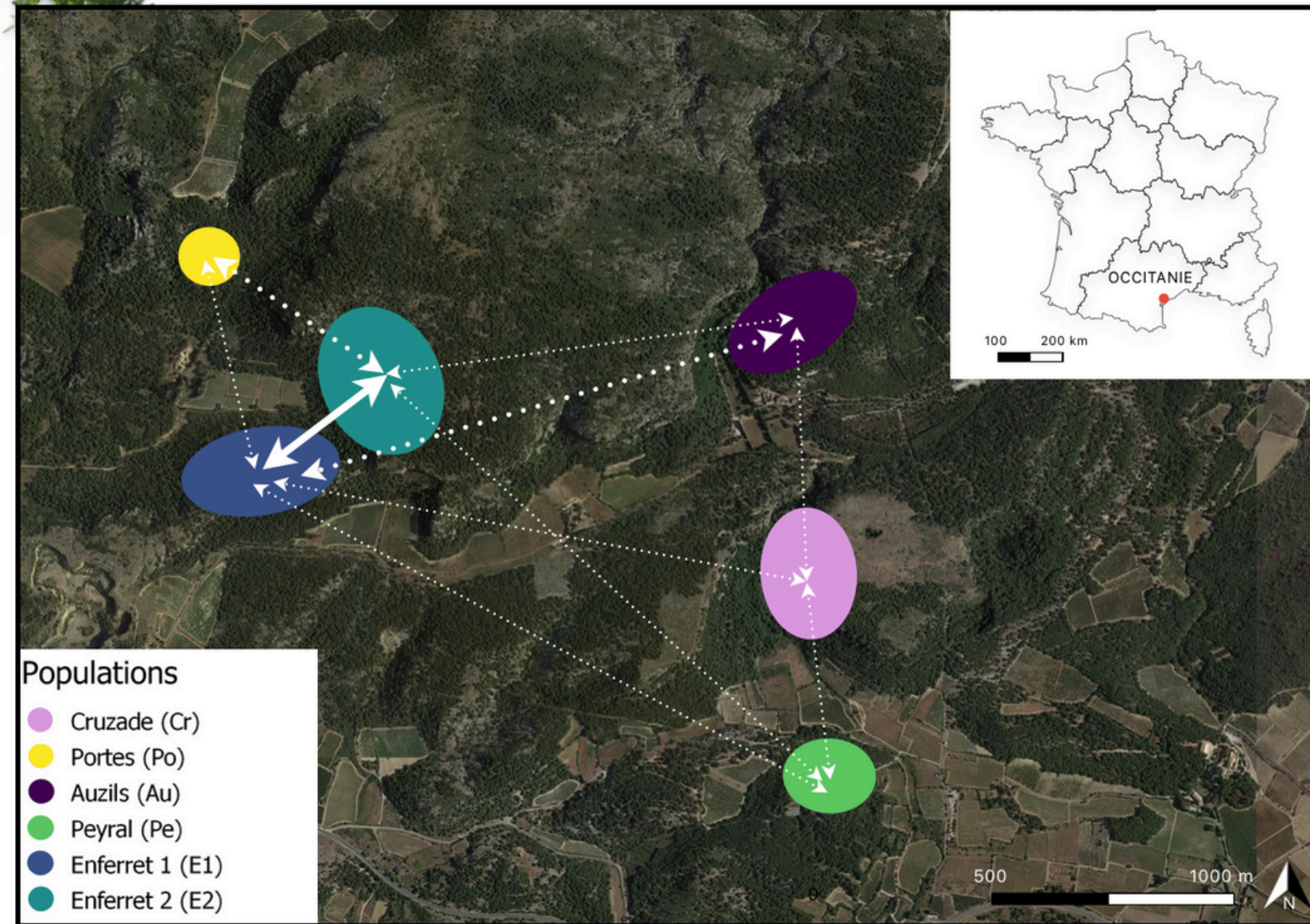
Comprendre la connectivité récente de *C. corymbosa*



Centaurea corymbosa (Asteraceae)

VU

Fréville et al. 2004



Carte d'Échantillonnage (d'après Ducrettet, 2022)

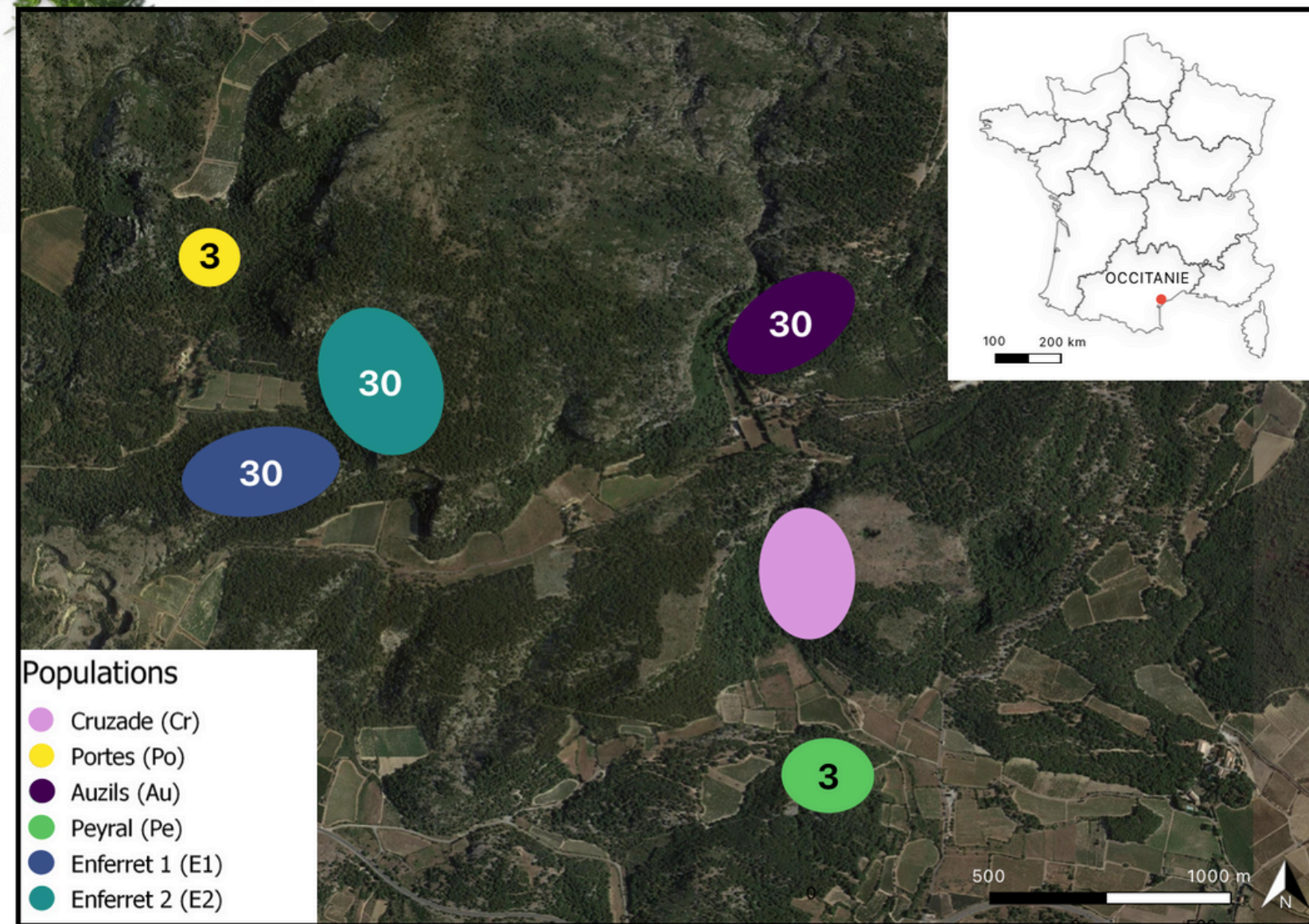
Pas de flux de gènes contemporains
Différenciation forte ($F_{ST} \sim 0.25$)

Les populations de *C. corymbosa*
ont-elles été reliées par un flux
génique dans un passé récent ?

Reconstruction de Génomes Phasés Complets

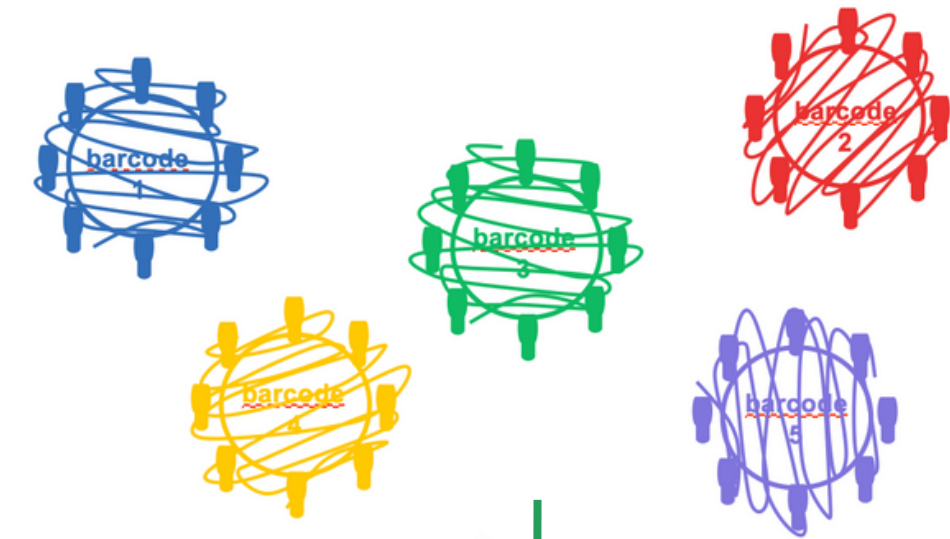
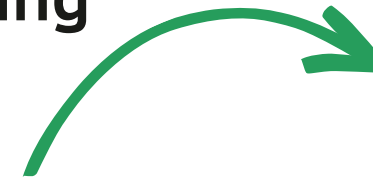


Centaurea corymbosa (Asteraceae)

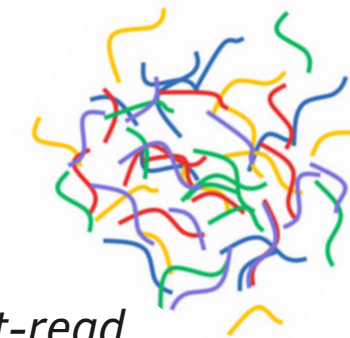


Carte d'Échantillonnage (d'après Ducretet, 2022)

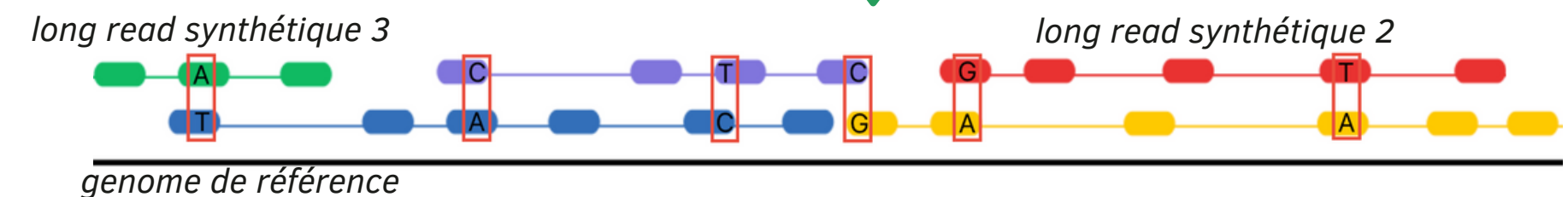
Haplotagging



Molécule d'ADN
liée à la bille



séquençage short-read
et alignement



A C T C G T

haplotype maternel

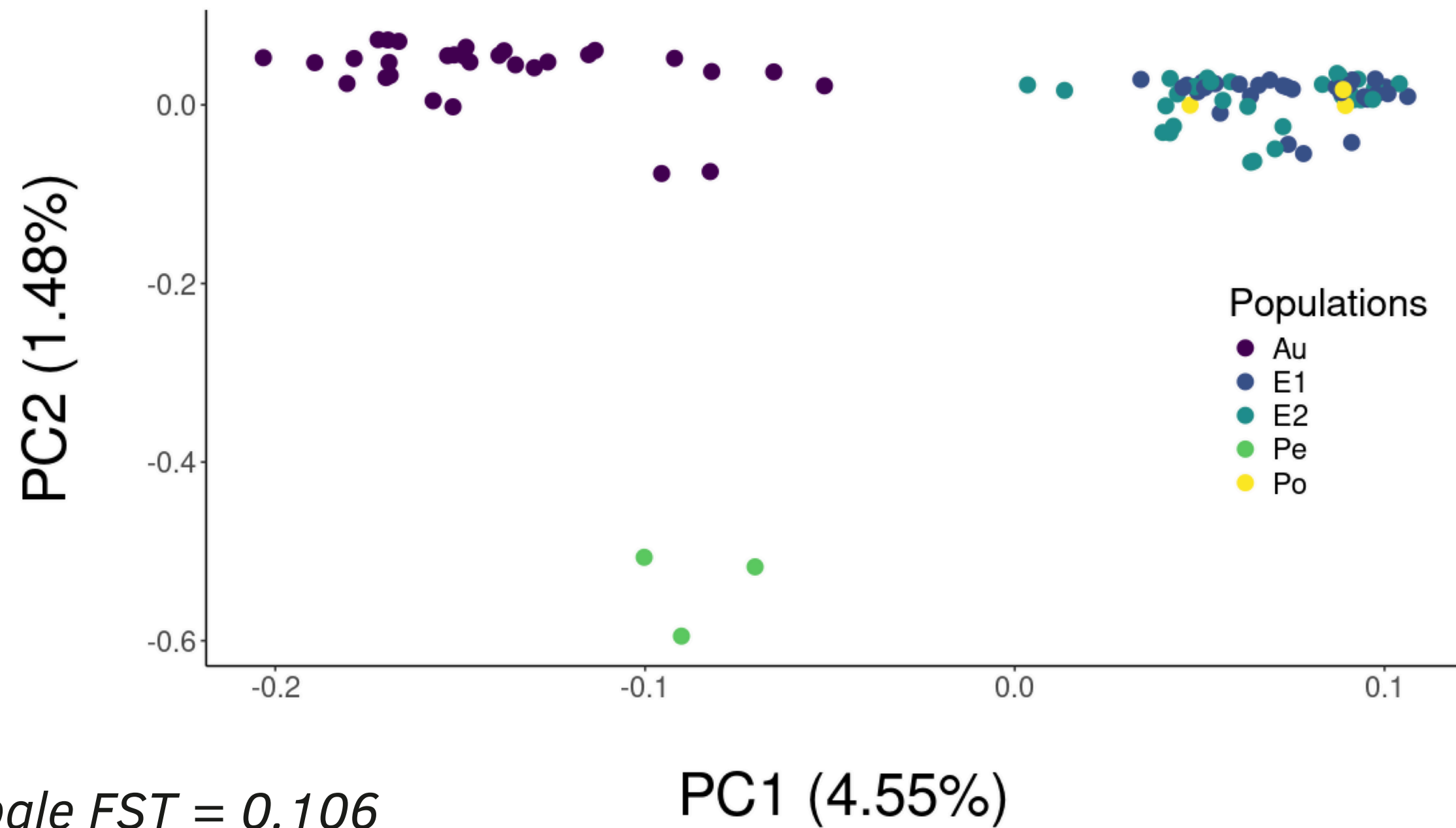
haplotype paternel

T A C G A A

Structure récente révélée par les segments IBD

basé sur les SNP

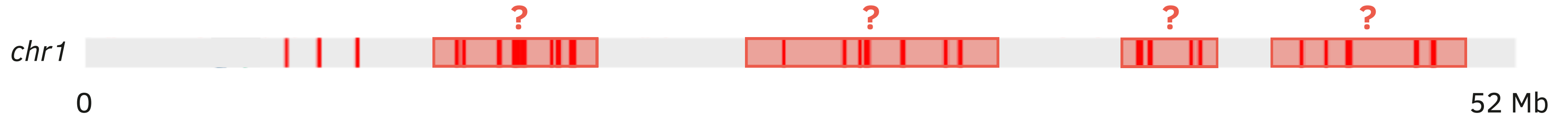
Structure sur le long terme



moyenne globale $F_{ST} = 0.106$

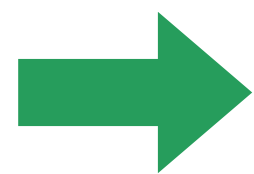
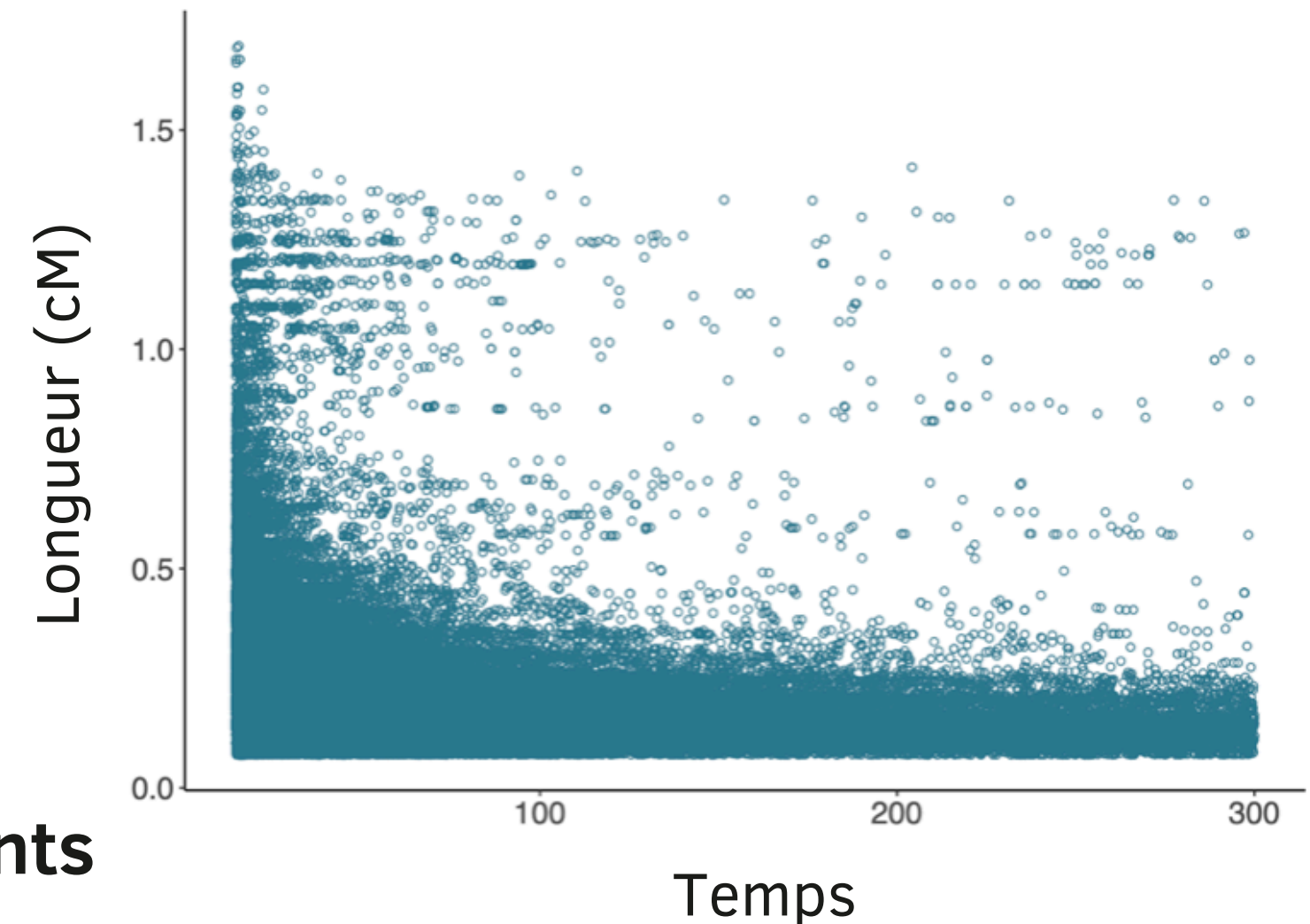
Structure récente révélée par les segments IBD

Signes de fragmentation des segments IBD



Filtration de Qualité :

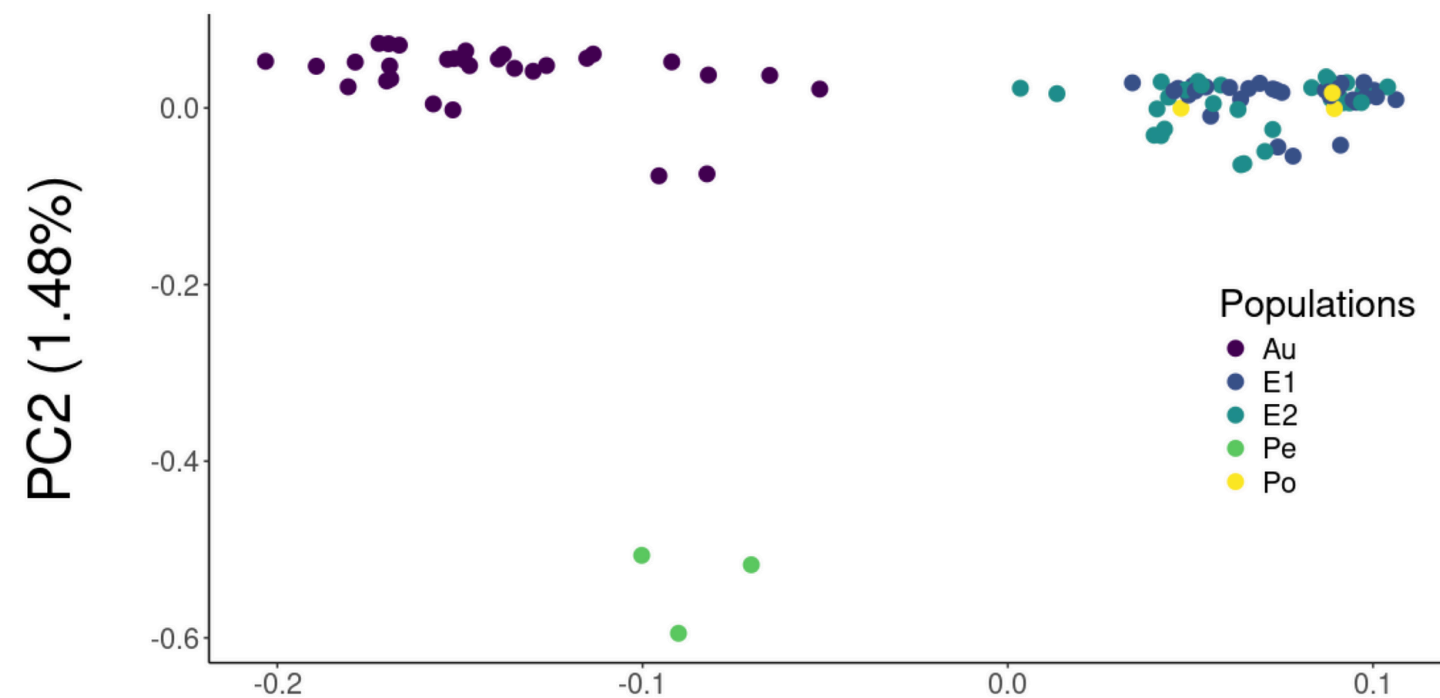
- score de qualité IBD > 0.7
- segment IBD > 75 Mb
 < 300 générations



Segments IBD détectés sont des **fragments**
de **vrais segments IBD** plus longs

Structure récente révélée par les segments IBD

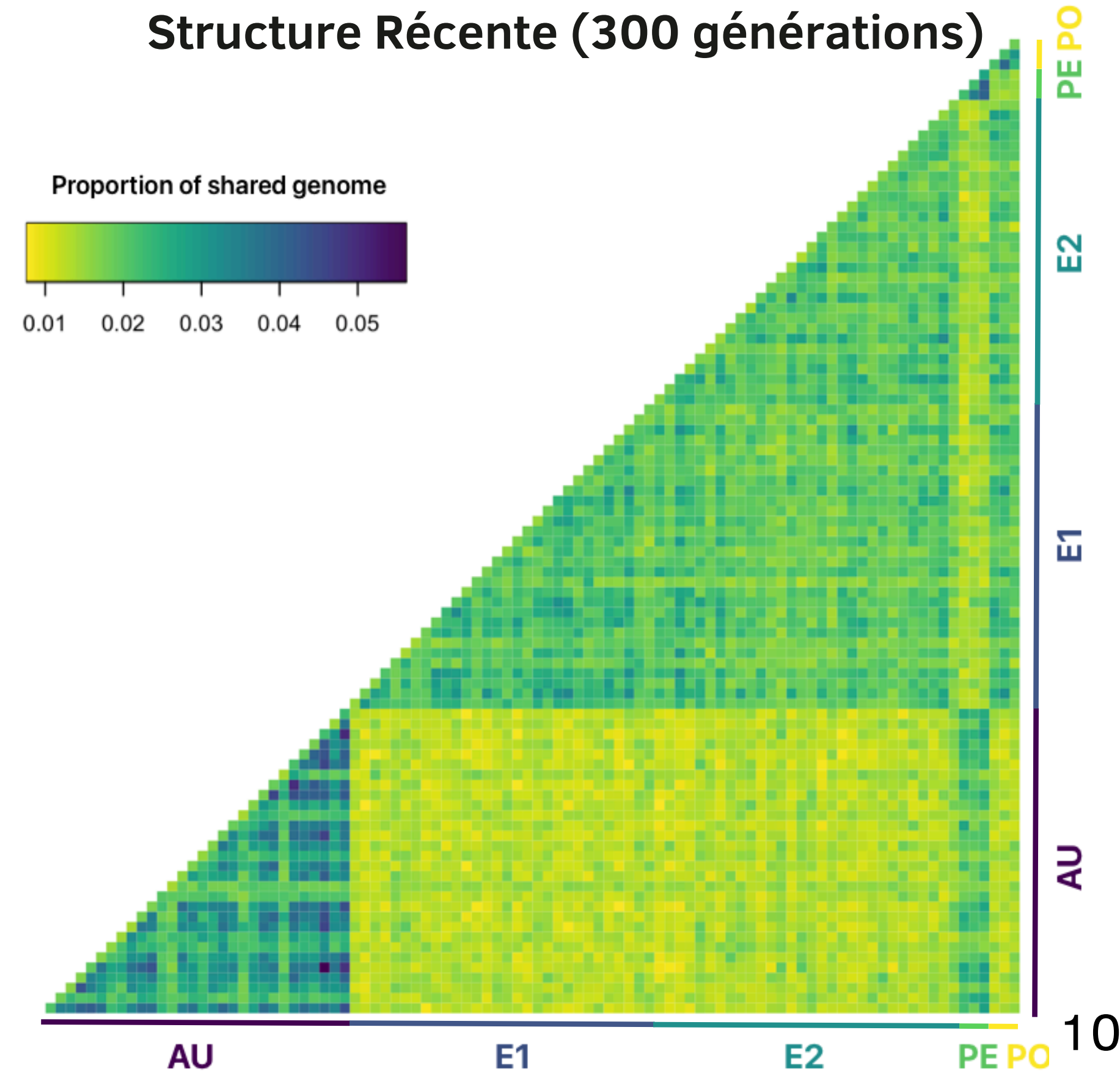
basé sur les SNP
Structure sur le long terme



moyenne globale $F_{ST} = 0.106$

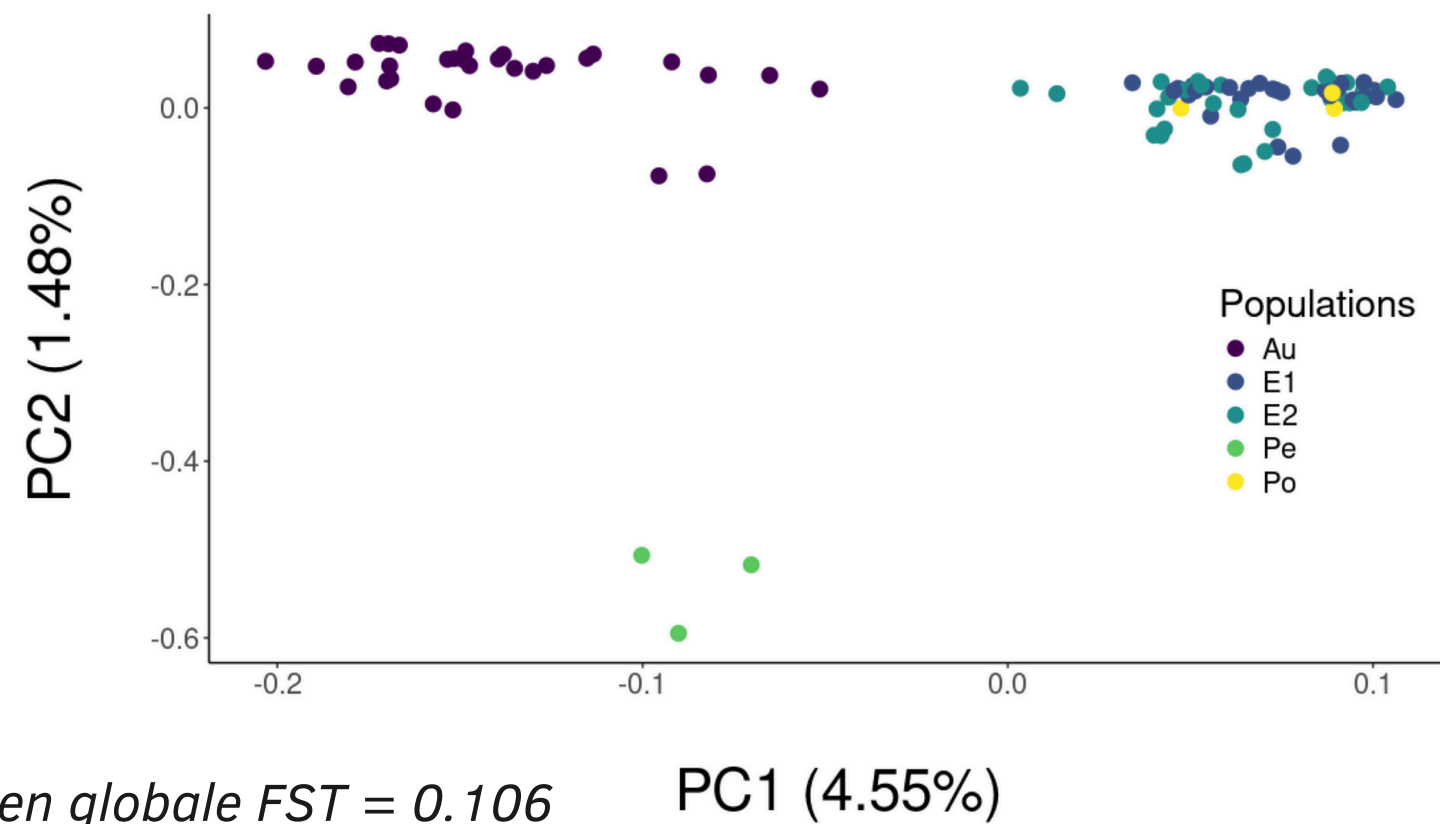
PC1 (4.55%)

basé sur les segments IBD
Structure Récente (300 générations)

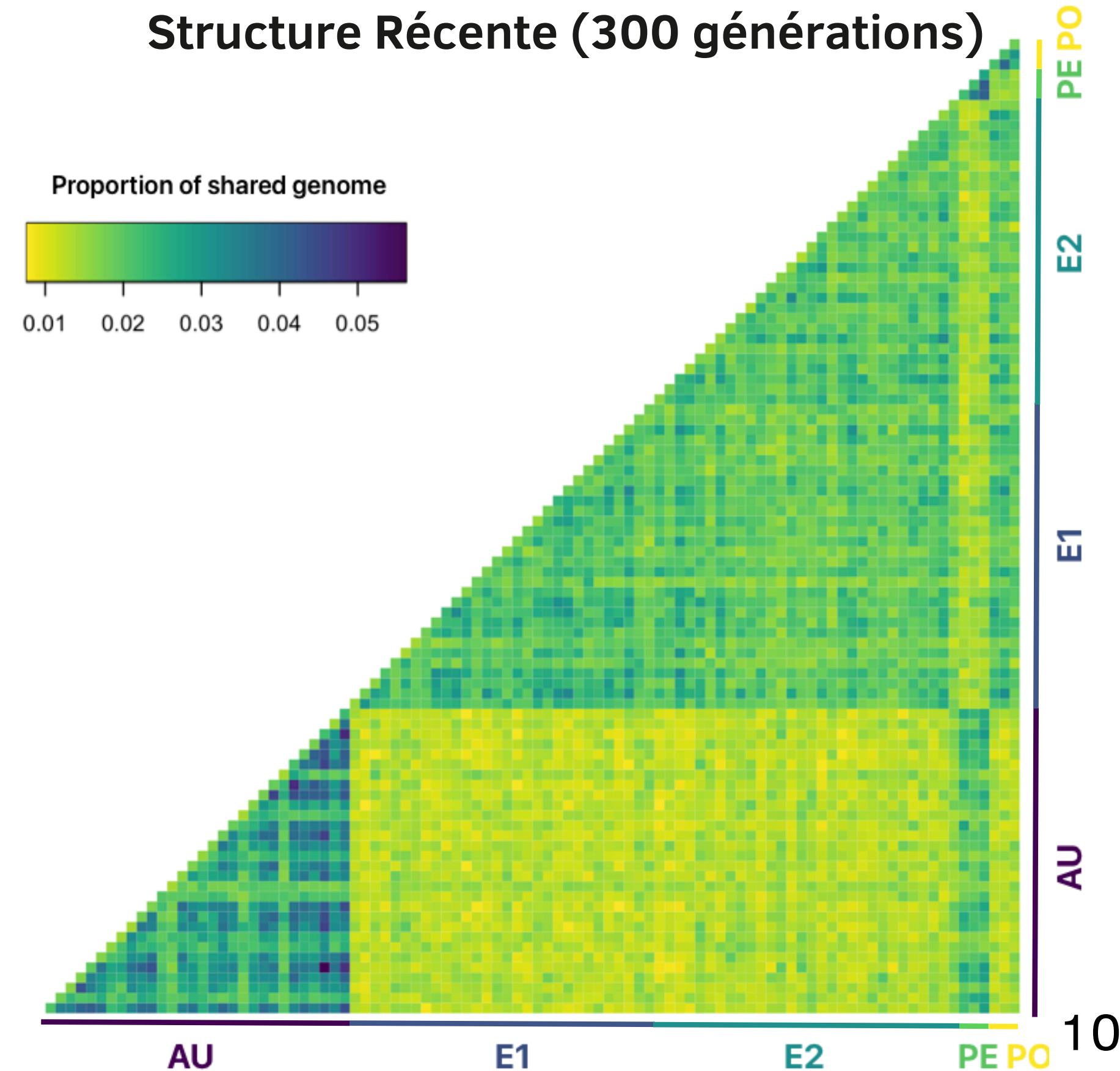


Structure récente révélée par les segments IBD

basé sur les SNP
Structure sur le long terme



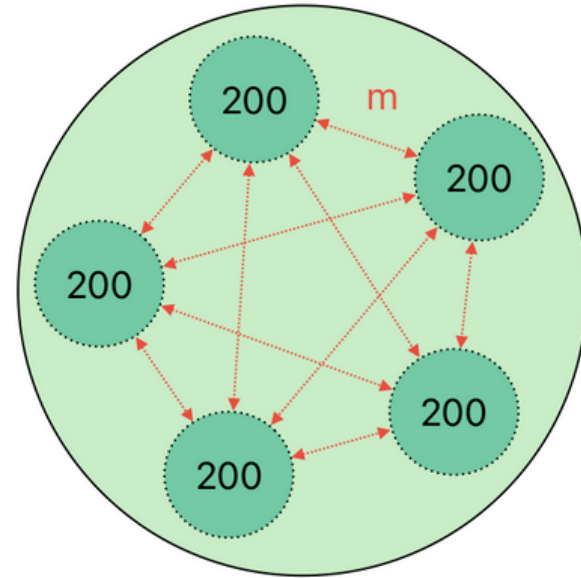
basé sur les segments IBD
Structure Récente (300 générations)



Quand le flux génique a stoppé ?

Dynamique de la Connectivité dans le Passé Récent

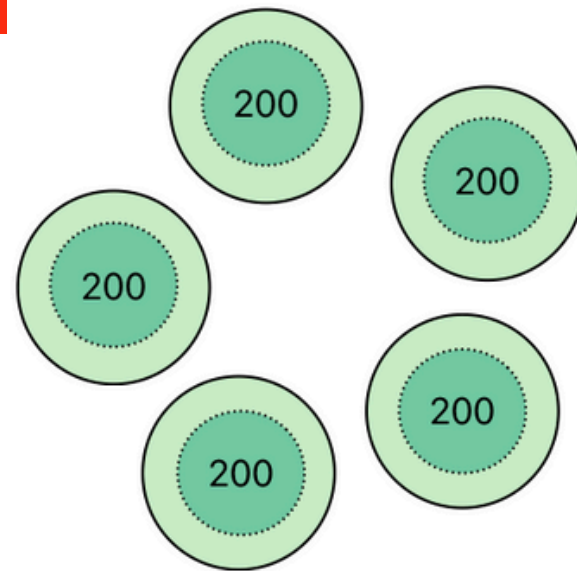
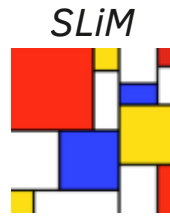
ms prime *simulation coalescente*



$m > 0$

Fragmentation

simulation en avant



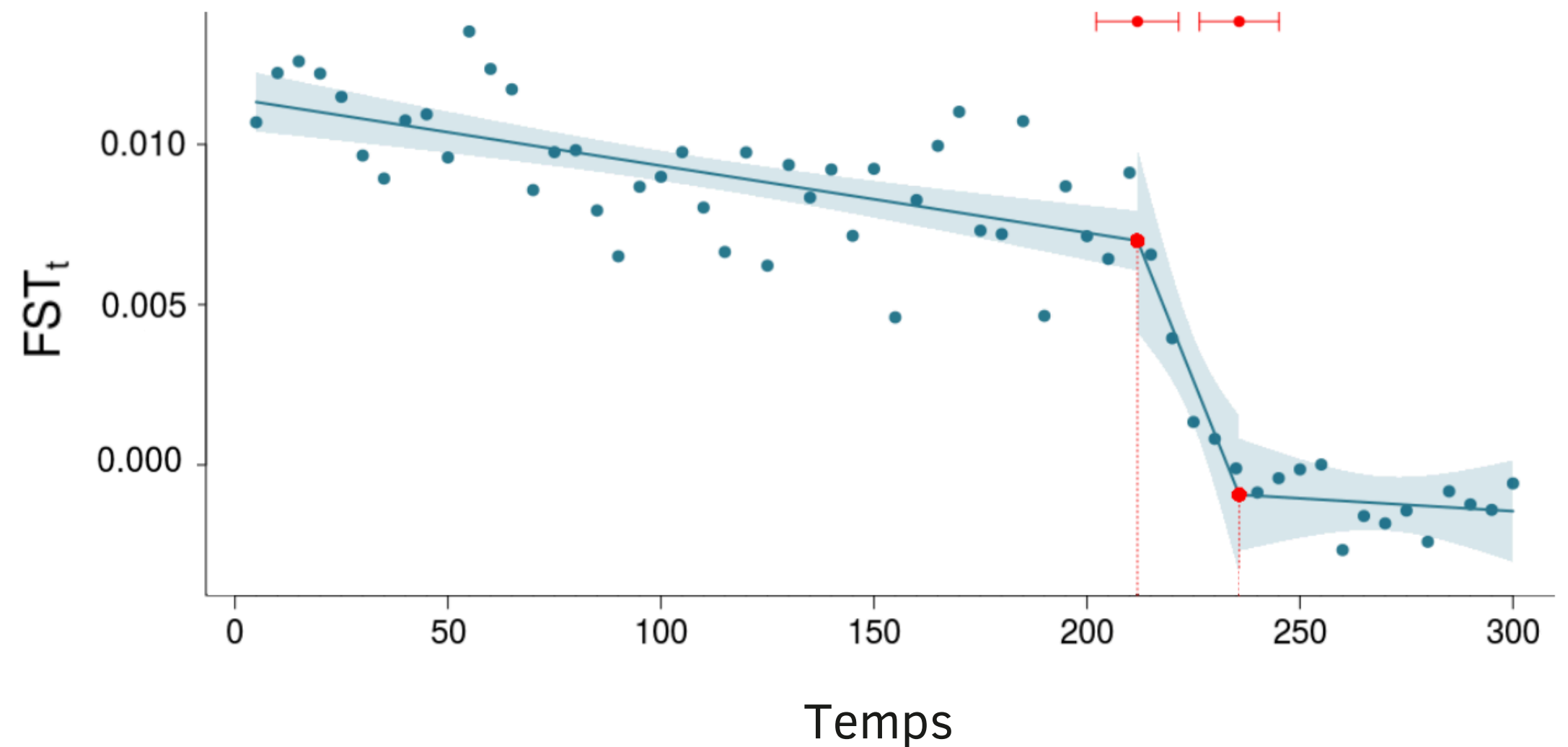
$m = 0$



détection des segments IBD

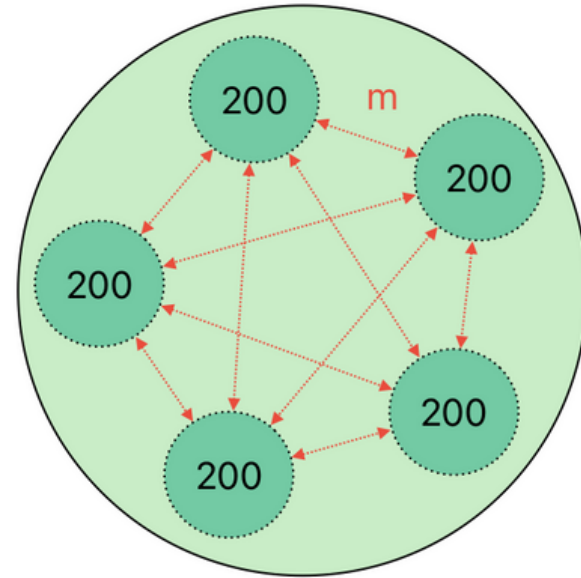
Estimation de l'âge d'arrêt du flux génique

$$FST_t = \frac{Q_{within} - Q_{between}}{1 - Q_{between}}$$

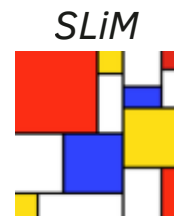


Dynamique de la Connectivité dans le Passé Récent

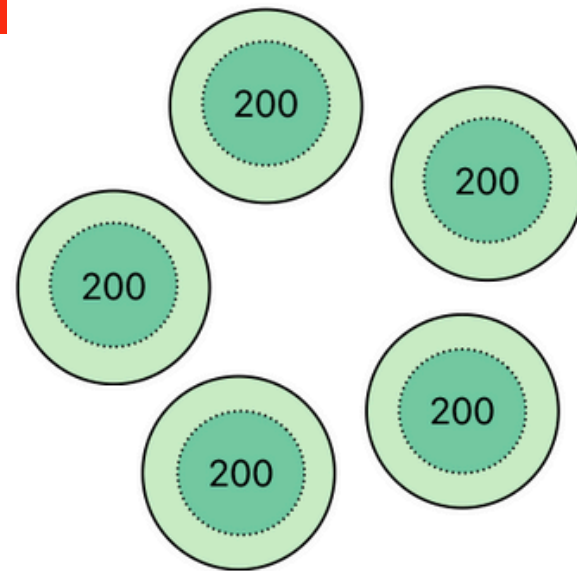
ms prime simulation coalescente



$m > 0$



simulation en avant



$m = 0$

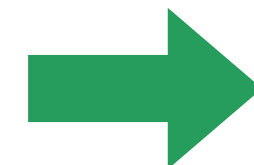
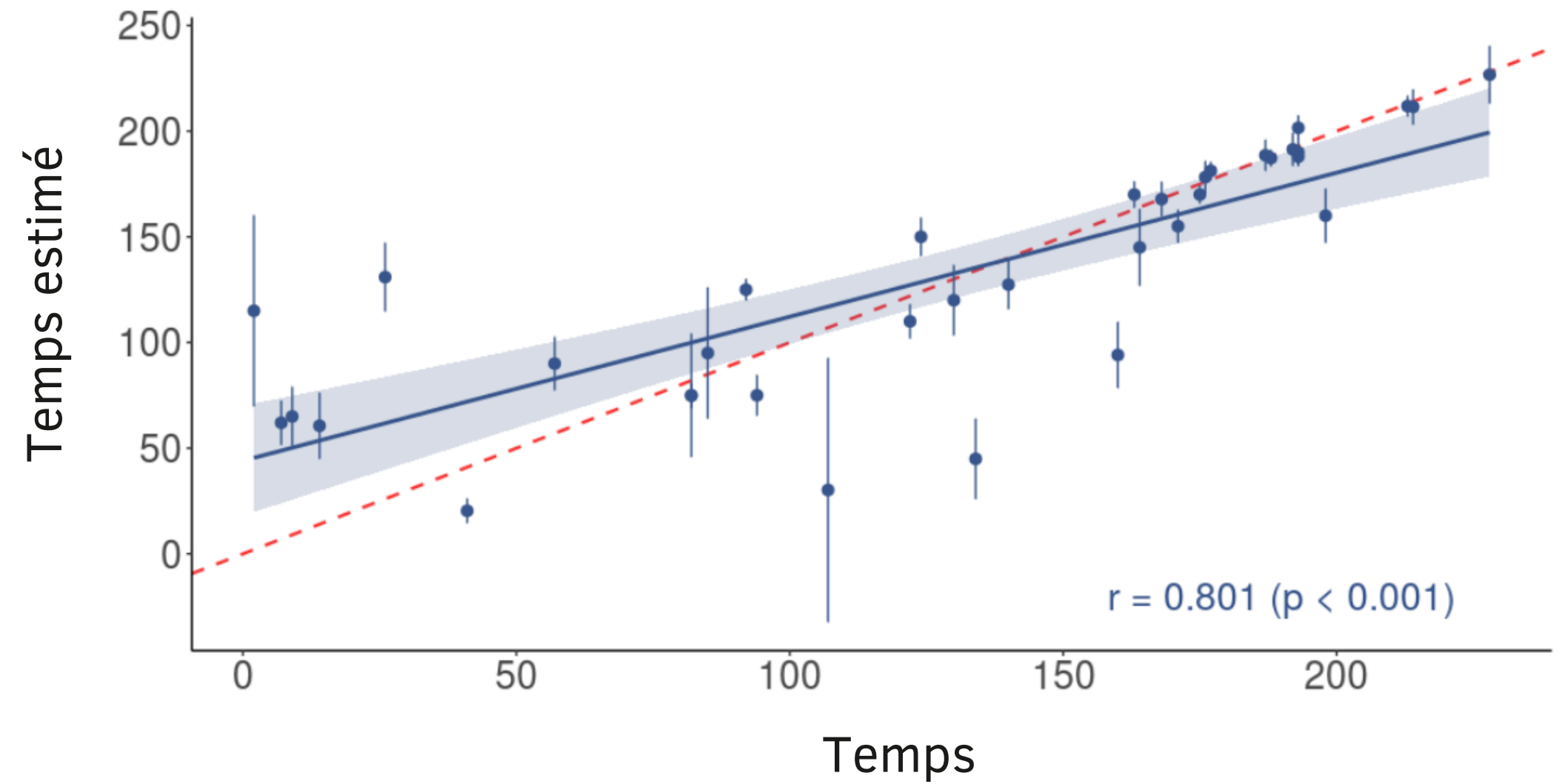
Fragmentation



détection des segments IBD

Estimation de l'âge d'arrêt du flux génique

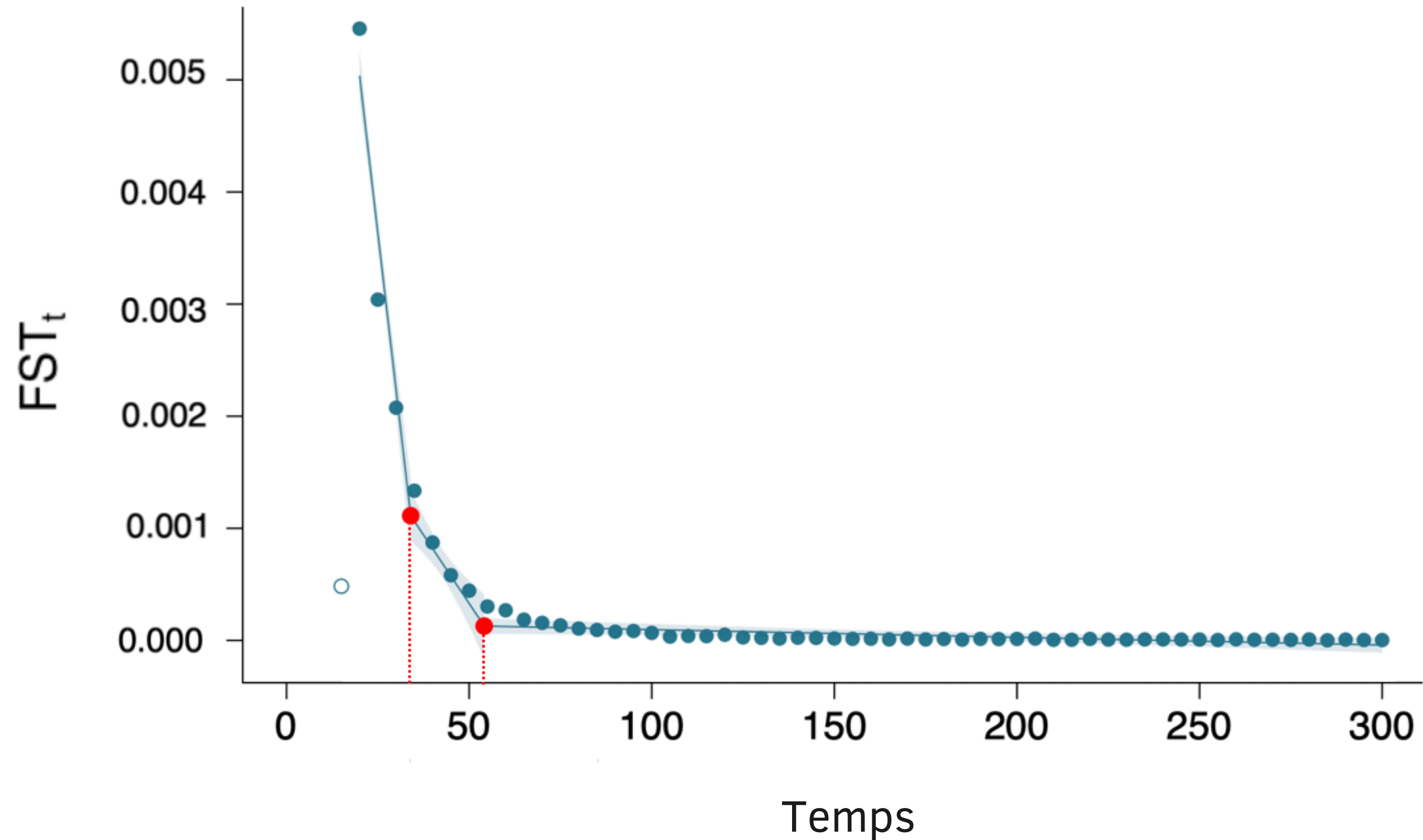
$$FST_t = \frac{Q_{within} - Q_{between}}{1 - Q_{between}}$$



Bonne **approximation** par le temps estimé

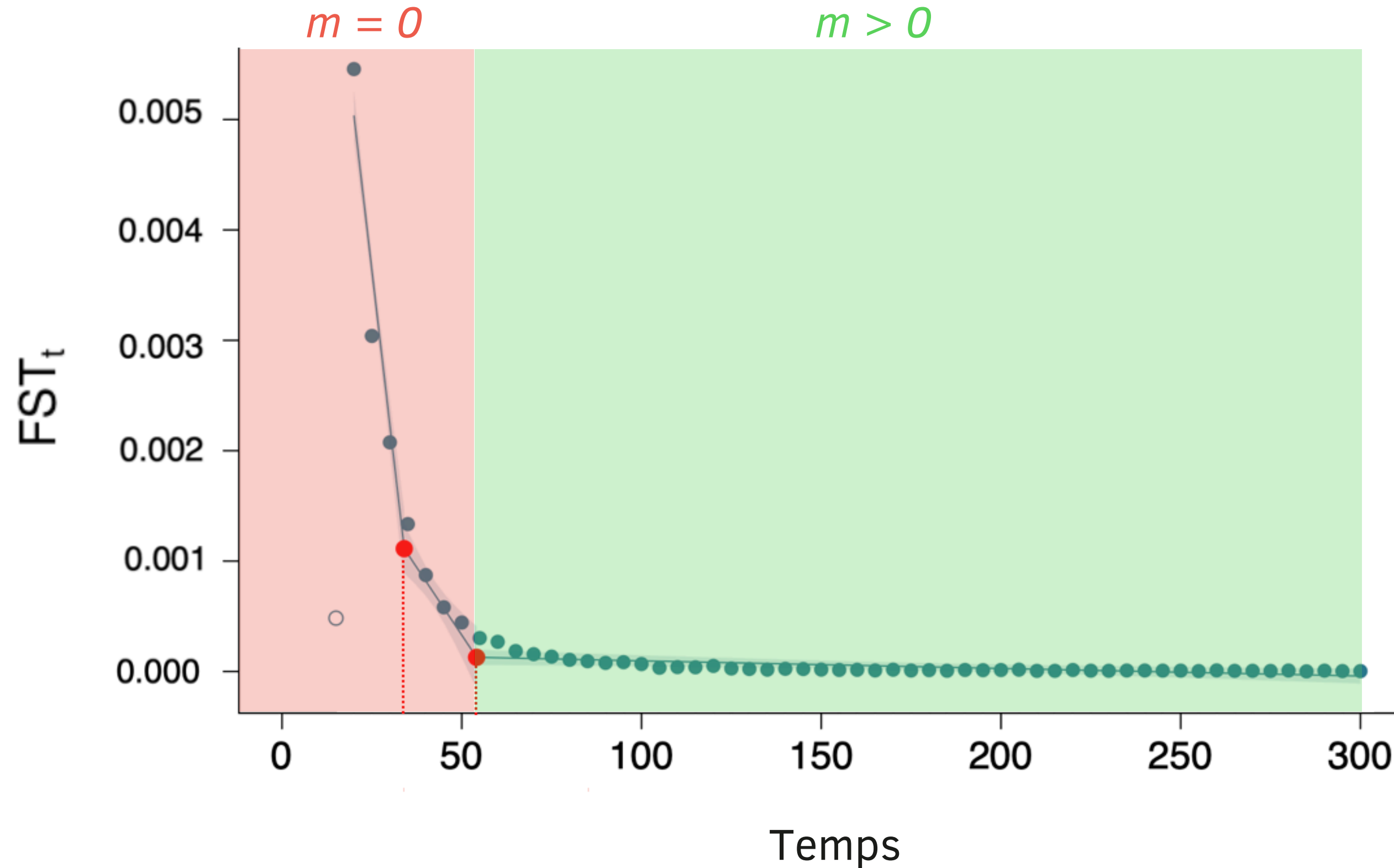
Évidence d'un isolement récent ?

$$FST_t = \frac{Q_{within} - Q_{between}}{1 - Q_{between}}$$



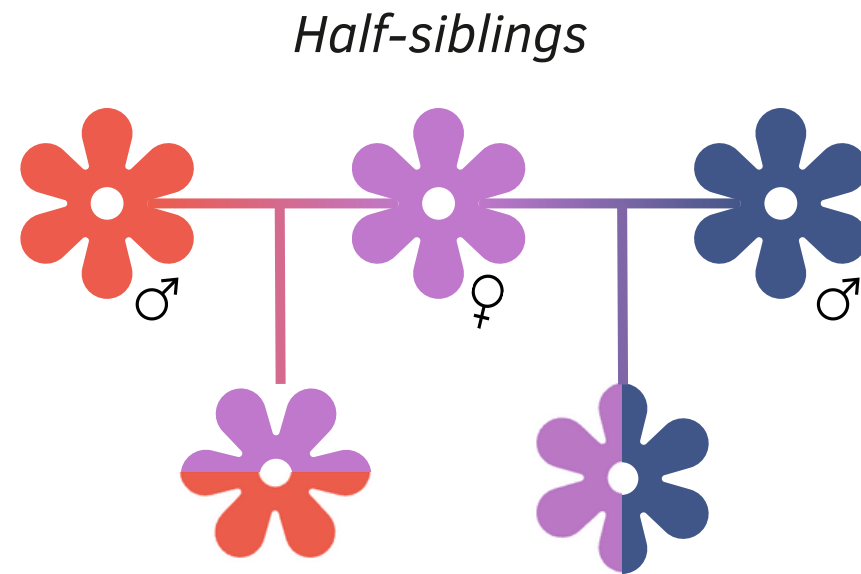
Évidence d'un isolement récent ?

$$FST_t = \frac{Q_{within} - Q_{between}}{1 - Q_{between}}$$



Sous-apparemment évident...

Cas concret des Half-siblings connus

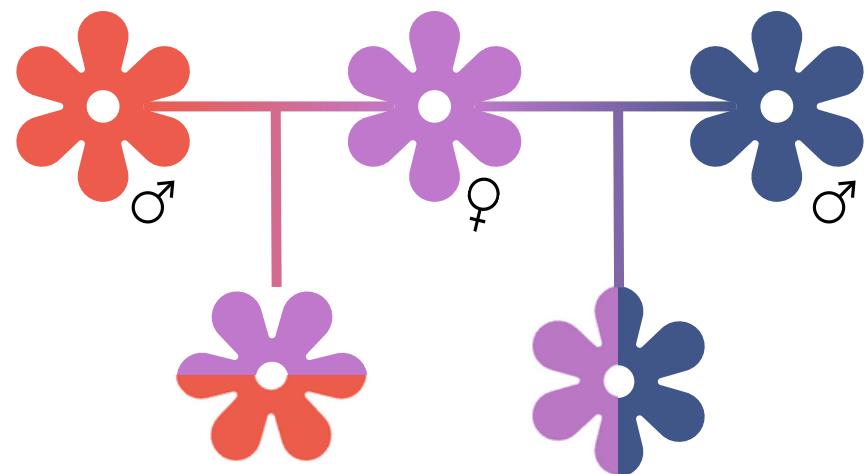


*Attendu : 25% du génome
partagé sous la forme IBD*

Sous-apparemment évident...

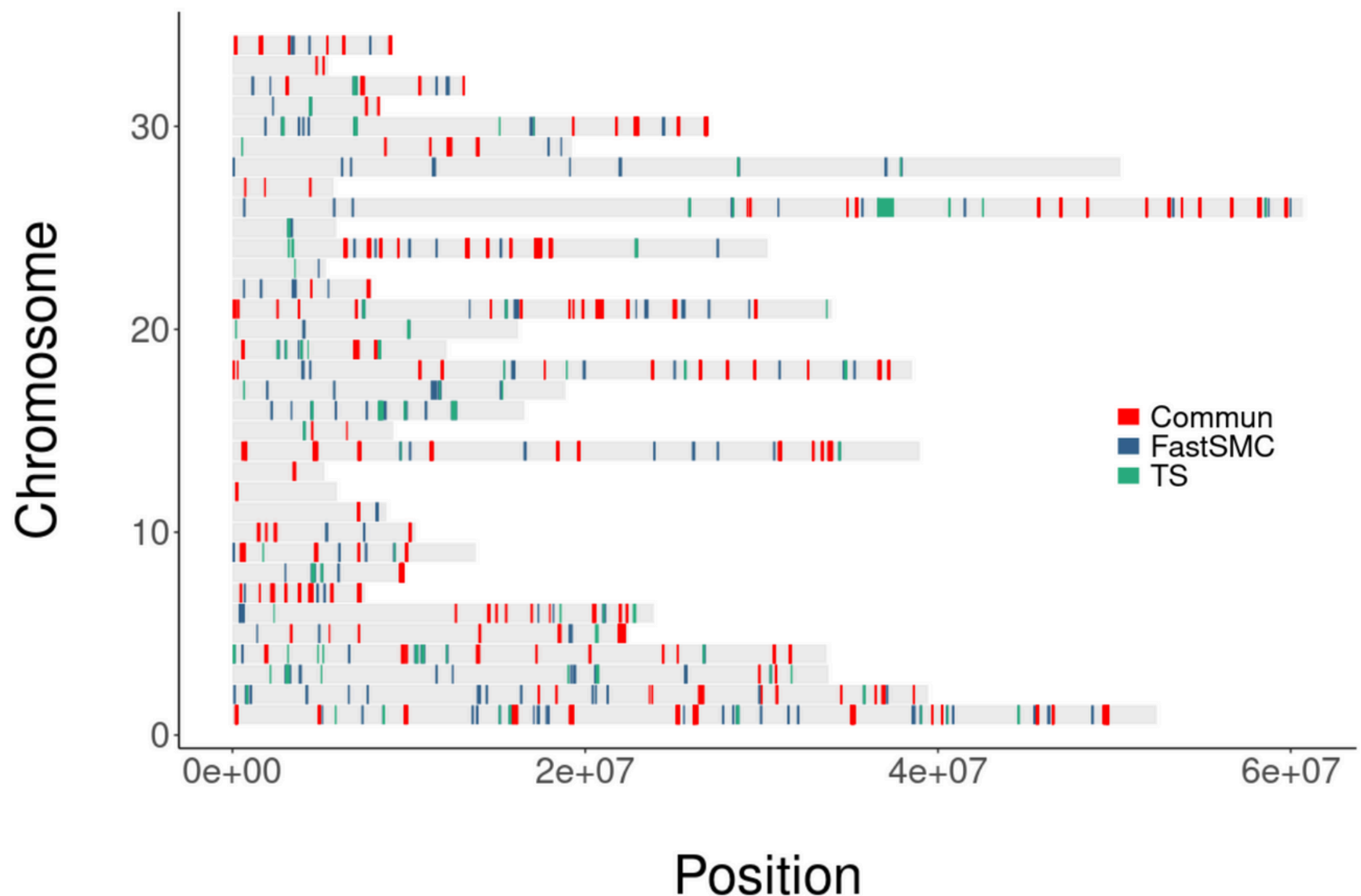
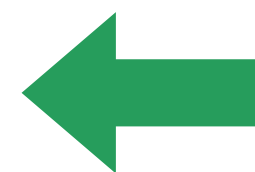
Cas concret des Half-siblings connus

Half-siblings



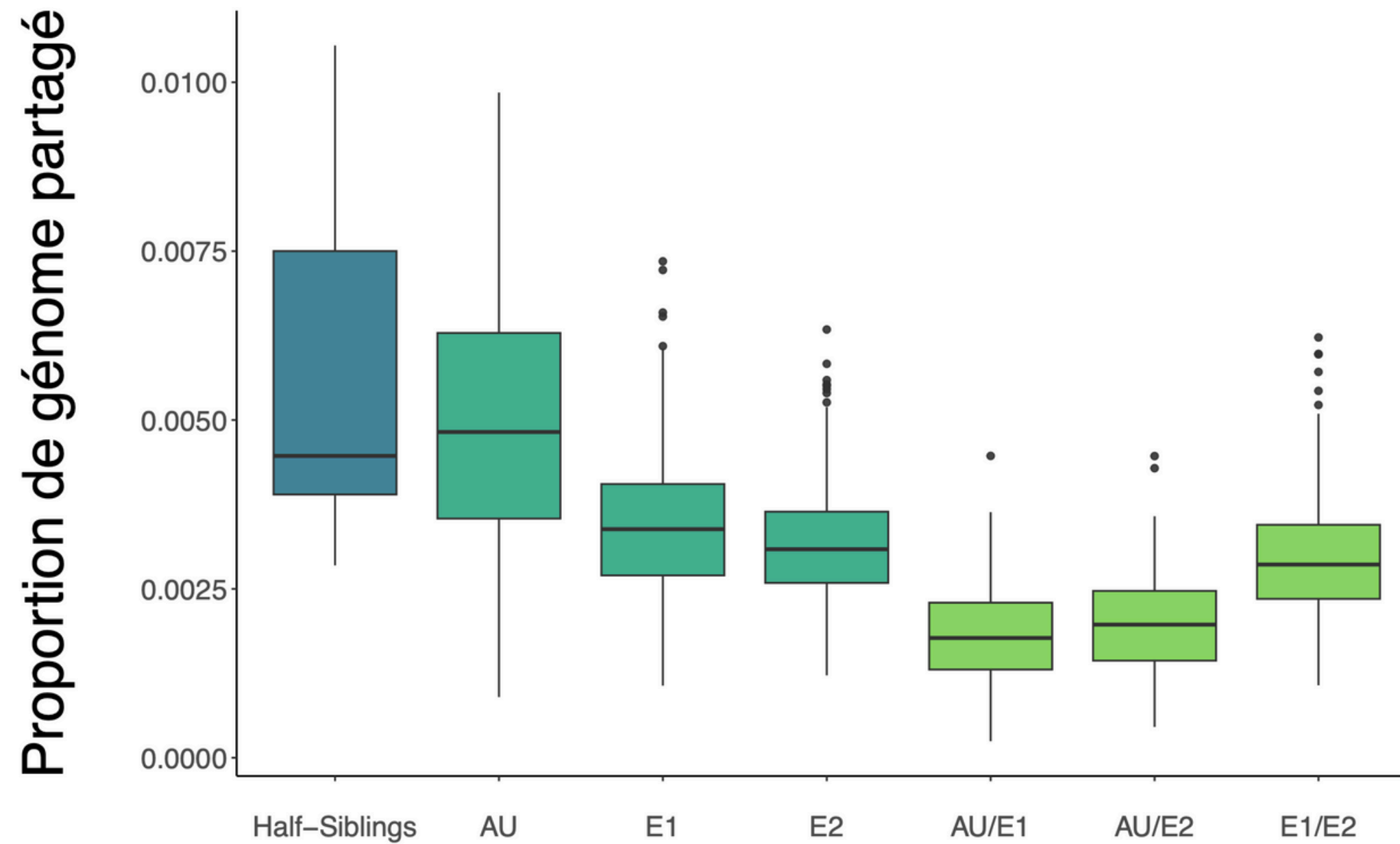
*Attendu : 25% du génome
partagé sous la forme IBD*

En réalité, seulement
5% du génome trouvé
sous la forme d'IBD...



Sous-apparemment évident...

Concordance des niveaux d'apparentement entre catégories



Half-Siblings > Même population > Populations différentes

Sous-apparentement évident...

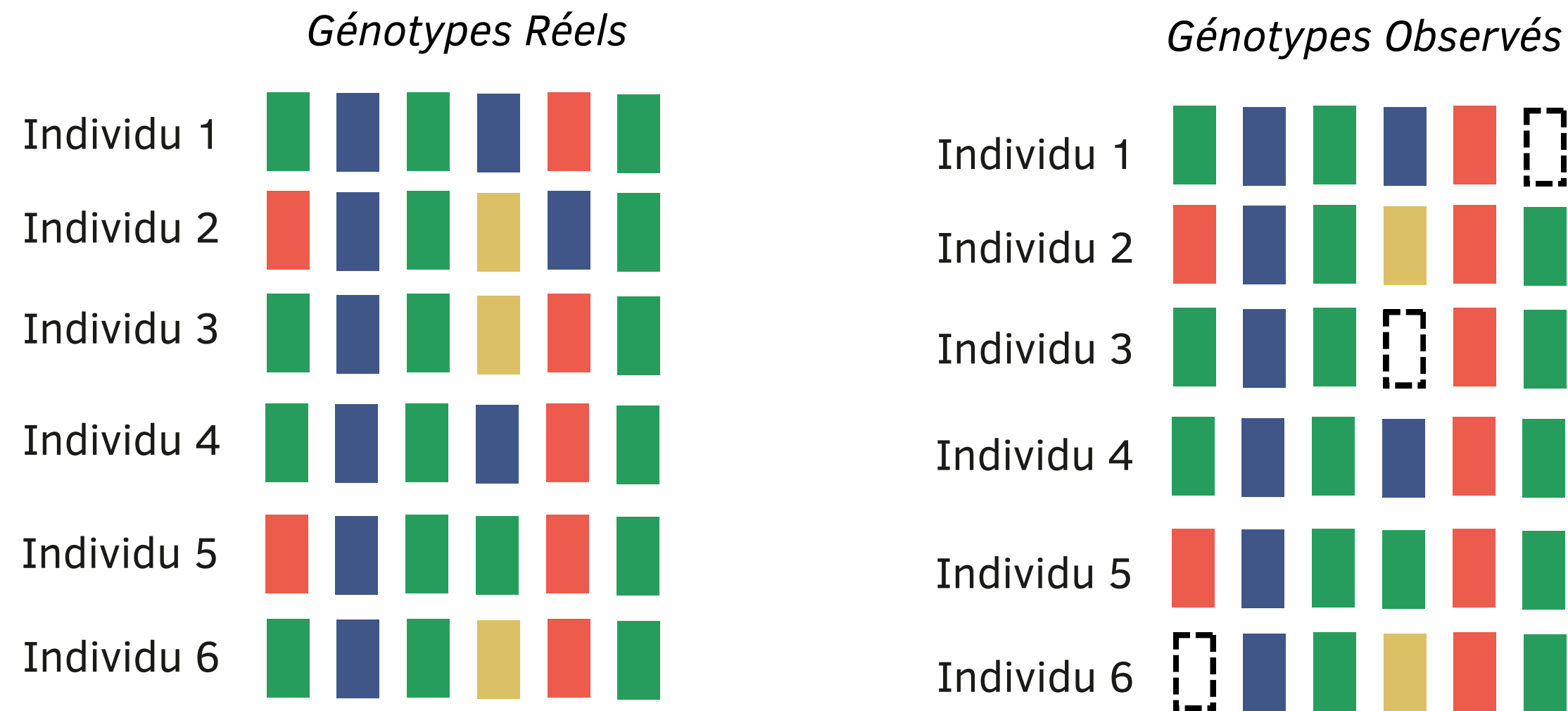
Signes de fragmentation des segments IBD

(1) Imputation des données par STITCH

Sous-apparentement évident...

Signes de fragmentation des segments IBD

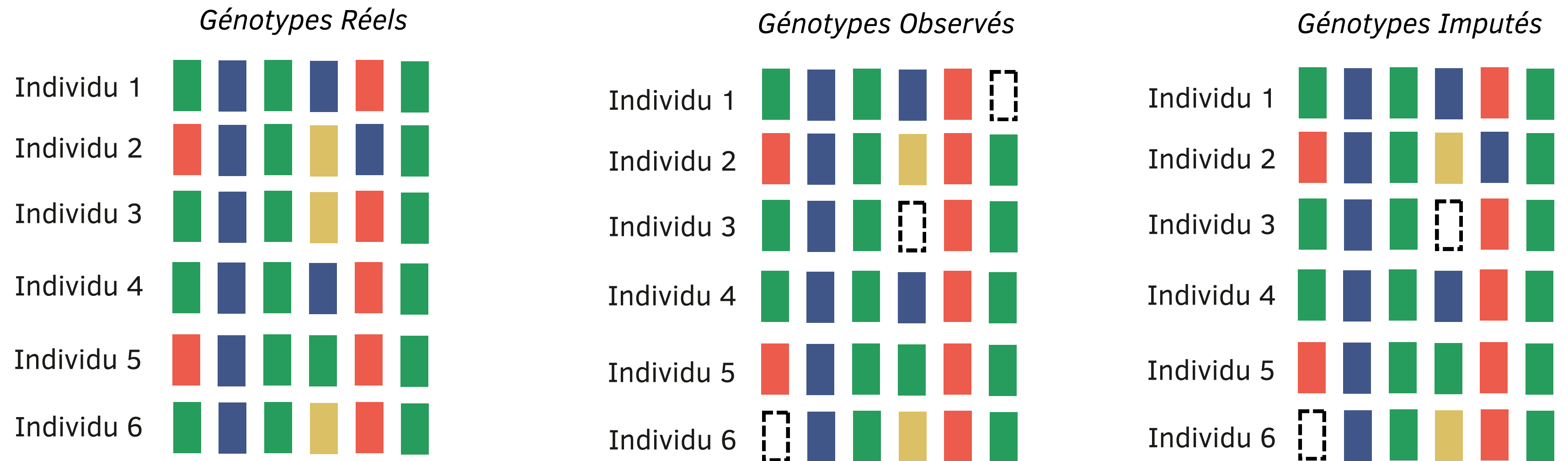
(1) Imputation des données par STITCH



Sous-apparemment évident...

Signes de fragmentation des segments IBD

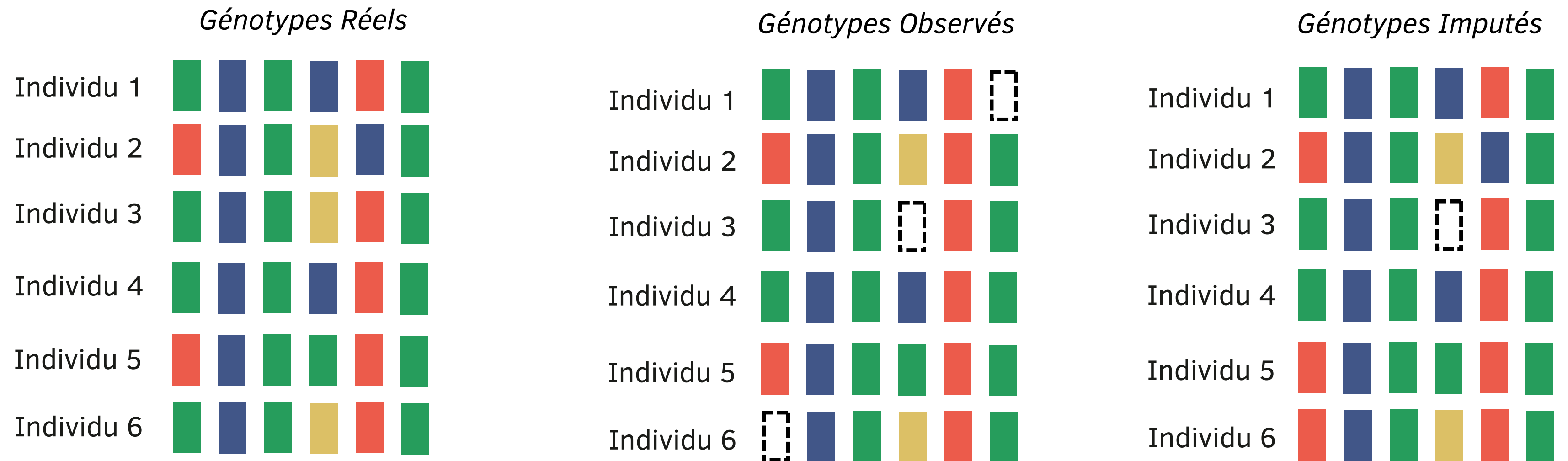
(1) Imputation des données par STITCH



Sous-apparemment évident...

Signes de fragmentation des segments IBD

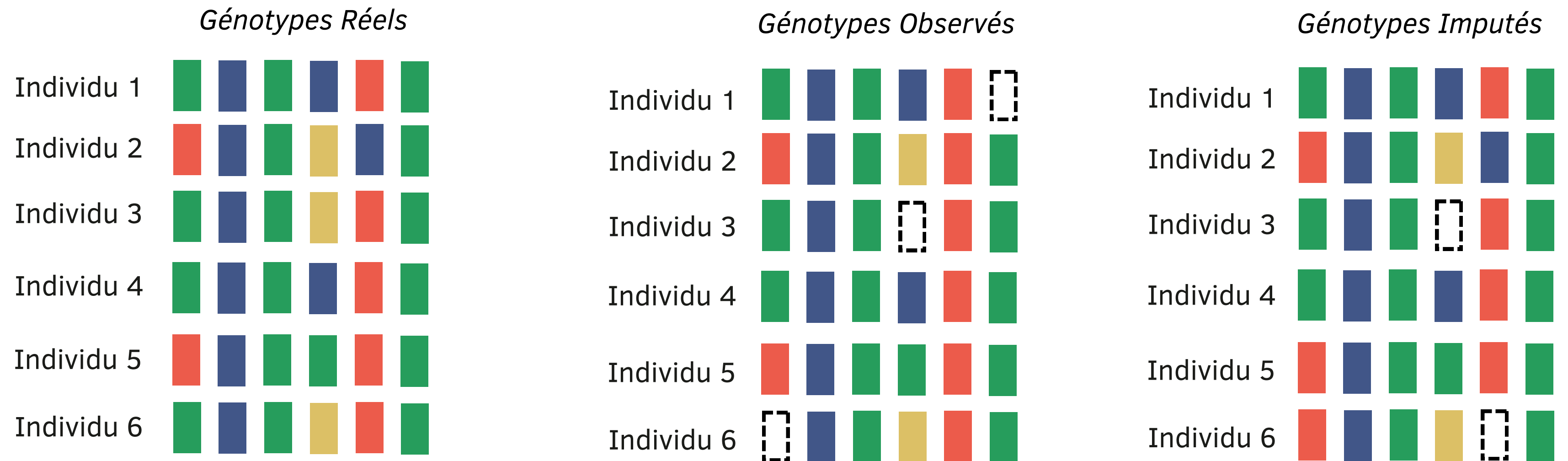
(1) Imputation des données par STITCH



Sous-apparemment évident...

Signes de fragmentation des segments IBD

(1) Imputation des données par STITCH



Sous-apparemment évident...

Signes de fragmentation des segments IBD

(1) Imputation des données par STITCH

Individu 3 


segment IBD

Individu 6 

Individu 3 

Individu 3 


segment IBD

Individu 6 

Individu 6 

Sous-apparemment évident...

Signes de fragmentation des segments IBD

(1) Imputation des données par STITCH

(2) Fragmentation du génome

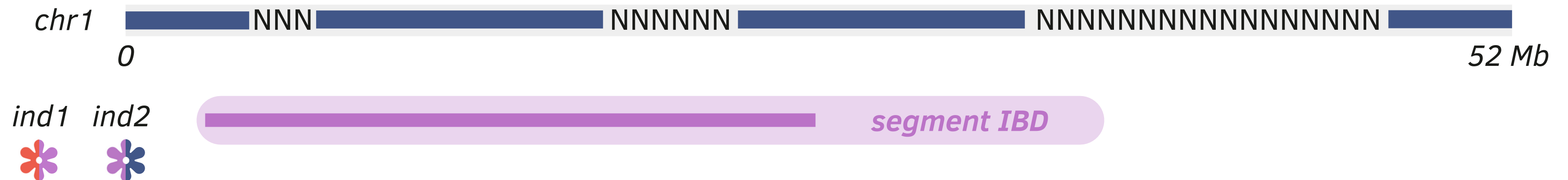


Sous-apparemment évident...

Signes de fragmentation des segments IBD

(1) Imputation des données par STITCH

(2) Fragmentation du génome

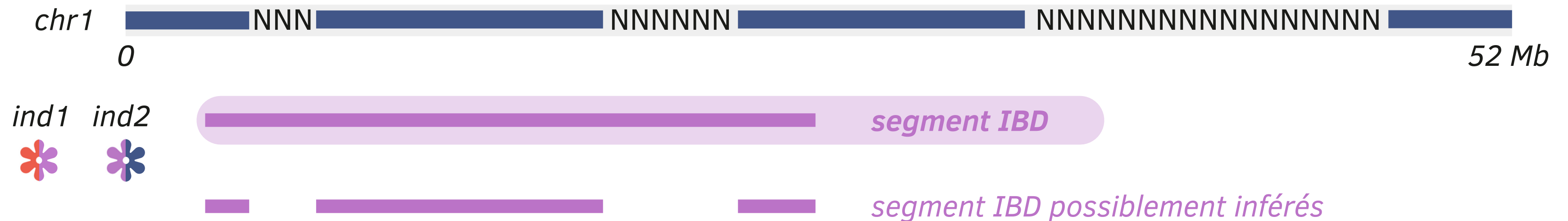


Sous-apparementement évident...

Signes de fragmentation des segments IBD

(1) Imputation des données par STITCH

(2) Fragmentation du génome



Sous-apparemment évident...

Signes de fragmentation des segments IBD

(1) Imputation des données par STITCH

(2) Fragmentation du génome

(3) Faible profondeur de séquençage ~ 2.5X

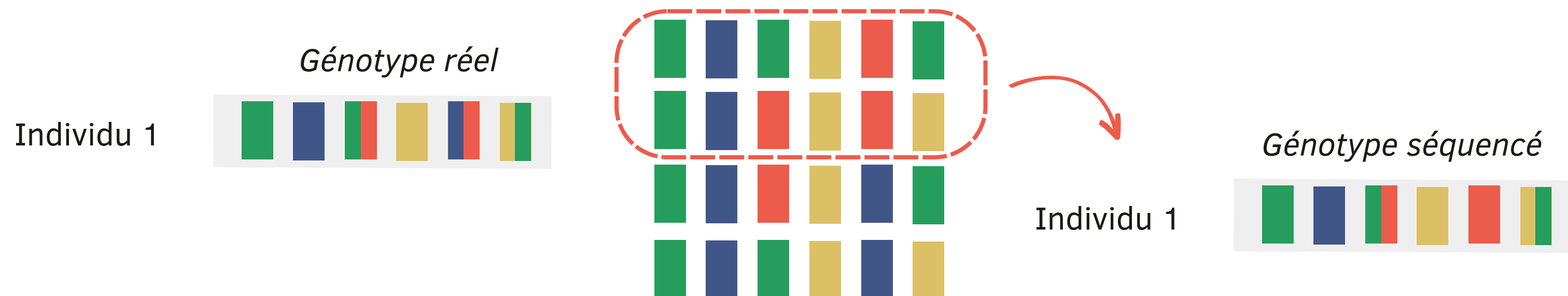
Sous-apparemment évident...

Signes de fragmentation des segments IBD

(1) Imputation des données par STITCH

(2) Fragmentation du génome

(3) Faible profondeur de séquençage $\sim 2.5X$



Sous-apparemment évident...

Signes de fragmentation des segments IBD

(1) Imputation des données par STITCH

(2) Fragmentation du génome

(3) Faible profondeur de séquençage ~ 2.5X

Sous-apparentement évident...

Signes de fragmentation des segments IBD

(1) Imputation des données par STITCH

(2) Fragmentation du génome

(3) Faible profondeur de séquençage

Sous-apparemment évident...

Différents tests sur l'étape d'imputation STITCH

Comparaison du N50 post-phasage = 50 % de la somme totale des longueurs d'haplotypes est portée par des haplotypes dont la longueur est \geq à cette valeur

Sous-apparemment évident...

Différents tests sur l'étape d'imputation **STITCH**

Comparaison du N50 post-phasage = 50 % de la somme totale des longueurs d'haplotypes est portée par des haplotypes dont la longueur est \geq à cette valeur

Scénario de référence : GridWindowSize 50000 K30 = **59827**

	PARSEC
N50	84087

Sous-apparemment évident...

Différents tests sur l'étape d'imputation **STITCH**

Comparaison du N50 post-phasage = 50 % de la somme totale des longueurs d'haplotypes est portée par des haplotypes dont la longueur est \geq à cette valeur

Scénario de référence : GridWindowSize 50000 K30 = **59827**

	PARSEC	SNPs-based
N50	84087	82288

Sous-apparemment évident...

Différents tests sur l'étape d'imputation STITCH

Comparaison du N50 post-phasage = 50 % de la somme totale des longueurs d'haplotypes est portée par des haplotypes dont la longueur est \geq à cette valeur

Scénario de référence : GridWindowSize 50000 K30 = **59827**

	PARSEC	SNPs-based	GridWindowSize 10000 100000
N50	84087	82288	61620 59421

Sous-apparemment évident...

Différents tests sur l'étape d'imputation STITCH

Comparaison du N50 post-phasage = 50 % de la somme totale des longueurs d'haplotypes est portée par des haplotypes dont la longueur est \geq à cette valeur

Scénario de référence : GridWindowSize 50000 K30 = **59827**

	PARSEC	SNPs-based	GridWindowSize 10000 100000	GridWindowSize 50000 + Buffer
N50	84087	82288	61620 59421	85906

Sous-apparemment évident...

Différents tests sur l'étape d'imputation STITCH

Comparaison du N50 post-phasage = 50 % de la somme totale des longueurs d'haplotypes est portée par des haplotypes dont la longueur est \geq à cette valeur

Scénario de référence : GridWindowSize 50000 K30 = **59827**

	PARSEC	SNPs-based	GridWindowSize 10000 100000	GridWindowSize 50000 + Buffer	K10
N50	84087	82288	61620 59421	85906	54322

Sous-apparemment évident...

Différents tests sur l'étape d'imputation STITCH

Comparaison du N50 post-phasage = 50 % de la somme totale des longueurs d'haplotypes est portée par des haplotypes dont la longueur est \geq à cette valeur

Scénario de référence : GridWindowSize 50000 K30 = **59827**

	PARSEC	SNPs-based	GridWindowsize 10000 100000	GridWindowSize 50000 + Buffer	K10	Filtration en amont
N50	84087	82288	61620 59421	85906	54322	89202

Sous-apparemment évident...

Différents tests sur l'étape d'imputation STITCH

Comparaison du N50 post-phasage = 50 % de la somme totale des longueurs d'haplotypes est portée par des haplotypes dont la longueur est \geq à cette valeur

Scénario de référence : GridWindowSize 50000 K30 = **59827**

	PARSEC	SNPs-based	GridWindowSize 10000 100000	GridWindowSize 50000 + Buffer	K10	Filtration en amont	Filtration en aval
N50	84087	82288	61620 59421	85906	54322	89202	53863

Sous-apparemment évident...

Différents tests sur l'étape d'imputation STITCH

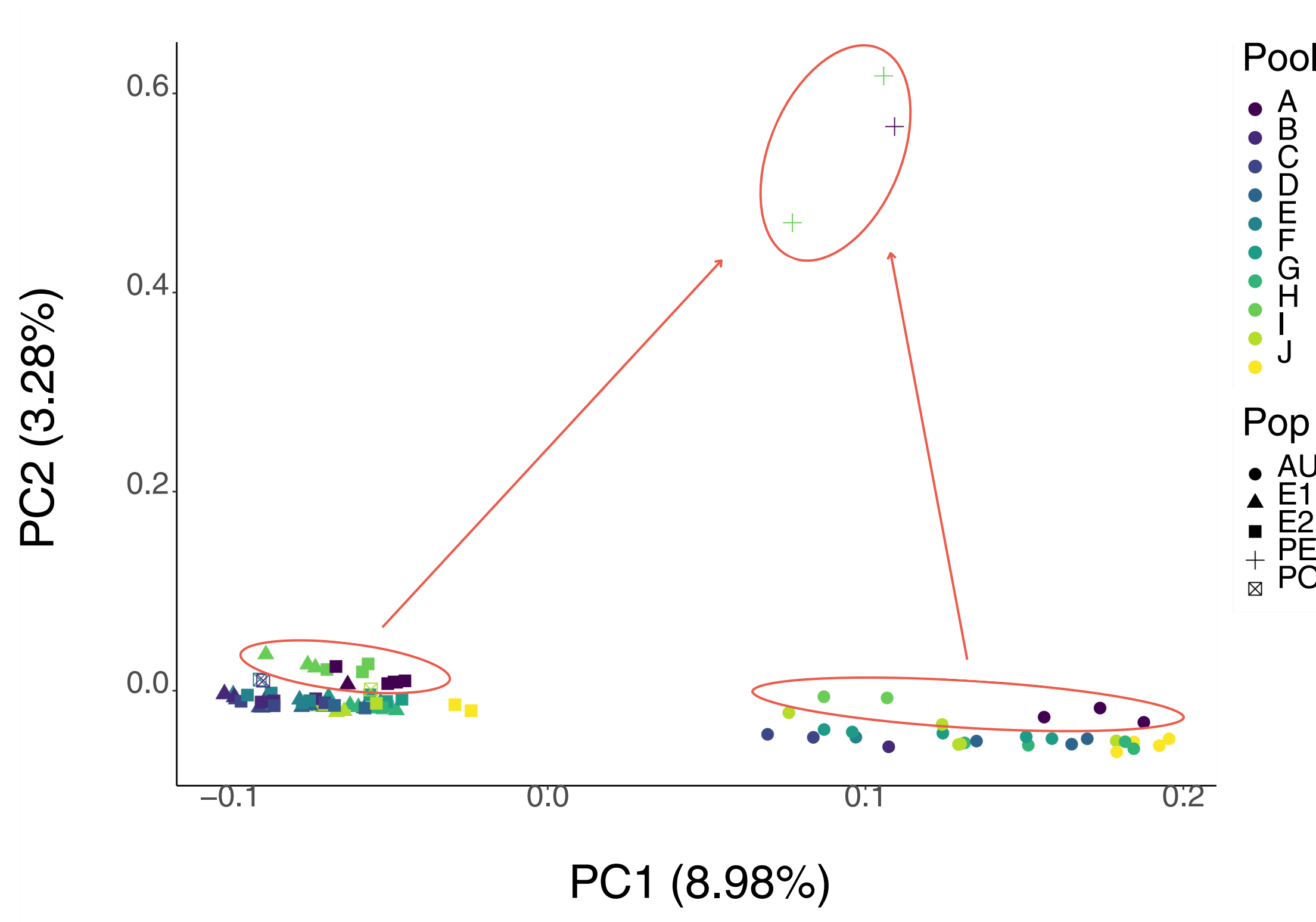
Comparaison du N50 post-phasage = 50 % de la somme totale des longueurs d'haplotypes est portée par des haplotypes dont la longueur est \geq à cette valeur

Scénario de référence : GridWindowSize 50000 K30 = **59827**

	PARSEC	SNPs-based	GridWindowsize 10000 100000	GridWindowSize 50000 + Buffer	K10	Filtration en amont	Filtration en aval	Sans STITCH
N50	84087	82288	61620 59421	85906	54322	89202	53863	93820

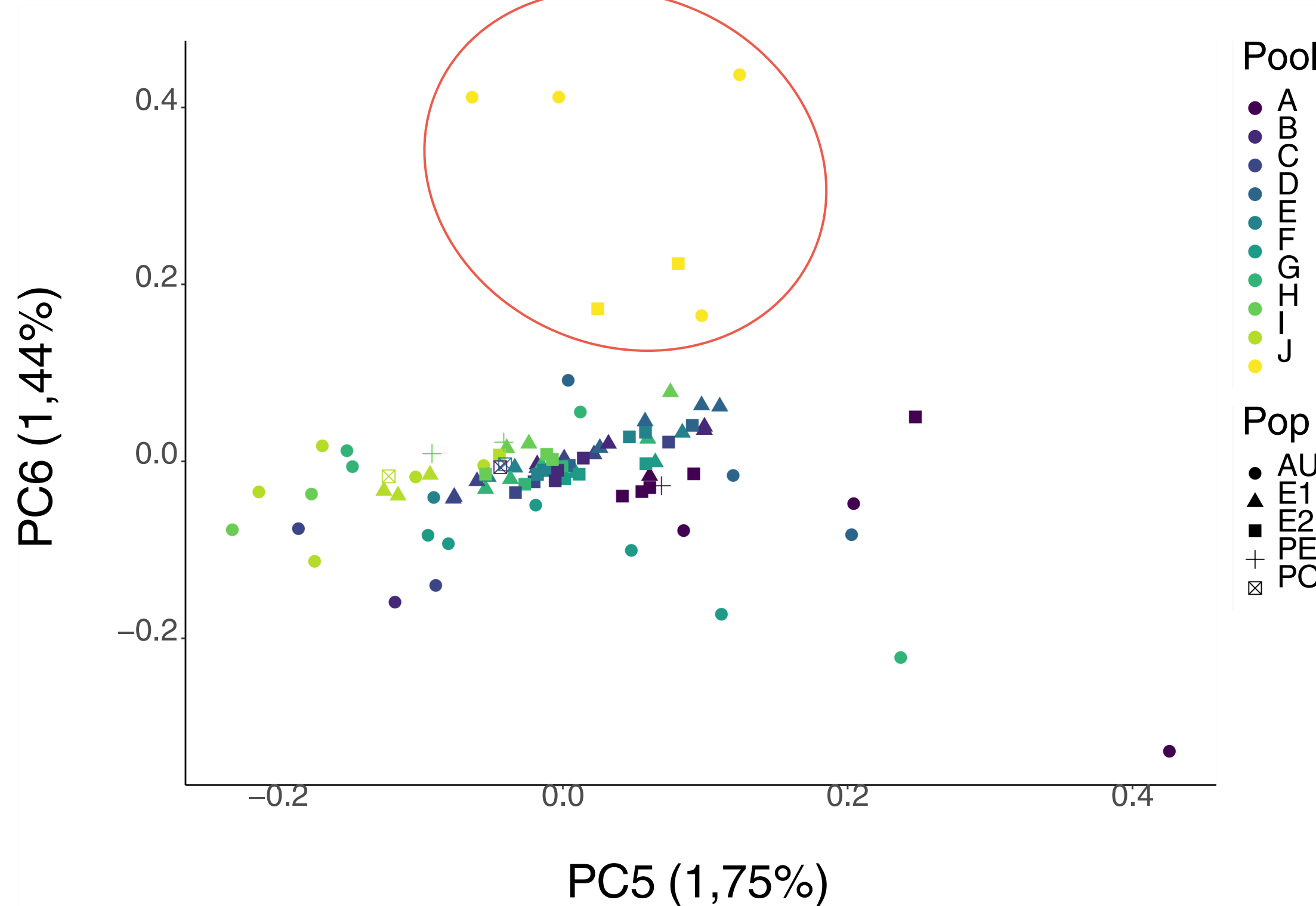
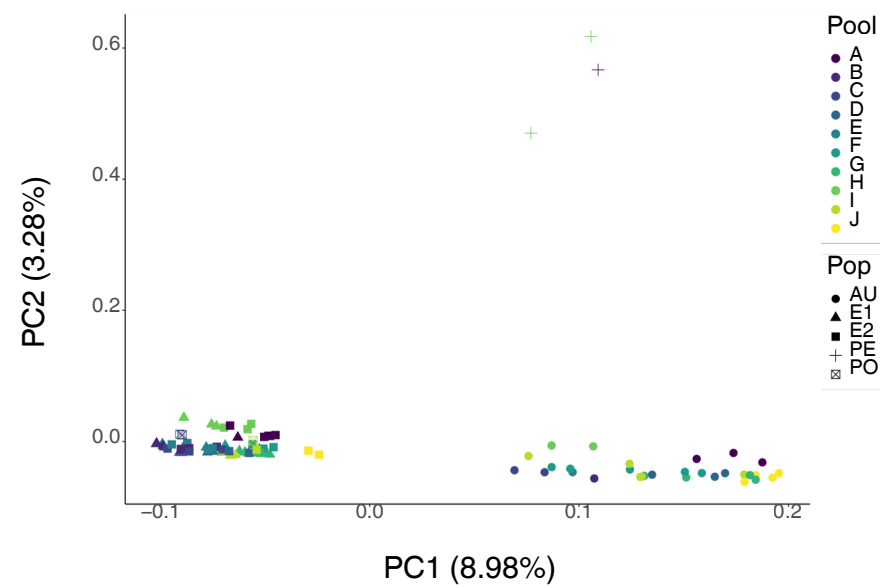
Sous-apparemment évident...

Contamination lors de la PCR poolé



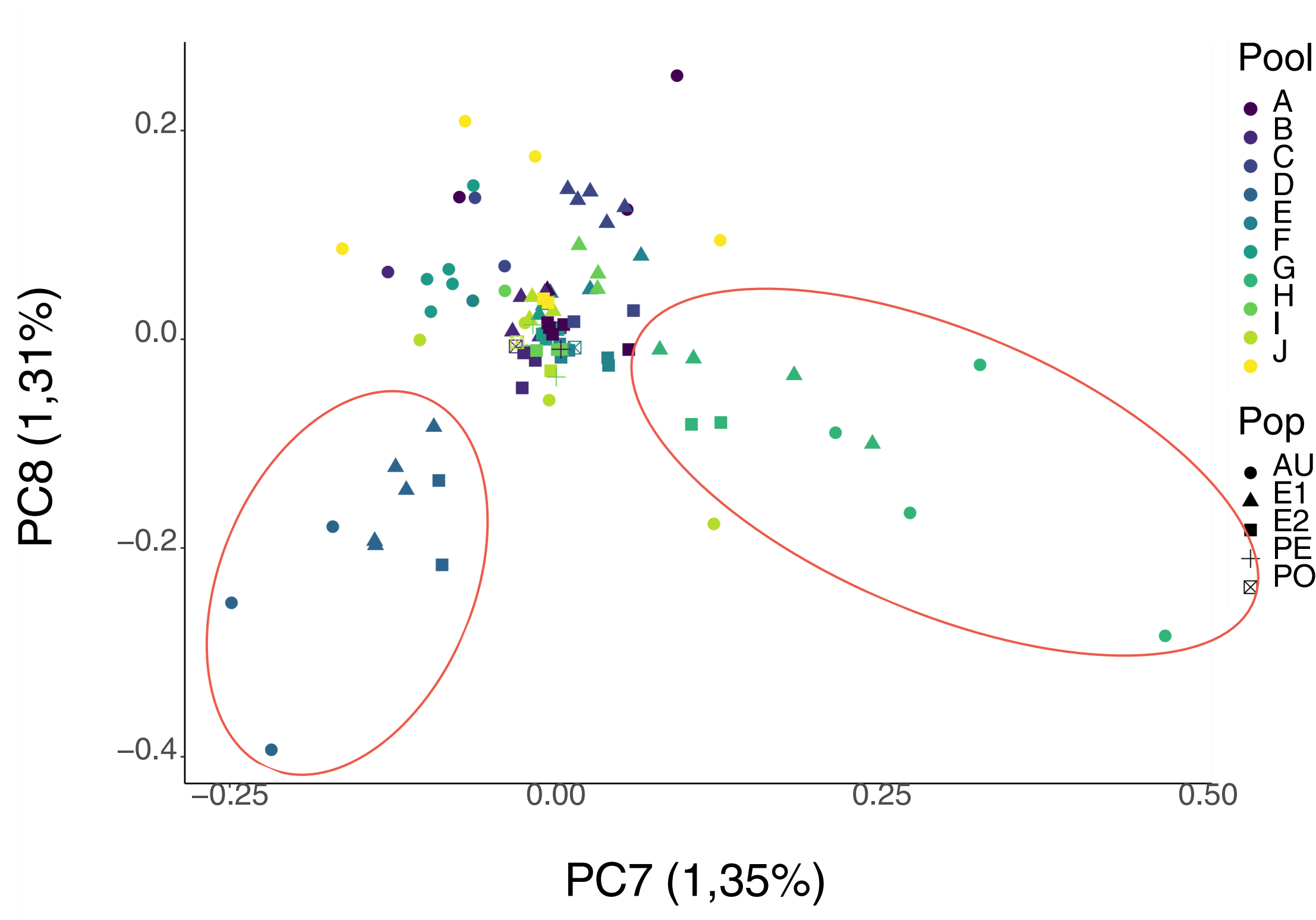
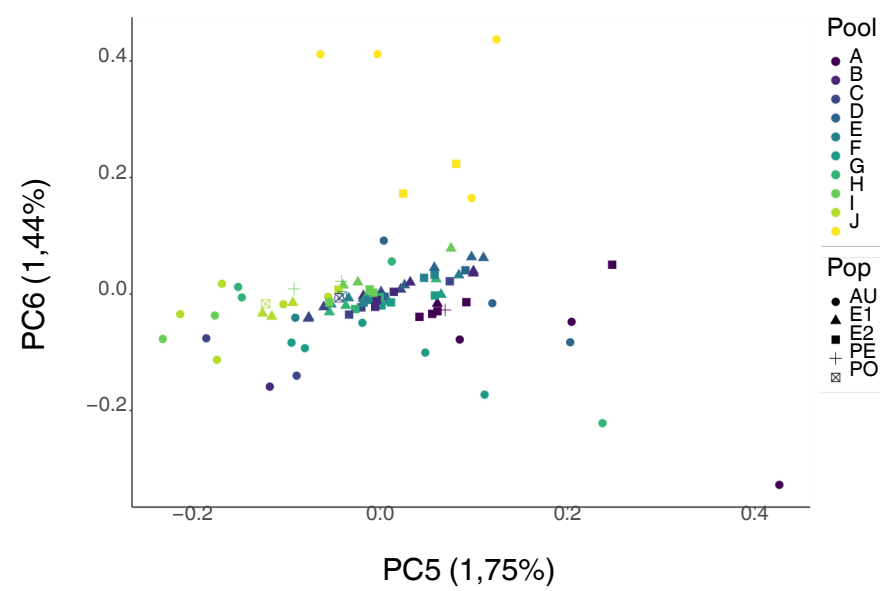
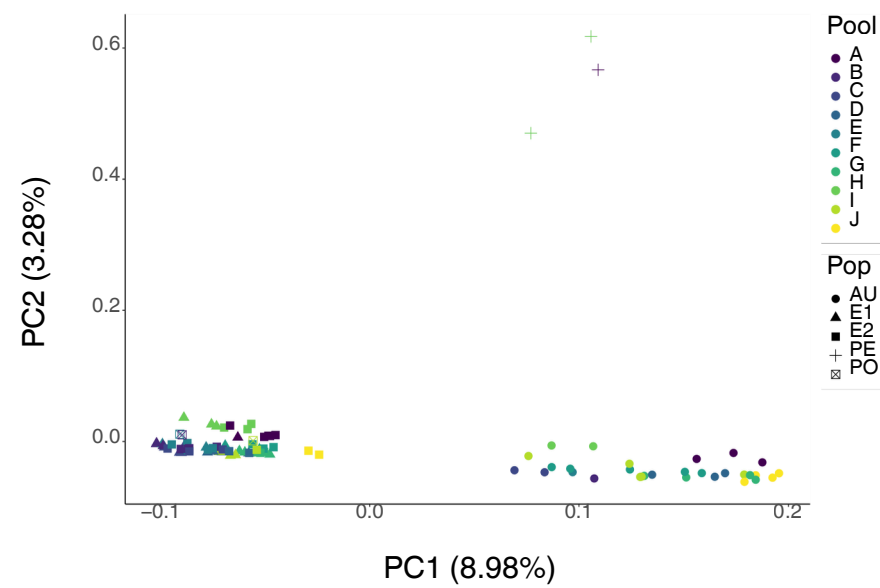
Sous-apparemment évident...

Contamination lors de la PCR poolé



Sous-apparemment évident...

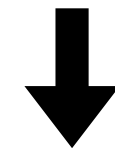
Contamination lors de la PCR poolé



Sous-apparemment évident...

Filtration jeu de données No Singletons

Avant filtration : 16 770 198 SNPs

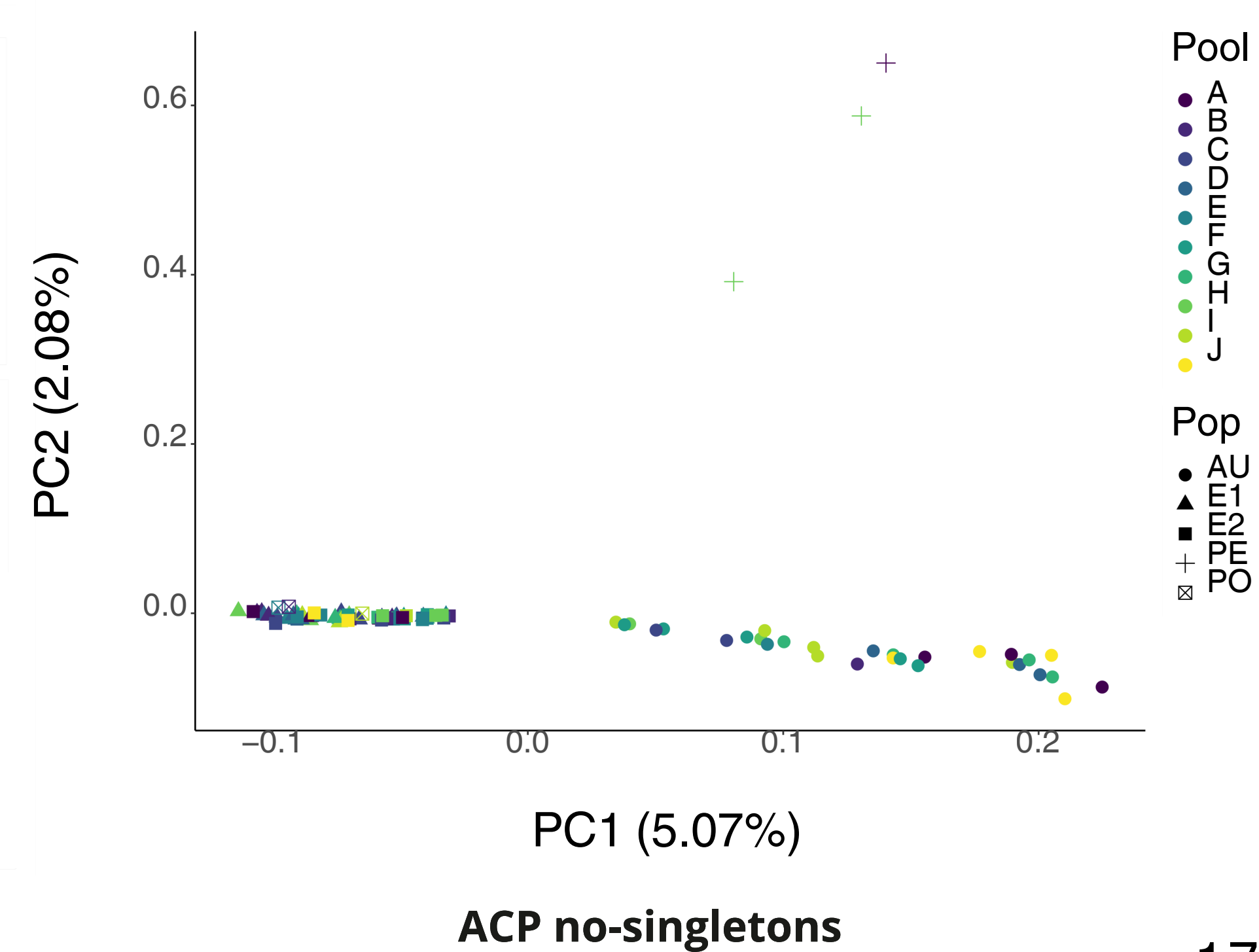
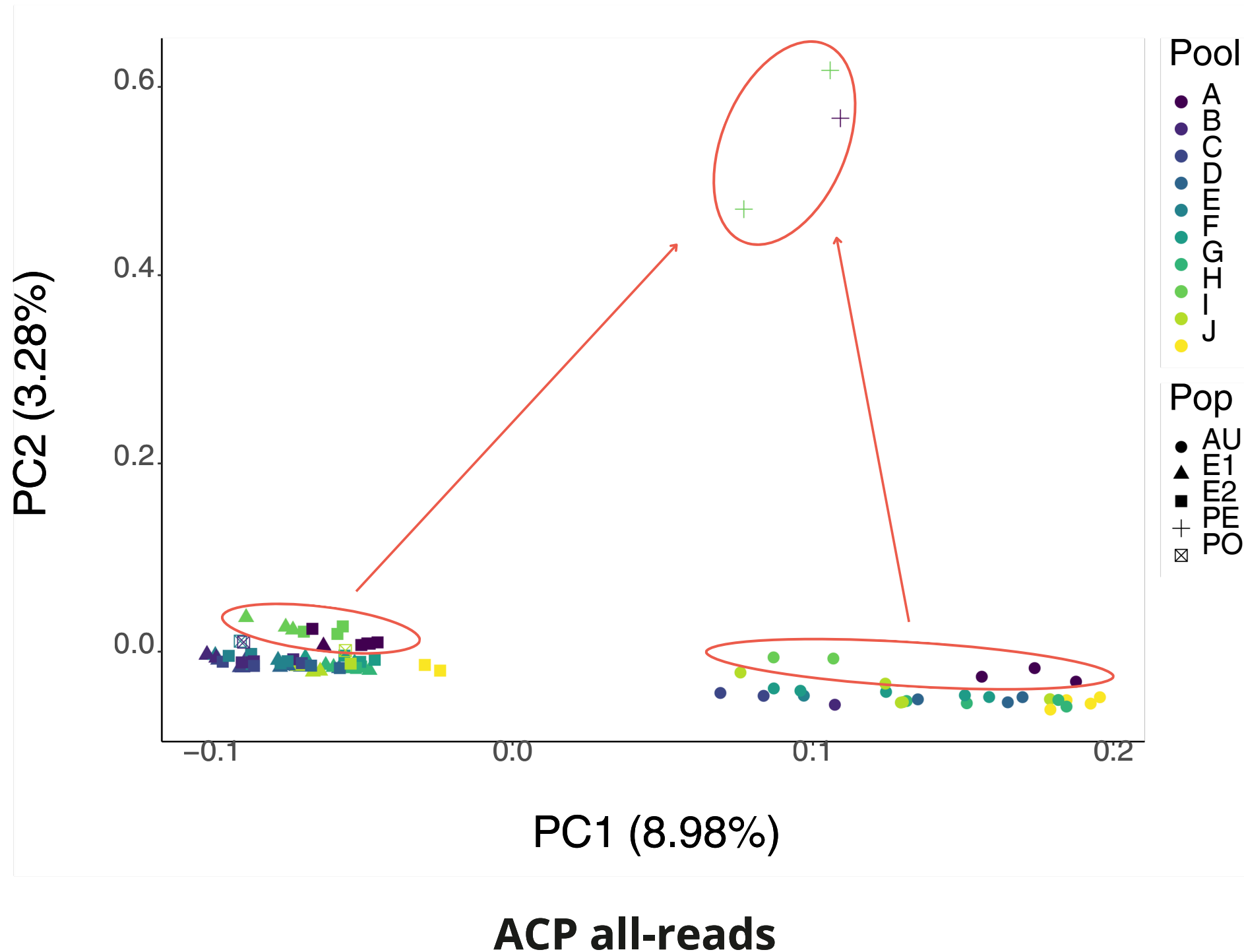


élimination des singletons (reads non liés)

Après filtration : 10 295 555 SNPs

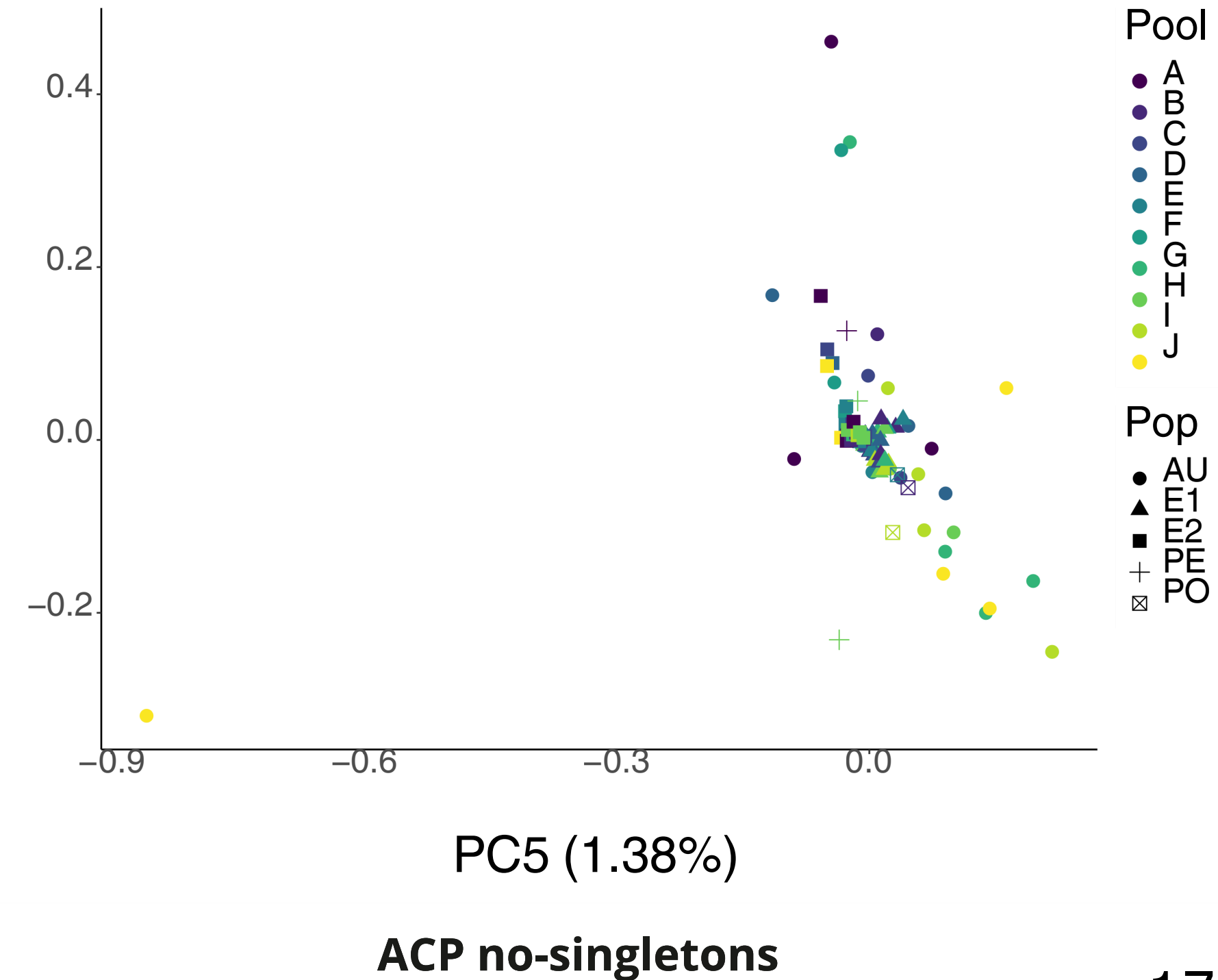
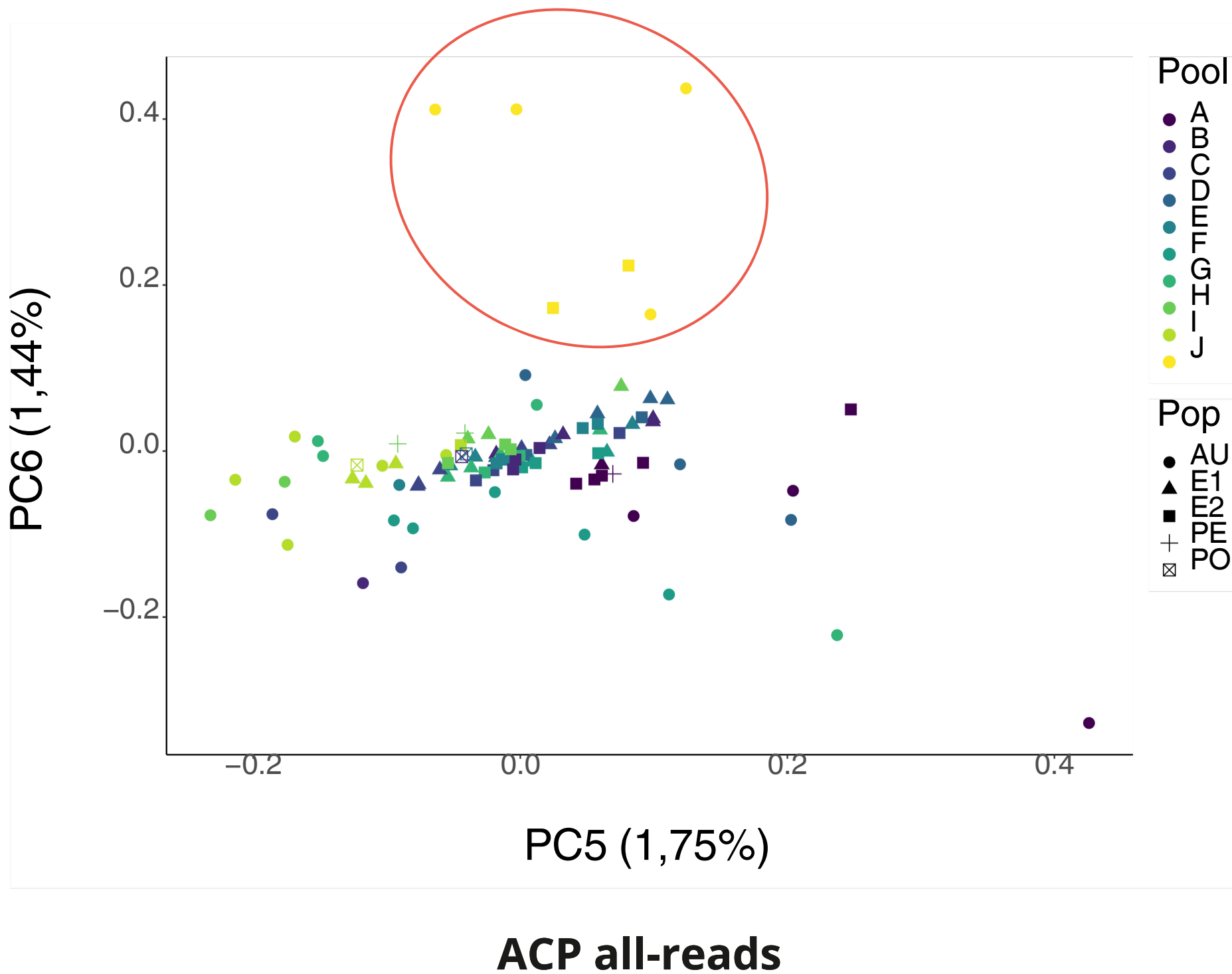
Sous-apparemment évident...

Filtration jeu de données No Singletons



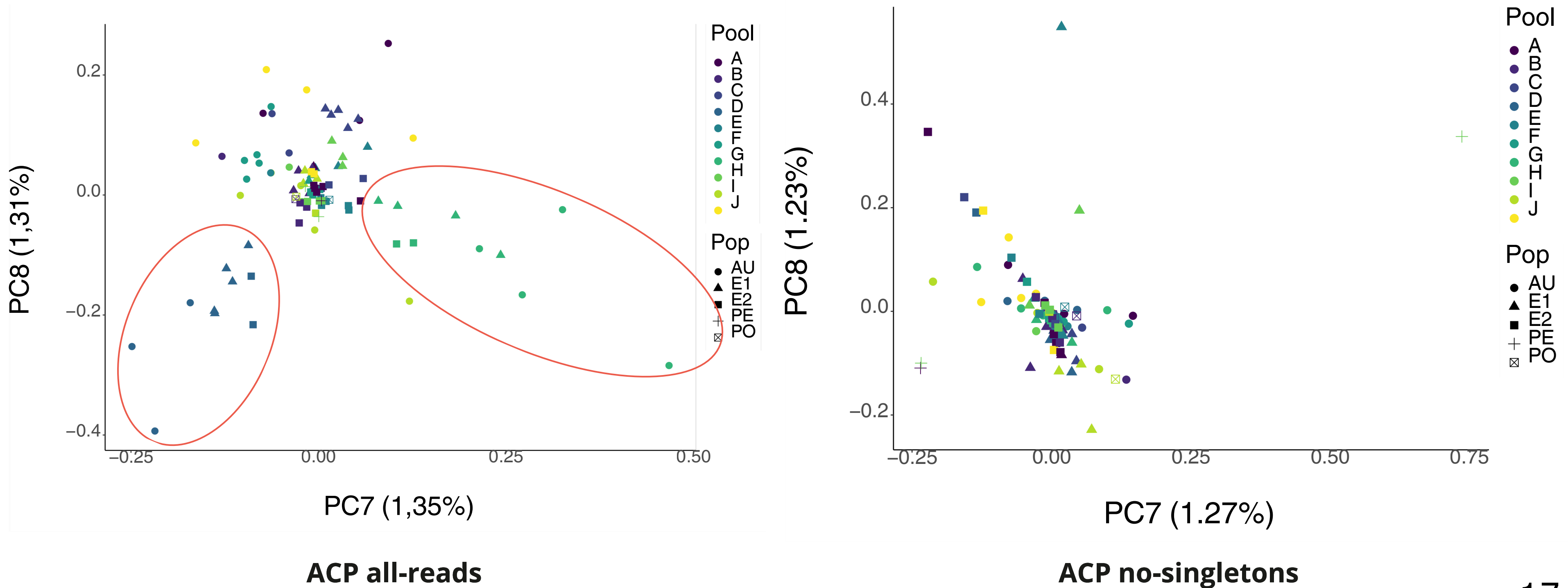
Sous-apparemment évident...

Filtration jeu de données No Singletons



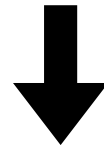
Sous-apparemment évident...

Filtration jeu de données No Singletons

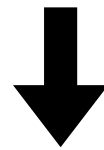


Sous-appareusement évident...

VCF no Singletons



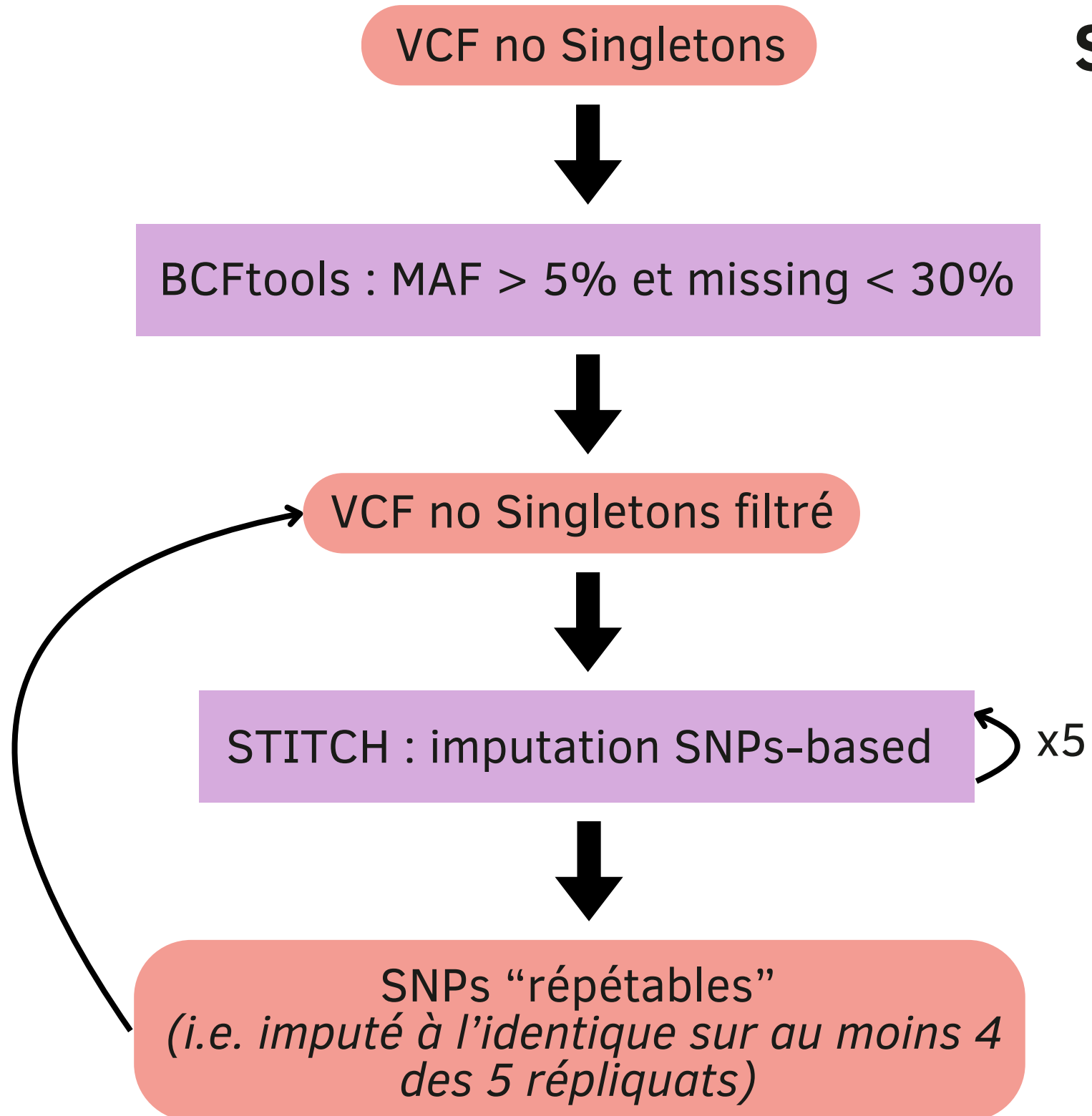
BCFtools : $MAF > 5\%$ et $missing < 30\%$



VCF no Singletons filtré

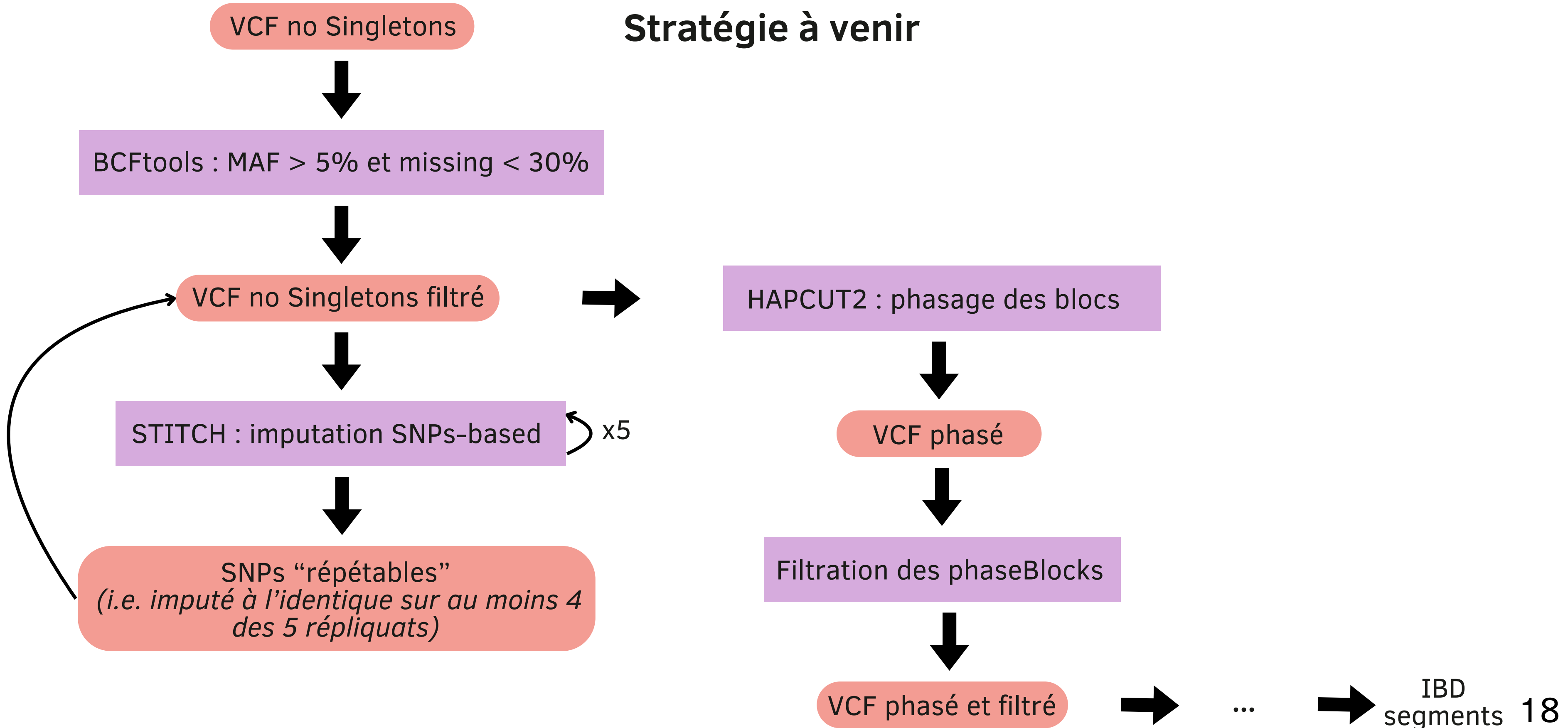
Stratégie à venir

Sous-apparemment évident...



Stratégie à venir

Sous-apparementement évident...



Conclusions & Perspectives

(1) Utilisation de l'haplotagging pour détecter les segments IBD

 augmenter la **profondeur de séquençage** > 10X

Conclusions & Perspectives

(1) Utilisation de l'haplotagging pour détecter les segments IBD

→ augmenter la **profondeur de séquençage** > 10X

(2) Flux de gènes assisté

→ peut être bénéfique pour la survie de *Centaurea corymbosa*



Conclusions & Perspectives

(1) Utilisation de l'haplotagging pour détecter les segments IBD

→ augmenter la **profondeur de séquençage** > 10X

(2) Flux de gènes assisté

→ peut être bénéfique pour la survie de *Centaurea corymbosa*

(3) Potentiels futurs en génomique de la conservation

- inférer la **démographie** dans le **passé très récent**
- reconstruire l'histoire **récente** de la **connectivité** dans les **métapopulations** fragmentées





Remerciements

Encadrants de stages :

Ophélie RONCE, Éric IMBERT & Pierre-Alexandre GAGNAIRE

Partenaires du projet DevOCGen :

Simon BOITARD & Raphaël LEBLOIS

Préparation des banques d'haplotagging :

Elodie FLAVEN, Frédérique CERQUEIRA & plateforme GenSeq

Financements :



Equipe BEM :

