Genome scans (TD)

Véronique Jorge

TD GENOME SCANS

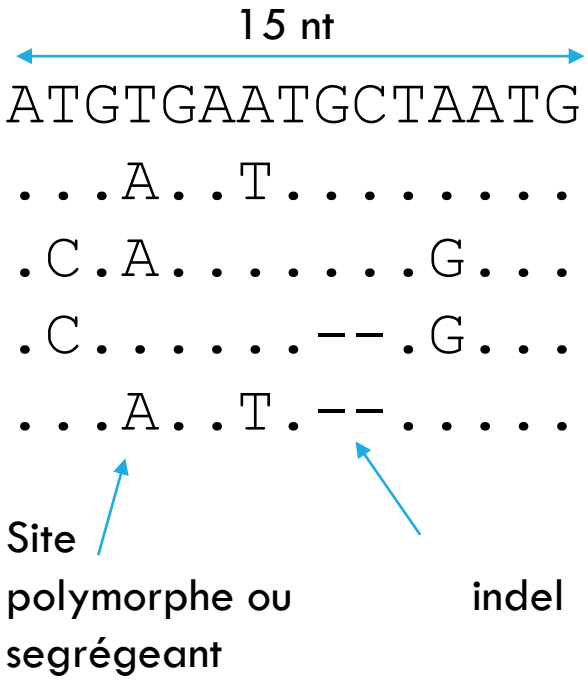
Véronique Jorge

UE génomique des populations

SOMMAIRE

1. **Rappels de génétique des populations : statistiques de diversité appliquées aux séquences**
2. Rappels effets de la sélection
3. Rappel outils / méthodes de détection de sélection
4. Formats des données
5. Logiciels

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

$$\theta_\pi = \sum_{i,j} \frac{\Pi_{ij}}{\frac{N \times (N-1)}{2}}$$

$$\theta_W = \frac{S}{\sum_{i=1}^{N-1} \frac{1}{i}} \text{ ou } \theta_s$$

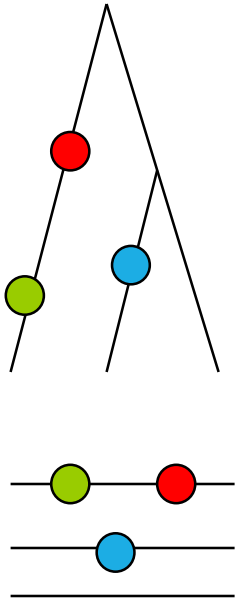
Dénominateur : nb total de paires de sequences

N: taille de l'échantillon

Π_{ij} nb de différences entre les haplotypes i et j

S : nb de sites polymorphes

Dénominateur : nombre harmonique



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

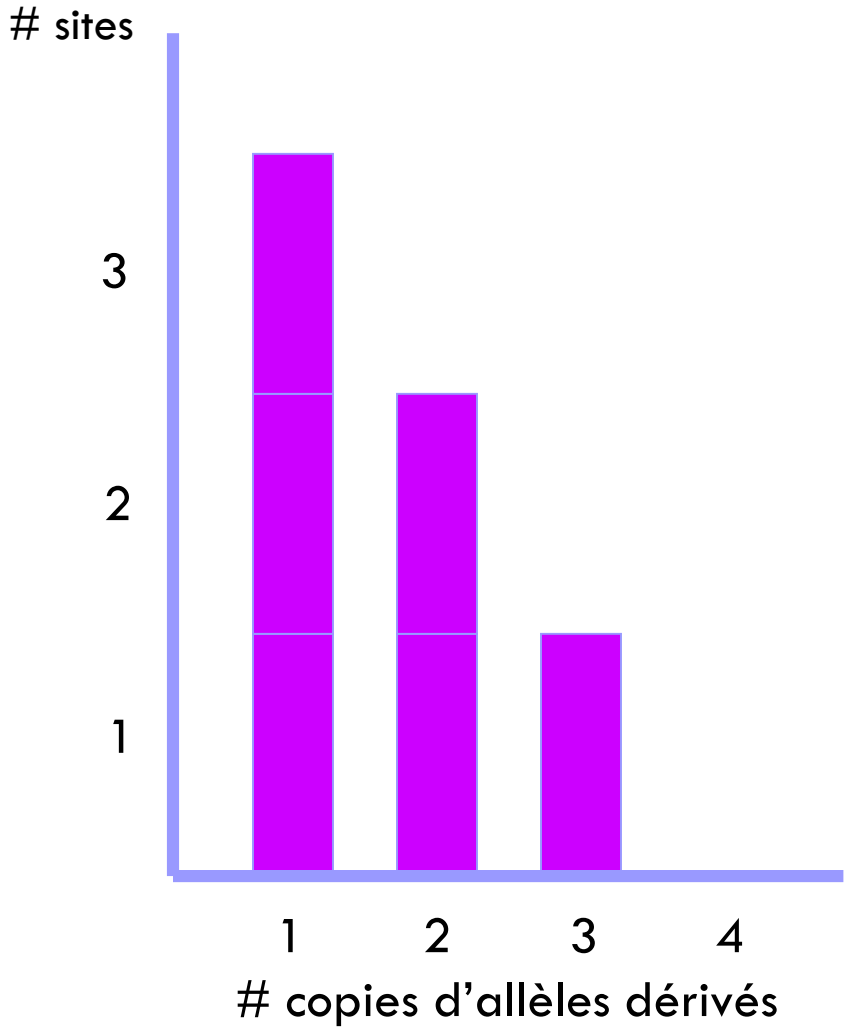
$\pi_{silent} = 0.038$

$\pi_{noncoding} = 0.023$

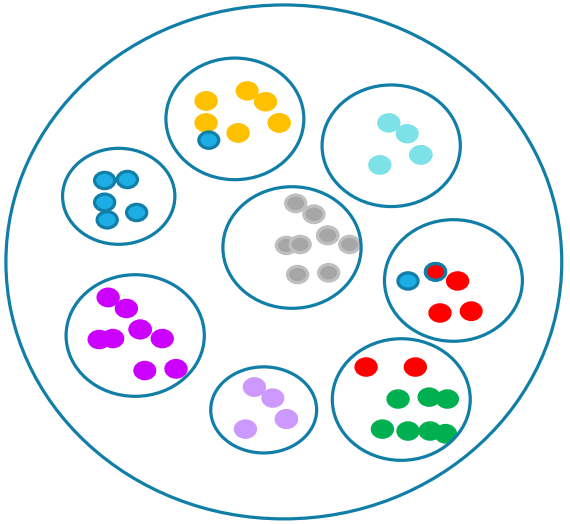
(D'après G. McVean 2001)

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



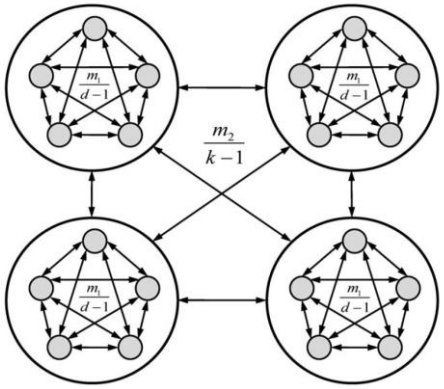
Les population naturelles sont subdivisées à cause de :

- 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.

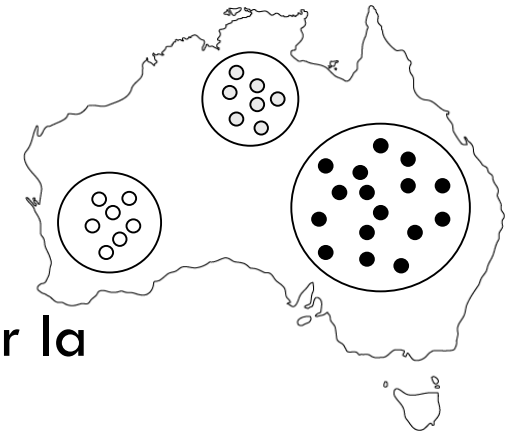
Les subdivisions peuvent être hiérarchisées.



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

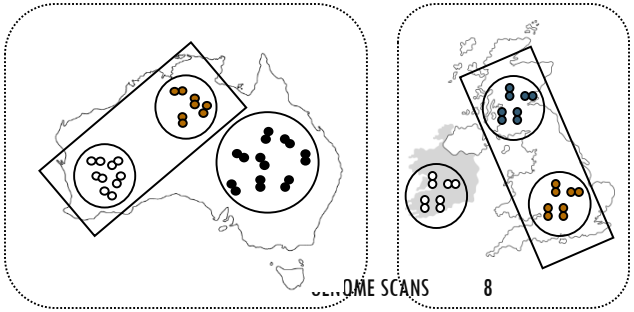
$$= \frac{H_T - \overline{H}_S}{H_T}$$ → (Nei's Gst)

Significativité testée par permutation

Les statistiques *F* sont hiérarchisée (indices de fixation)

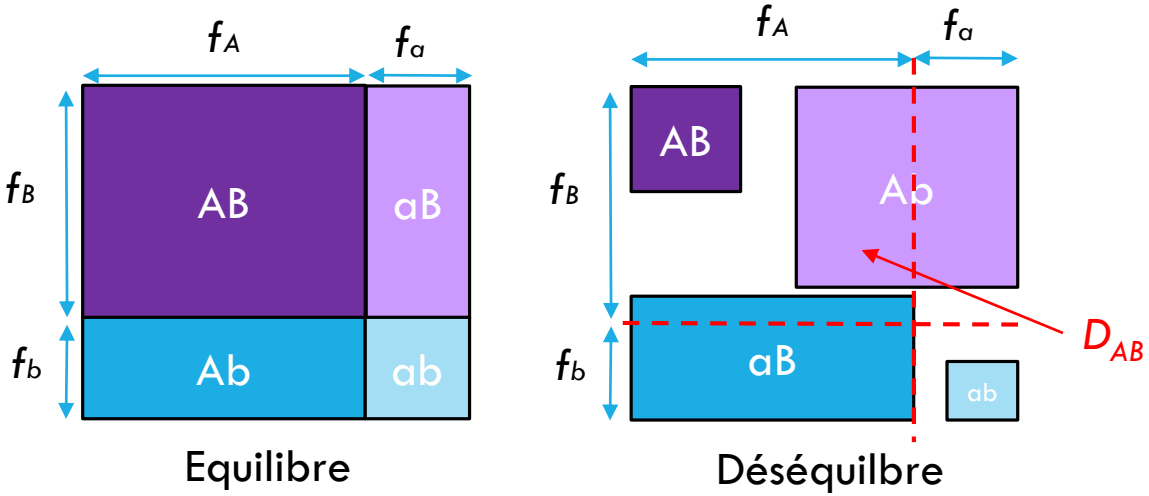
$$H_{Individual} < H_{Subpopulation} < H_{Population} < H_{Region} < H_{Total}$$

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

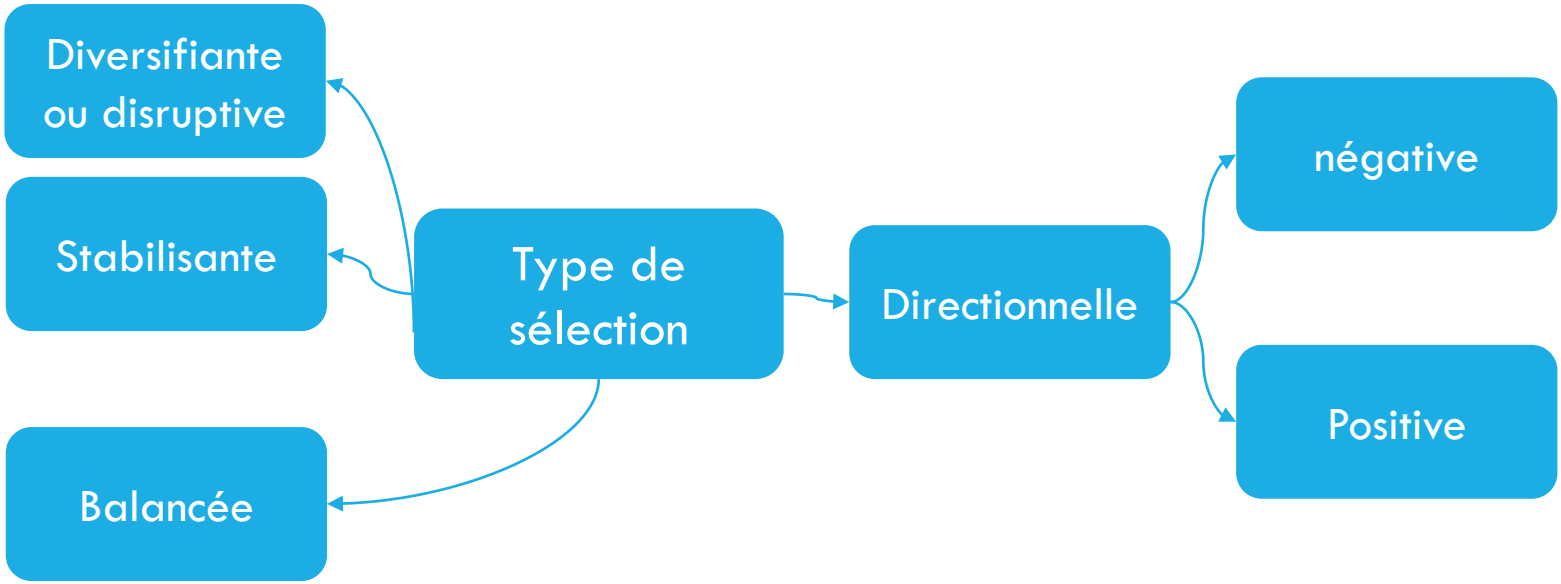
SOMMAIRE

1. Rappels de génétique des populations : statistiques de diversité appliquées aux séquences
2. **Rappels effets de la sélection**
3. Rappel outils / méthodes de détection de sélection
4. Formats des données
5. Logiciels

LES DIFFÉRENTS TYPES DE SÉLECTION

Définition générale

Propagation différentielle non aléatoire d'un allèle.

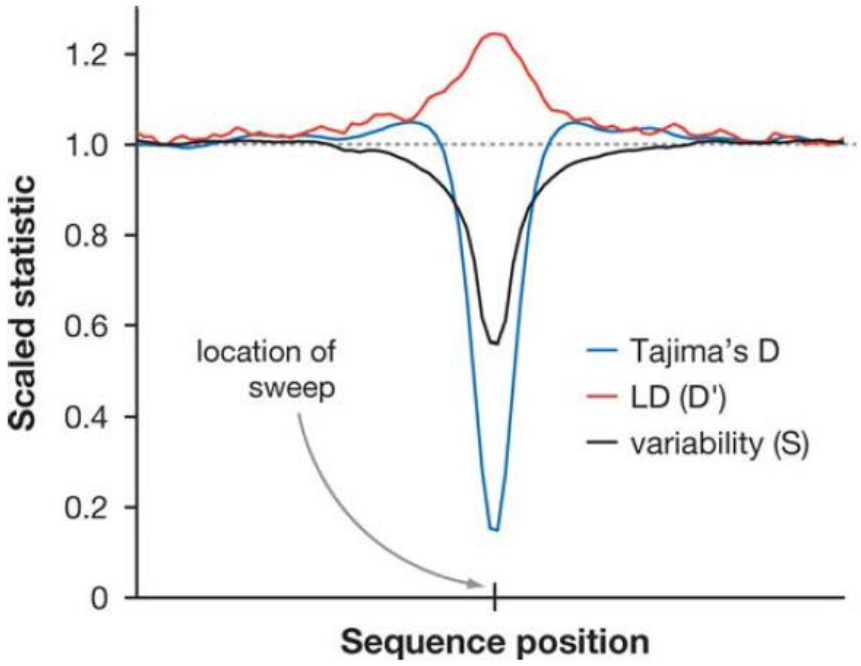


Organismes diploïdes

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

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

EFFET DE LA SÉLECTION SUR LES PARAMÈTRES DE DIVERSITÉ



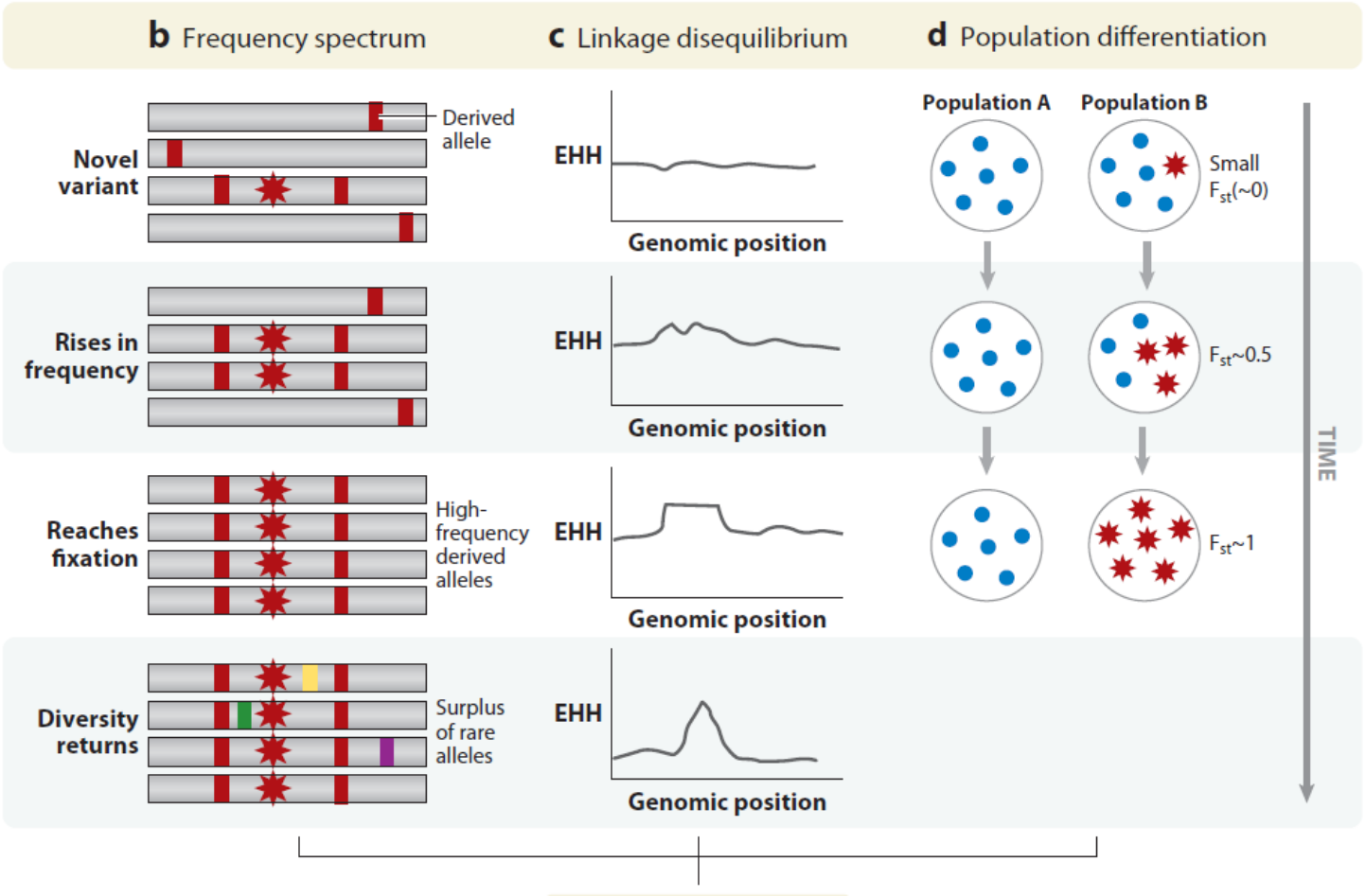
Selective sweep / balayage sélectif

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

SOMMAIRE

1. Rappels de génétique des populations : statistiques de diversité appliquées aux séquences
2. Rappels effets de la sélection
3. **Rappel outils / méthodes de détection de sélection**
4. Formats des données
5. Logiciels

TESTS BASÉS SUR LA MICROÉVOLUTION



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

TESTS BASÉS SUR LA MICROÉVOLUTION

Test du D de Tajima (1989)

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

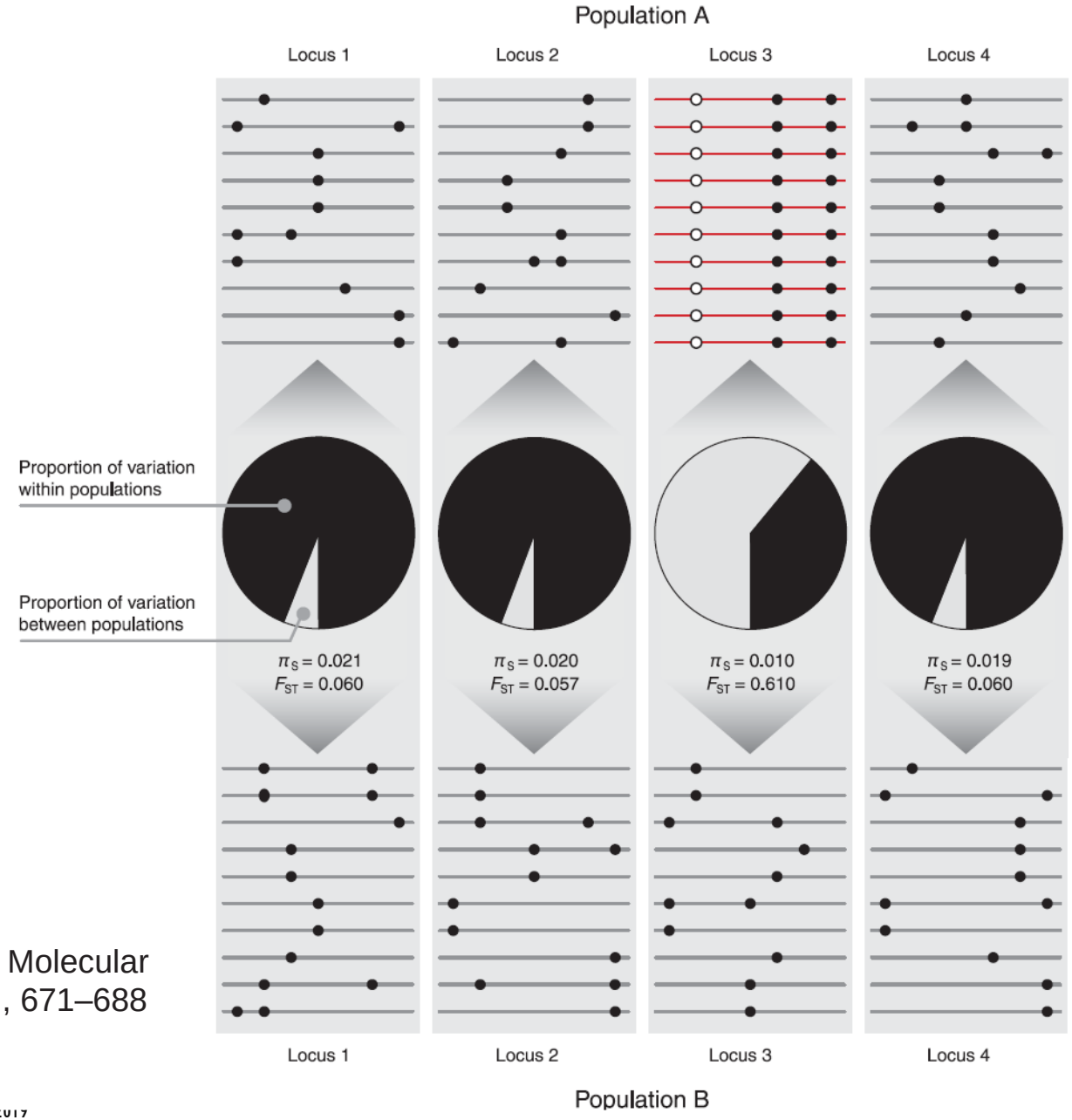
- ϑ_S, basé sur le nb de sites polymorphes **S**
- ϑ_π, basé sur le nb moyen de différences entre paires d'haplotypes (séquences) → hétérozygotie au niveau nucléotidique.

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

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.

OUTILS POUR LA DÉTECTION DE TRACES DE SÉLECTION



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

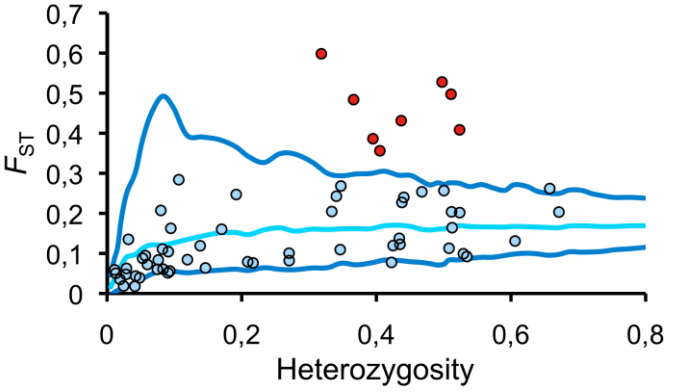
TESTS BASÉS SUR LA MICROÉVOLUTION

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



SOMMAIRE

1. Rappels de génétique des populations : statistiques de diversité appliquées aux séquences
2. Rappels effets de la sélection
3. Rappel outils / méthodes de détection de sélection
4. **Formats des données**
5. Logiciels

LE FORMAT FASTA

Séquençage SANGER

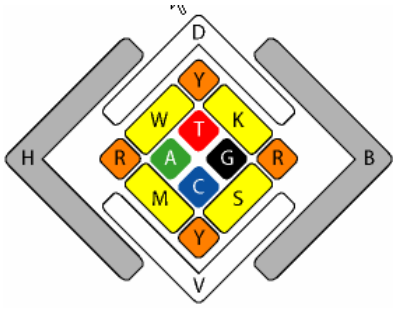
Nom de la séquence incluant le nom de l'individu

Signe supérieur

Séquence de l'individu

```
>cSYR_VV00037B
???GTGTTGTCGTCAACTCGTAAC---TGTCATGTCGTTGTGCGTATGC
>eVPE_VV00037B
??TGTGTTGTCGTCAACTCGGAAC---TGTCACGTCGTTGT-CGTATGC
>cPNI_VV00037B
ATTGTGTTGTCGTCAACTCGKAAC---TGTCAYGTCGTTGT-CGTATGC
>cMAK_VV00037B
ATTGTGTTGTCGTCAACTCGGAACGTATGTCATGTCGTTGT-CGTATGC
>cEST_VV00037B
?TTGTGTTGTCGTCAACTCGGAACGCATGTCACGTCGTTGTGCGTATGC
>sLAS_VV00037B
?TTGTGTTGTCGTCAACTCGTAACGTATGTCACGTCGTTGTGCGTATGC
```

Code IUPAC




```
ATTGTGTTGTCGTCAACTCG[T/G]AAC*[T/C]*TGTCA[T/C]GTCGTTGT*CGTATGC
```

LE FORMAT VCF

Le séquençage nouvelle génération (NGS) génère un tel volume de données que les formats classiques ne sont plus adaptés

```
##fileformat=VCFv4.2
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS     ID        REF    ALT     QUAL FILTER INFO                              FORMAT      NA00001        NA00002        NA00003
20     14370   rs6054257 G      A       29   PASS   NS=3;DP=14;AF=0.5;DB;H2           GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:.,.
20     17330   .         T      A       3    q10    NS=3;DP=11;AF=0.017               GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3   0/0:41:3
20     1110696 rs6040355 A      G,T     67   PASS   NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2   2/2:35:4
20     1230237 .         T      .       47   PASS   NS=3;DP=13;AA=T                   GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20     1234567 microsat1 GTC    G,GTCT  50   PASS   NS=3;DP=9;AA=G                    GT:GQ:DP    0/1:35:4       0/2:17:2       1/1:40:3
```

+ exemple VCF avec info de SNP (un peu différent)

```
##contig=<ID=10,length=100000000>
#CHROM  POS ID  REF ALT QUAL    FILTER  INFO    FORMAT  ind1    ind2    ind3    ind4    ind5    ind6    ind7    ind8
1   1894    .   G   T   100 PASS    AC=2;AF=1;AN=10 GT  0/0 1/1 1/1 1/1 0/0 0/0 0/0 1/1
1   2631    .   C   T   100 PASS    AC=2;AF=1;AN=10;EFF=SYNONYMOUS_CODING(LOW|SILENT|tcC/tcT|S156|lg000200|mRNA||lg000200.1|Exon_
1   2874    .   G   A   100 PASS    AC=2;AF=1;AN=10 GT  ./. 1/0 0/0 ./. 0/0 ./. ./. ./.
1   3196    .   C   T   100 PASS    AC=2;AF=1;AN=10 GT  0/0 0/1 ./. ./. 0/0 ./. ./. 0/0
1   9700    .   T   A   100 PASS    AC=2;AF=1;AN=10 GT  ./. 0/0 ./. ./. ./. 1/1 ./. ./.
1   9717    .   A   T   100 PASS    AC=2;AF=1;AN=10 GT  ./. 0/0 ./. ./. ./. 1/1 ./. ./.
1   9724    .   C   T   100 PASS    AC=2;AF=1;AN=10 GT  ./. 0/0 ./. ./. ./. 1/1 ./. ./.
1   9729    .   C   T   100 PASS    AC=2;AF=1;AN=10 GT  ./. 0/0 ./. ./. ./. 1/1 ./. ./.
1   12324   .   A   C   100 PASS    AC=2;AF=1;AN=10;EFF=SYNONYMOUS_CODING(LOW|SILENT|acT/acG|T343|lg000210|mRNA||lg000210.1|Exon_
1   22564   .   C   T   100 PASS    AC=2;AF=1;AN=10 GT  ./. 1/1 ./. ./. ./. ./. ./. ./.
1   26950   .   C   T   100 PASS    AC=2;AF=1;AN=10;EFF=INTRON(MODIFIER||||lg000220|mRNA||lg000220.1|)  GT  0/0 1/1 0/0 0/0 0/0 .
```

AUTRES FORMATS

Beaucoup d'autres formats …

Genepop
Plink
Staden
….

⇒ Recodage des allèles
⇒ Matrice de codage
⇒ Individus x marqueurs (le plus souvent)
⇒ Option : Info populations

Marqueurs

individus

```
A , 0101 0202 0201 0201 0102 0201 0101 0101 0202 0101 0000 0000 0000 0101 0201 0101 0101 0202 0101 0000 0101 0202 0101 0
1 , 0101 0202 0202 0201 0202 0101 0101 0101 0102 0101 0202 0202 0201 0101 0202 0101 0102 0202 0202 0101 0202 0202 0202 0
2 , 0101 0202 0201 0201 0102 0201 0101 0202 0102 0101 0202 0000 0201 0101 0202 0101 0101 0202 0102 0102 0201 0202 0202 0
3 , 0101 0202 0201 0201 0000 0101 0201 0101 0102 0101 0201 0202 0202 0101 0201 0102 0101 0202 0101 0202 0101 0202 0101 0
4 , 0101 0202 0202 0201 0202 0101 0101 0102 0202 0101 0101 0202 0202 0101 0101 0101 0101 0202 0101 0202 0101 0202 0101 0
5 , 0101 0202 0201 0201 0102 0101 0101 0102 0202 0101 0201 0201 0202 0101 0202 0101 0101 0202 0101 0202 0101 0202 0000 0
6 , 0102 0202 0101 0101 0000 0201 0101 0102 0202 0101 0202 0201 0202 0101 0202 0101 0202 0202 0102 0102 0101 0202 0000 0
7 , 0101 0202 0201 0201 0102 0000 0101 0202 0202 0101 0201 0201 0202 0101 0202 0101 0102 0201 0102 0101 0201 0201 0101 0
8 , 0101 0202 0202 0201 0202 0101 0101 0101 0202 0101 0202 0201 0202 0101 0202 0101 0102 0202 0101 0202 0101 0202 0201 0
9 , 0101 0202 0202 0201 0202 0101 0101 0102 0202 0101 0202 0201 0201 0101 0202 0101 0101 0201 0102 0102 0201 0201 0101 0
10 , 0101 0202 0202 0201 0202 0101 0101 0101 0102 0101 0201 0201 0202 0101 0202 0101 0000 0202 0102 0102 0201 0202 0202
11 , 0101 0202 0202 0201 0202 0101 0101 0101 0102 0101 0201 0201 0202 0101 0202 0101 0102 0202 0202 0101 0202 0201 0000
12 , 0101 0202 0201 0000 0102 0101 0201 0101 0202 0101 0202 0201 0201 0101 0000 0000 0101 0202 0102 0102 0101 0202 0201
13 , 0101 0202 0201 0201 0102 0101 0101 0101 0102 0101 0201 0201 0202 0101 0202 0101 0102 0202 0202 0101 0202 0202 0000
14 , 0101 0202 0202 0201 0202 0101 0101 0102 0202 0101 0202 0101 0202 0101 0202 0101 0102 0202 0202 0101 0202 0202 0000
15 , 0101 0202 0202 0201 0102 0101 0101 0102 0202 0101 0000 0000 0202 0101 0202 0101 0101 0101 0102 0102 0201 0202 0201
16 , 0101 0202 0201 0201 0102 0101 0101 0102 0000 0101 0202 0101 0000 0101 0202 0101 0202 0202 0202 0101 0201 0202 0000
17 , 0101 0202 0101 0101 0101 0201 0101 0202 0202 0201 0201 0202 0202 0101 0201 0101 0202 0202 0102 0102 0101 0202 0101
18 , 0101 0202 0202 0201 0202 0101 0101 0102 0202 0101 0202 0202 0202 0101 0201 0101 0101 0202 0202 0101 0201 0202 0101
19 , 0101 0202 0202 0201 0202 0101 0101 0102 0202 0101 0202 0101 0202 0101 0202 0101 0102 0201 0102 0101 0101 0202 0101
20 , 0101 0202 0201 0201 0102 0101 0101 0101 0202 0101 0202 0202 0202 0101 0201 0101 0101 0202 0202 0101 0101 0202 0000
21 , 0101 0202 0101 0101 0101 0201 0201 0202 0102 0101 0202 0201 0202 0101 0202 0101 0102 0201 0102 0102 0201 0202 0101
22 , 0101 0202 0202 0201 0202 0101 0101 0101 0102 0101 0201 0201 0201 0101 0202 0101 0101 0202 0101 0202 0101 0202 0101
23 , 0101 0202 0201 0201 0102 0201 0101 0102 0202 0101 0000 0000 0201 0101 0202 0101 0101 0202 0101 0202 0101 0202 0201
24 , 0101 0202 0202 0201 0202 0101 0101 0101 0102 0101 0201 0202 0202 0101 0202 0101 0102 0202 0102 0102 0000 0202 0202
25 , 0101 0202 0101 0101 0000 0101 0101 0101 0202 0101 0202 0201 0202 0101 0202 0101 0101 0202 0102 0102 0201 0202 0101
26 , 0102 0201 0201 0201 0102 0201 0101 0202 0202 0101 0202 0101 0202 0101 0202 0101 0101 0201 0202 0101 0101 0202 0101
27 , 0101 0202 0202 0201 0202 0101 0101 0202 0102 0101 0201 0201 0202 0101 0202 0101 0102 0202 0101 0102 0101 0202 0101
28 , 0101 0202 0000 0201 0202 0101 0101 0202 0202 0101 0202 0101 0202 0101 0202 0101 0101 0101 0101 0101 0101 0202 0202
29 , 0101 0202 0201 0201 0102 0101 0101 0202 0102 0101 0202 0101 0202 0101 0202 0101 0101 0201 0202 0101 0201 0202 0101
30 , 0101 0202 0201 0201 0102 0101 0101 0101 0102 0101 0101 0202 0202 0102 0201 0101 0102 0202 0202 0101 0201 0202 0201
31 , 0101 0202 0202 0201 0202 0101 0101 0102 0202 0101 0202 0101 0101 0101 0202 0101 0101 0202 0102 0102 0101 0202 0101
32 , 0101 0202 0000 0201 0102 0101 0201 0101 0102 0101 0201 0201 0202 0101 0202 0101 0101 0202 0102 0102 0201 0202 0101
33 , 0000 0202 0202 0201 0202 0101 0101 0102 0202 0101 0202 0202 0201 0101 0202 0101 0102 0202 0102 0102 0201 0202 0101
POP
```

SOMMAIRE

1. Rappels de génétique des populations : statistiques de diversité appliquées aux séquences
2. Rappels effets de la sélection
3. Rappel outils / méthodes de détection de sélection
4. Formats des données
5. **Logiciels**
    1. Pipeline Sniplay
    2. Detsel

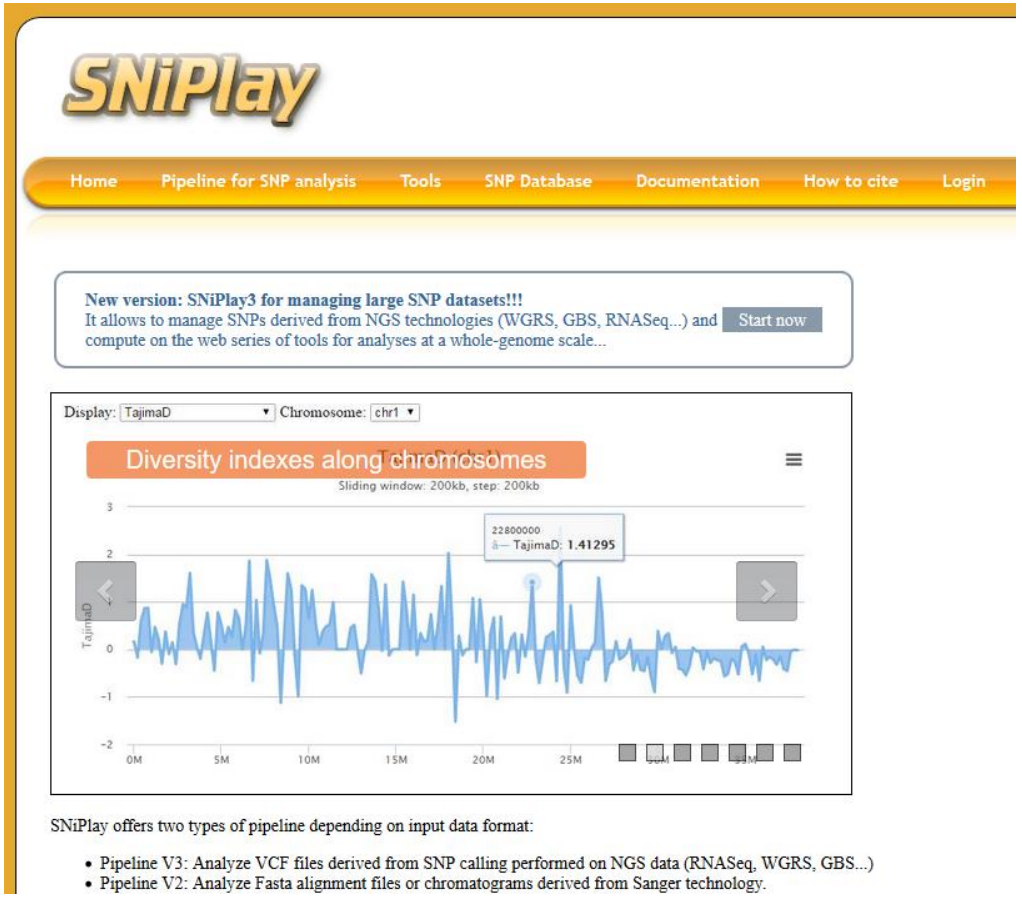
SNIPLAY

Plateforme intégrant une suite de logiciels pour les analyses de diversité basées sur les SNP

http://sniplay.southgreen.fr/cgi-bin/home.cgi

Dédié aux plantes

« Pipelines »
= enchaînement de différents programmes / logiciels



SNIPLAY V2

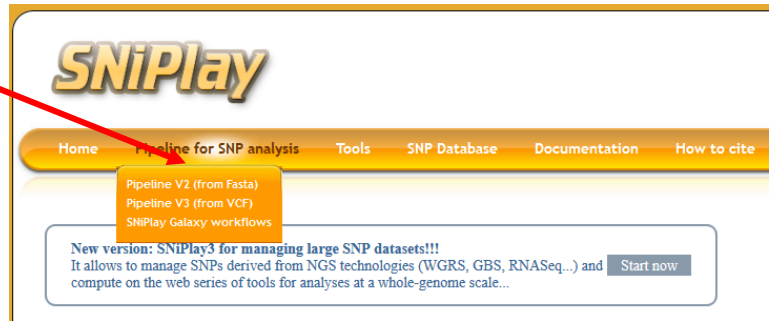
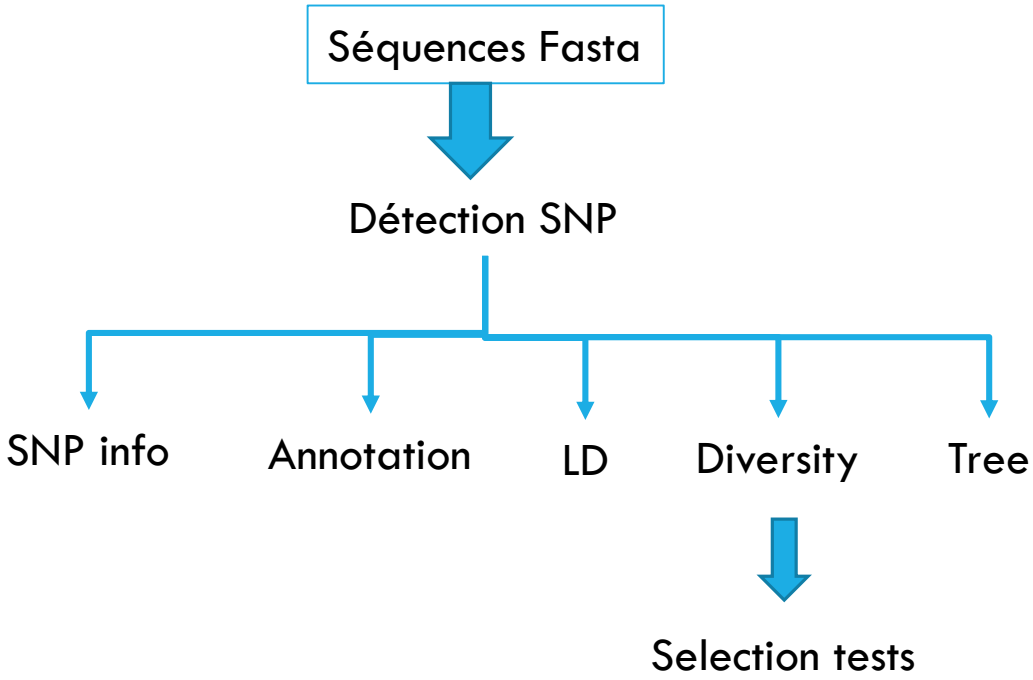


Schéma général du pipeline
(étapes abordées lors du TD)



SNIPLAY V2

Données

Veget Hist Archaeobot (2017) 26:345–356
DOI 10.1007/s00334-016-0597-4

ORIGINAL ARTICLE

Potential of combining morphometry and ancient DNA information to investigate grapevine domestication

Roberto Bacilieri[1] · Laurent Bouby[2] · Isabel Figueiral[2,3] · Caroline Schaal[4] · Jean-Frédéric Terral[2] · Catherine Breton[2] · Sandrine Picq[5] · Audrey Weber[1] · Angela Schlumbaum[6]

D'après https://archeorient.hypotheses.org/7096

77 accessions :

- 29 Vitis vinifera domestiquées
- 29 V. sylvestris
- 16 Vitis spp.
- 3 accessions => Pépins retrouvés sur 1 site archéo (1-2ème siècle, Romains)

Séquences : 6 fragments de gènes

VSVV007.fas
VSVV010.fas
VV00743A.fas
VV02879A.fas
VV04275A.fas
VVC6002B.fas

Gènes liés à la domestication

Fichier supplémentaire

Info_pop_Bacilieri.txt

SNIPLAY V2

Questions

SNIPLAY V3

Chargement des données à partir de la base de données

SNIPLAY V3

Nicolas *et al. BMC Plant Biology* (2016) 16:74
DOI 10.1186/s12870-016-0754-z

BMC Plant Biology

RESEARCH ARTICLE Open Access

Genetic diversity, linkage disequilibrium and power of a large grapevine (*Vitis vinifera* L) diversity panel newly designed for association studies

398 accessions de la collection Vassal (Centre de Ressources Biologiques INRA Montpellier)

https://www6.montpellier.inra.fr/vassal

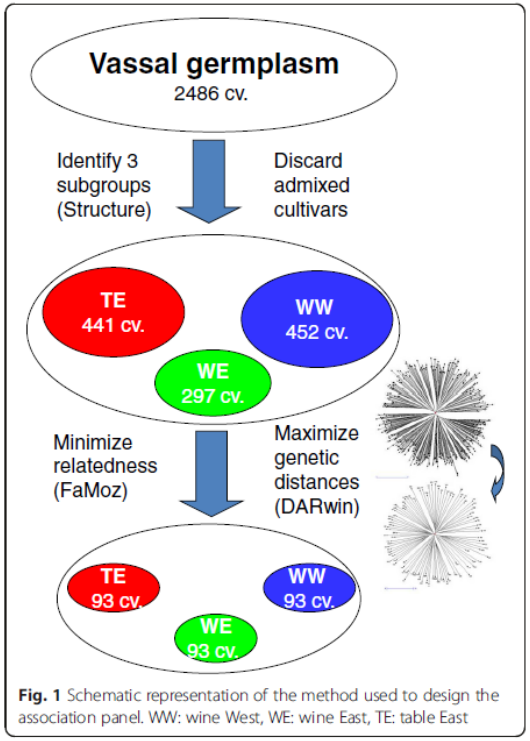
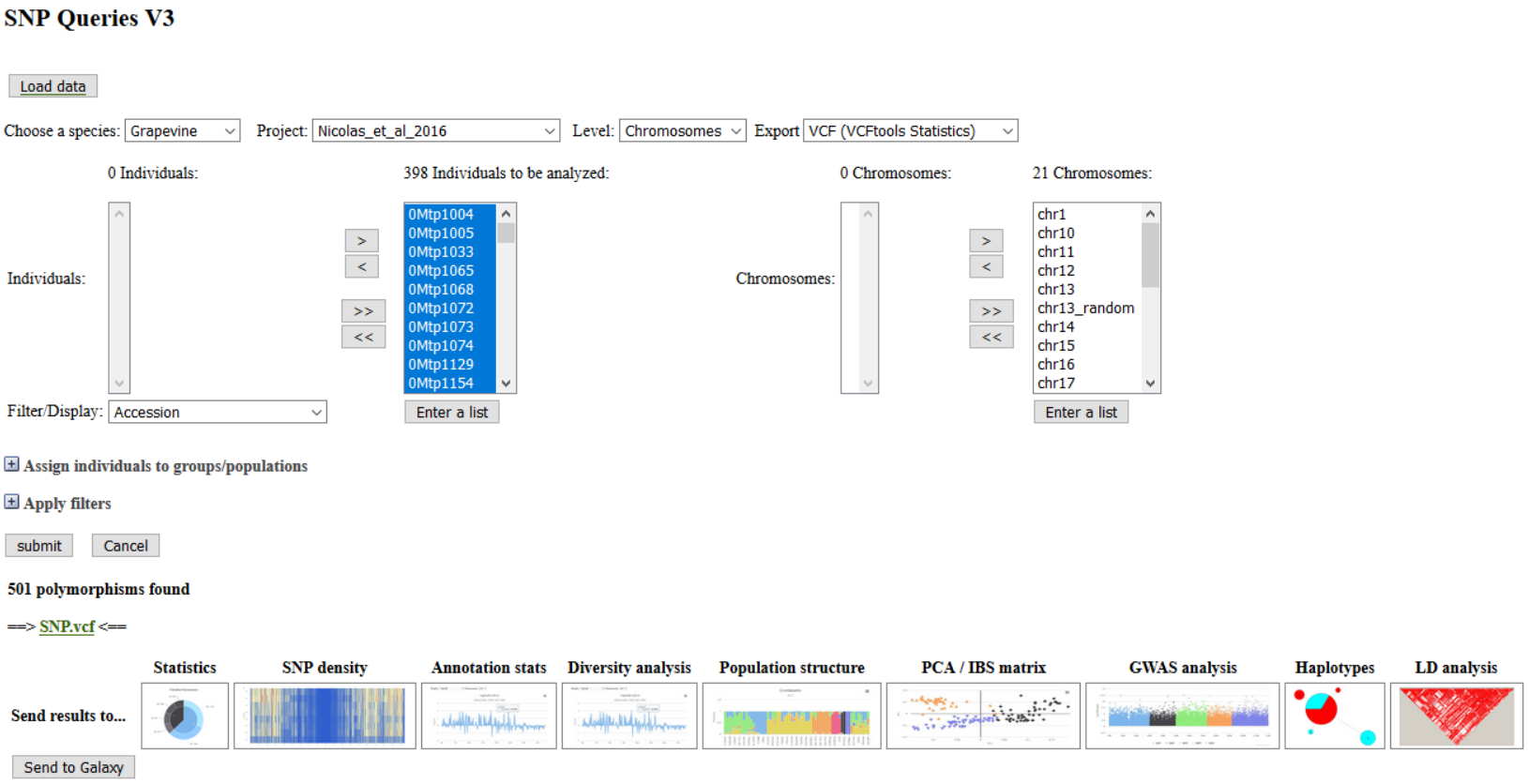


Fig. 1 Schematic representation of the method used to design the association panel. WW: wine West, WE: wine East, TE: table East

SNIPLAY V3



Regarder tableau plus bas dans la page avant de lancer les autres étapes

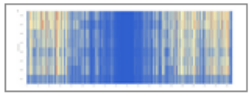
SNIPLAY V3

Questions

Statistics

Que peut-on dire à propos de la distribution des MAF (Minor Allele Frequency) ?
Et la distribution du % d'hétérozygotie ?
Qu'est-ce une transversion/transition ?

SNP density



(rq : sliding window)
Pic sur certains chromosomes ?

Diversity analysis

Interpréter TajimaD ? (attention à la densité de SNP !)

Population structure

(sNMF) quel est le nombre de population le plus probable ?

DETSEL

Vitalis R, (2014) DetSel: An R-Package to Detect Marker Loci Responding to Selection

10.1007/978-1-61779-870-2_16

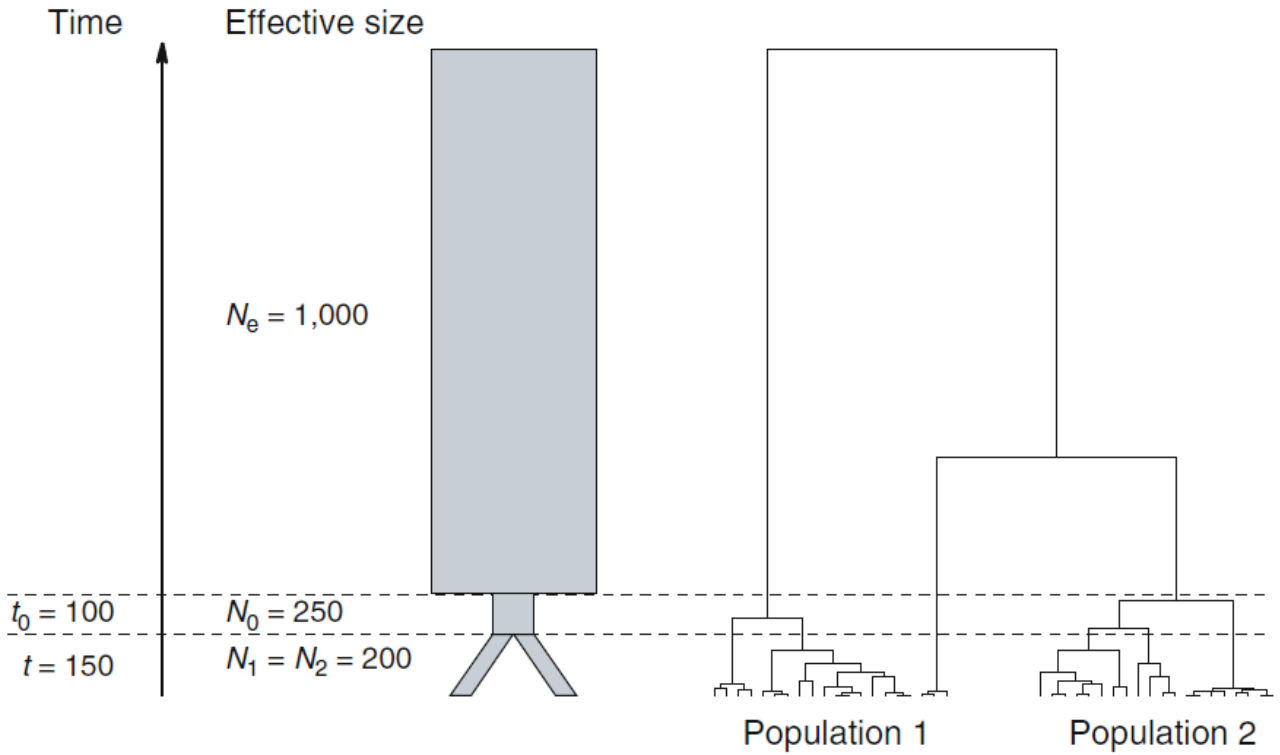


Fig. 1. DETSEL underlying demographic model and illustrated example of a simulated genealogy. The parameter values of this example are provided along the time scale on the left-hand side of the graph, for the different periods of the demographic model. The demographic model is schematized in *gray*, in the middle of the graph. On the right-hand side of the graph, the genealogy of a sample of 2×20 genes is shown.

DETSEL

Calcul par simulation de la distribution conjointe de F_1 et F_2
(pour les populations 1 et 2 respectivement) :

$$F_{\mathrm{i}} = \frac{Q_{\mathrm{w,i}} - Q_{\mathrm{b}}}{1 - Q_{\mathrm{b}}}, \quad \text{for } i \in \{1,2\}$$

$Q_{\mathrm{w,i}}$ Proba d'identité par état 2 SNP identique tirés dans la même pop

Q_{b} Proba d'identité par état 2 SNP tirés dans 2 pop différentes

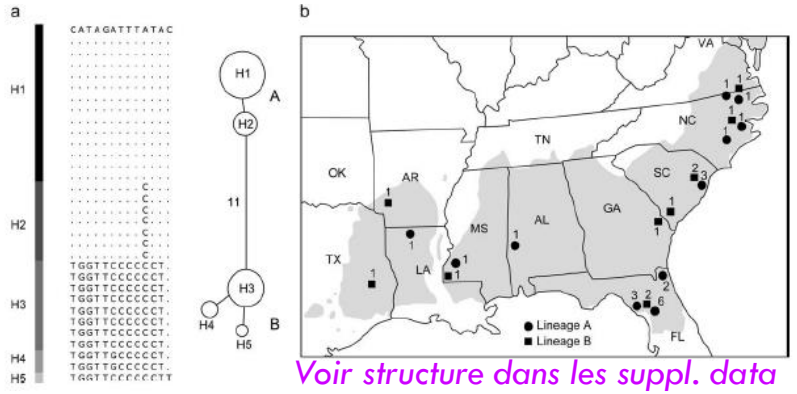
(Fst calculable à partir de ces composantes « Fi »)

Fi dépendent du temps écoulé depuis la divergence,
le temps = paramètre de la simulation :

$$F_{\mathrm{i}} \approx 1 - \exp\left(-T_{\mathrm{i}}\right)$$

Vitalis (2014)

DETSEL



Voir structure dans les suppl. data

Jeu de données d'entrée :

55 SNPs issus du séquençage de gènes candidats
(formation du bois et la résistance à la sécheresse)
Génotypés dans 3 populations de *Pinus taeda*
(33 north-east, 30 central-east, 25 west of Mississippi Valley).

Une partie des données de : González-Martínez, S.C., Ersoz, E., Brown, G.R., Wheeler, N.C., and Neale, D.B. (2006). DNA Sequence Variation and Selection of Tag Single-Nucleotide Polymorphisms at Candidate Genes for Drought-Stress Response in Pinus taeda L. Genetics *172*, 1915–1926.

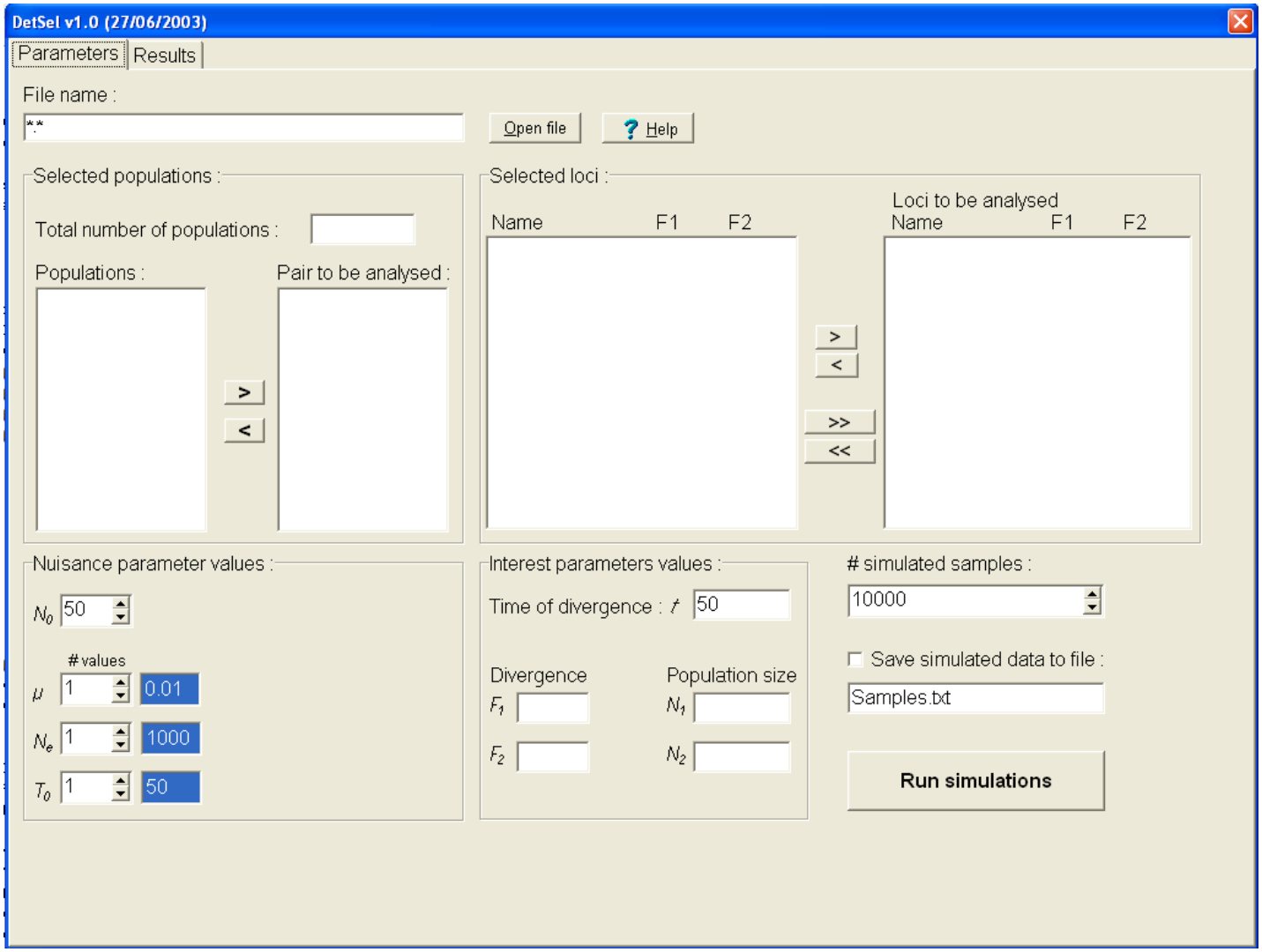
Format : GenePop (Raymond and Rousset 1995) => **in-taeda_55SNPs.txt**

```
Title          Description of the data
m1             Locus names
m2
…
POP
IND1 , 0101 0202 0201 …      Multilocus genotype individual1, pop1
IND2 , 0102 0101 0201 …      Multilocus genotype individual2, pop1
…
POP
IND1 , 0101 0202 0201 …      Multilocus genotype individual1, pop2
IND2 , 0102 0101 0201 …      Multilocus genotype individual2, pop2
```

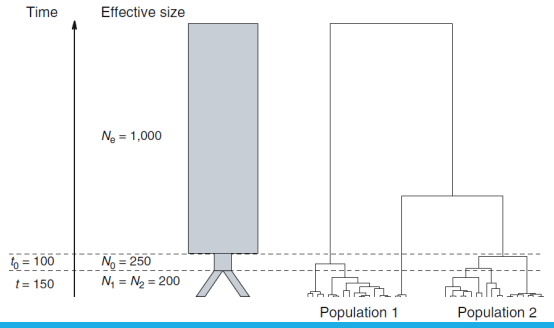

TD d'après S. Gonzalez-Martinez

DETSEL

detsel.exe



DETSEL



Paramètres suggérés (pour commencer):

Paramètre	Valeur	Commentaire
Nb de simulations	10000	Valeur / défault, mais temps de calcul disponible faible en TD !
m	0.000001	Taux de mutation *
Ne	1500	Taille efficace de la population (peut-être x 2 pour le pin taeda)
T_0	400	Temps (en générations) depuis le dernier goulot d'étranglement (**10 000 ans** et 25 années / générations)
N_0	150	Taille de la population au goulot
t	100	Temps de divergence (en générations)

* Calcul rapide, sachant que : $\theta = 4\,Ne\,\mu$

Et les valeurs de théta (ϑ_w ou ϑ_π) données dans Gonzalez-Martinez et al (2006) (TABLE 2)

Calculer le taux de mutation μ

DETSEL

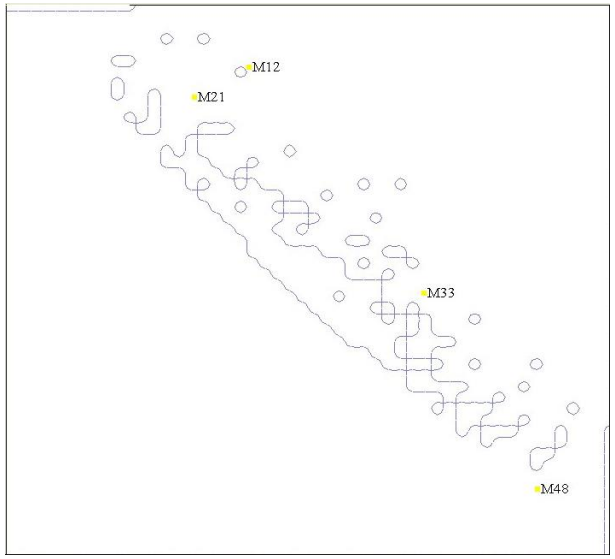
Questions

- Quels sont les SNP « outliers » ?

- Augmenter le temps de divergence, que se passe-t-il ?
(Comment les flux de gènes peuvent perturber la simulation ?)

- Que se passe-t-il quand on réduit encore le goulot ?

Pop 1 vs 2



Pop 2 vs 3