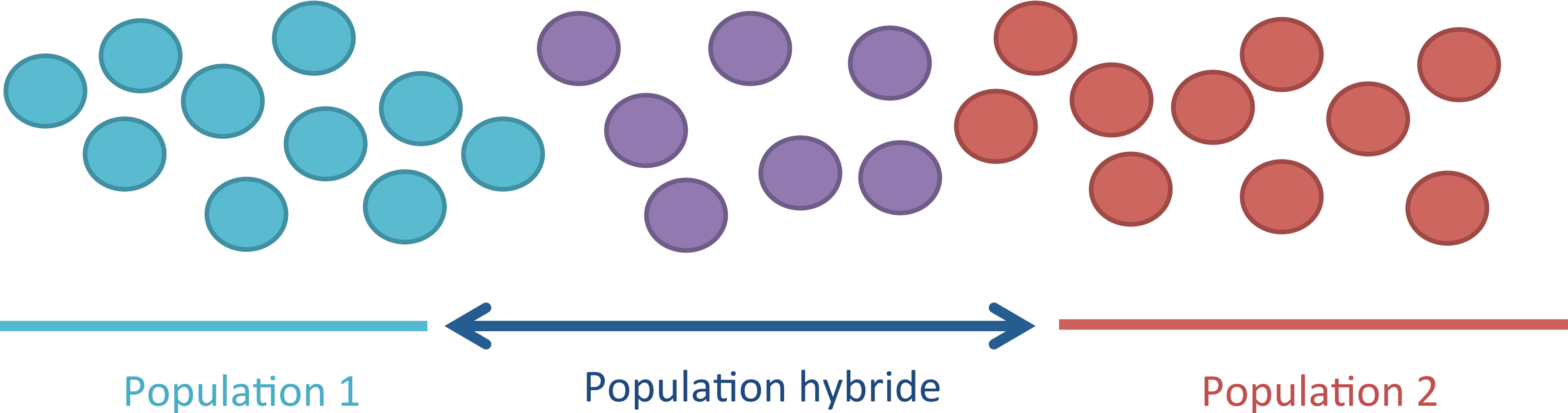
Du design aux analyses bio-informatiques : de nouveaux marqueurs pour une espèce protégée, l'hippocampe moucheté *Hippocampus guttulatus*

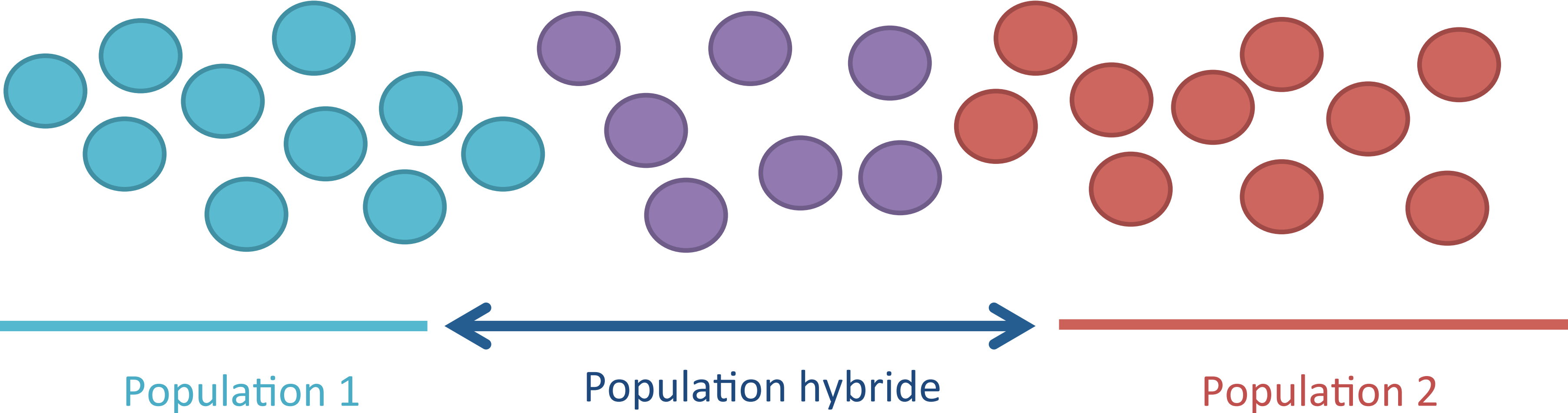
Florentine Riquet, Cathy Liautard-Haag & Nicolas Bierne
Institut des Sciences de l'Evolution
Université de Montpellier CNRS
Station marine de Sète - OSU OREME

Contact: flo.riquet@gmail.com & nicolas.bierne@univ-montp2.fr

Zone hybride: région où des populations **génétiquement différenciées** se **rencontrent**, se **reproduisent** et produisent des populations **hybrides** (Barton & Hewitt 1985)

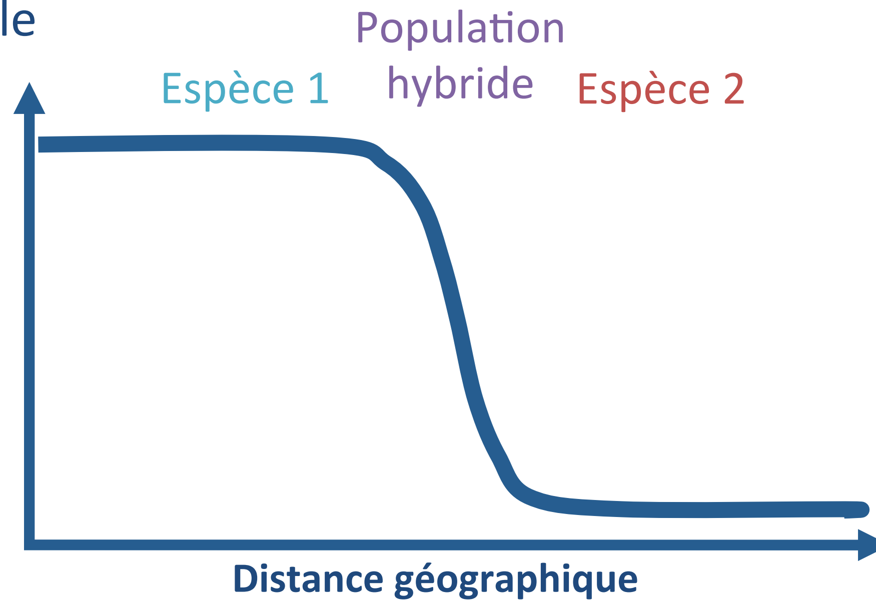


Zone hybride: région où des populations **génétiquement différenciées** se **rencontrent**, se **reproduisent** et produisent des populations **hybrides** (Barton & Hewitt 1985)



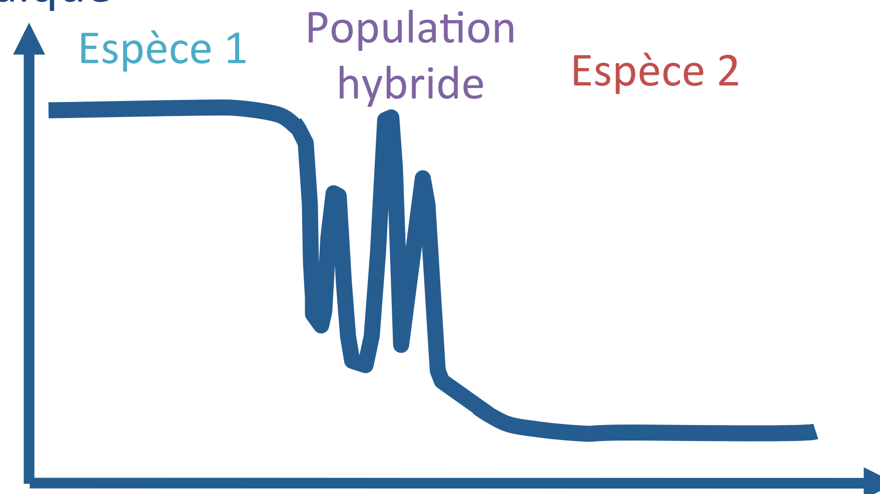
Contexte spatial particulier: **zone de contact entre espèces naissantes**

- Zone hybride clinale



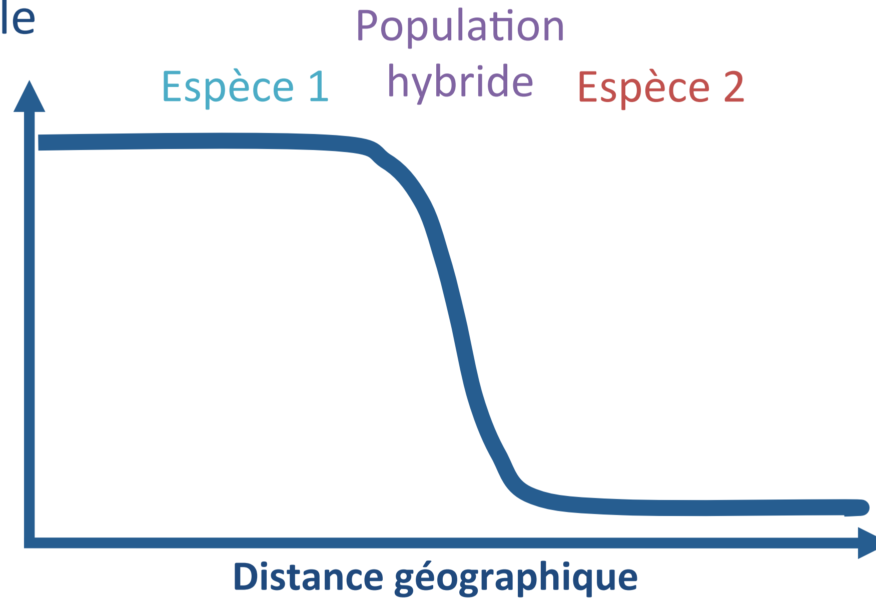
e.g. le cas des sauterelles (Butlin 1998), des souris (Raufaste 2001)

- Zone hybride mosaïque



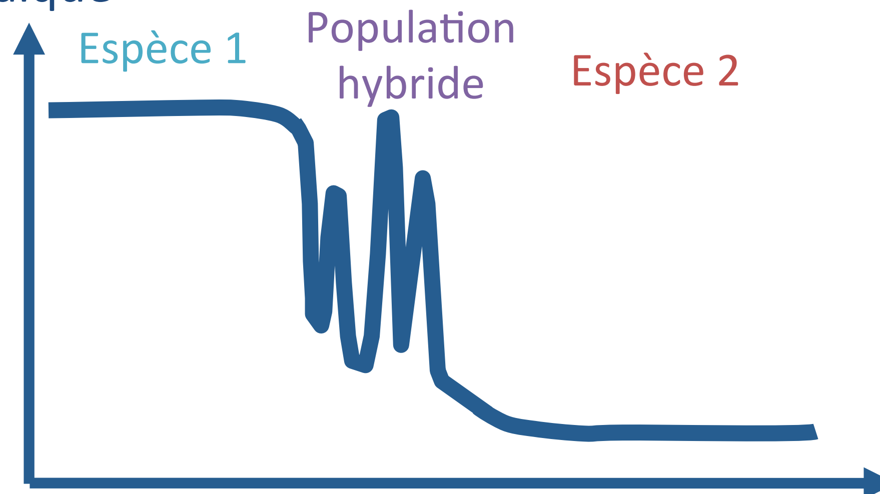
e.g. le cas des criquets (Howard et al. 1998) et moules(Bierne et al. 2002)

- Zone hybride clinale



Contact secondaire & Isolement reproductif intrinsèque

- Zone hybride mosaïque



Spéciation sympatrique & Isolement reproductif écologique (adaptation locale)

Zones hybrides multifactorielles

Ex. de *Bombina bombina* et *B. variegata* (Szymura & Barton 1986)

bombina

N

10 Km

variegata

Le cas de l'hippocampe moucheté *Hippocampus guttulatus*

- **Faibles capacités de dispersion** (peu mobile, territorial, fertilisation interne, phase dispersive courte)

- **Rare** : petite taille de population

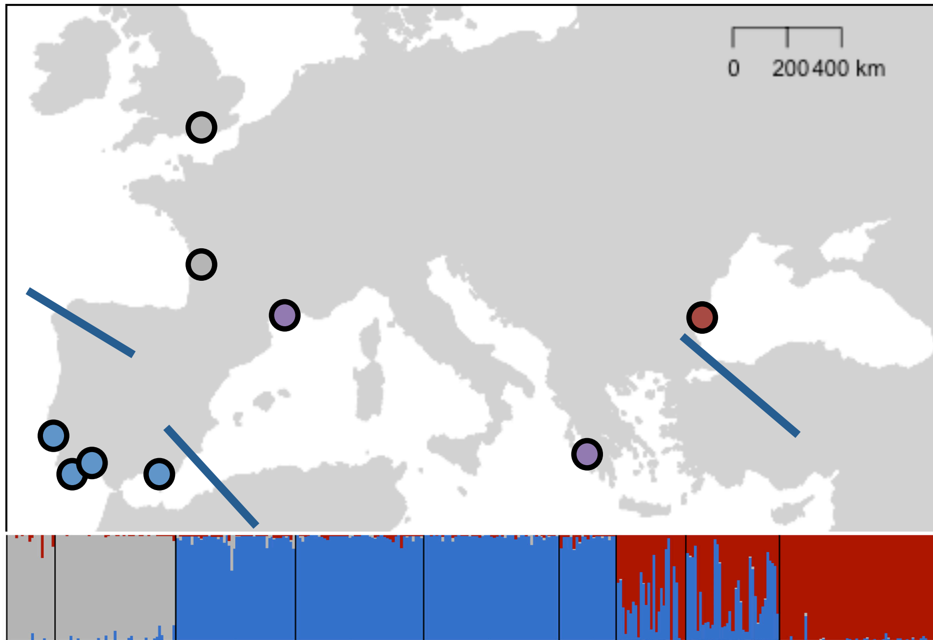
Le cas de l'hippocampe moucheté *Hippocampus guttulatus*

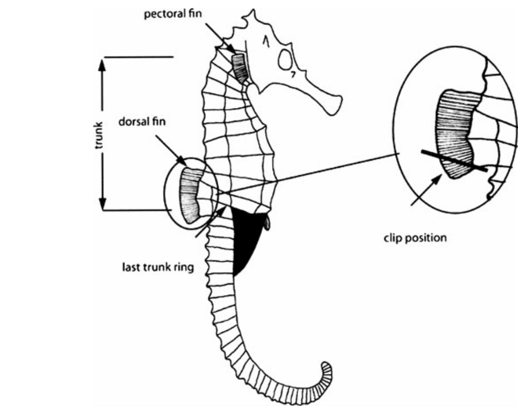
- **Faibles capacités de dispersion** (peu mobile, territorial, fertilisation interne, phase dispersive courte)

- **Rare** : petite taille de population

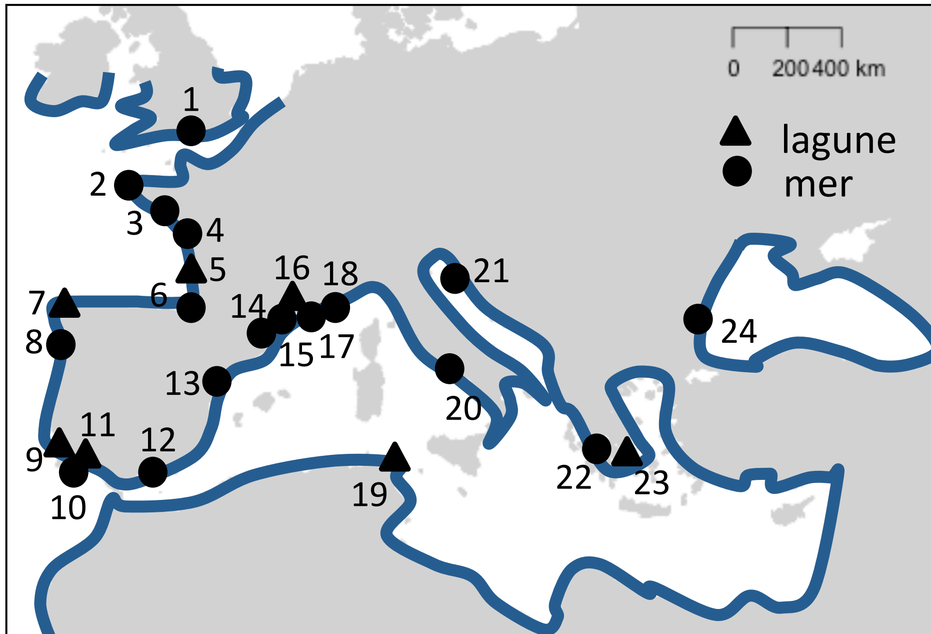
une forte différenciation génétique entre des populations isolées et de petites tailles

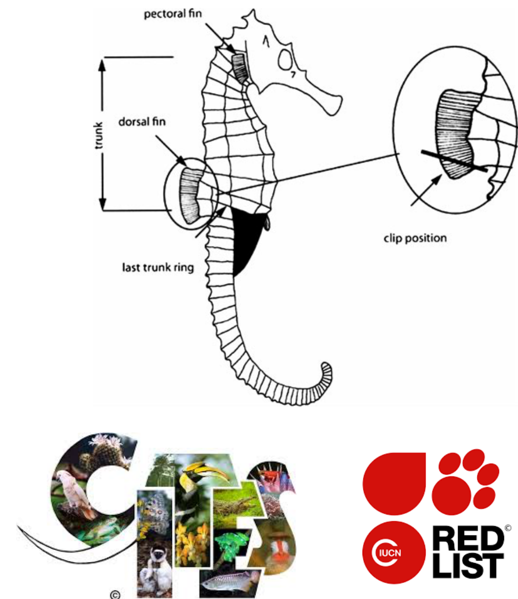
Le cas de l'hippocampe moucheté *Hippocampus guttulatus*

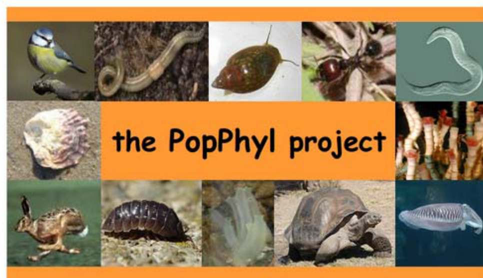




Le cas de l'hippocampe moucheté *Hippocampus guttulatus*





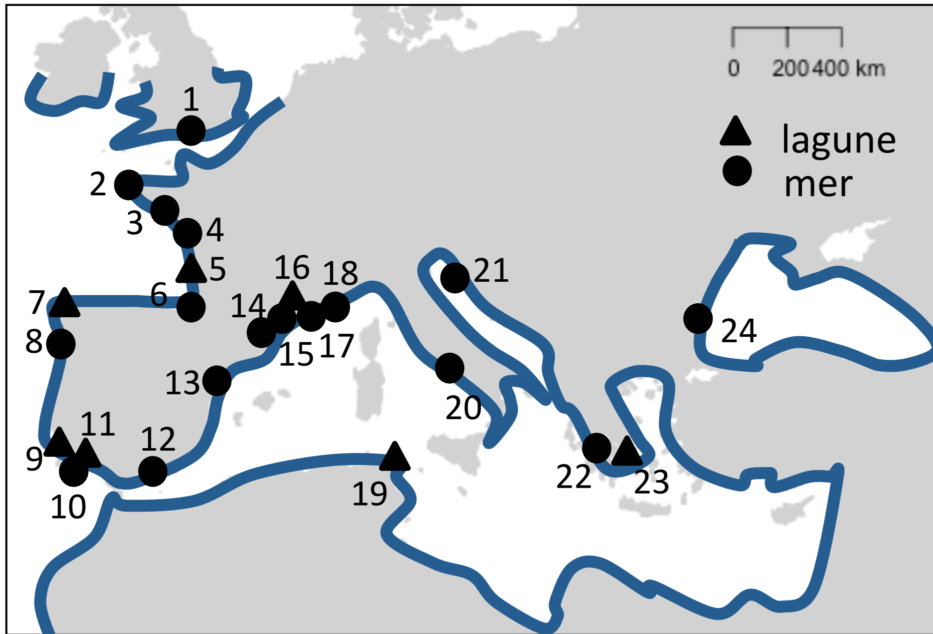


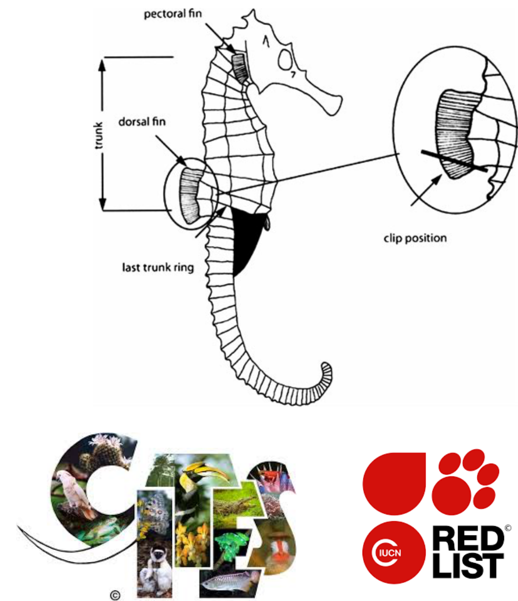
Linking molecular evolution to species biology and ecology.

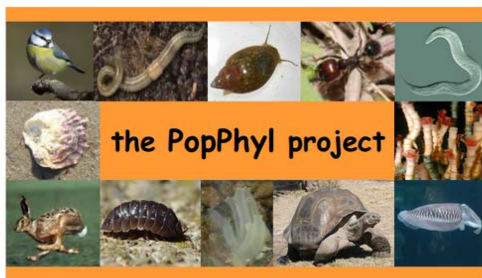
Transcriptome complet de 6 *H. guttulatus*
de 3 localités (Croisic, Faro and Thau)
(Rominguier et al. 2014)

→ **286 SNPs BeadXpress**

Le cas de l'hippocampe moucheté *Hippocampus guttulatus*

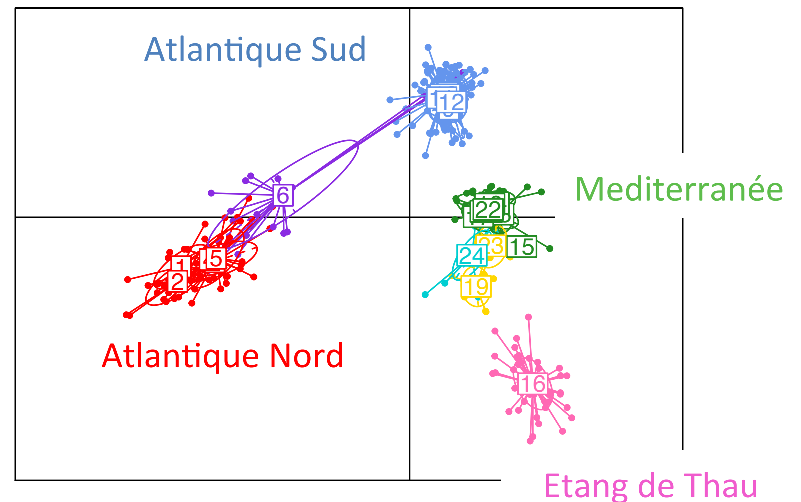






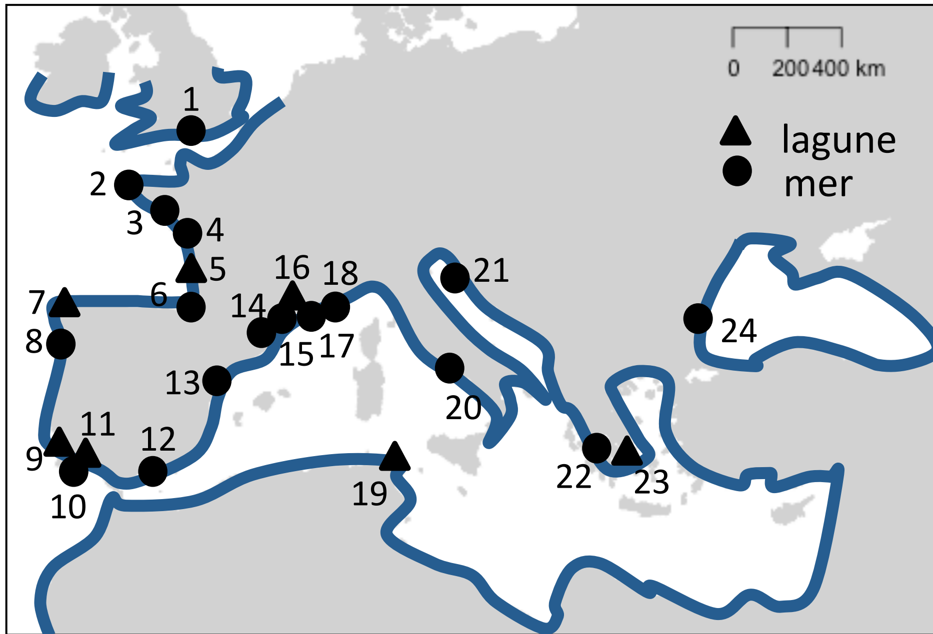
Linking molecular evolution to species biology and ecology.

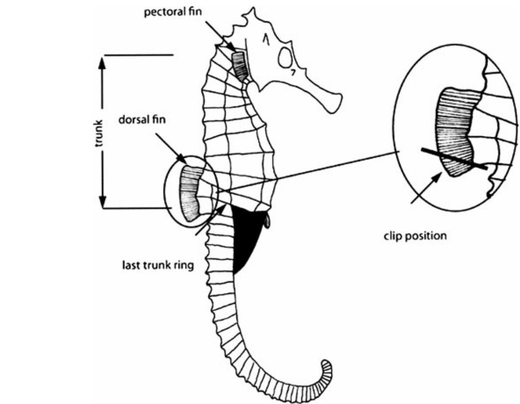
Transcriptome complet de 6 *H. guttulatus*
de 3 localités (Croisic, Faro and Thau)
(Rominguier et al. 2014)

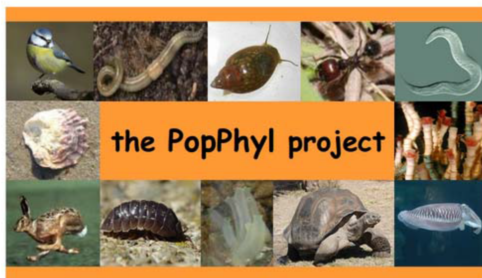


→ 286 SNPs BeadXpress

Le cas de l'hippocampe moucheté *Hippocampus guttulatus*





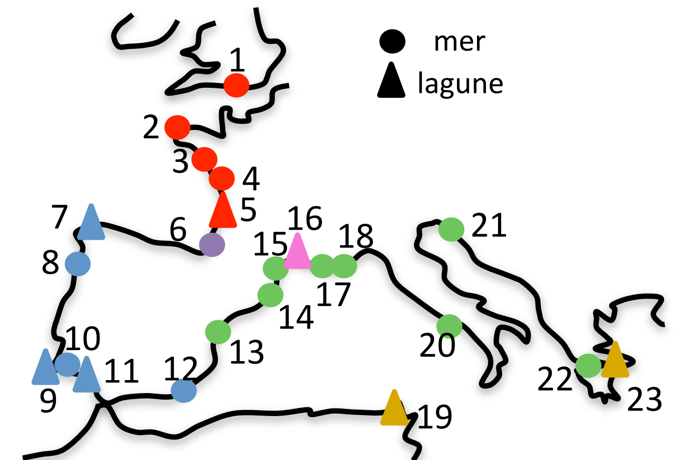


Linking molecular evolution to species biology and ecology.

Transcriptome complet de 6 *H. guttulatus* de 3 localités (Croisic, Faro and Thau)

(Rominguier et al. 2014)

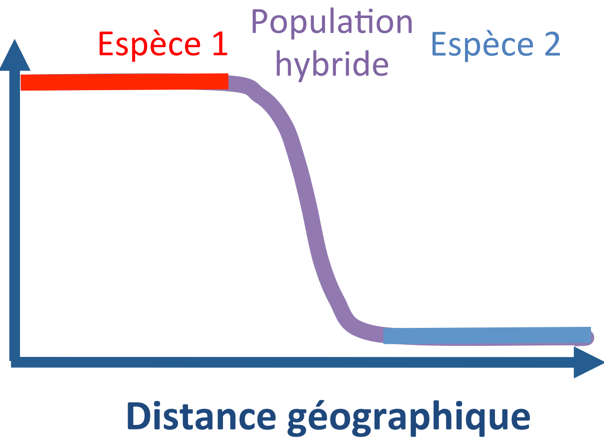
286 SNPs BeadXpress

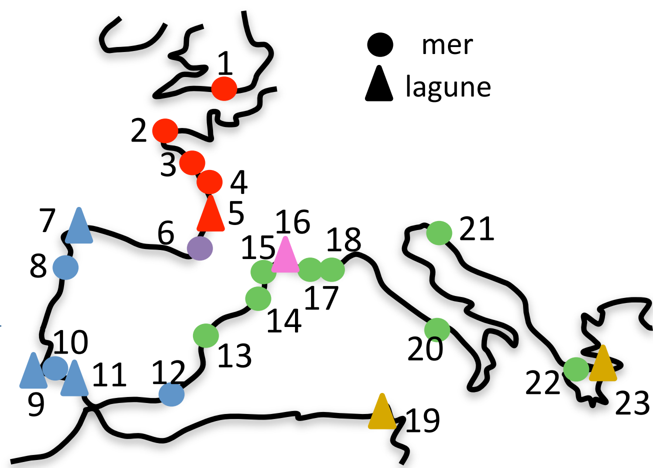


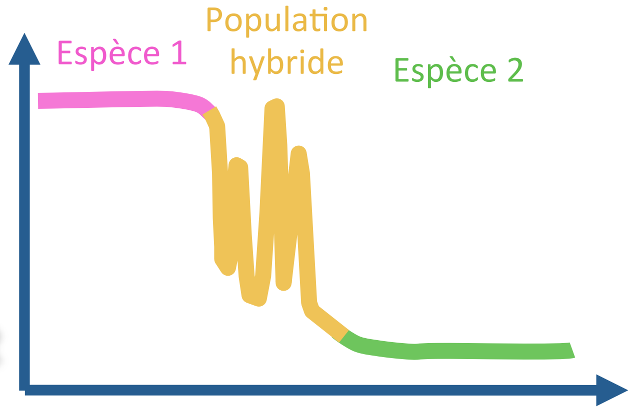
4 espèces identifiées

isolement reproducteur très fort

Développer (« suffisamment ») de nouveaux marqueurs chez une espèce non modèle

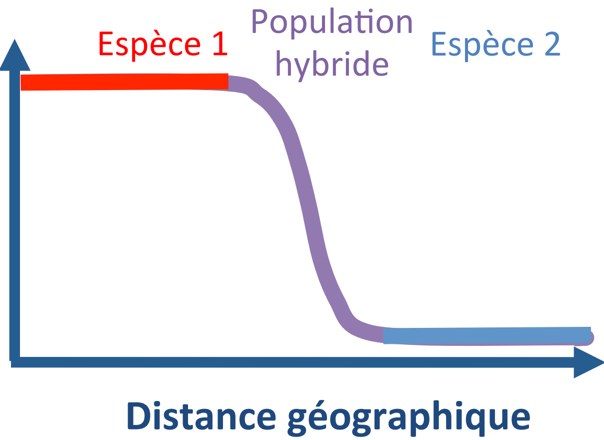


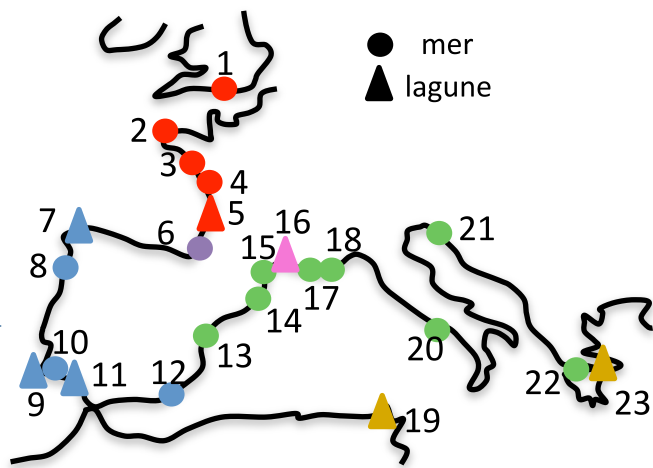


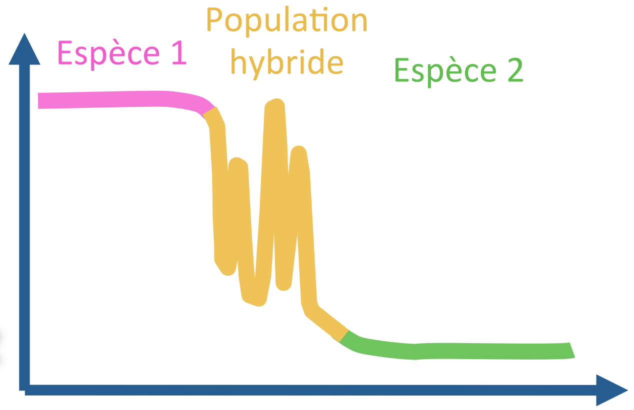


4 espèces identifiées
isolement reproducteur très fort
Identification de 14 outliers

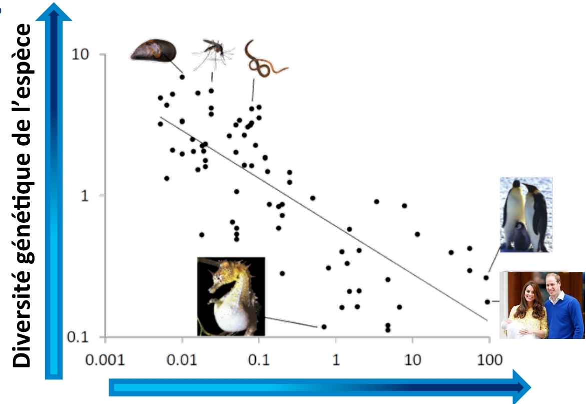
Développer (« suffisamment ») de nouveaux marqueurs chez une espèce non modèle







4 espèces identifiées
isolement reproducteur très fort
Identification de 14 outliers



Investissement parental

Objectifs:

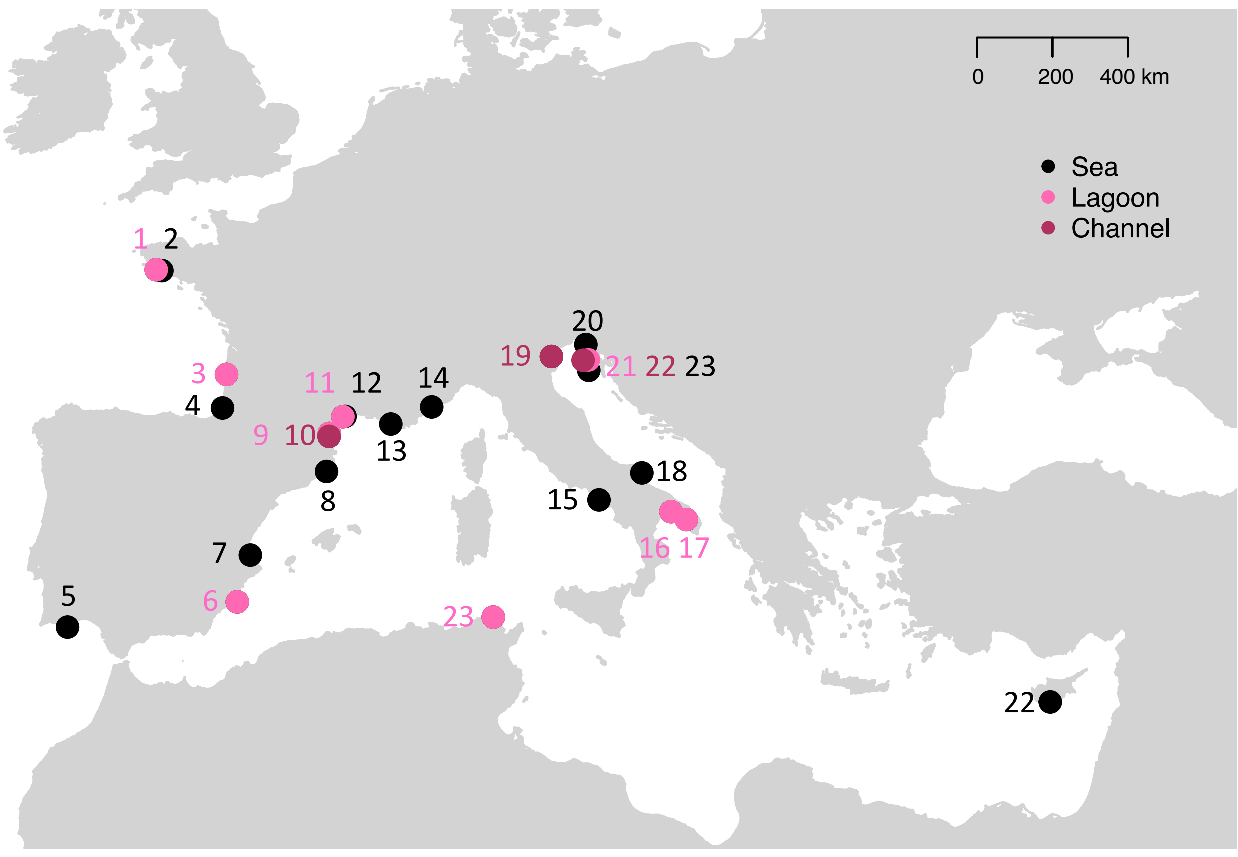
- De nouveaux sites lagunaires et maritimes
- De nouveaux marqueurs

chez *Hippocampus guttulatus*, espèce non modèle & peu polymorphe

Développer (« suffisamment ») de nouveaux marqueurs chez une espèce non modèle

Objectifs:

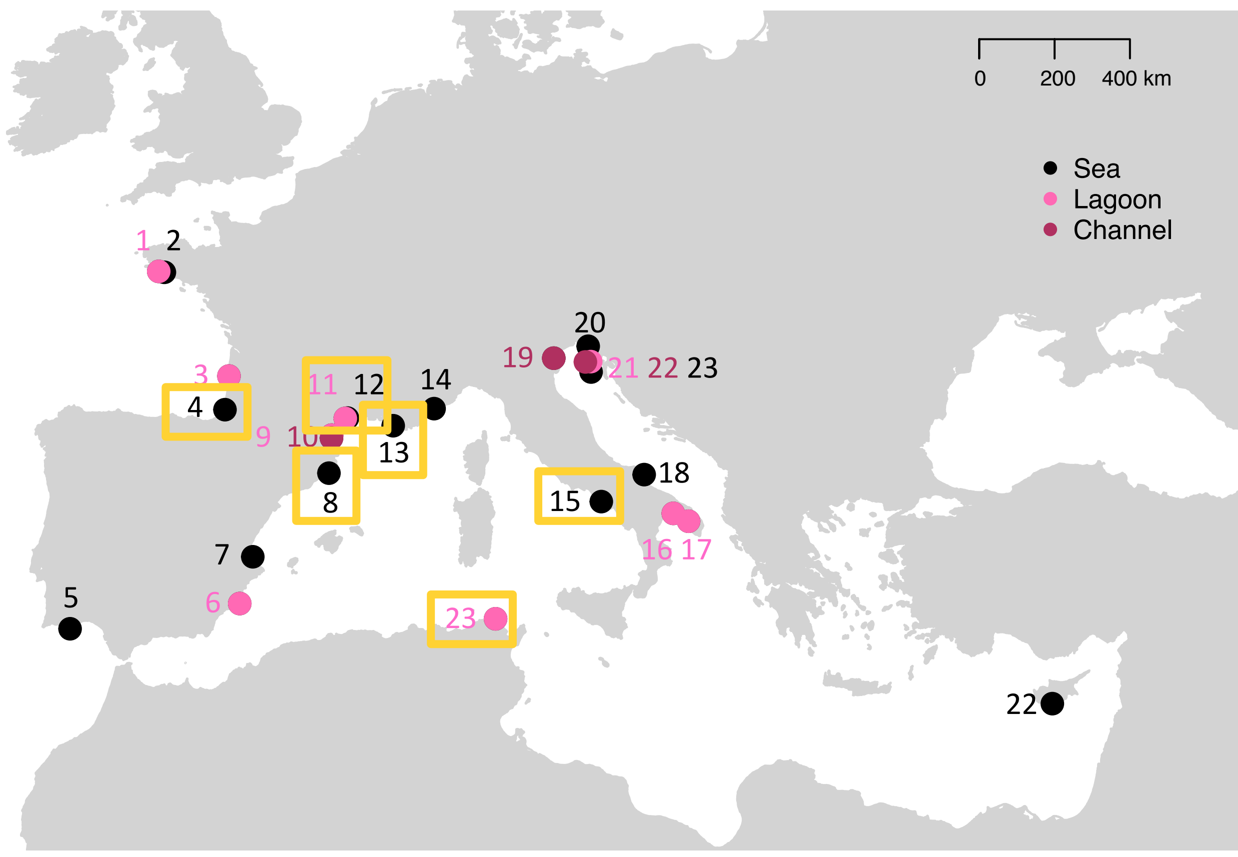
- De nouveaux sites lagunaires et maritimes ✔
- De nouveaux marqueurs



Développer (« suffisamment ») de nouveaux marqueurs chez une espèce non modèle

Objectifs:

- De nouveaux sites lagunaires et maritimes ✔
- De nouveaux marqueurs

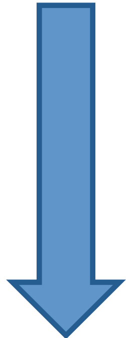


19 individus en commun typés BeadXpress

Développer (« suffisamment ») de nouveaux marqueurs chez une espèce non modèle

Objectifs:

- De nouveaux sites lagunaires et maritimes ✔
- De nouveaux marqueurs



700 marqueurs – 96 individus – 2 jours de paillasse – 1 run MiSeq

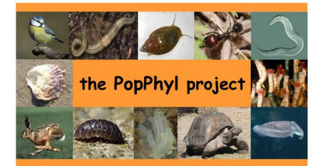
I. **Design des sondes**

II. **Préparation des librairies**

III. **Séquencage**

IV. **Analyses bio-informatiques**

I. Design des sondes



7151 SNPs

Tri selon leur fréquence allélique et leur distance aux limites exoniques

↓

2082 SNPs

Tri selon les résultats du blast

↓

1910 SNPs

↓

Priorité 1
12 SNPs
montrant une différenciation extrême

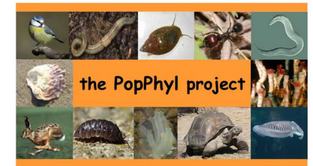
Priorité 2
246 SNPs
BeadXpress
90 pb min et 25pb de part et d'autre du SNP

Priorité 3
60 SNPs
Fixé dans un échantillon RNAseq de 2 individus

Priorité 4
642 SNPs
au min 50 pb de part et d'autre du SNP

960 SNPs

I. Design des sondes



Linking molecular evolution to species biology and ecology.

<u>Conditions d'Illumina:</u>
- Séquences de min 300pb avec le SNP au milieu, i.e. 150pb de chaque coté
- Potentiellement peuvent faire des amplicons de 100 voire 75pb
- Garantie 80% des positions avec au moins 0.2x la couverture moyenne visée

Design de 915 sondes possibles
Test de 176 séquences sur les 960 SNPs ciblés

Priorité 1

12 SNPs

11 SNPs

Priorité 2

246 SNPs

215 SNPs

Priorité 3

60 SNPs

59 SNPs

Priorité 4

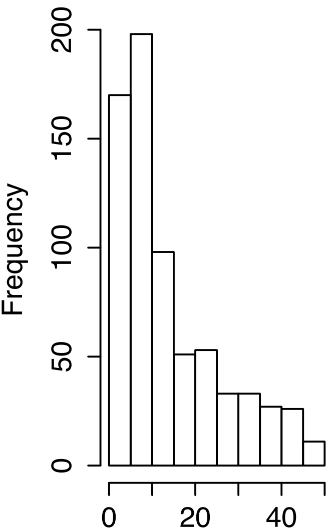
642 SNPs

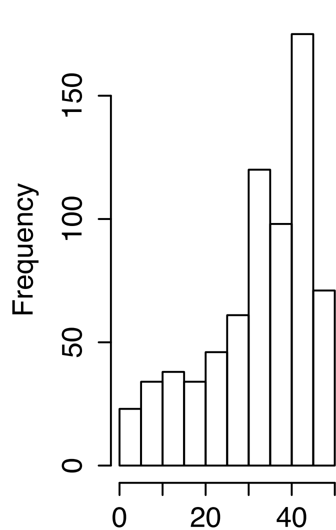
415 SNPs
Exclusion des SNPs les plus proches des sondes

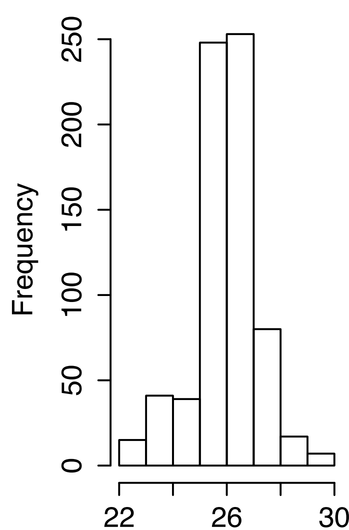
700 SNPs

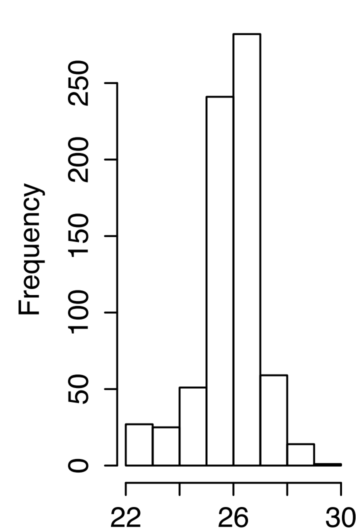
I. Design des sondes

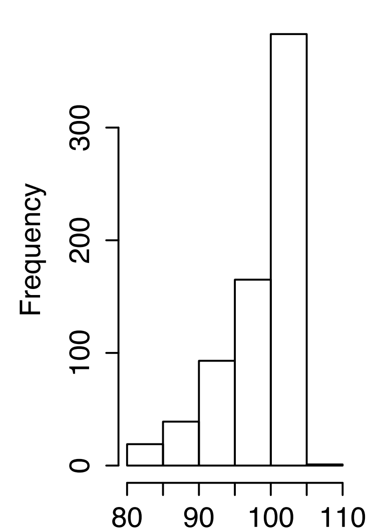
- 700 paires de sondes designés => 700 SNPs ciblés mais 890 SNPs attendus
    - Séquences entre 80 et 106bp
    - Position du SNP ciblé : entre 26ème bp -76ème bp (sans les sondes : 1-49)
    - Longueur des sondes : 22 – 30 bp
    - Chevauchement : entre 44 et 70 bp entre les 2 reads











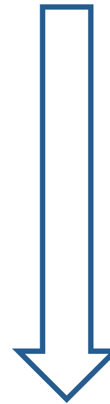
I. Design des sondes

II. Préparation des librairies

III. Séquencage

IV. Analyses bio-informatiques

II. Préparation des librairies



700 marqueurs
96 individus
2 jours de manip

II. Préparation des librairies

Quantify and Dilute DNA
Hands-on: 10 minutes
Total: 10 minutes
Reagents: gDNA, RS1, SS1

> 10 ng d'ADNg

II. Préparation des librairies

Quantify and Dilute DNA
Hands-on: 10 minutes
Total: 10 minutes
Reagents: gDNA, RS1, SS1

Hybridize Oligo Pool
Hands-on: 15 minutes
Total: 1 hour 50 minutes
Reagents: gDNA, CAT, OHS2, 2800M, ACP3, RS1, also ELB, ELE
Plate: HYP

> 10 ng d'ADNg

II. Préparation des librairies

1. **Quantify and Dilute DNA**
   Hands-on: 10 minutes
   Total: 10 minutes
   Reagents: gDNA, RS1, SS1

2. **Hybridize Oligo Pool**
   Hands-on: 15 minutes
   Total: 1 hour 50 minutes
   Reagents: gDNA, CAT, OHS2, 2800M, ACP3, RS1, also ELB, ELE
   Plate: HYP

3. **Remove Unbound Oligos**
   Hands-on: 20 minutes
   Total: 20 minutes
   Reagents: SW1, SPB, Fresh 60% EtOH
   Plate: HYP

> 10 ng d'ADNg

II. Préparation des librairies

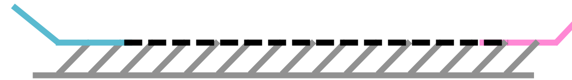
1. **Quantify and Dilute DNA**
   Hands-on: 10 minutes
   Total: 10 minutes
   Reagents: gDNA, RS1, SS1

2. **Hybridize Oligo Pool**
   Hands-on: 15 minutes
   Total: 1 hour 50 minutes
   Reagents: gDNA, CAT, OHS2, 2800M, ACP3, RS1, also ELB, ELE
   Plate: HYP

3. **Remove Unbound Oligos**
   Hands-on: 20 minutes
   Total: 20 minutes
   Reagents: SW1, SPB, Fresh 60% EtOH
   Plate: HYP

4. **Extend-Ligate Bound Oligos**
   Hands-on: 20 minutes
   Total: 1 hour 10 minutes
   Reagents: ELB/ELE mixture, also EDP, EMM
   Plate: HYP

> 10 ng d'ADNg



II. Préparation des librairies

1 Quantify and Dilute DNA
Hands-on: 10 minutes
Total: 10 minutes
Reagents: gDNA, RS1, SS1

> 10 ng d'ADNg

2 Hybridize Oligo Pool
Hands-on: 15 minutes
Total: 1 hour 50 minutes
Reagents: gDNA, CAT, OHS2, 2800M, ACP3, RS1, also ELB, ELE
Plate: HYP

3 Remove Unbound Oligos
Hands-on: 20 minutes
Total: 20 minutes
Reagents: SW1, SPB, Fresh 60% EtOH
Plate: HYP

4 Extend-Ligate Bound Oligos
Hands-on: 20 minutes
Total: 1 hour 10 minutes
Reagents: ELB/ELE mixture, also EDP, EMM
Plate: HYP

6 Amplify Libraries
Safe Stopping Point
Hands-on: 30 minutes
Total: 2 hours—2 hours 15 minutes
Reagents: EDP/EMM mixture, i5 adapters, i7 adapters
Plate: HYP

II. Préparation des librairies

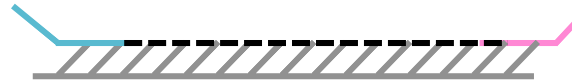
1. Quantify and Dilute DNA
   Hands-on: 10 minutes
   Total: 10 minutes
   Reagents: gDNA, RS1, SS1

> 10 ng d'ADNg

2. Hybridize Oligo Pool
   Hands-on: 15 minutes
   Total: 1 hour 50 minutes
   Reagents: gDNA, CAT, OHS2, 2800M, ACP3, RS1, also ELB, ELE
   Plate: HYP

3. Remove Unbound Oligos
   Hands-on: 20 minutes
   Total: 20 minutes
   Reagents: SW1, SPB, Fresh 60% EtOH
   Plate: HYP

4. Extend-Ligate Bound Oligos
   Hands-on: 20 minutes
   Total: 1 hour 10 minutes
   Reagents: ELB/ELE mixture, also EDP, EMM
   Plate: HYP



6. Amplify Libraries
   Hands-on: 30 minutes
   Total: 2 hours—2 hours 15 minutes
   Reagents: EDP/EMM mixture, i5 adapters, i7 adapters
   Plate: HYP

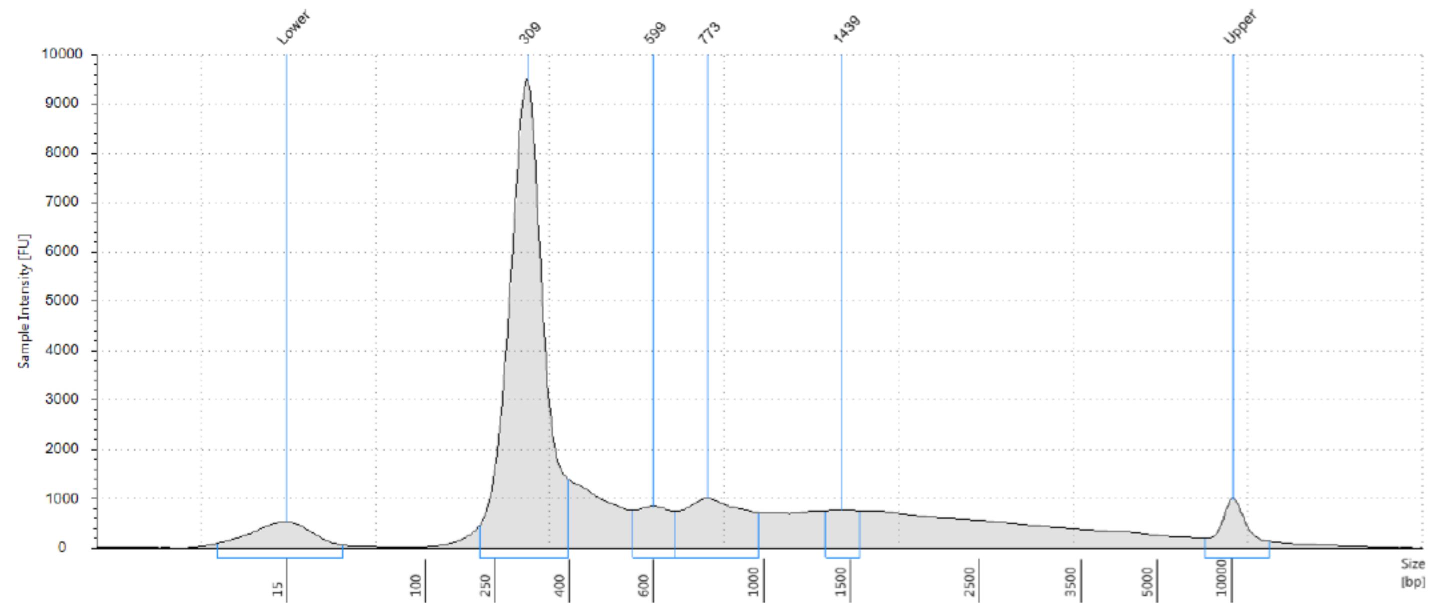
Safe Stopping Point

7. Clean Up Libraries
   Hands-on: 20 minutes
   Total: 30 minutes
   Reagents: RSB, SPB, Fresh 80% EtOH
   Plate: CLP, LNP

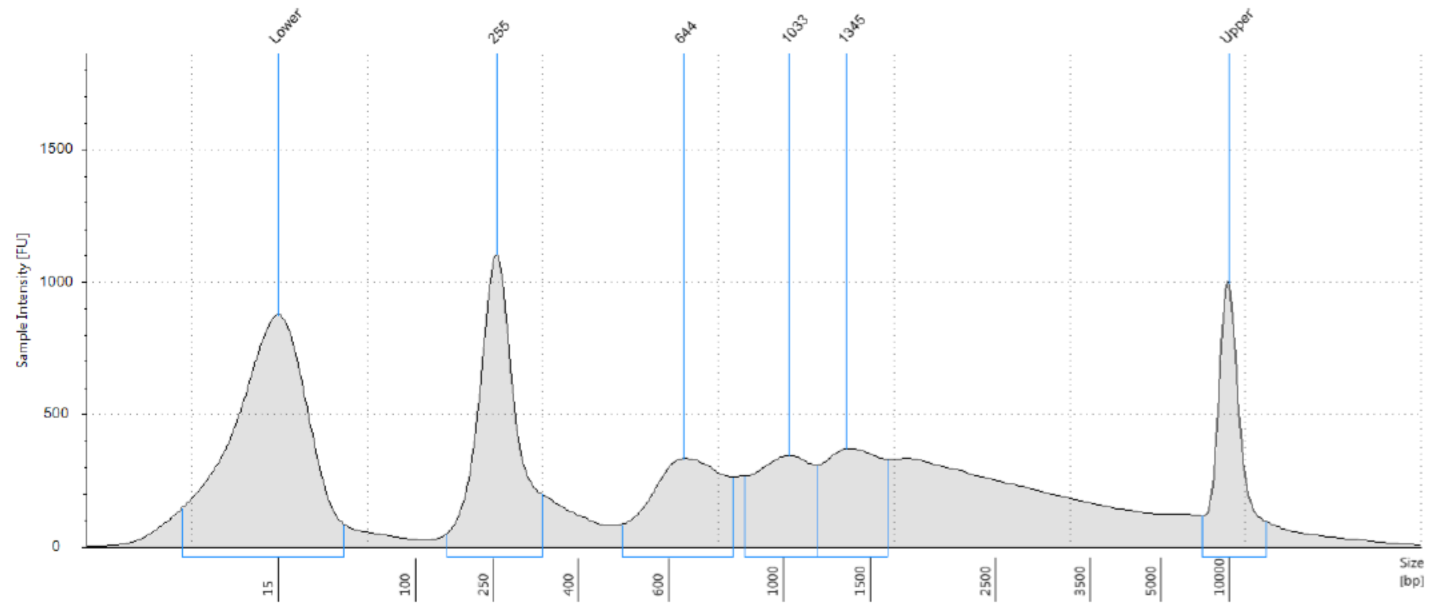
Safe Stopping Point

> 10 nM

II. Préparation des librairies



II. Préparation des librairies



II. Préparation des librairies

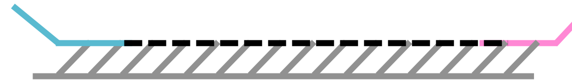
1 Quantify and Dilute DNA
Hands-on: 10 minutes
Total: 10 minutes
Reagents: gDNA, RS1, SS1

> 10 ng d'ADNg

2 Hybridize Oligo Pool
Hands-on: 15 minutes
Total: 1 hour 50 minutes
Reagents: gDNA, CAT, OHS2, 2800M, ACP3, RS1, also ELB, ELE
Plate: HYP

3 Remove Unbound Oligos
Hands-on: 20 minutes
Total: 20 minutes
Reagents: SW1, SPB, Fresh 60% EtOH
Plate: HYP

4 Extend-Ligate Bound Oligos
Hands-on: 20 minutes
Total: 1 hour 10 minutes
Reagents: ELB/ELE mixture, also EDP, EMM
Plate: HYP



6 Amplify Libraries
Safe Stopping Point
Hands-on: 30 minutes
Total: 2 hours—2 hours 15 minutes
Reagents: EDP/EMM mixture, i5 adapters, i7 adapters
Plate: HYP

7 Clean Up Libraries
Safe Stopping Point
Hands-on: 20 minutes
Total: 30 minutes
Reagents: RSB, SPB, Fresh 80% EtOH
Plate: CLP, LNP

> 10 nM

8 Normalize Libraries
Safe Stopping Point
Hands-on: 30 minutes
Total: 50 minutes
Reagents: LNA1, LNB1 LNW1, LNS2, NaOH
Plate: SGP

II. Préparation des librairies

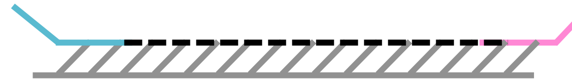
Quantify and Dilute DNA
Hands-on: 10 minutes
Total: 10 minutes
Reagents: gDNA, RS1, SS1

> 10 ng d'ADNg

Hybridize Oligo Pool
Hands-on: 15 minutes
Total: 1 hour 50 minutes
Reagents: gDNA, CAT, OHS2, 2800M, ACP3, RS1, also ELB, ELE
Plate: HYP

Remove Unbound Oligos
Hands-on: 20 minutes
Total: 20 minutes
Reagents: SW1, SPB, Fresh 60% EtOH
Plate: HYP

Extend-Ligate Bound Oligos
Hands-on: 20 minutes
Total: 1 hour 10 minutes
Reagents: ELB/ELE mixture, also EDP, EMM
Plate: HYP



Amplify Libraries
Safe Stopping Point
Hands-on: 30 minutes
Total: 2 hours—2 hours 15 minutes
Reagents: EDP/EMM mixture, i5 adapters, i7 adapters
Plate: HYP

Clean Up Libraries
Safe Stopping Point
Hands-on: 20 minutes
Total: 30 minutes
Reagents: RSB, SPB, Fresh 80% EtOH
Plate: CLP, LNP

> 10 nM

Normalize Libraries
Safe Stopping Point
Hands-on: 30 minutes
Total: 50 minutes
Reagents: LNA1, LNB1 LNW1, LNS2, NaOH
Plate: SGP

Pool Libraries
Safe Stopping Point
Hands-on: 5 minutes
Total: 5 minutes
Reagents: HT1
Tube: PAL

Pre-PCR · Post-PCR

I. Design des sondes

II. Préparation des librairies

III. Séquencage

IV. Analyses bio-informatiques

III. Séquencage

- N= contemporains et N= 33 secs
  - 1ère manip 96 ind dont 3 contrôles, 60 contemporains et 33 secs
  - 2ème manip: 47 ind dont 16 contemporains et 31 secs

2X75 cycles sur MiSeq Reagent kit MiSeq Reagent kit v3

- Nombre de reads : 25M
- Output max. 3.8 Gb

=> 500X et 1000X attendues

RUN 02 (81 normalisés manuellement et 15 TSCA)

Dénaturation, dilution et mélange : PhiX & Banque

Tube PhiX	**Tube Banque**
5µl PhiX à 4nM + 5µl NaOH 0.2N	20µl Banque **1nM** + 20µl NaOH 0.2N
(à ce moment le PhiX est à 2nM)	(à ce moment la banque est à 0,33nM)

Vortexer-spin

Incubation 5 min à température ambiante.

Ajouter 990µl de HT1 froid et mélange par inversion.	Ajouter 20µL 200mM Tris-HCl pH7 puis Vortex-spin
	Ajouter 940µL de HT1 froid et mélanger par inversion

Mettre dans la glace.

(à ce moment le PhiX est à 20pM)	(à ce moment la banque est à 20pM)
Il peut se conserver 3 semaines à -20°C.	

Tube Banque à 8 pM

240µl banque Denat + 360 µl HT1

Tube Final (banque + PhiX 20%)

472µl de banque à 8pmol + 8µL pool2+120µl de PhiX à 20pM

RUN 03 (31 normalisés manuellement et 16 TSCA)

Dénaturation, dilution et mélange : PhiX & Banque

Tube PhiX

5µl PhiX à 4nM + 5µl NaOH 0.2N

(à ce moment le PhiX est à 2nM)

Tube Banque

20µl Banque **1nM** + 20µl NaOH 0.2N

(à ce moment la banque est à 0,33nM)

Vortexer-spin

Incubation 5 min à température ambiante.

Ajouter 990µl de HT1 froid et mélange par inversion.

Ajouter 20µL 200mM Tris-HCl pH7 puis Vortex-spin

Ajouter 940µL de HT1 froid et mélanger par inversion

Mettre dans la glace.

(à ce moment le PhiX est à 20pM)

(à ce moment la banque est à 20pM)

Il peut se conserver 3 semaines à -20°C.

Tube Banque à 8 pM

240µl banque Denat + 360 µl HT1

Tube Final (banque + PhiX 5%)

562 de banque à 8pmol + 8µL pool2 + 30µl de PhiX à 20pM

III. Séquencage

	RUN 02 (20% Phi X)	RUN 03 (5% Phi X)
Cluster Densité au mm^2 (K)	1255	980
Cluster passing filter (%)	90.71	91.3
Q30 final (%)	94.2	95.8
Read1 (%)	96.7	97.4
Read2 (%)	80.9	89.8
Read3 (%)	93.3	92.1
Read4 (%)	93.4	95.2
Indexing	entre 0 et 4%: - 1 échantillon (POS2) à 14% - 81 échantillons < 1% - 14 échantillons entre 1 et 4%	- 28 échantillons de 0,001 à 0,74% - 18 échantillons de 1,18% à 6,61%
Rendement (G)	4.2	3.4
% alignement	18.37	6.97

I. Design des sondes

II. Préparation des librairies

III. Séquencage

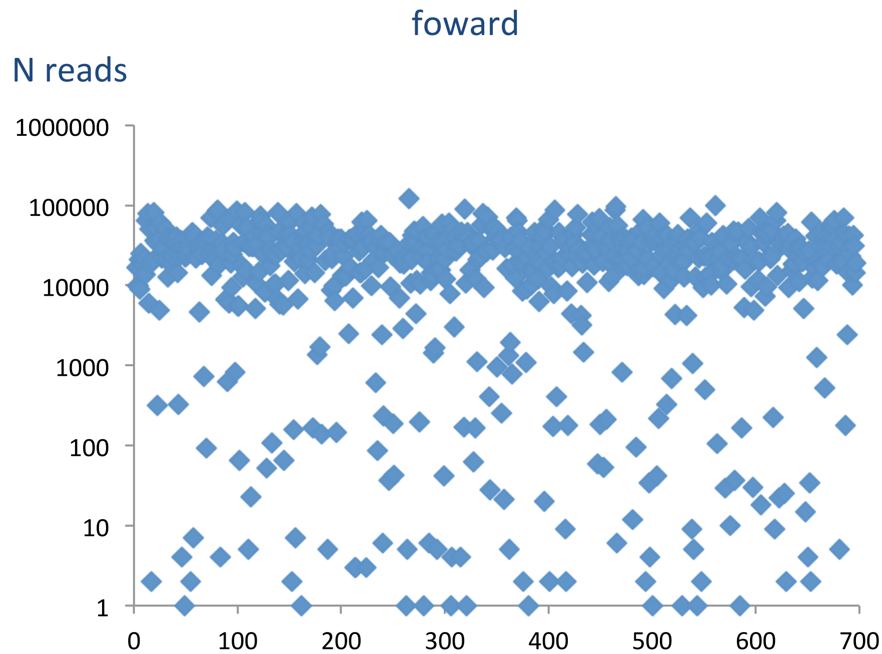
IV. Analyses bio-informatiques

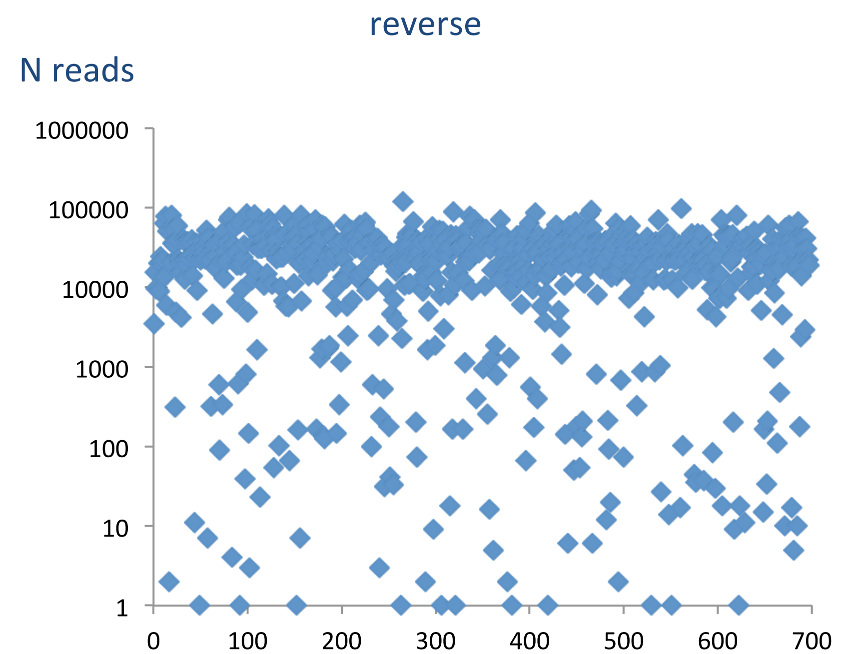
IV. Analyses bio-informatiques

Blast local des données brutes sur les 700 séquences attendues

IV. Analyses bio-informatiques

Blast local des données brutes sur les 700 séquences attendues

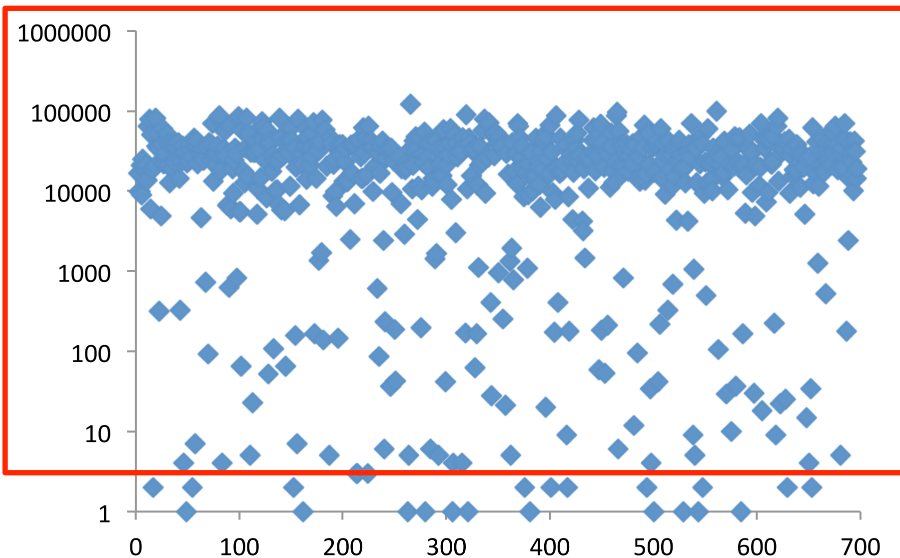


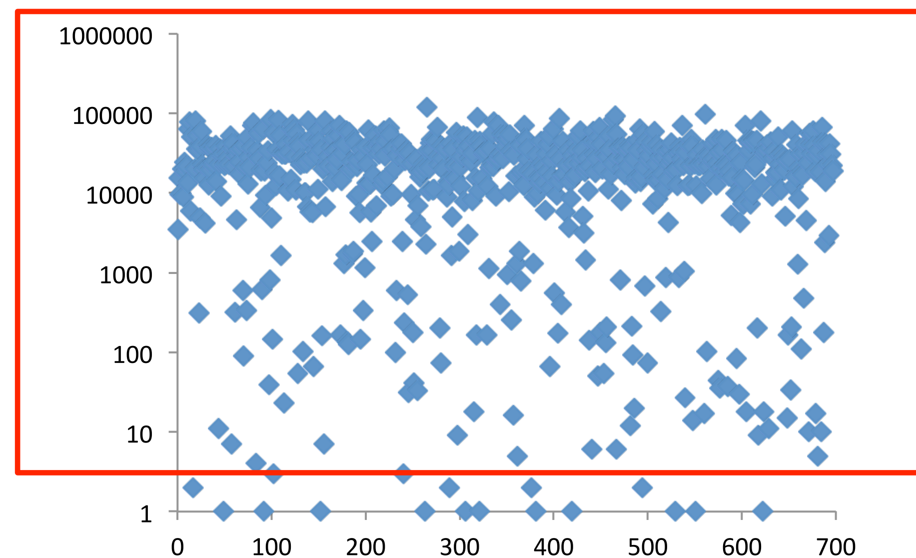


- 3 séquences retrouvées 11.10^6, 8.10^6 et $0.3.10^6$
- Nombre de reads : 16.3M & 16.4M
- 77 % de reads mappés
- 100 000X par marqueurs

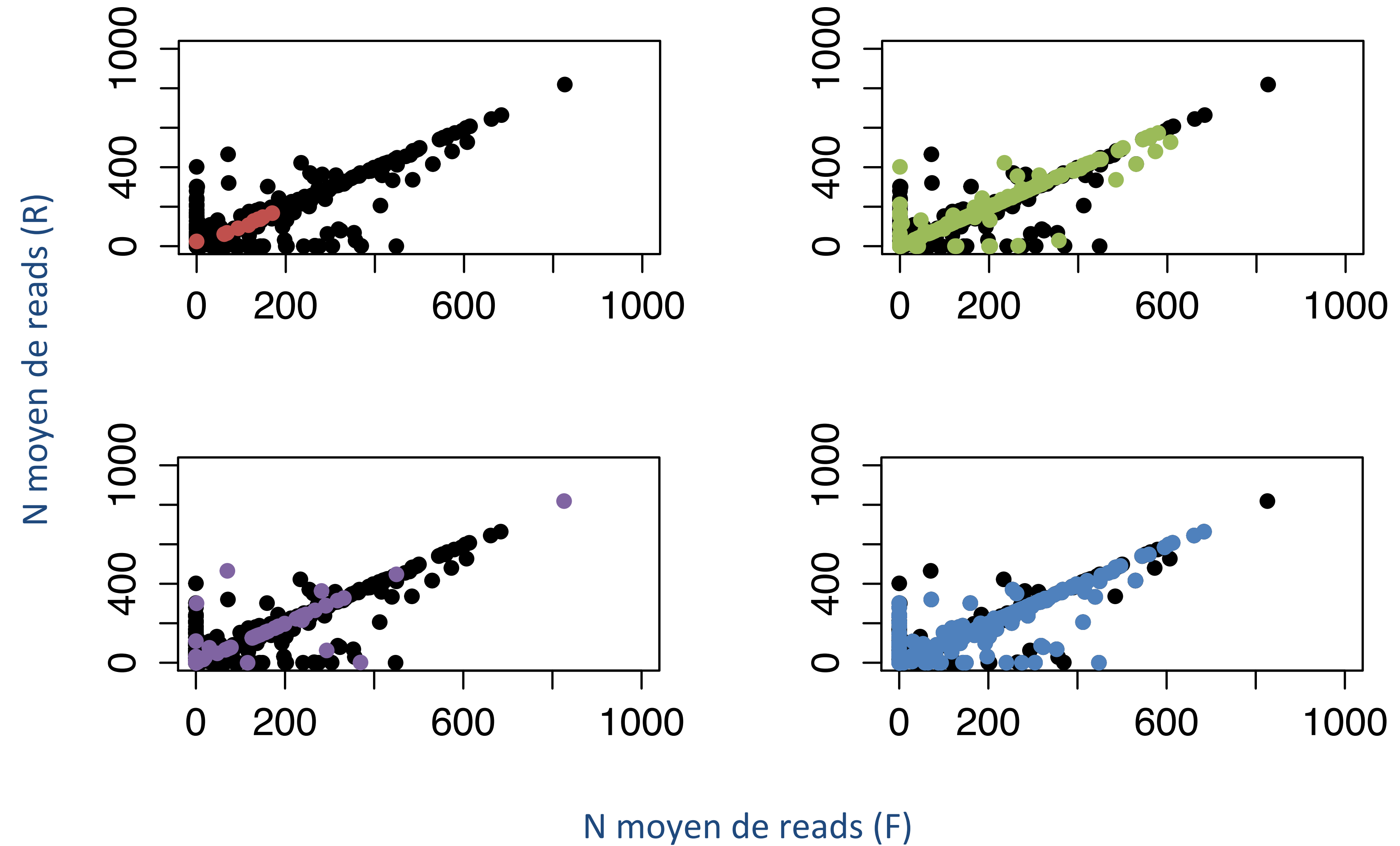
IV. Analyses bio-informatiques

Blast local des données brutes sur les 700 séquences attendues



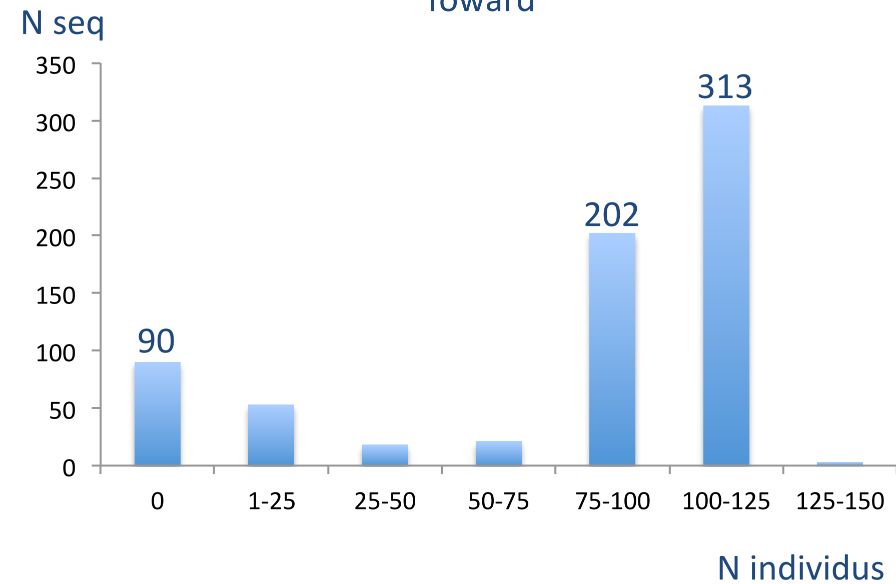


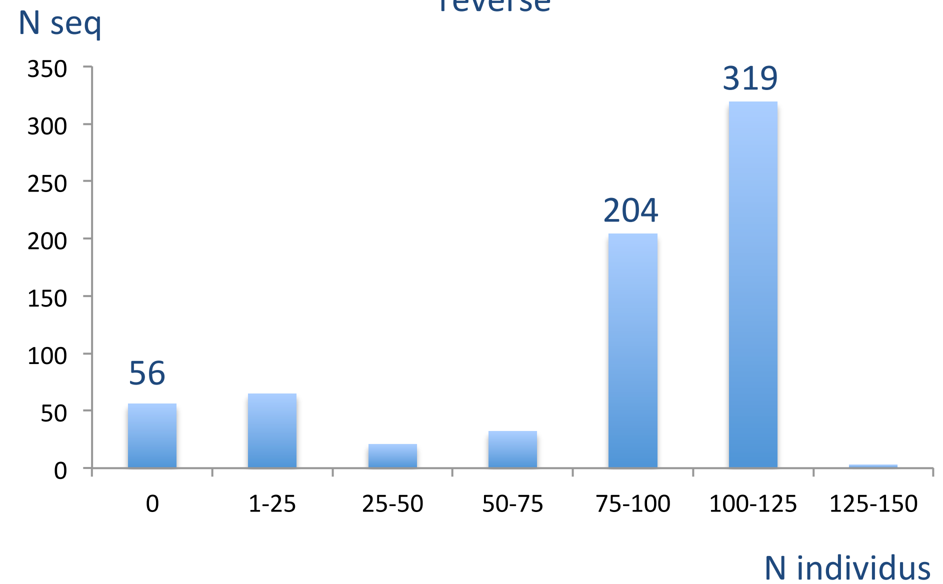
- 3 séquences retrouvées 11.10^6, 8.10^6 et $0.3.10^6$
- Nombre de reads : 16.3M & 16.4M
- 77 % de reads mappés
- 100 000X par marqueurs


N moyen de reads (R)
N moyen de reads (F)
0
200
600
1000
0
400
1000

IV. Analyses bio-informatiques

Blast local des données brutes sur les 700 séquences attendues – seuil de 5 X retenu

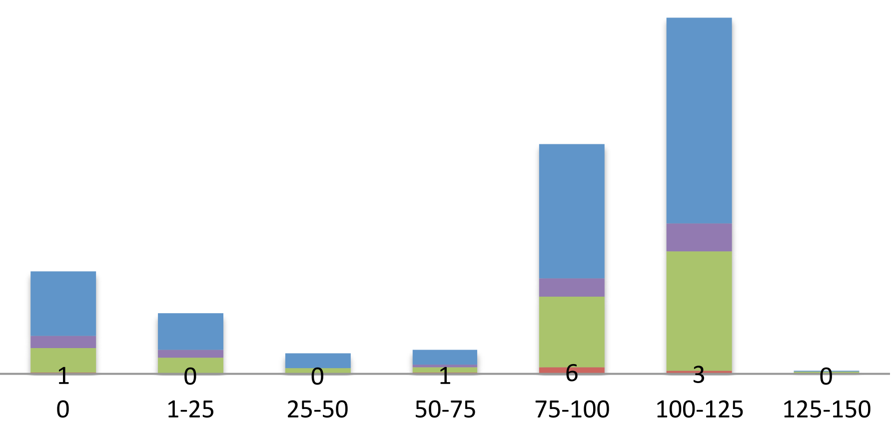


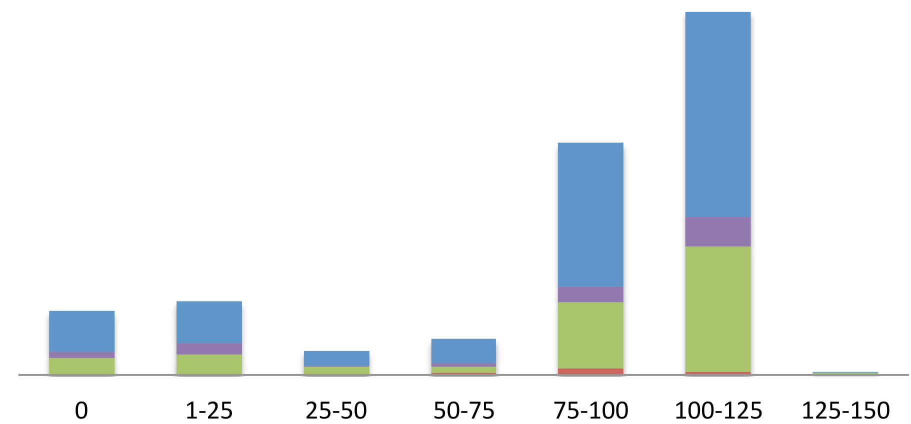


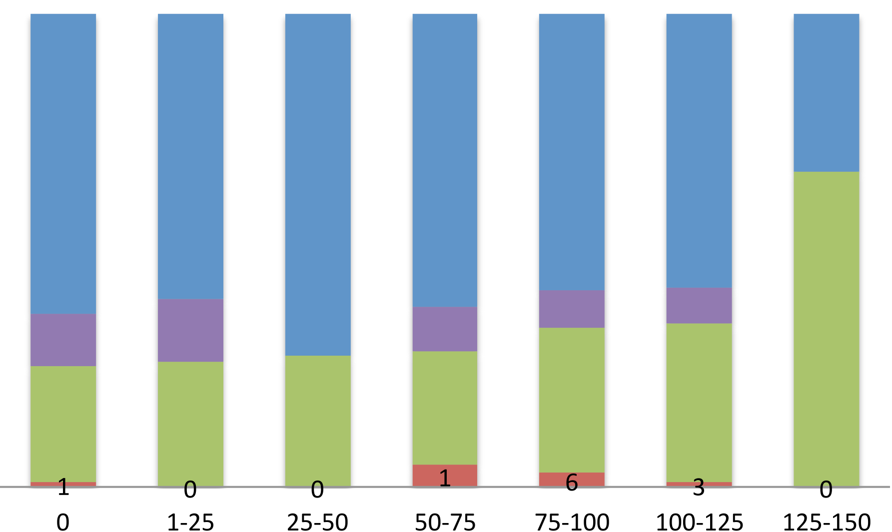
N reads moyen/seq varie 0 – 80015 (mean=359.1)

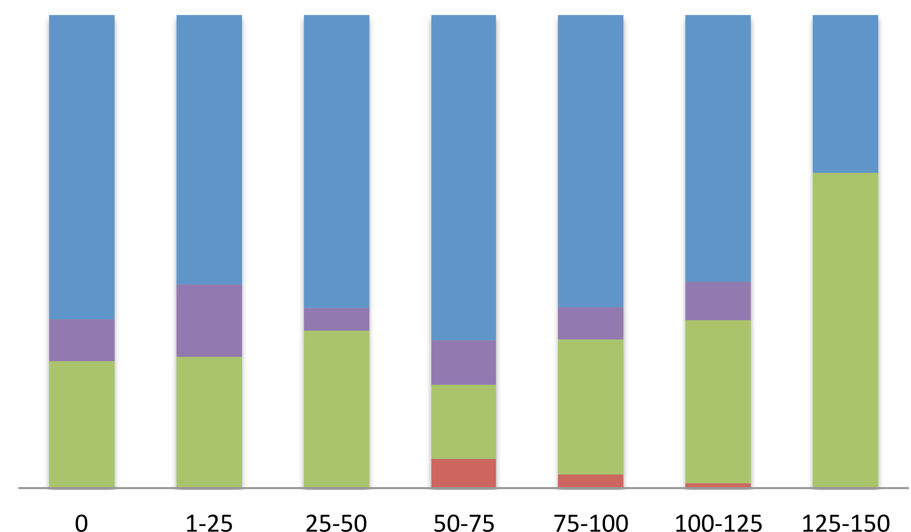
N reads moyen/seq varie 0 – 79565 (mean=357.0)

IV. Analyses bio-informatiques

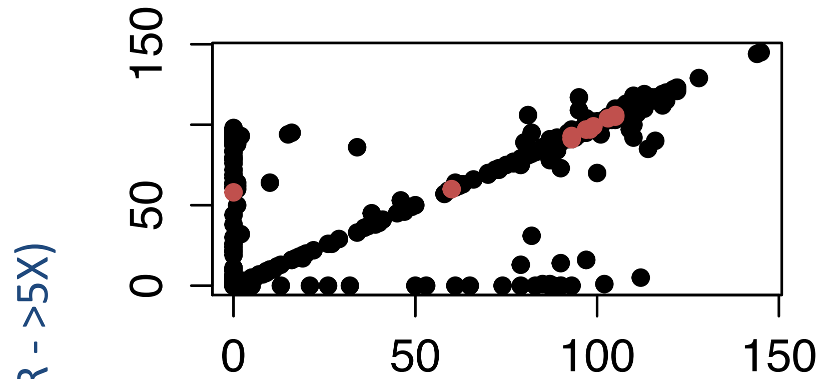


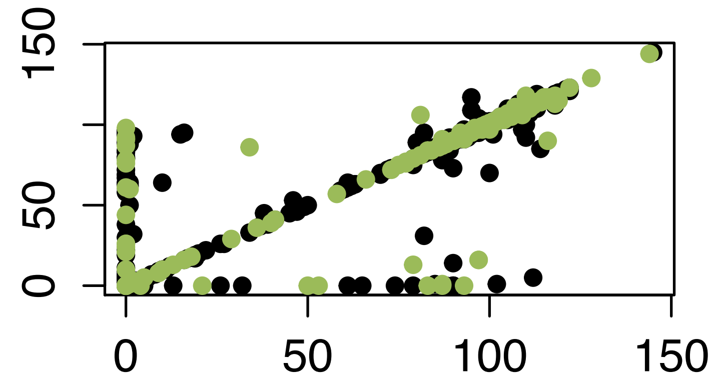


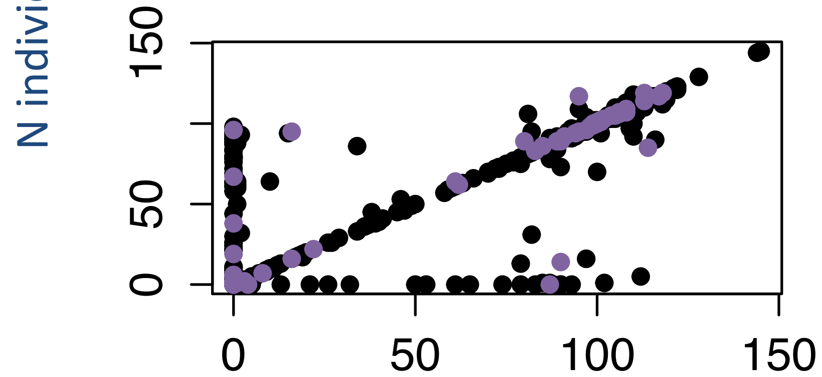


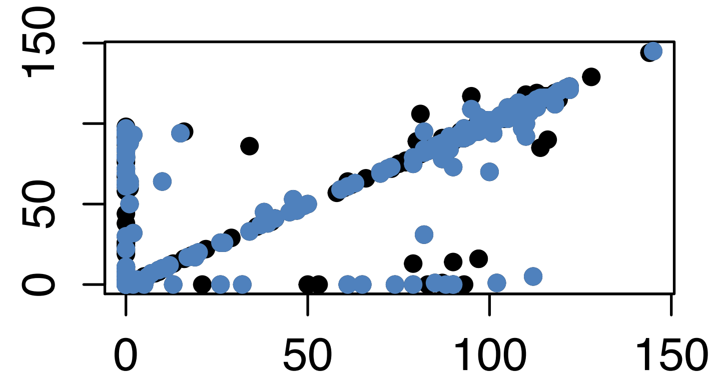


IV. Analyses bio-informatiques









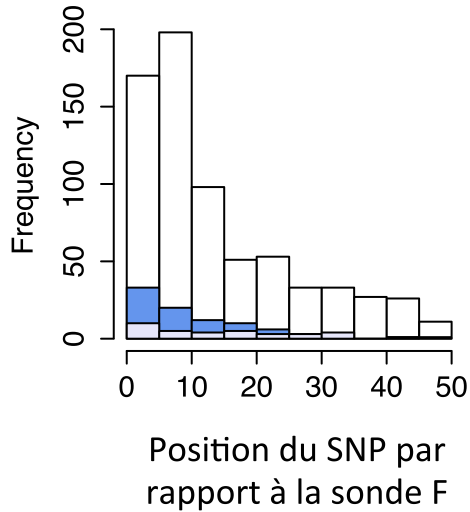
N individus (R - >5X)

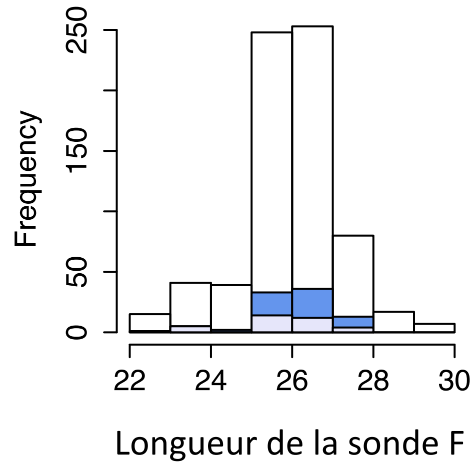
N individus (F - >5X)

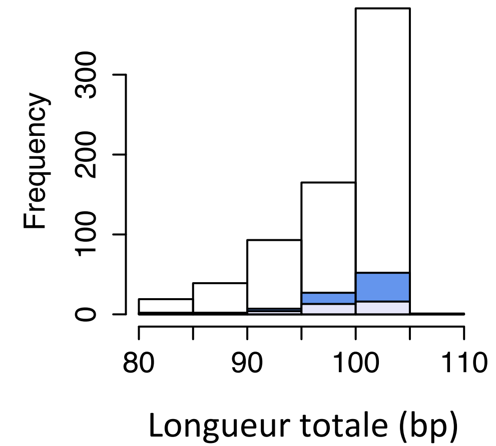
IV. Analyses bio-informatiques

Ceux qui n'ont pas marché

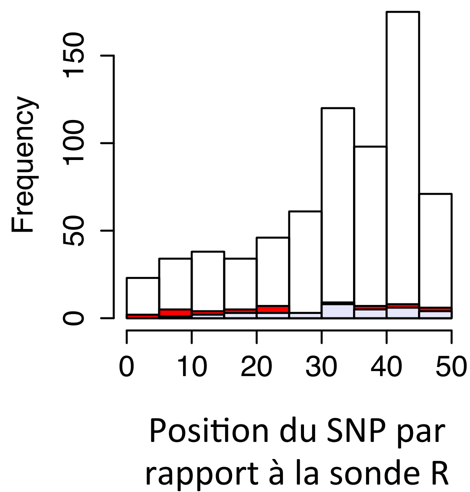
foward

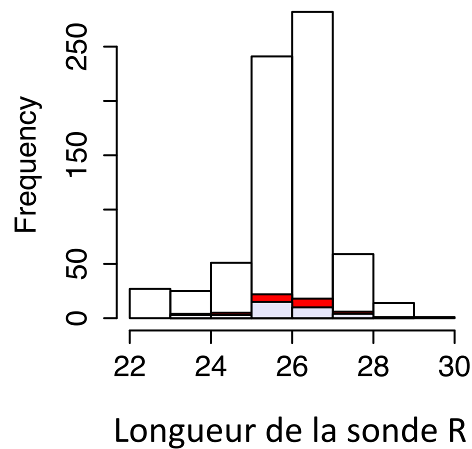


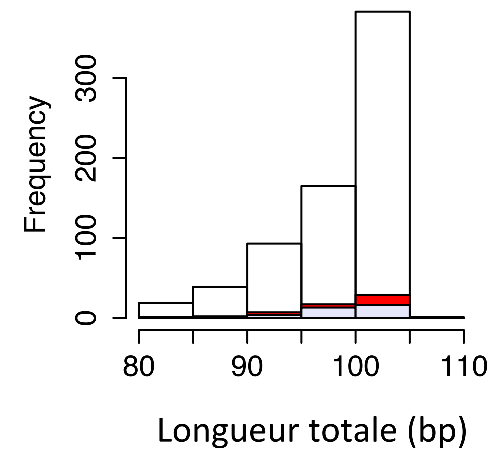




reverse



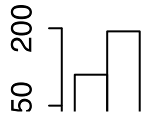


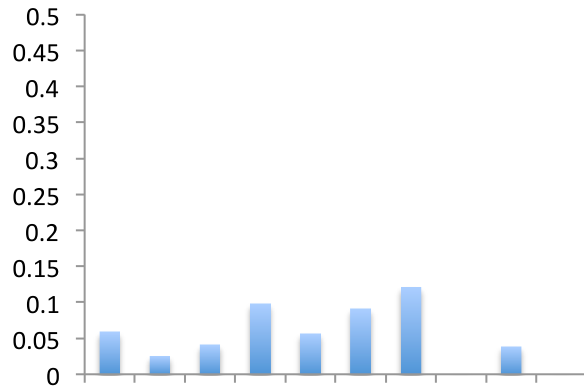


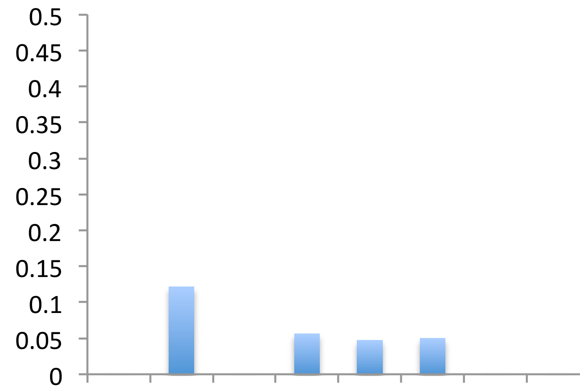
IV. Analyses bio-informatiques

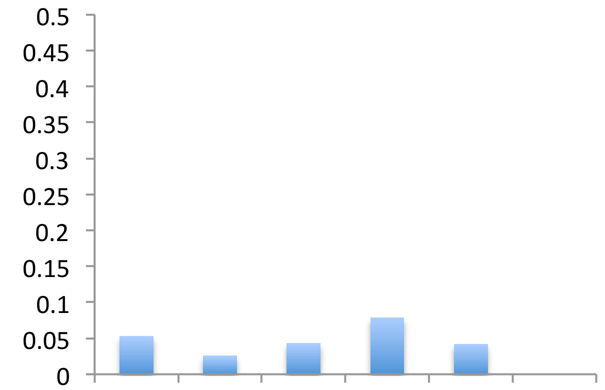
Ceux qui n'ont pas marché

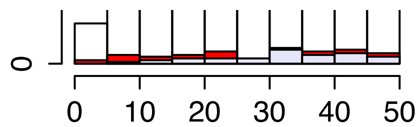
foward



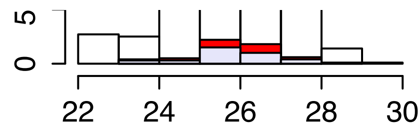




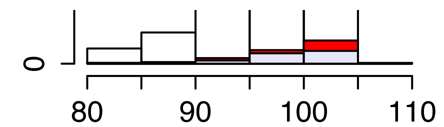




Position du SNP par rapport à la sonde R

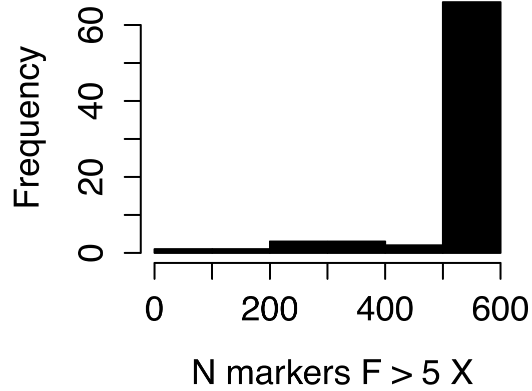


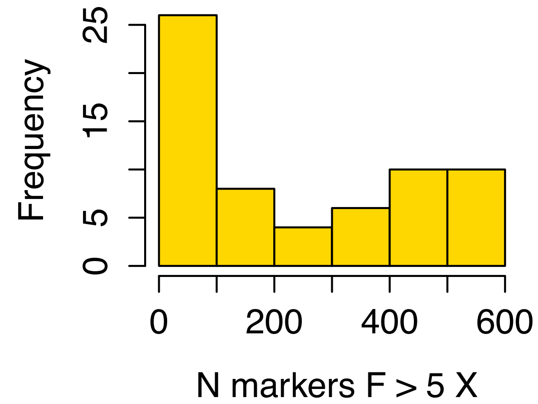
Longueur de la sonde R

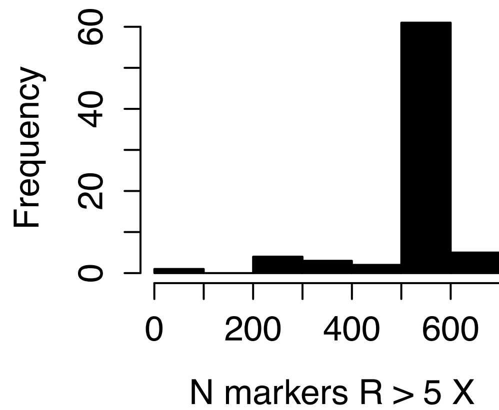


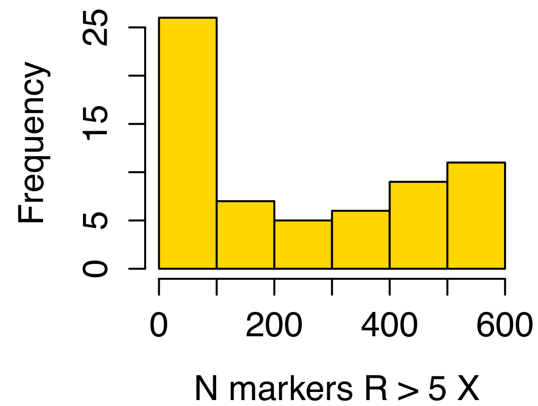
Longueur totale (bp)

IV. Analyses bio-informatiques



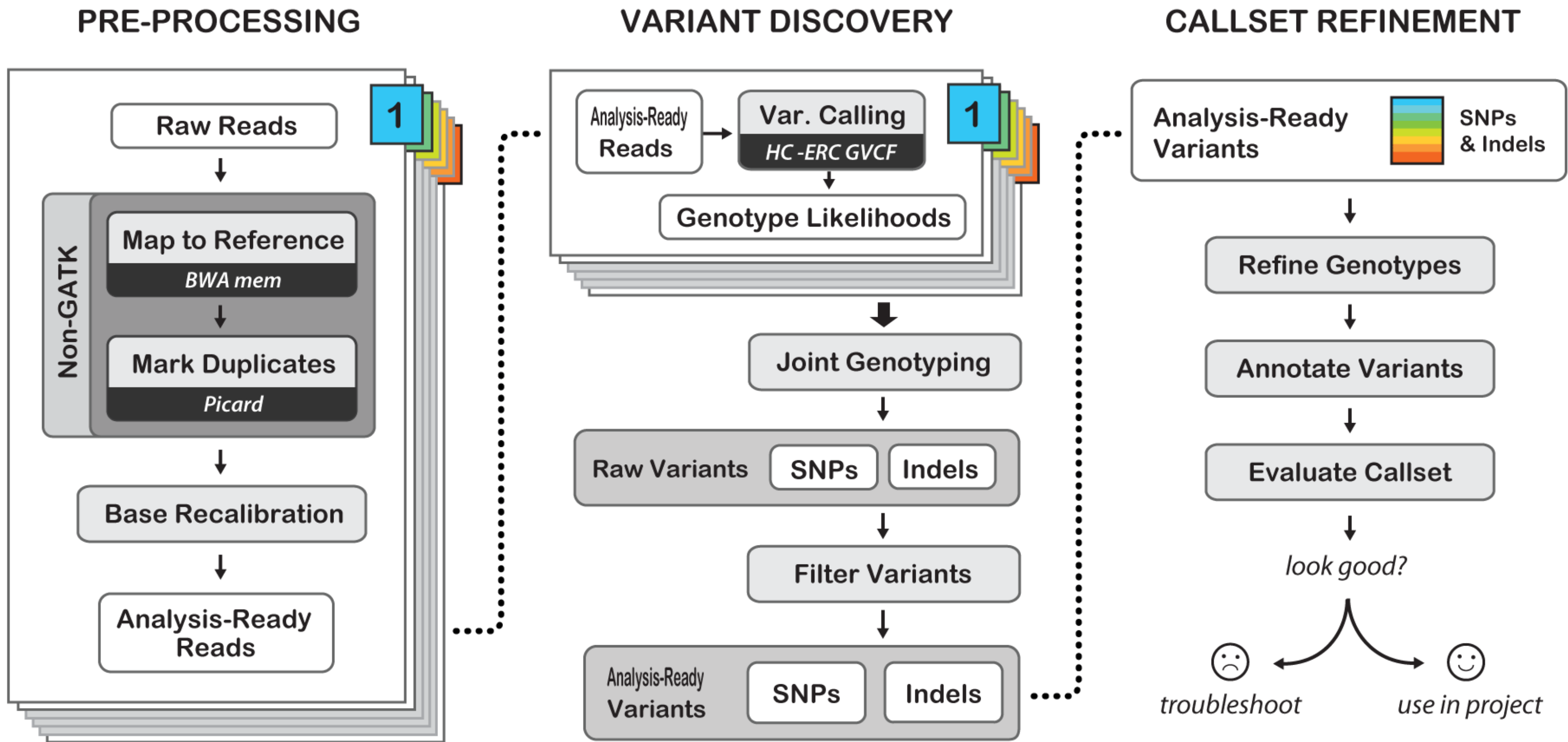






Ind: contemporains
sec

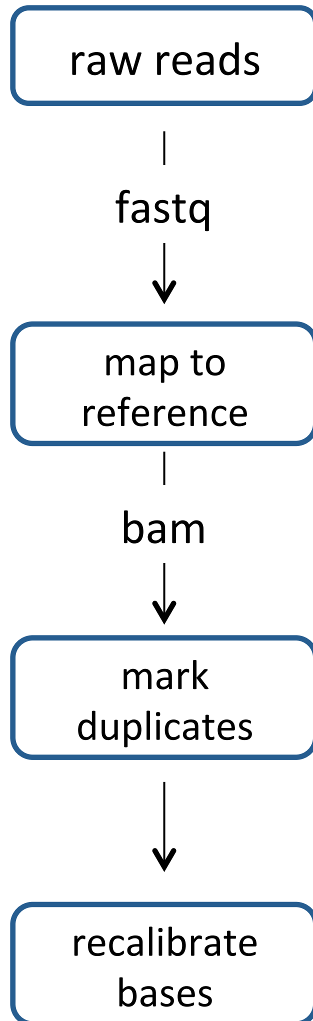
IV. Analyses bio-informatiques



Best Practices for Germline SNPs and Indels in Whole Genomes and Exomes - June 2016

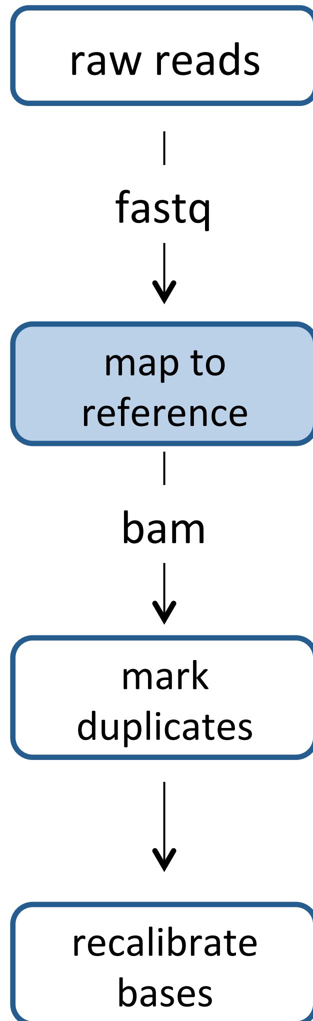
IV. Analyses bio-informatiques

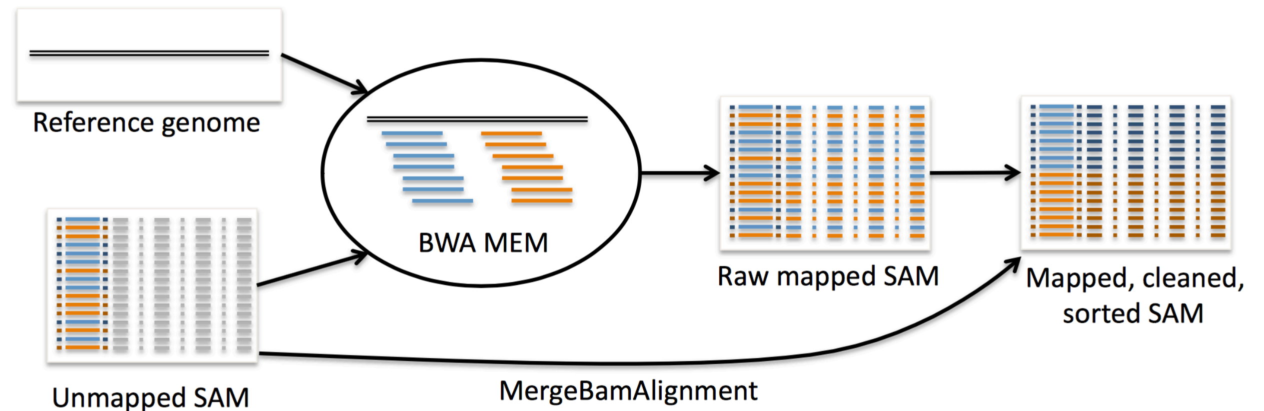
PRE-PROCESSING



IV. Analyses bio-informatiques

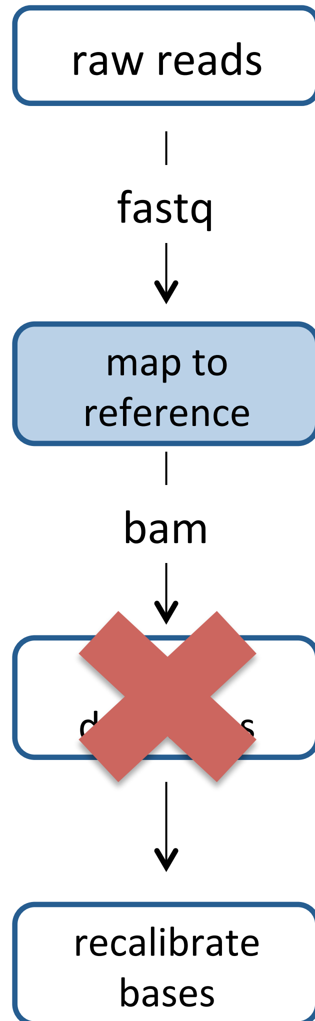
PRE-PROCESSING

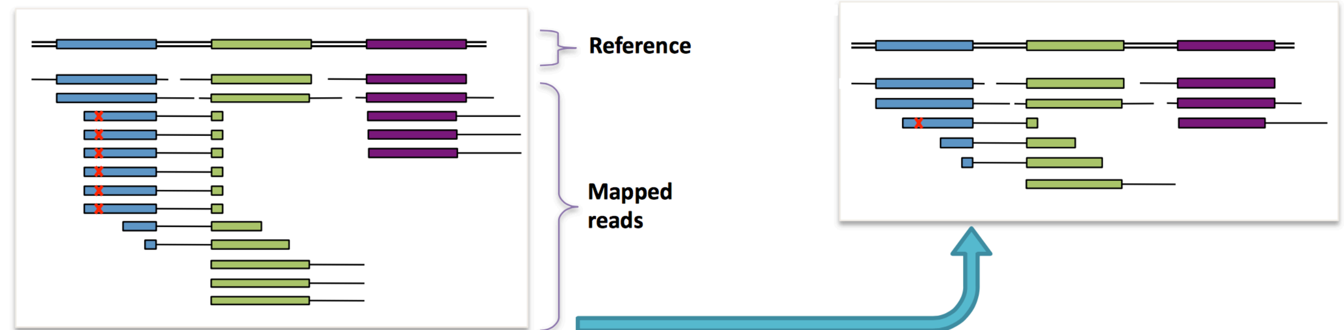


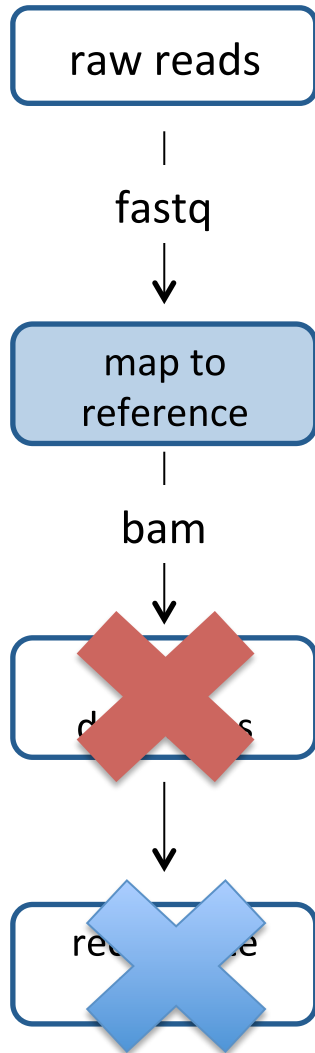


IV. Analyses bio-informatiques

PRE-PROCESSING







"Quality scores emitted by sequencing machine are biased and inacurate"

Recalibrate base quality scores in order to correct sequencing errors and other experimental artifacts.

BaseRecalibrator (1)

PrintReads

BaseRecalibrator (2)

AnalyzeCovariates

PROCESSED DATA

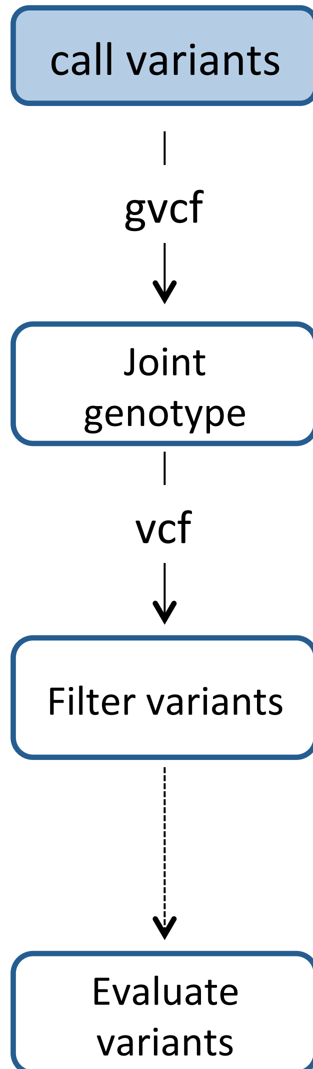
BEFORE

AFTER

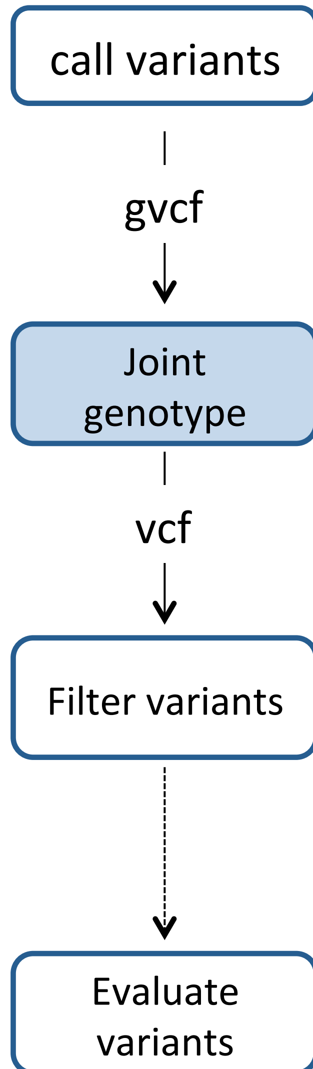
PLOTS

IV. Analyses bio-informatiques

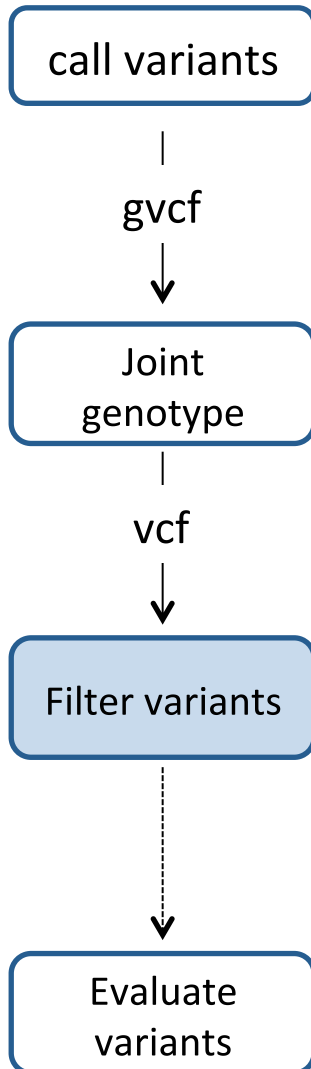
VARIANT DISCOVERY



VARIANT DISCOVERY

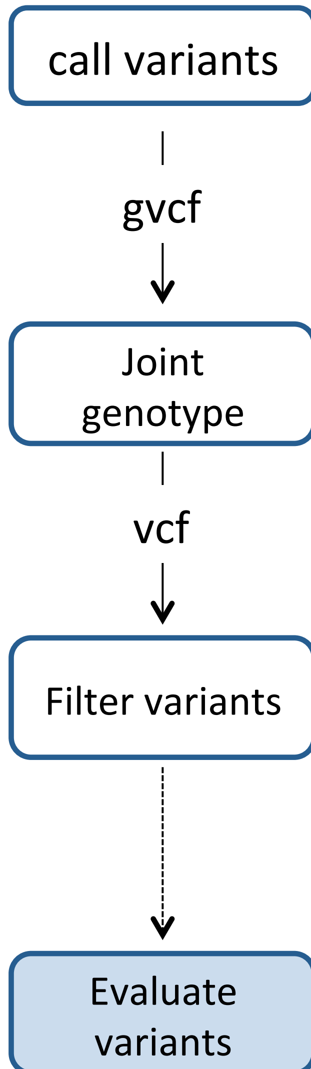


VARIANT DISCOVERY



IV. Analyses bio-informatiques

VARIANT DISCOVERY



IV. Analyses bio-informatiques

VARIANT DISCOVERY

call variants

gvcf

Joint genotype

vcf

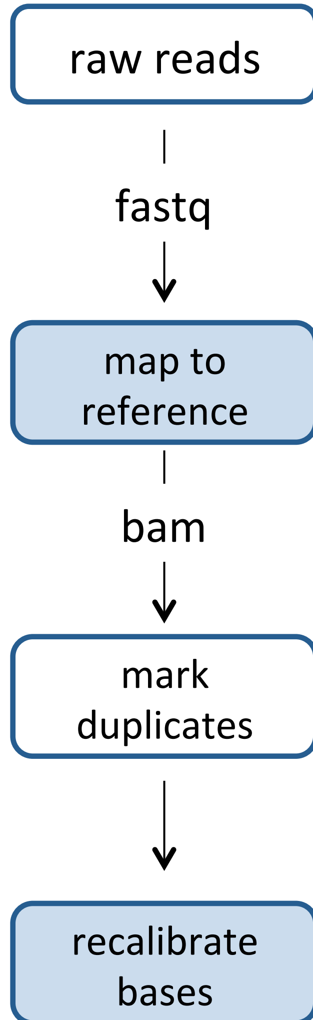
Filter variants

Evaluate variants

734 SNPs (645 séquences différentes)

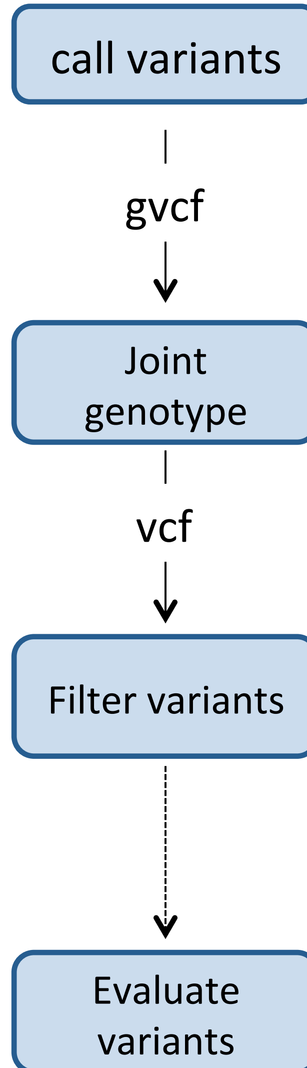
IV. Analyses bio-informatiques

PRE-PROCESSING



db: 734 SNPs

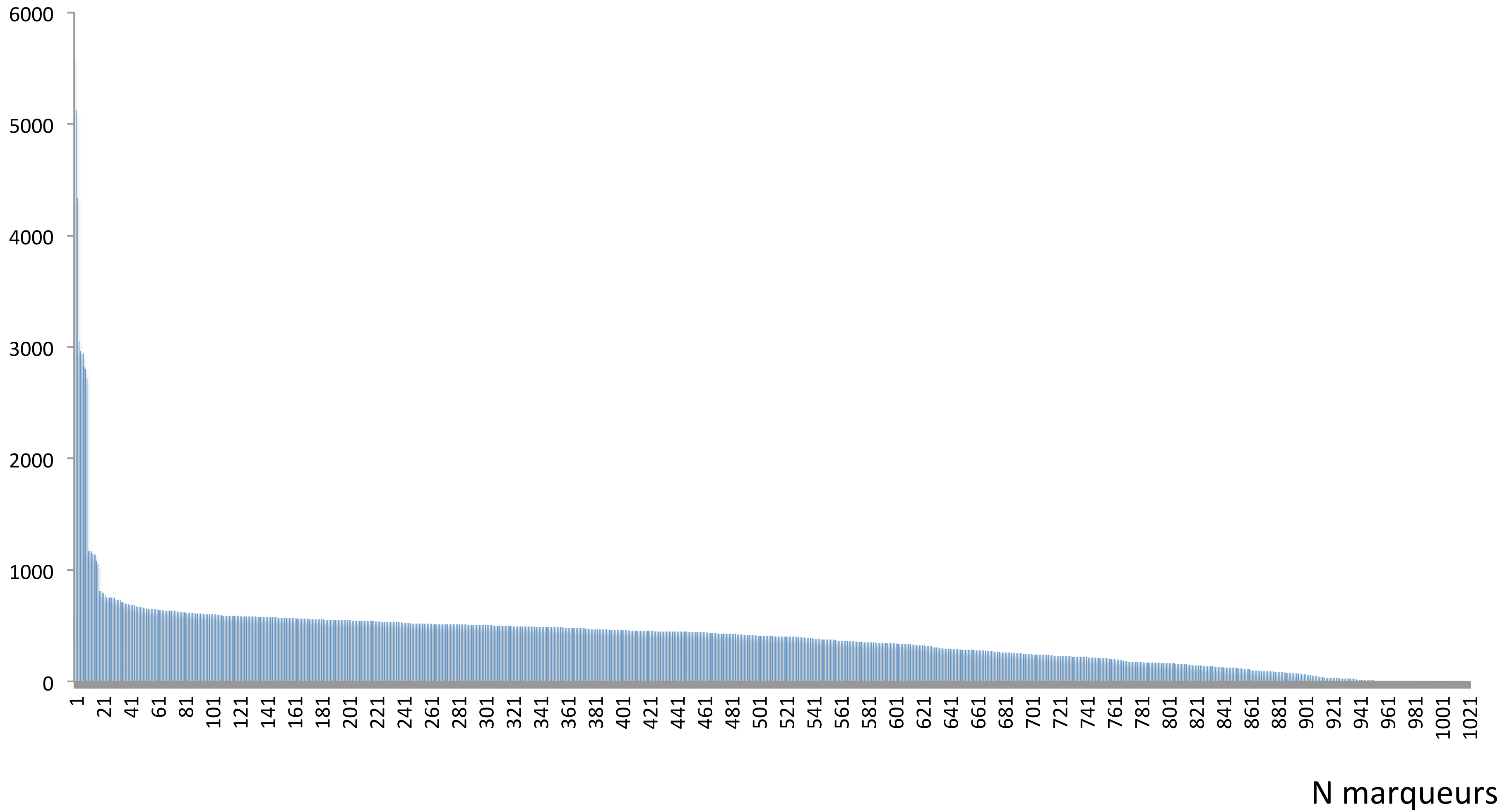
VARIANT DISCOVERY



1204 sites:
- 181 indels
- 1022 SNPs

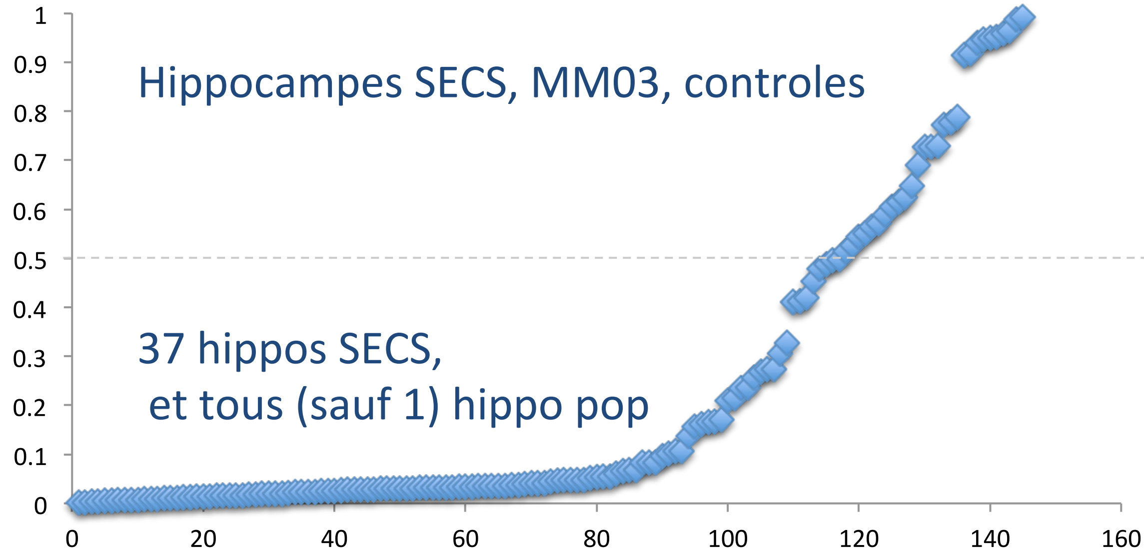
596 SNPs parmi les 890 connus dont 517 parmi les 700 ciblés

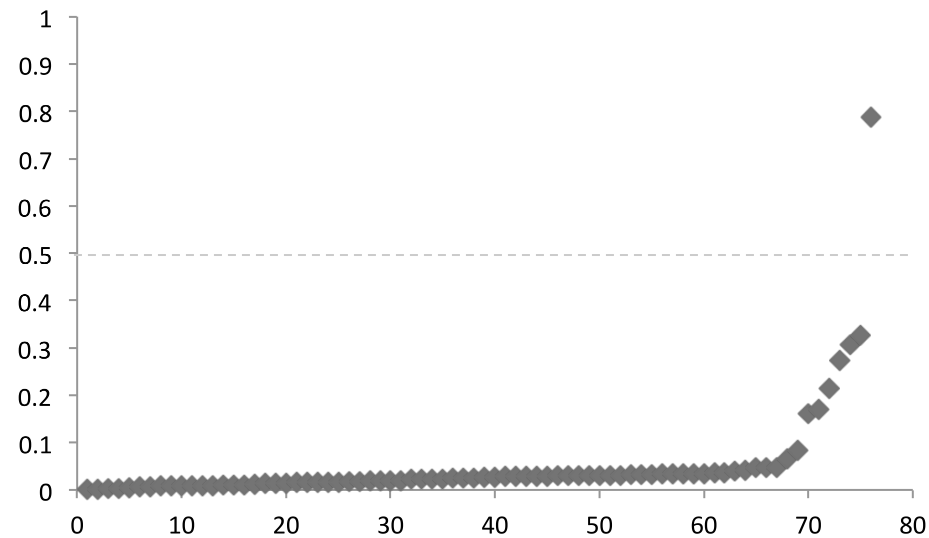
IV. Analyses bio-informatiques

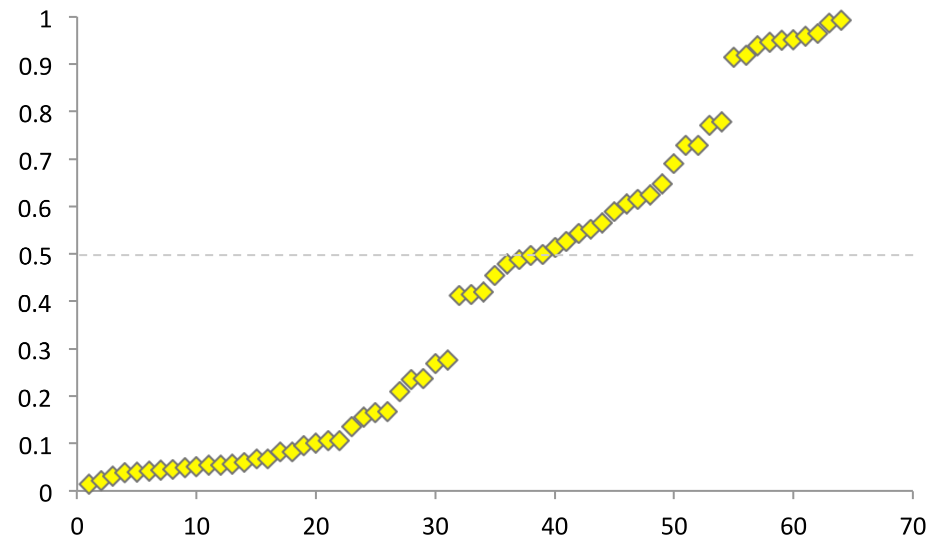


IV. Analyses bio-informatiques

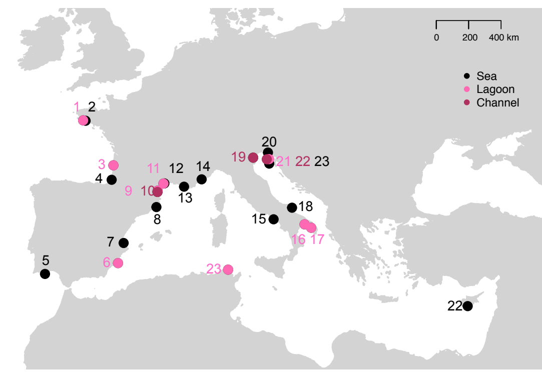
% missing data







IV. Analyses bio-informatiques

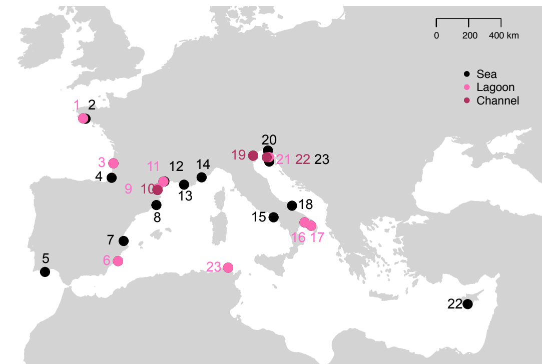


Profondeur

1000
900
800
700
600
500
400
300
200
100
0

individus

IV. Analyses bio-informatiques



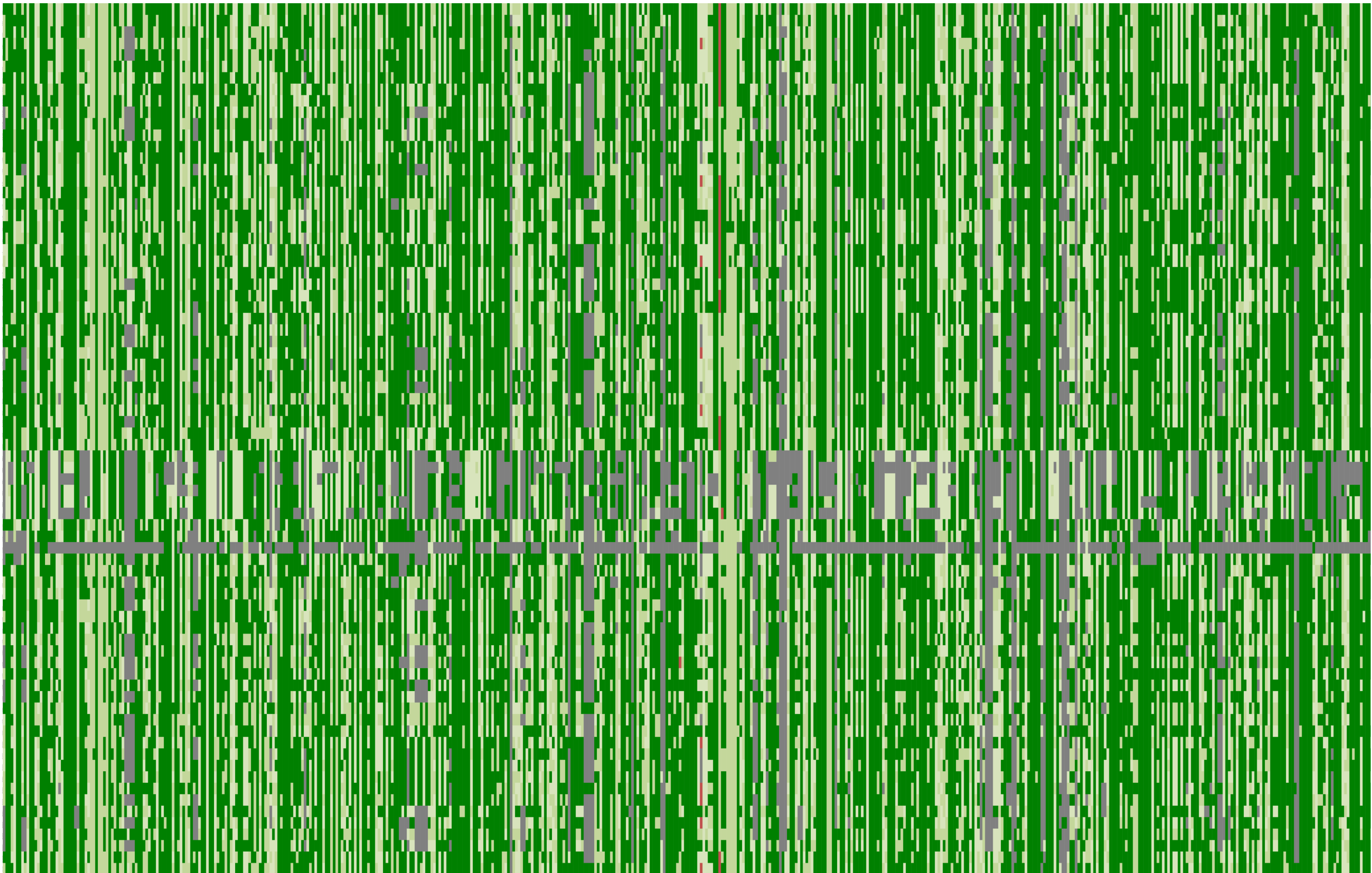
Profondeur

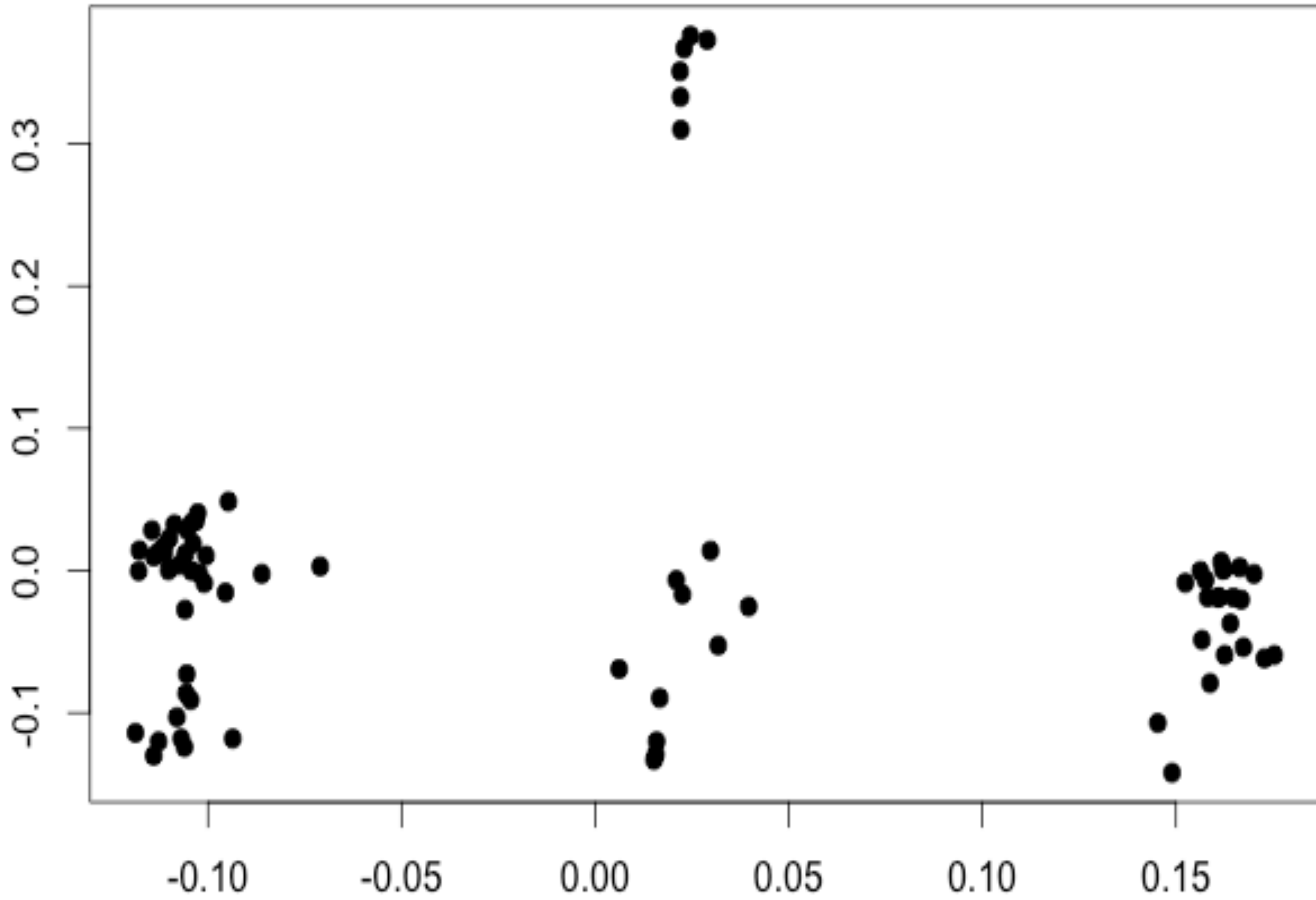
1000
900
800
700
600
500
400
300
200
100
0

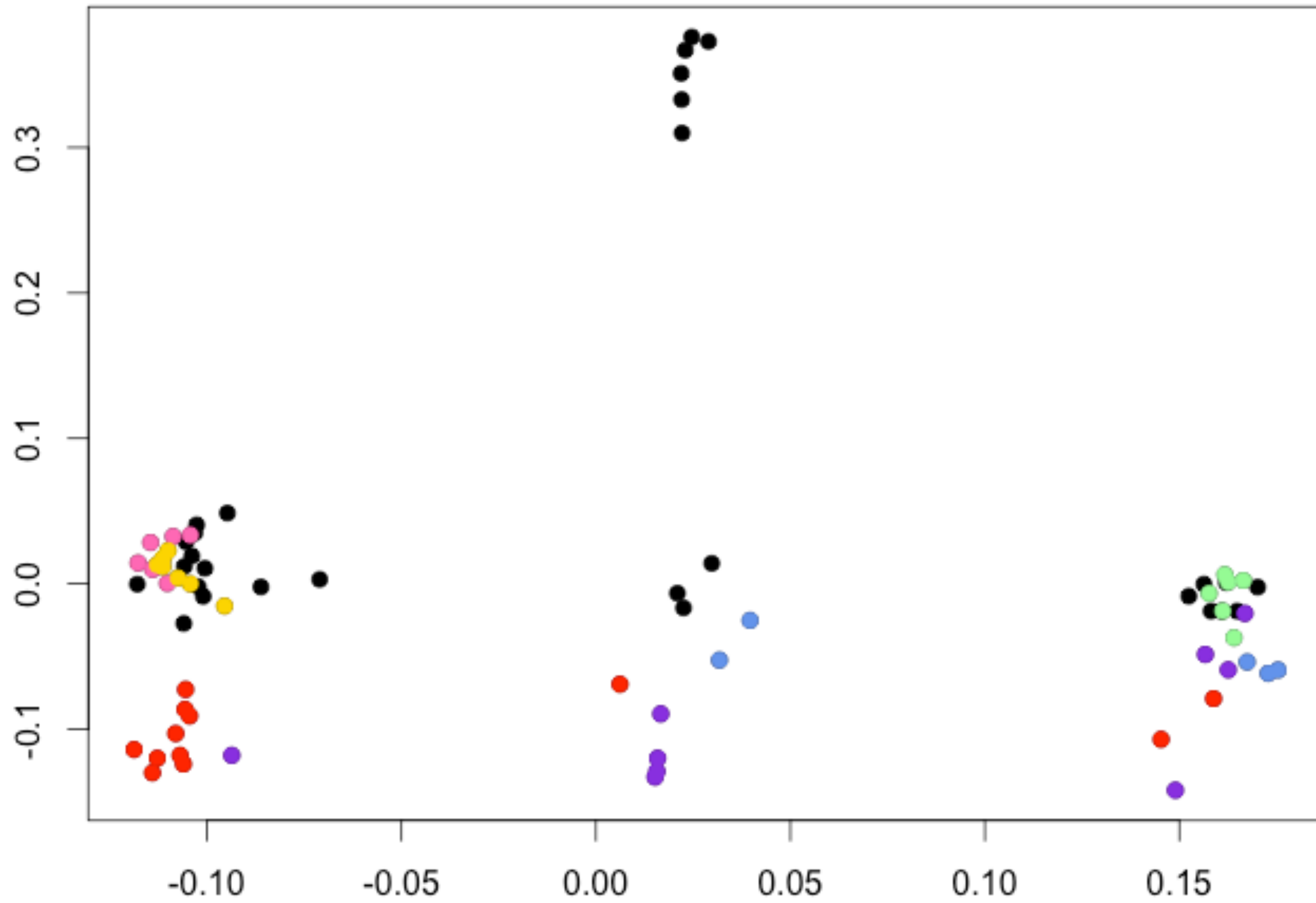
Extraction en 2016 sauf: 2012 2014 2014 * 2014 * 2014 *

*: Manip' BX – (RAD) - TSCA

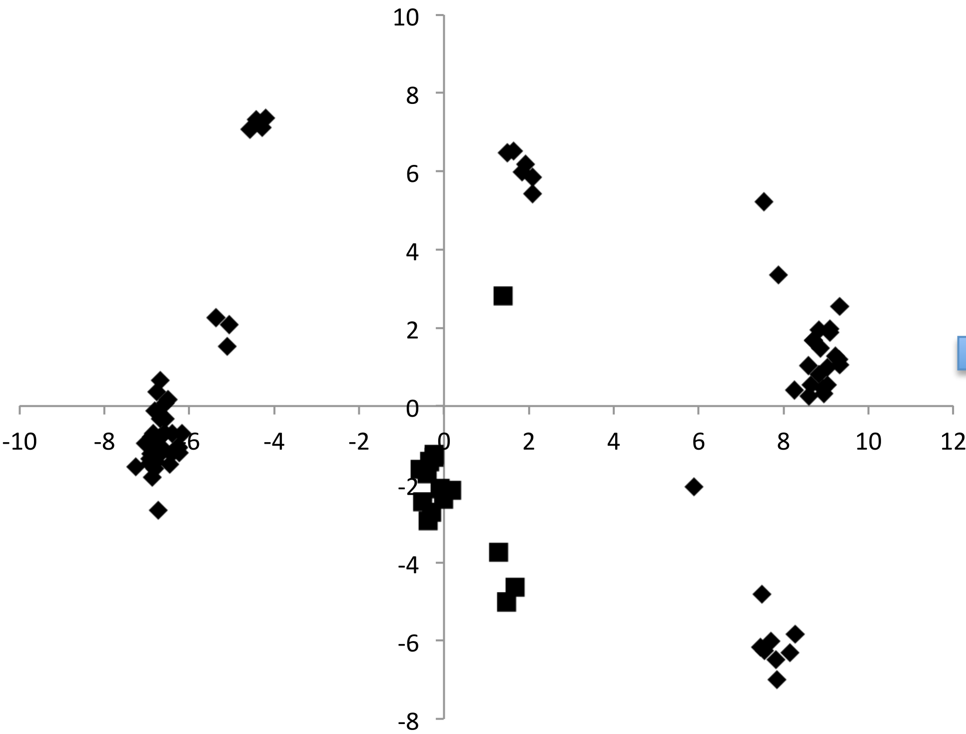
IV. Analyses bio-informatiques

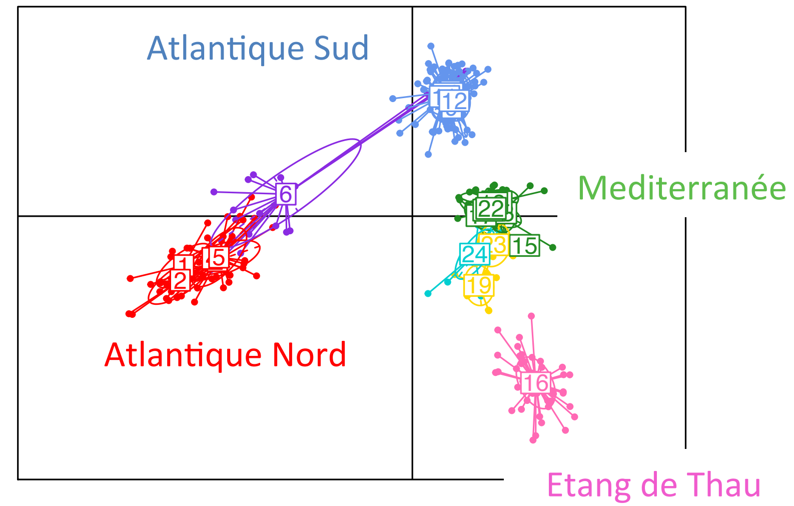



0.3
0.2
0.1
0.0
-0.1
-0.10
-0.05
0.00
0.05
0.10
0.15


0.3
0.2
0.1
0.0
-0.1
-0.10
-0.05
0.00
0.05
0.10
0.15

IV. Analyses bio-informatiques



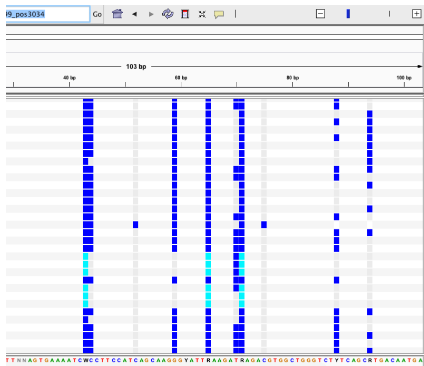


1ère ACP BeadXpress

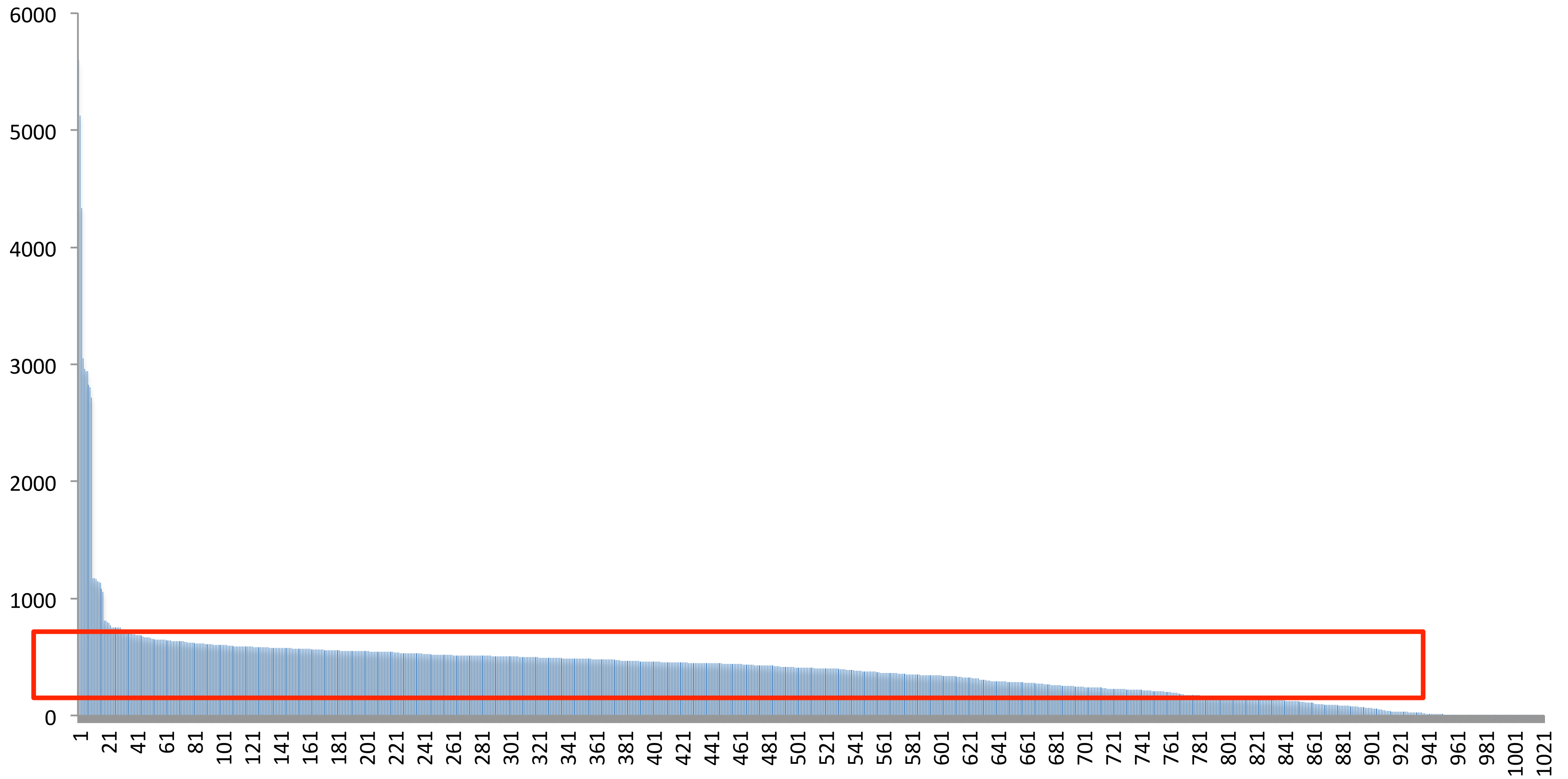
IV. Analyses bio-informatiques

Filtres:

- Bi allélisme **1012**
- Filtre sur qualité : PASS **964**
- Données manquantes 5% **958**
- Monomorphisme **775**
- Hétérozygotie identique **755**
- Fréquences alléliques maximale et minimale **589**
- Un SNP/séquence **455**
- Couverture: 200 – 700 .... **350**



IV. Analyses bio-informatiques



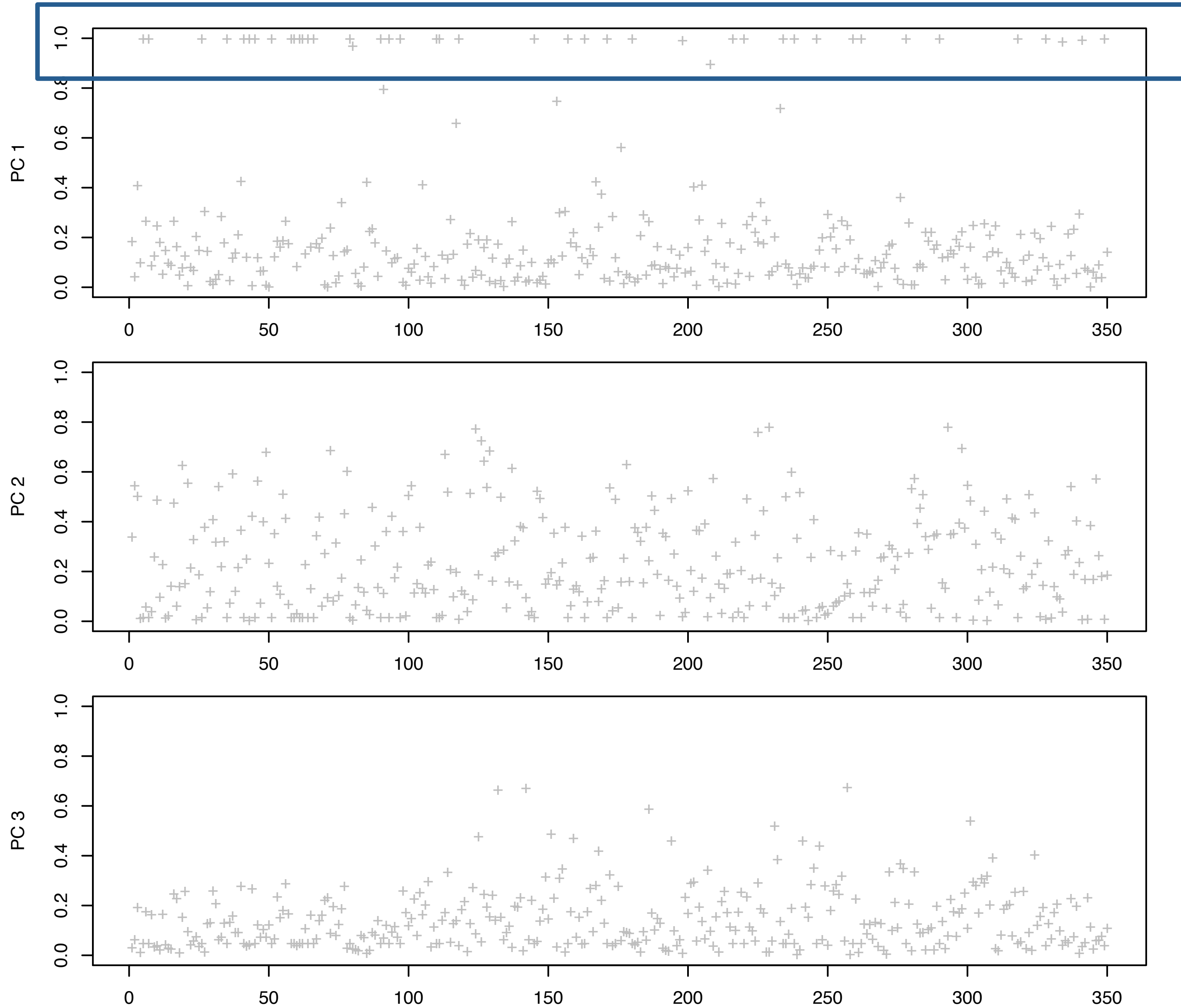
IV. Analyses bio-informatiques

Filtres:

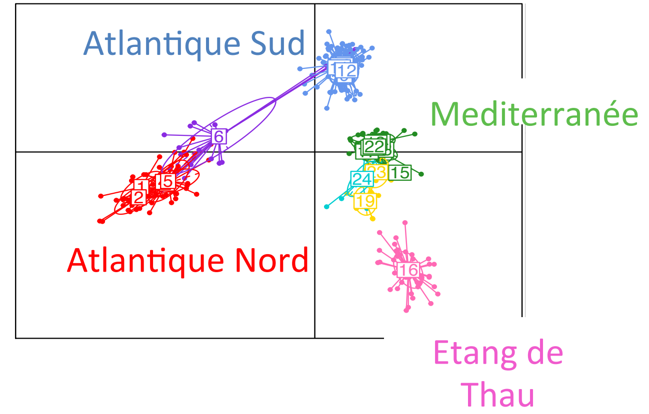
- Bi allélisme **1012**
- Filtre sur qualité : PASS **964**
- Données manquantes 5% **958**
- Monomorphisme **775**
- Hétérozygotie identique **755**
- Fréquences alléliques maximale et minimale **589**
- Un SNP/séquence **455**
- Couverture: 200 – 700 .... **350**
- Contribution aux axes (ACP) **307**

IV. Analyses bio-informatiques

Filtres:



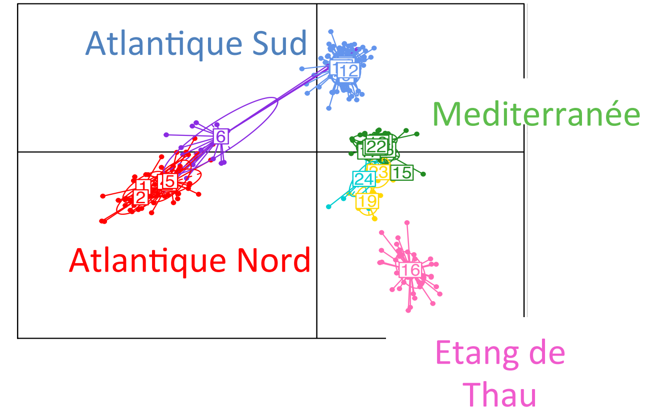
IV. Analyses bio-informatiques

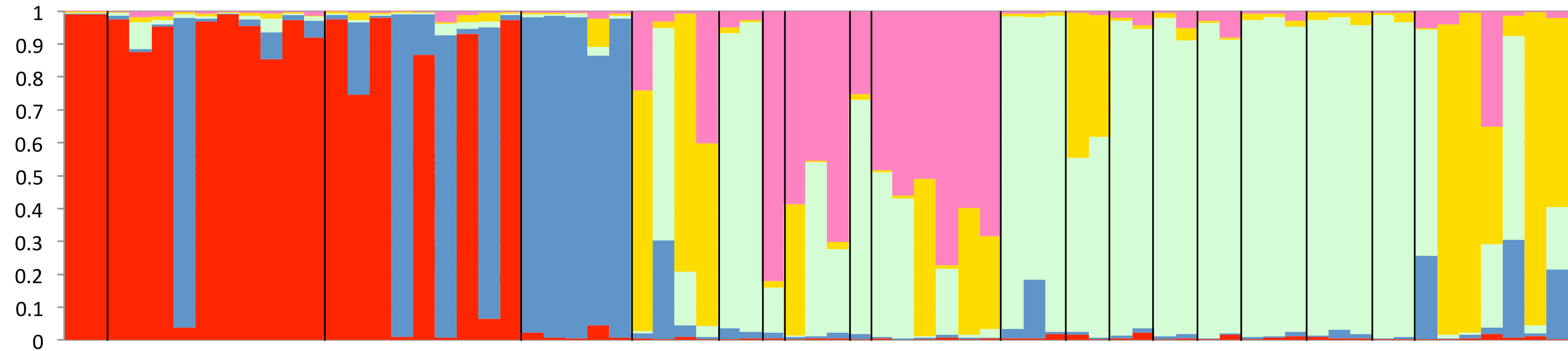


0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3

-0.3 -0.25 -0.2 -0.15 -0.1 -0.05 0 0.05 0.1 0.15 0.2

IV. Analyses bio-informatiques





isem
Institut des Sciences de l'Evolution-Montpellier

cnrs

IRD
Institut de recherche pour le développement

École Pratique des Hautes Études

cirad