

Séminaire présenté au CBGP le 27 mai 2014

Tetranychus urticae : Gènes responsables de la capacité d'adaptation à la plante hôte ?

J.-E. Longueville

P. Auger, S. Fellous, A. Migeon, M. Navajas, L. Sauné

en collaboration avec M. Gauthier et R. Vitalis

- 1 Introduction

- 2 Explications du pipeline d'analyse des données FASTQ → YY NN

- Snp Calling

- 3 Analyses et Résultats

- ACP
- Écriture des fichiers Kimtree
- Selestim

- 4 Conclusion

Introduction

Cadre de travail

ANR GENEVOLSPE 2011-2014 : GÉNétique et ÉVOLution de la SPÉcialisation

chez des acariens phytophages ou prédateurs. dirigé par Isabelle Olivieri et Maria Navajas

Question :

Comment la diversité génétique permet aux acariens de contrecarrer les défenses des plantes ? quels

sont les gènes en causes ?

Méthode choisie

Recherche de marqueurs sous sélection dans le cadre d'une expérience de sélection contrôlée.

TASK 1. CONSEQUENCES OF HOST-PLANT USE FOR NATURALLY OCCURRING GENETIC DIVERGENCE AMONG POPULATIONS OF SPIDER MITES

C
A
S
Is

Q
C

Task leader: CBGP, M. Navajas

d

Participants involved: Gersande Angot , Philippe Auger, Simon Fellous , Alain Migeon and Laure

M

Sauné

Rec
s é l

Introduction

Cadre de travail

ANR GENEVOLSPE 2011-2014 : GÉNétique et ÉVOLution de la SPÉcialisation

chez des acariens phytophages ou prédateurs. dirigé par Isabelle Olivieri et Maria Navajas

Question :

Comment la diversité génétique permet aux acariens de contrecarrer les défenses des plantes ? quels sont les gènes en causes ?

Méthode choisie

Recherche de marqueurs sous sélection dans le cadre d'une expérience de sélection contrôlée.

Introduction

Expérience de départ

Cette analyse fait suite d'une expérience de sélection contrôlée (Fellous et al. 2014)

Au départ 3 populations de terrain :

Citrus



^a

^a. Illustration contributed by Natural History Museum, London, U.K.

Pool

Un ensemble de reproductions croisées et contrôlées ont permis de produire une population pool (regroupant les trois

génomomes initiaux)

Introduction

Expérience de départ

Cette analyse fait suite à l'expérience de sélection contrôlée :

Tomate



a. Blackwell, E., A curious herbal, vol. 1 : t. 133 (1737) [E. Blackwell]

Pool

Un ensemble de reproductions croisées et contrôlées ont permis de produire une population pool (regroupant les trois

génomés initiaux)

Introduction

Expérience de départ

Cette analyse fait suite à l'expérience de sélection contrôlée

Nerium oleander



a

a. Botanical Register, vol. 1 : t. 74 (1815) [S. Edwards]

Introduction

Expérience de départ

Cette analyse fait suite à l'expérience de sélection contrôlée menée par Fellous et al. 2014.

Au départ 3 populations de terrain :

Citrus

Tomates

Nerium

Pool

Un ensemble de reproductions croisées et contrôlées ont permis de produire une population pool (regroupant les trois génomes initiaux).

Introduction

Resélection

Le pool est ensuite redivisé en plusieurs réplicats sur les plantes hôtes d'origine. Après 4 à 8 générations nous obtenons les populations suivantes :

Citrus (1 pop)

Nerium (2 pop)

Tomate (4 pop)

Séquençage

Ces populations sont ensuite séquencées en poolseq à l'aide d'un séquenceur Illumina, avec une profondeur de 50X

Introduction

Resélection

Le pool est ensuite redivisé en plusieurs réplicats sur les plantes hôtes d'origine. Après 4 à 8 générations nous obtenons les populations suivantes :

Citrus (1 pop)

Nerium (2 pop)

Tomate (4 pop)

Séquençage

Ces populations sont ensuite séquencées en poolseq à l'aide d'un séquenceur Illumina, avec une profondeur de 50X

Lab tests on
host-plant
acceptance/fitness
s



Host-associated
adaptation leading to
fitness trade-offs

Experimental
design
of
Cross
tests
Replicates
and
Controls



**LHT
measures**

fertility
juvenile fertility

LHT measures

T. urticae whole genome sequence project

Mity model: *Tetranychus urticae*, a candidate for chelicerate model organism

Miodrag Grbic,^{1*} Abderrahman Khila,¹ Kwang-Zin Lee,¹ Anica Bjelica,¹ Vojislava Grbic,¹ Jay Whistlecraft,³ Lou Verdon,³ Maria Navajas,⁴ and Lisa Nagy²

BioEssays 29:489–496, © 2007

The screenshot shows the JGI-DOE website interface. At the top left is the JGI logo with the text "DOE JOINT GENOME INSTITUTE AN DEPARTMENT OF ENERGY OFFICE OF SCIENCE". To the right are navigation links: "search jgi", "contact us", "site map", and "internal". A search bar contains the number "60". Below this is a horizontal menu with items: "HOME", "ABOUT US", "PROGRAMS", "SEQUENCING" (highlighted with an orange circle), "JGI SCIENCE", "MEETINGS", "NEWS", "EDUCATION", and "EMPLOYMENT". On the left side, a vertical navigation menu titled "sequencing" lists several options: "Sequencing Strategy", "Sequencing Plans and Progress", "Statistics", "Why Sequence Them?" (highlighted with an orange circle), "Information for Collaborators", "Protocols", and "Genome Portal Site". The main content area features the heading "Why Sequence the Two-Spotted Spider Mite?". To the left of the text is a photograph of a two-spotted spider mite. The text explains that *Tetranychus urticae* belongs to the second largest group of animals, the Chelicerata, and lists its scientific importance in areas like phylogenetics, developmental biology, evolution, ecology, and genomics. It also notes that spider mites and ticks are economically significant chelicerates and that spider mites are major agricultural pests while ticks are disease vectors.

sequencing

- Sequencing Strategy
- Sequencing Plans and Progress
- Statistics
- Why Sequence Them?**
- Information for Collaborators
- Protocols
- Genome Portal Site

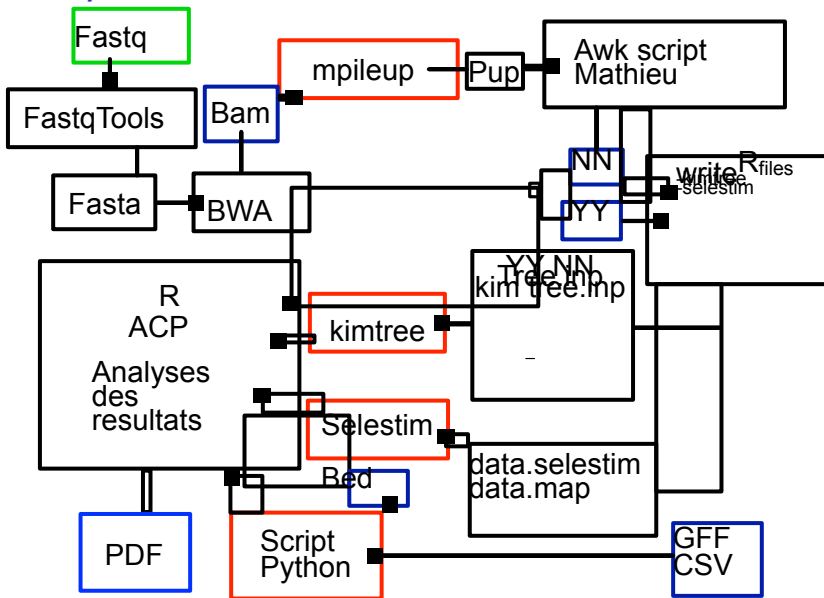
Why Sequence the Two-Spotted Spider Mite?



The two-spotted spider mite *Tetranychus urticae* belongs to the second largest group of animals, the Chelicerata, which includes spiders, scorpions, mites, and ticks. As representatives of this basal taxon of arthropods, spider mites are of special importance to several areas of science, including phylogenetics, developmental biology, evolution, ecology, and genomics. The most economically significant chelicerates are spider mites and ticks, which belong to the order Acari. Spider mites represent major pests in agriculture, while ticks are vectors of human diseases,

Genome sequenced
by JGI-DOE
(<http://www.jgi.doe.gov/sequencing/why/50028.html>)

Analyses



Des FastQ aux YY et NN

Le FastQ

Format de fichier brut, issu directement du séquenceur dans notre cas un Illumina

FastQ 4 lignes par lectures.

```
@HWI-ST539.131.1.1101.1331.1946#19@0/1
NTTTTCACTCCTACTTTTACTTTTGGTATAAAAAGGCTCATTTC AATTTTATCTCAAAA
TCATTCACAATCAGTTCAAACATTATTCAACATCTAAAAAACCA
+
#4=DFDFHHHGHIIJJJJJJJJJEHHIIJJJJJJJJJJJJJJJJGHIJJJJJJHIIJJJJJIH
FHGJIIIIJGHHGGFHHFFF@BEEED?BB
```

Snpcalling travail réalisé par Mathieu Gauthier

Étape 1 : Élagage de qualité des séquences

On élimine les dernières bases des lectures qui ont une trop faible qualité.

Étape 2 : Alignement des lectures sur le Génome de référence

A l'aide de bowtie ou de Bwa nous alignons les lectures.

Étape 3 : Identification de SNP -

```
samtools mpile up -f $ref -d 5000 -B file.bam > out.pup
```


Snpcalling travail réalisé par Mathieu Gauthier

Étape 1 : Élagage de qualité des séquences

On élimine les dernières bases des lectures qui ont une trop faible qualité.

Étape 2 : Alignement des lectures sur le Génome de référence

A l'aide de bowtie ou de Bwa nous alignons les lectures.

Étape 3 : Identification de SNP -

```
samtools mpile up -f $ref -d 5000 -B file.bam > out.pup
```

SnpCalling travail réalisé par Mathieu Gauthier

Étape 1 : Élagage de qualité des séquences

On élimine les dernières bases des lectures qui ont une trop faible qualité.

Étape 2 : Alignement des lectures sur le Génome de référence

A l'aide de bowtie ou de Bwa nous alignons les lectures.

Étape 3 : Identification de SNP -

```
samtools mpile up -f $ref -d 5000 -B file.bam > out.pup
```

SnpCalling suite et fin

Étape 4 : Édition des fichiers YY et NN

Deux scripts awk qui permettent de ne garder que les SNPs bialléliques et organisent en deux tableaux :

allèle principal → YY toutes les lectures → NN

SNP

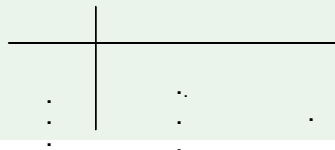
populations
nb lectures

...

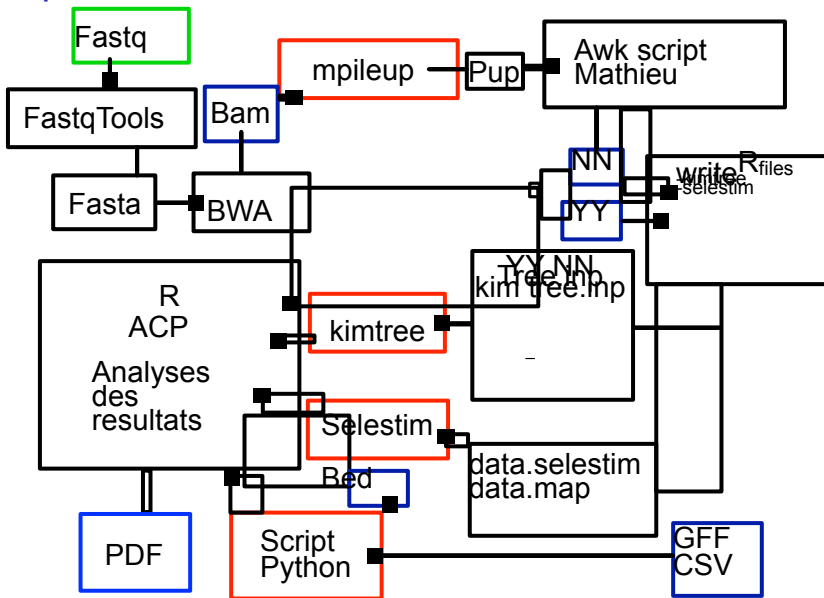
...

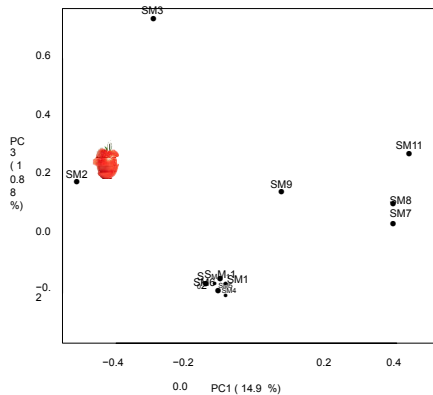
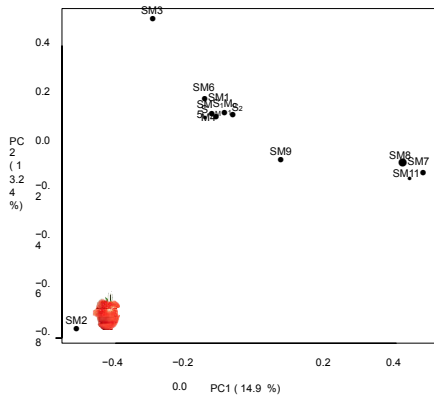
.

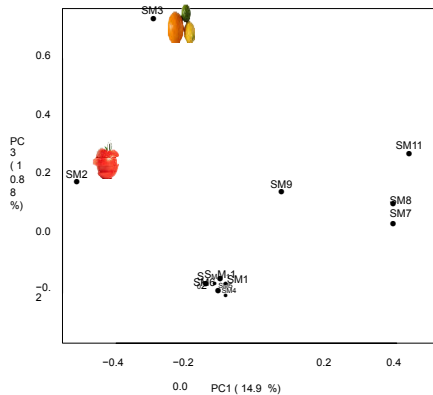
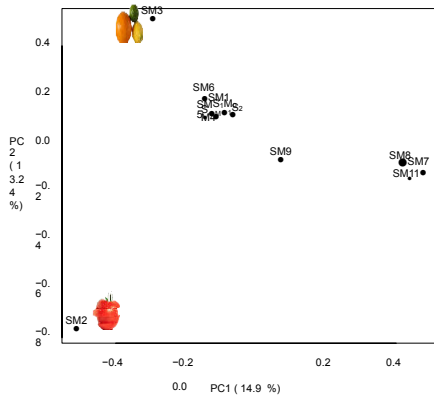
.

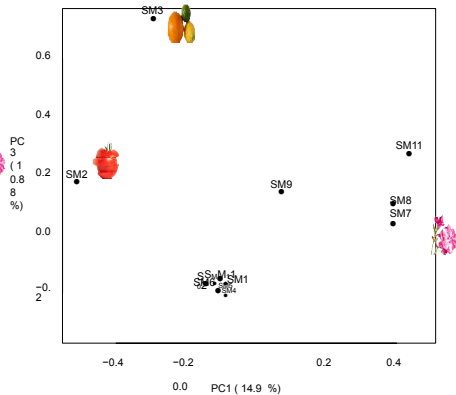
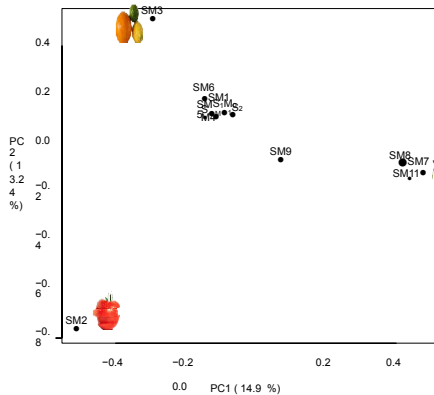


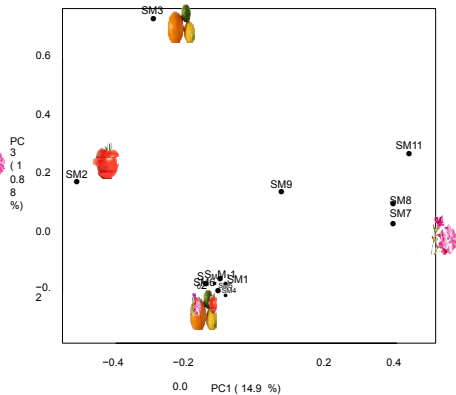
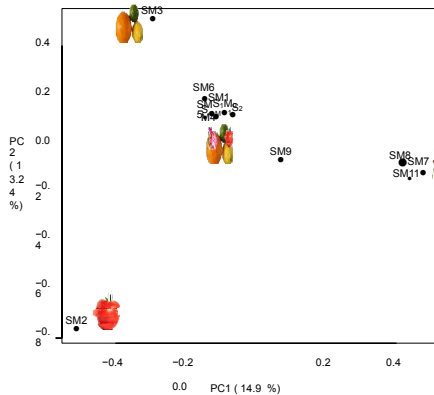
Pipeline







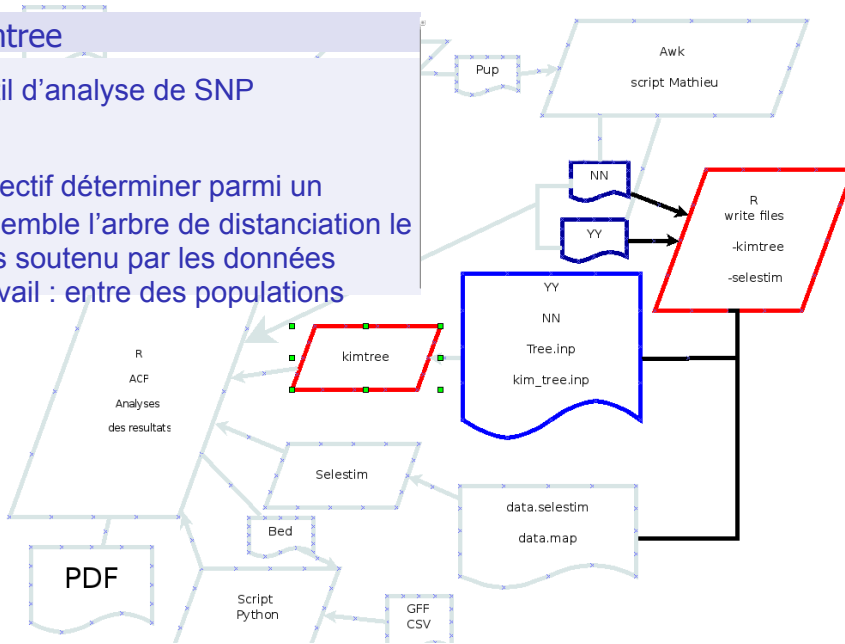




Kimtree

Outil d'analyse de SNP

Objectif déterminer parmi un ensemble l'arbre de distanciation le plus soutenu par les données
Travail : entre des populations

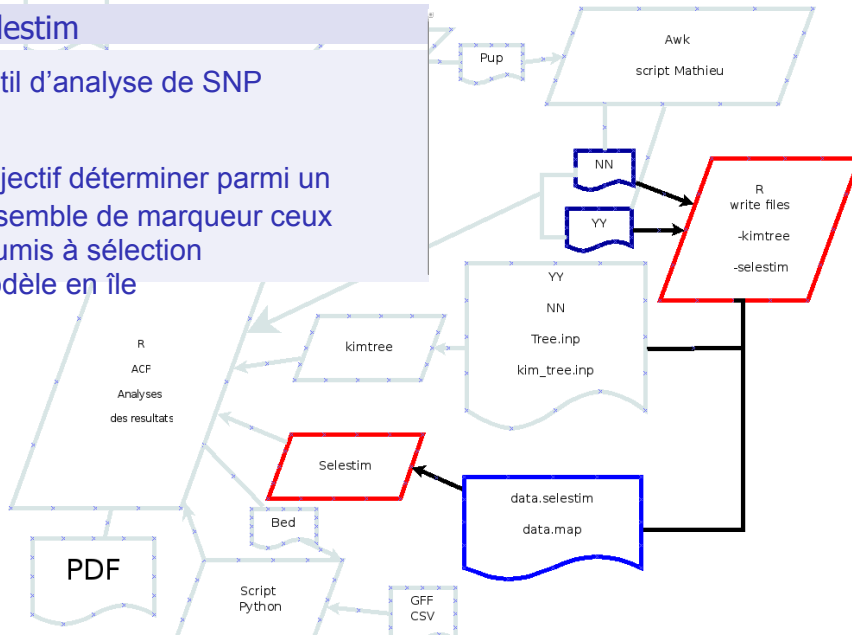


SeEstim

Selestim

Outil d'analyse de SNP

Objectif déterminer parmi un ensemble de marqueur ceux soumis à sélection
Modèle en île



Étape 1 :

Sélection des SNP d'intérêt

SNP ayant une couverture minimale de 40 lectures

SNP ayant une couverture maximale inférieure au quantile 95% SNP ne relevant pas d'une erreur de séquençage.

-

-

Écriture des fichiers d'entrée Kimtree et Selestim

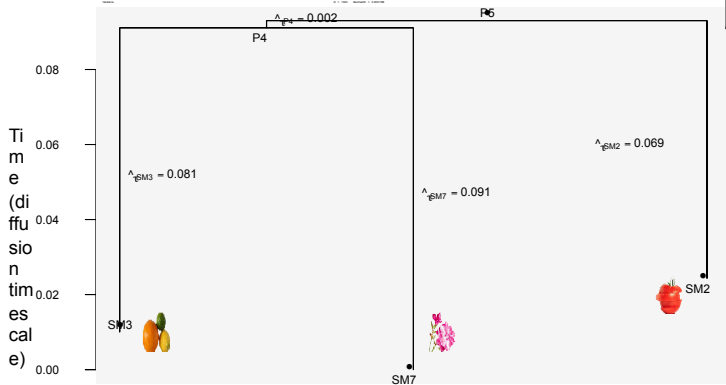
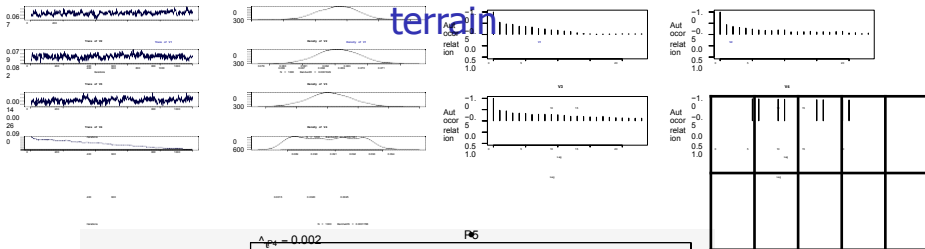
Étape 3 :

Écriture à proprement parlé des fichiers d'entrée de Selestim et de Kimtree

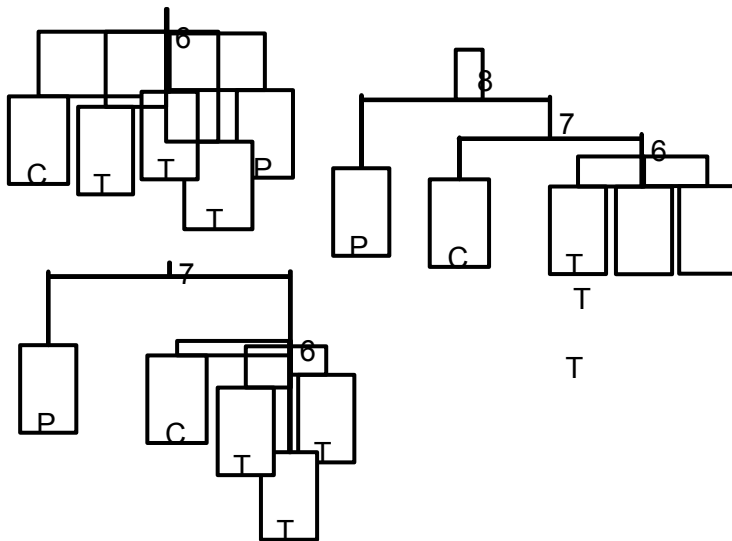
à l'aide de fonctions R développées par Mathieu

Puis lancement sur le cluster avec des scripts bash maison

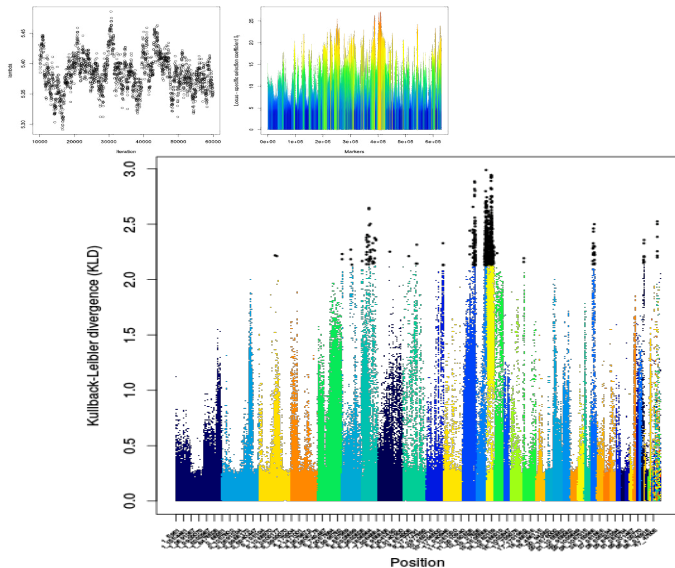
Analyse des résultats Kimtree sur les populations de terrain



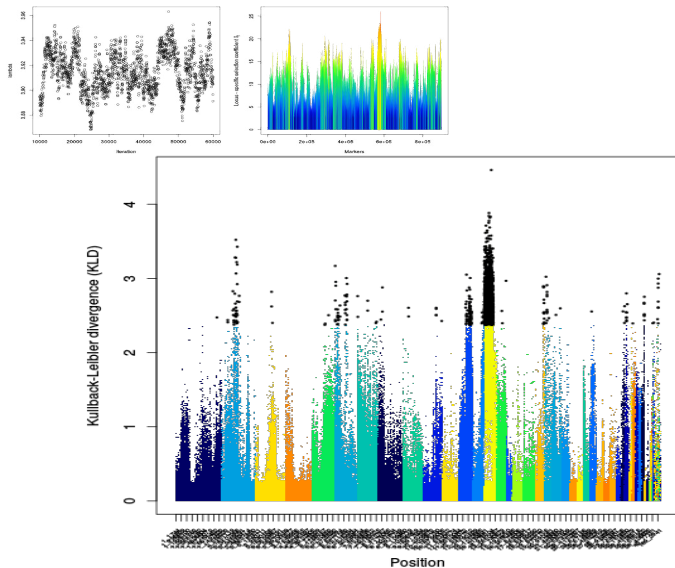
Arbres testés pour les populations reselectionnées



Résultats Selestim les populations reselectionnées sans le pool



Résultats Selestim les populations reselectionnées avec le pool



Nombre de marqueurs par scaffold

scaffold 2	38		
scaffold 3	3		
scaffold 5	3		
scaffold 6	36	scaffold 6	5
scaffold 7	6	scaffold 7	39
scaffold 8	3	scaffold 8	1
scaffold 9	2	scaffold 9	4
scaffold 10	5	scaffold 10	3
scaffold 12	78	scaffold	85
scaffold 13	65	scaffold 13	198
scaffold 14	1722	scaffold 14	335
scaffold 15	3	scaffold 15	1
scaffold 19	12	scaffold 18	2
scaffold 20	12		

Nombre de marqueurs par scaffold

Pool -	Nb Marqueurs	NoPool	Nb Marqueurs
scaffold_1	1	scaffold_3	2
scaffold_2	38	-	-
-	-	-	-
scaffold_3	3	-	-
-	-	-	-
scaffold_5	3	-	-
-	-	-	-
scaffold_6	36	scaffold_6	5
scaffold_7	6	scaffold_7	39
scaffold_8	3	scaffold_8	1
scaffold_9	2	scaffold_9	4
scaffold_10	5	scaffold_10	3
scaffold_12	78	scaffold_12	85
scaffold_13	65	scaffold_13	198

Quelques chiffres concernant les marqueurs

Nombres de snp sous sélection :

Populations re sélectionnées + Pool : 2022 Populations

resélectionnées - Pool : 710

Nombres de snp sous sélection dans une région dont la fonction est connue :

Populations re sélectionnées + Pool : 1142

Populations re sélectionnées - Pool : 440

Quelques chiffres concernant les fonctions associées

Nombres de fonction connues différentes sous
sélection :

Populations resélectionnées + Pool : 153

Populations resélectionnées - Pool : 142

Conclusion

- Mise en évidence de marqueur SNP candidats à la sélection
- Reste associer une protéine / famille à chaque marqueur
- Risque d'illusion graphique
- Trouver une méthode pour bien identifier la sélection d'un bruit.

Conclusion

- Mise en évidence de marqueur SNP candidats à la sélection
- Reste associer une protéine / famille à chaque marqueur
- Risque d'illusion graphique
- Trouver une méthode pour bien identifier la sélection d'un bruit.

Conclusion

- Mise en évidence de marqueur SNP candidats à la sélection
- Reste associer une protéine / famille à chaque marqueur
- **Risque d'illusion graphique**
- Trouver une méthode pour bien identifier la sélection d'un bruit.

Conclusion

- Mise en évidence de marqueur SNP candidats à la sélection
- Reste associer une protéine / famille à chaque marqueur
- Risque d'illusion graphique
- Trouver une méthode pour bien identifier la sélection d'un bruit.

Remerciements

- Franck Dorkel
- Bernhard Gschloessl
- Jean-Pierre Rossi

LES ACARIENS !
ILS SONT
PARTOUT !



PARTOUT !

ON EST
REPÉRÉS !

C'EST
DU BLUFF,
IL NOUS
VOIT PAS
""