

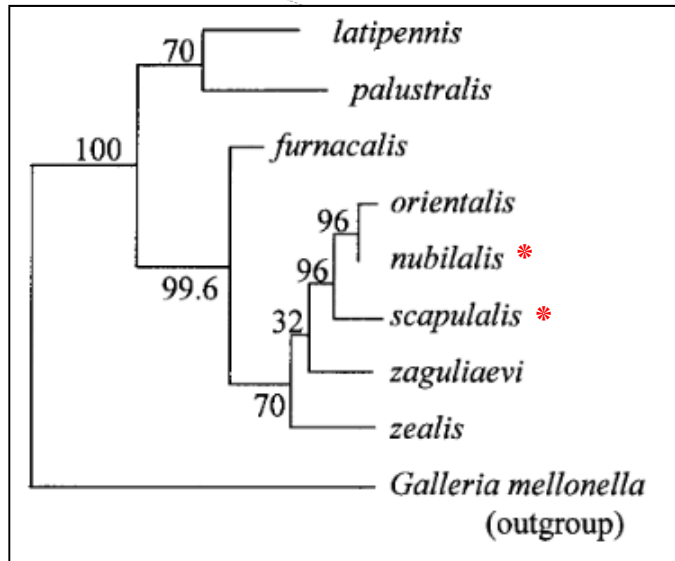
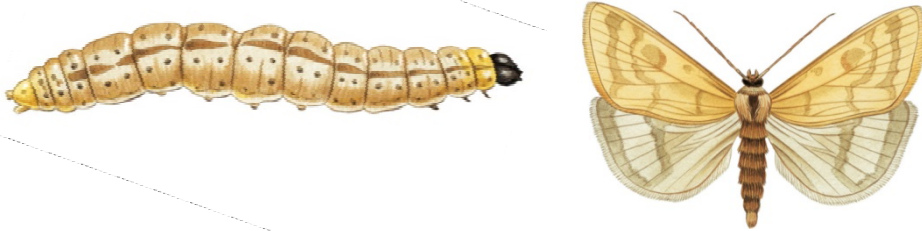


# Génomique de la spécialisation à la plante hôte chez deux espèces sœurs de pyrales (*Ostrinia nubilalis* et *O. scapularis*)

Louise Brousseau



Ostrinia (Lepidoptera) : vingtaine d'espèces dont 2 eu Europe → *O. nubilalis* et *O. scapularis*



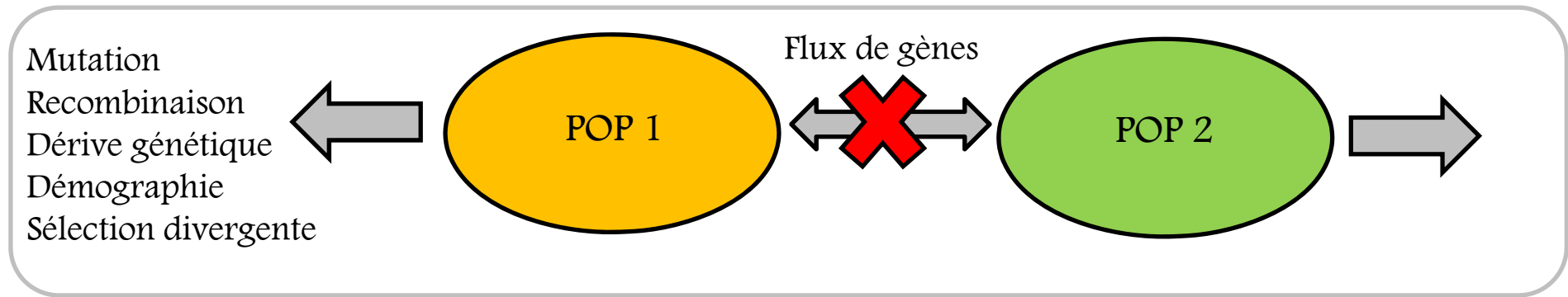
(Ischikawa *et al.* 1999)



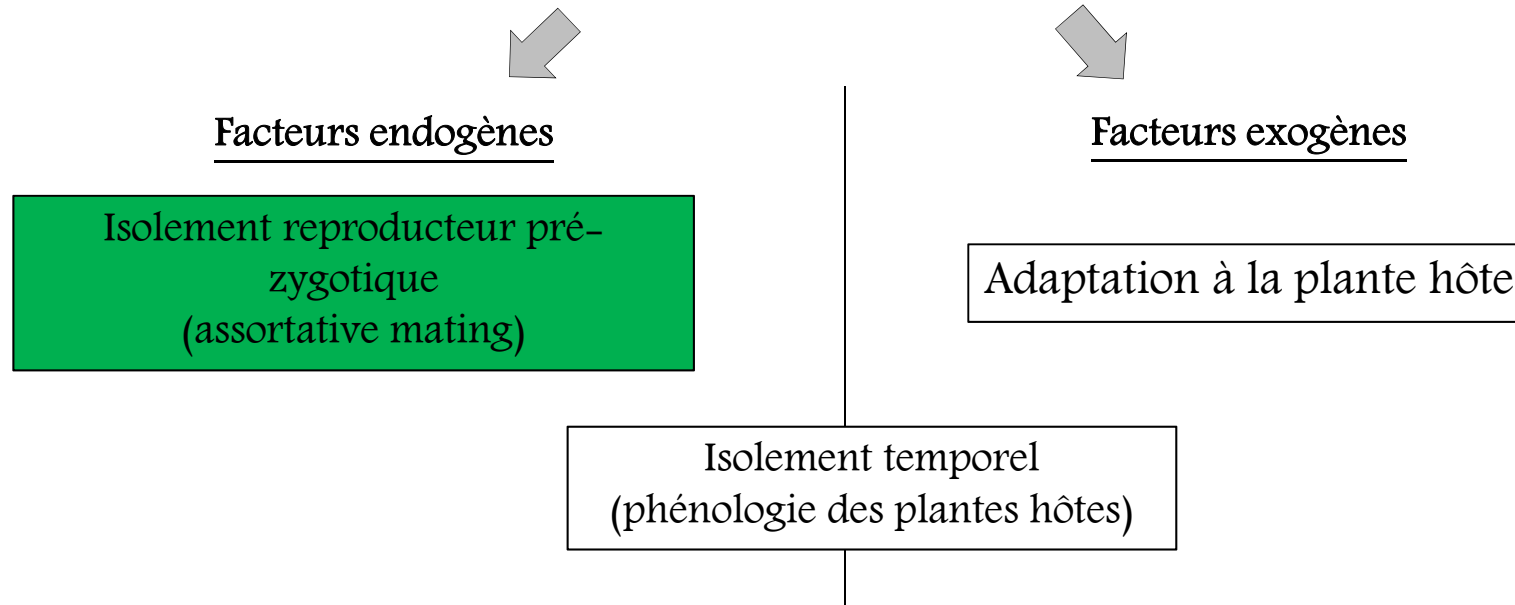
- *O. scapularis* (sensu nov.)
- *O. nubilalis* (sensu nov.)
- *O. furnacalis*

Histoire de la divergence entre espèces méconnue

Frolov *et al.* 2007



*Ostrinia* : Barrières au flux de gènes multi-factorielles



Isolement reproducteur pre-zygotique :

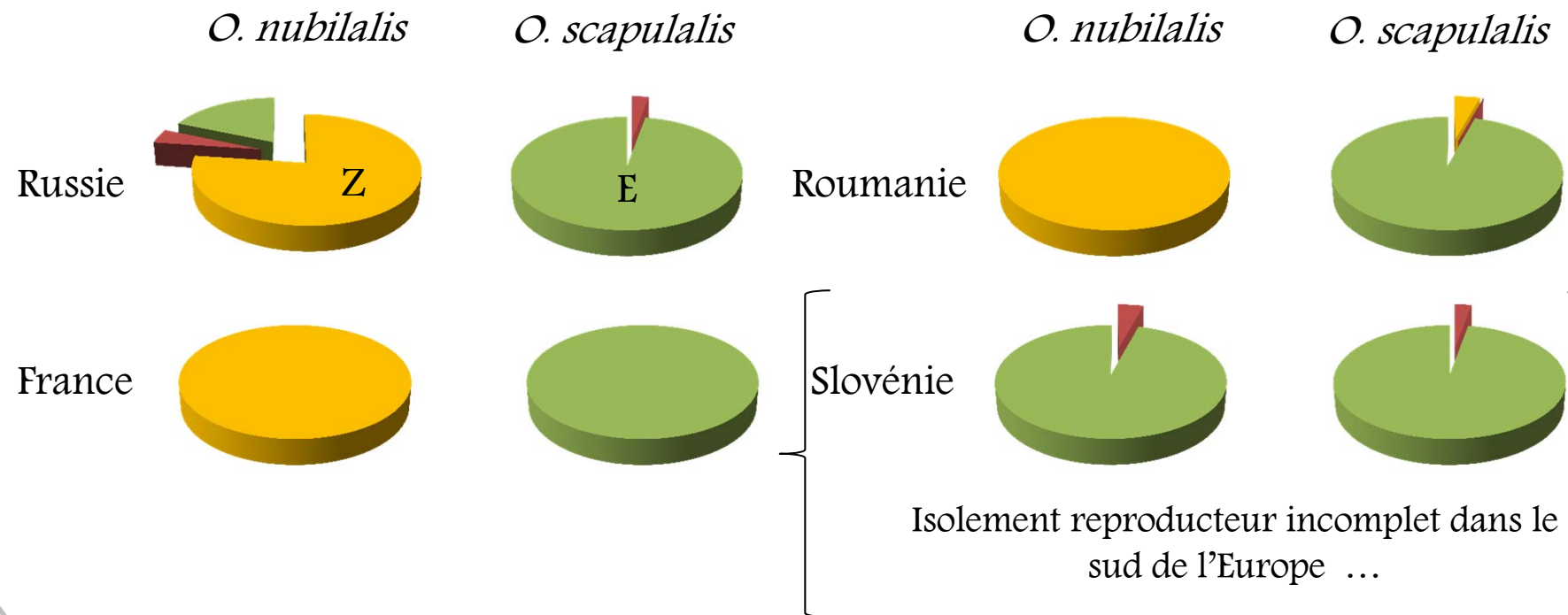
« assortative mating » dû à une différenciation phéromonale entre espèces (Frolov *et al.* 2007) + autre facteurs? (Pelozuelo *et al.* 2007)

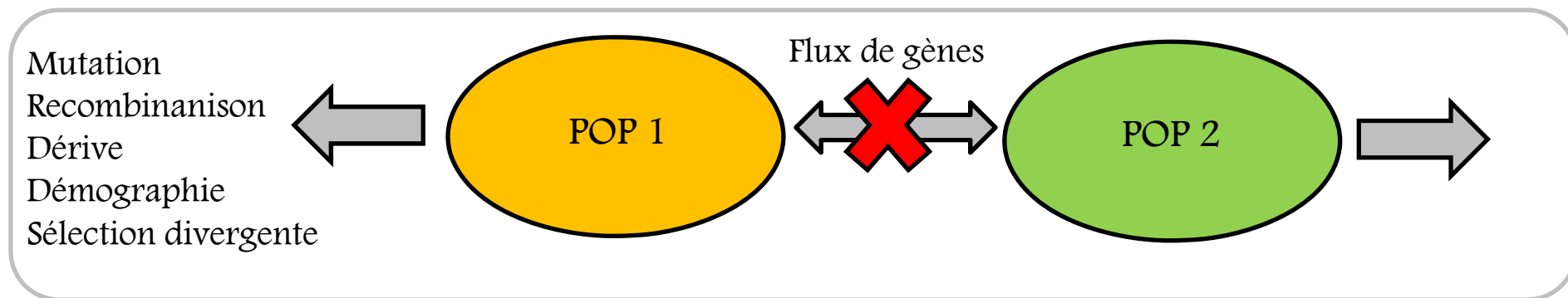
2 races phéromonales E et Z (gène *pgFAR*):

Z

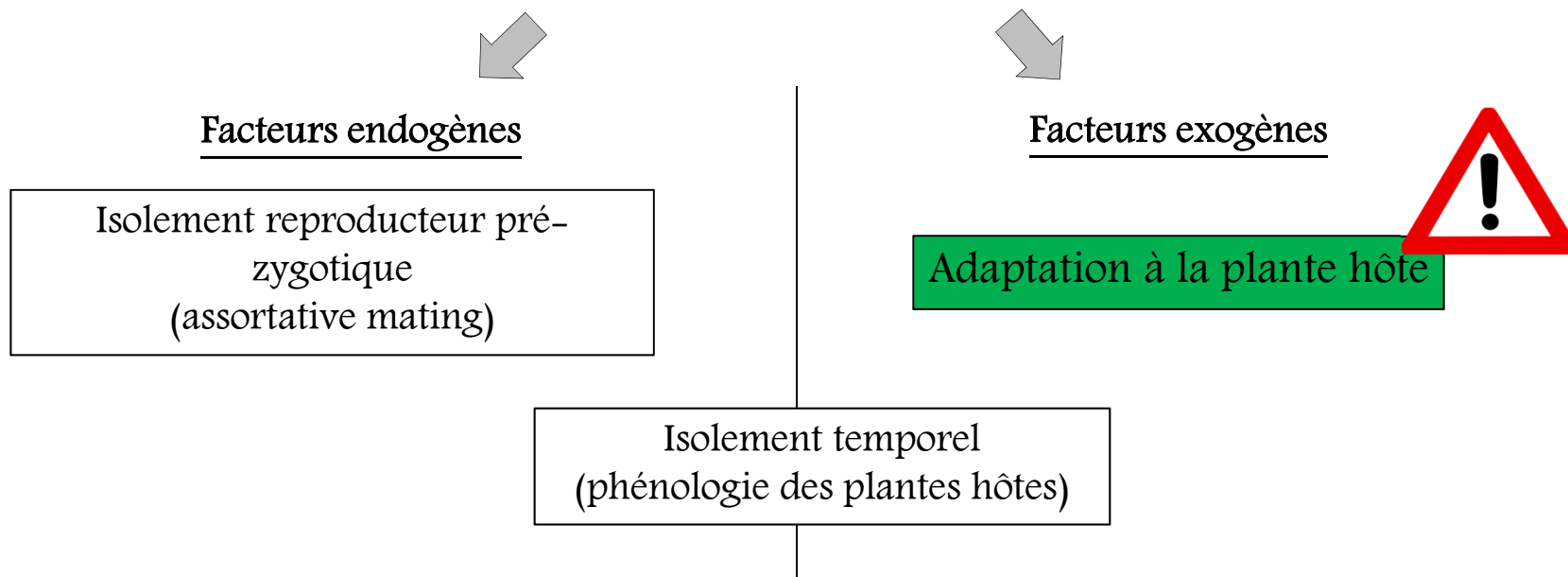
H

E





Ostrinia : Barrières au flux de gènes multi-factorielles

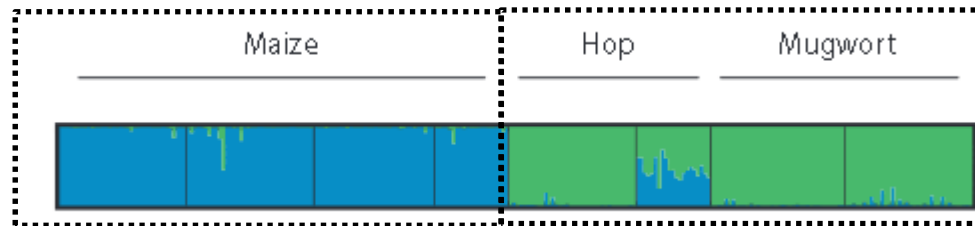


Adaptation à la plante hôte:

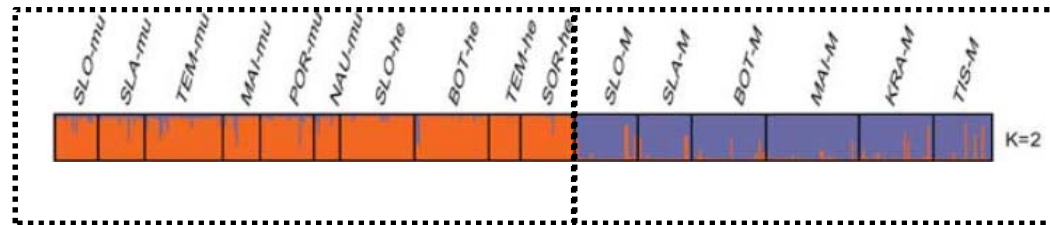
□ Différenciation génétique entre les deux espèces en lien avec la plante hôte



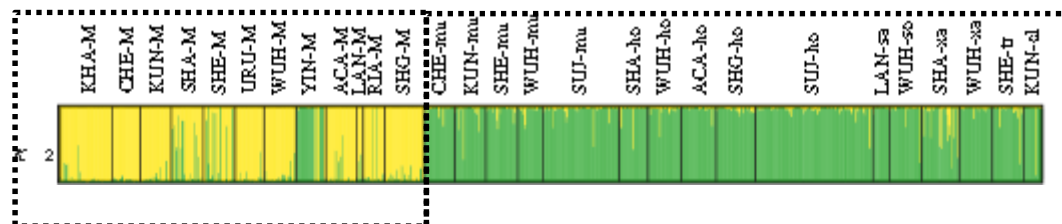
**FRANCE**  
Midamegbe *et al.* 2011.



**RUSSIE**  
Frolov *et al.* 2012.

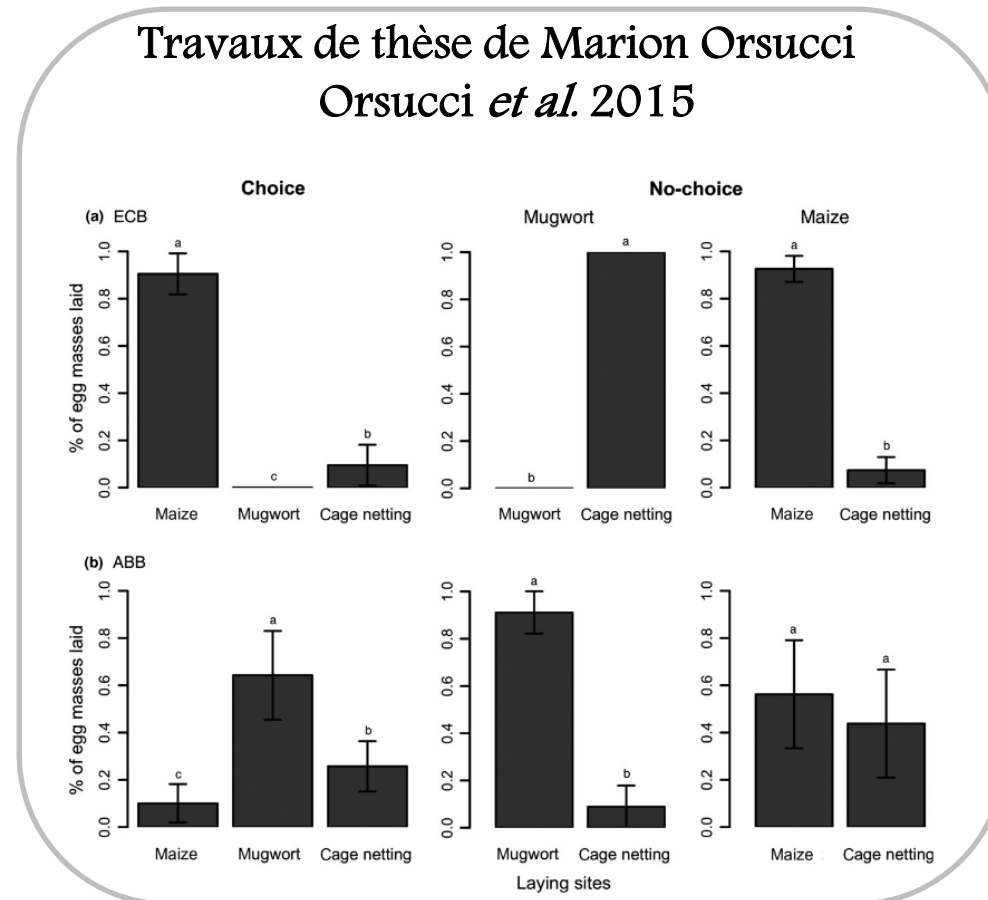


**CHINE**  
Bourguet *et al.* 2013.



## Adaptation à la plante hôte:

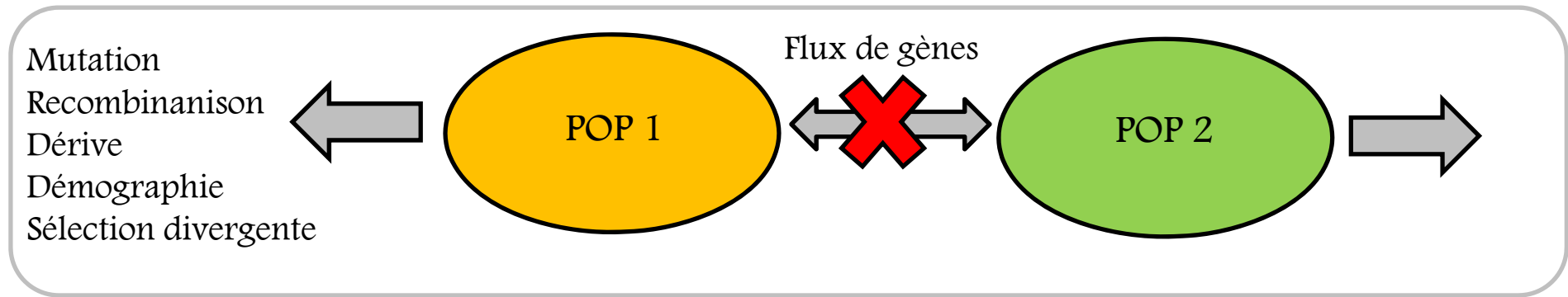
- Différences de performance et d'attraction vis-à-vis des plantes hôtes



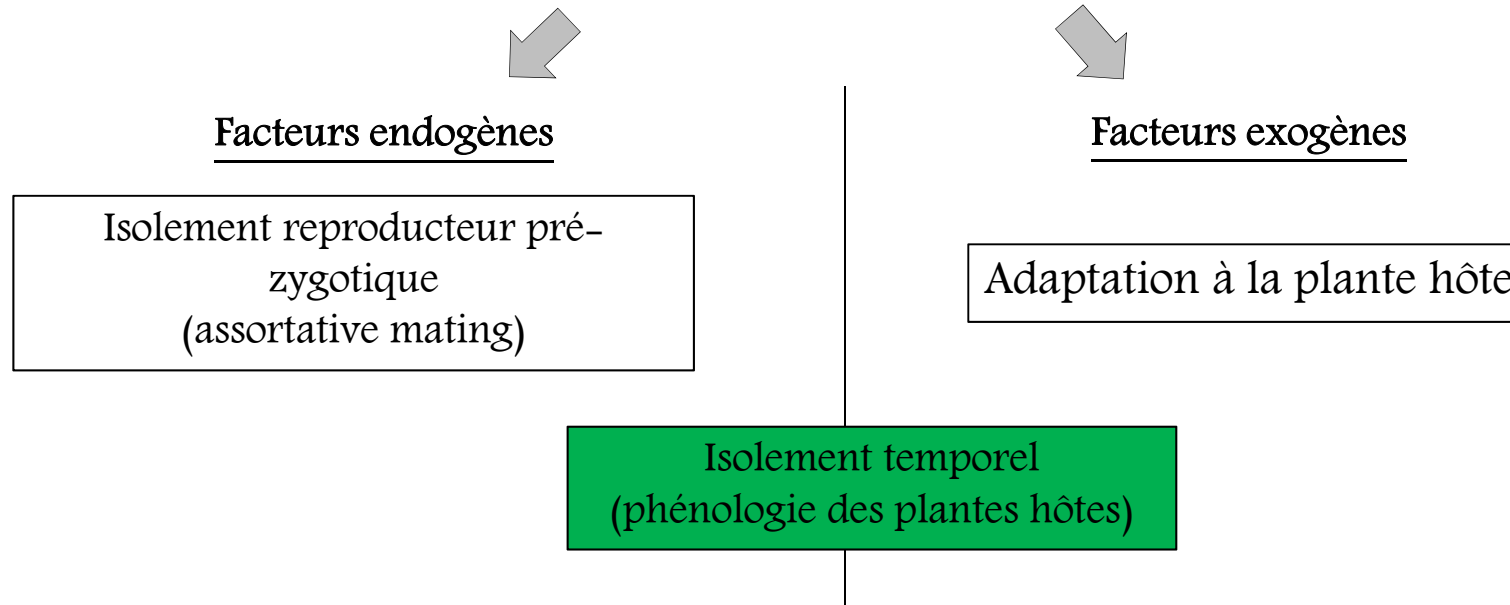
→ Rôle des plantes hôte dans le processus de spéciation?

Hyp : *O. scapulalis* originellement sur dicotylédones sauvages

Spéciation par changement de plante hôte : émergence d' *O. nubilalis* sur maïs



*Ostrinia* : Barrières au flux de gènes multi-factorielles





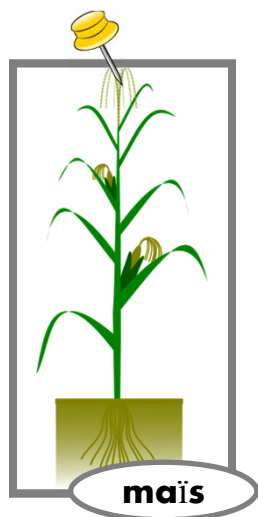
## Génomique de l'adaptation à la plante hôte

- Les deux espèces présentent-elles des traces d'adaptation différentielles (dues à des pressions de sélection divergentes entre plantes hôtes) dans leur génome?
- Quelles sont les bases génomiques de la spécialisation à la plante hôte?

Référence génomique  
~419 Mb / ~51000 scaffolds  
Gschloessl *et al.* (en prep.)

Etude de l'adaptation par une approche  
« PoolSeq »

Annotation de l'hétérochromosome Z



maïs



dicot



*O. nubilalis*

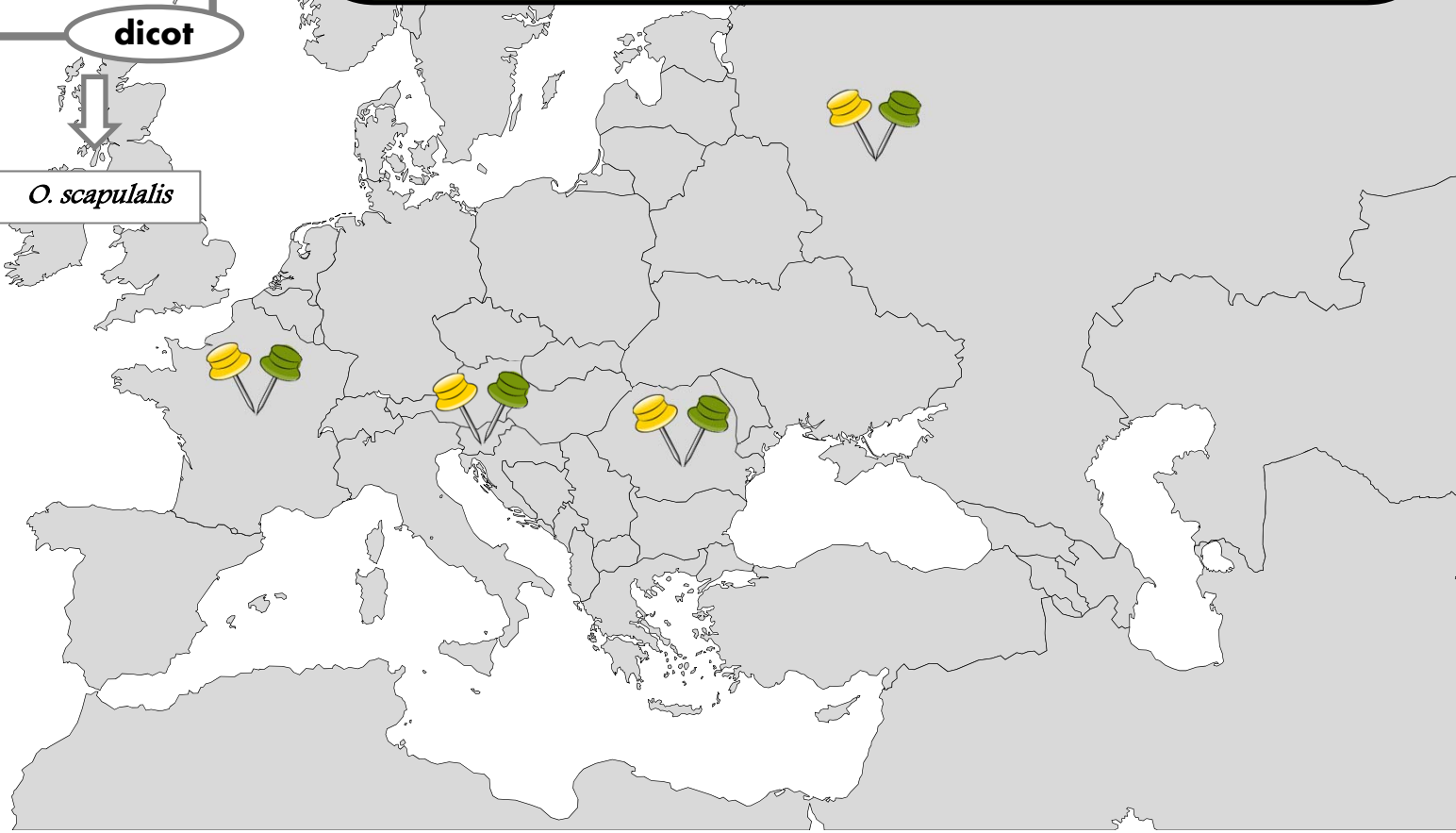



*O. scapularis*

### Design expérimental & Séquençage

- 8 pools (70 ♂ équimolaires)
- Protocole Illumina « TruSeq Nano DNA »
- Séquençage des 8 librairies en 2x125 nt sur 8 lignes d'un séquenceur Illumina HiSeq 2500

→ Sabine Nidelet + MGX



Pipeline bioinformatique

## Reads cleaning

PhiX removal

bowtie2

Reads purity filtering

perl script

Individual bases masking

fastq\_masker :  $q \geq 25$ 

Reads trimming &amp; Orphan reads exclusion

sickle :  $q \geq 25$

	Library							
Location	France	France	Romania	Romania	Slovenia	Slovenia	Russia	Russia
Host plant	Dicot	Maize	Dicot	Maize	Dicot	Maize	Dicot	Maize
Lib name	FR_Dicot	FR_Mais	RO_Dicot	RO_Mais	SL_Dicot	SL_Mais	RU_Dicot	RU_Mais
N raw reads	260,014,650	240,413,375	190,385,275	208,626,713	213,859,622	221,372,807	198,396,698	223,230,031
N cleaned reads	241,987,909	225,600,176	181,351,536	198,200,467	203,219,052	209,382,522	184,495,730	188,327,946
Percentage of raw reads (%)	93.06702872	93.83844638	<b>95.25502222</b>	95.0024396	95.02450724	94.58366854	92.9933471	<b>84.3649688</b>
N mapped reads	152,325,954	141,793,884	115,055,722	126,586,985	129,360,880	133,230,200	110,044,362	112479674
Percentage of cleaned reads (%)	62.94775414	62.8518499	63.44347809	63.86815678	63.65588203	63.63004836	59.64602108	59.72542917
Total bases mapped (mapping length)	18,044,553,291	16,781,087,502	13,602,770,367	14,975,315,499	15,306,177,640	15,763,379,652	13,012,150,314	13,294,844,231
Mismatches	441463114	425152203	348416015	380093930	386551425	401364668	338504155	345263906
Mismatches / bases mapped	0.024465173	0.0253352	0.025613607	0.025381364	0.025254602	0.025461841	0.026014467	0.025969759
Depth (mean $\pm$ d)	48.10 X $\pm$ 48.48	44.86 X $\pm$ 47.28	36.27 X $\pm$ 40.18	40.00 X $\pm$ 43.27	40.85 X $\pm$ 43.85	42.02 X $\pm$ 44.62	34.48 X $\pm$ 39.27	35.64 X $\pm$ 40.17

Pipeline bioinformatique

## Reads cleaning

PhiX removal

bowtie2

Reads purity filtering

perl script

Individual bases masking

fastq\_masker :  $q \geq 25$ 

Reads trimming &amp; Orphan reads exclusion

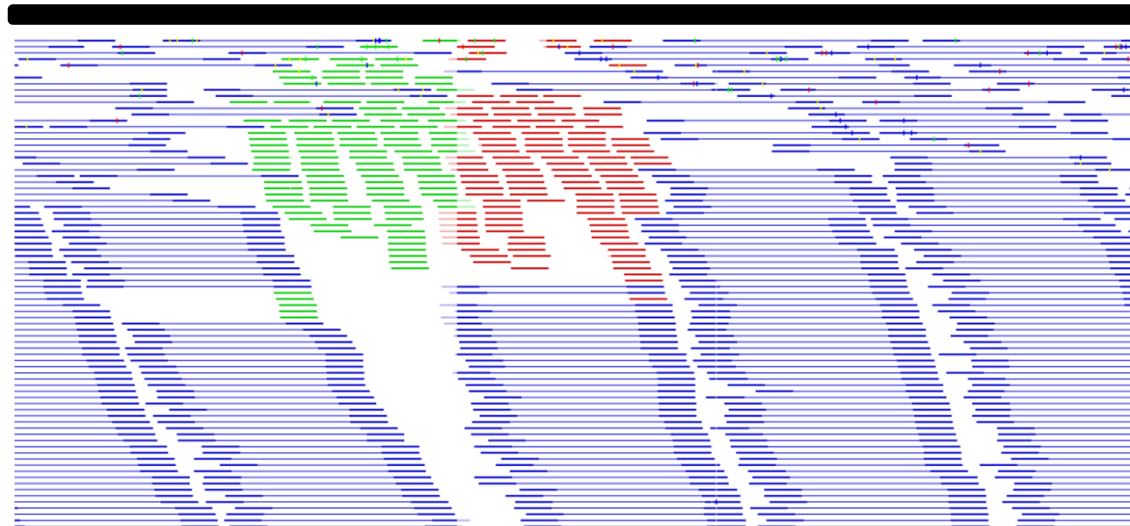
sickle :  $q \geq 25$ 

## Mapping

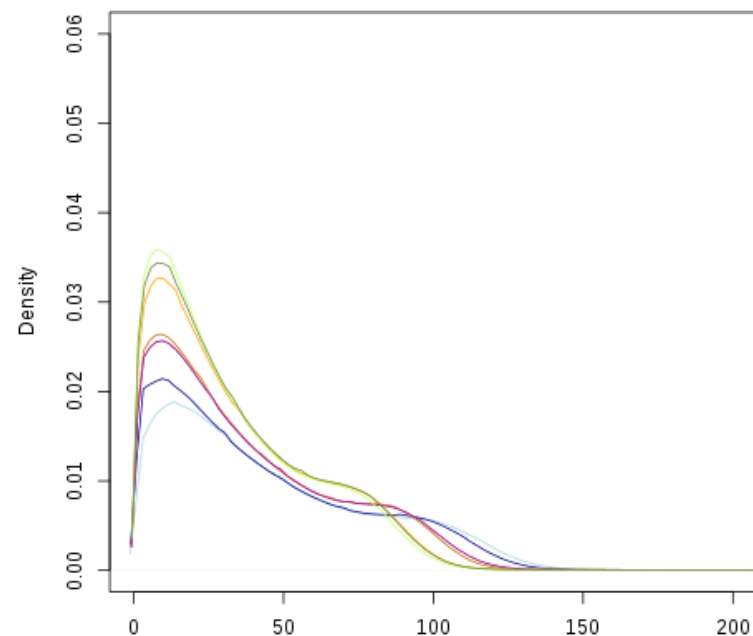
Mapping of paired reads against osca 1.2

bwa,  $q \geq 25$  M=10

référence



	Library							
Location	France	France	Romania	Romania	Slovenia	Slovenia	Russia	Russia
Host plant	Dicot	Maize	Dicot	Maize	Dicot	Maize	Dicot	Maize
Lib name	FR_Dicot	FR_Mais	RO_Dicot	RO_Mais	SL_Dicot	SL_Mais	RU_Dicot	RU_Mais
N raw reads	260,014,650	240,413,375	190,385,275	208,626,713	213,859,622	221,372,807	198,396,698	223,230,031
N cleaned reads	241,987,909	225,600,176	181,351,536	198,200,467	203,219,052	209,382,522	184,495,730	188,327,946
Percentage of raw reads (%)	93.06702872	93.83844638	95.25502222	95.0024396	95.02450724	94.58366854	92.9933471	84.3649688
N mapped reads	152,325,954	141,793,884	115,055,722	126,586,985	129,360,880	133,230,200	110,044,362	112479674
Percentage of cleaned reads (%)	62.94775414	62.8518499	63.44347809	<b>63.86815678</b>	63.65588203	63.63004836	<b>59.64602108</b>	59.72542917
Total bases mapped (mapping length)	18,044,553,291	16,781,087,502	13,602,770,367	14,975,315,499	15,306,177,640	15,763,379,652	13,012,150,314	13,294,844,231
Mismatches	441463114	425152203	348416015	380093930	386551425	401364668	338504155	345263906
Mismatches / bases mapped	0.024465173	0.0253352	0.025613607	0.025381364	0.025254602	0.025461841	0.026014467	0.025969759
Depth (mean $\pm$ d)	<b>48.10 X <math>\pm</math> 48.48</b>	44.86 X $\pm$ 47.28	36.27 X $\pm$ 40.18	40.00 X $\pm$ 43.27	40.85 X $\pm$ 43.85	42.02 X $\pm$ 44.62	<b>34.48 X <math>\pm</math> 39.27</b>	35.64 X $\pm$ 40.17



Profondeur (X)

Pipeline bioinformatique

## Reads cleaning

PhiX removal

bowtie2

Reads purity filtering

perl script

Individual bases masking

fastq\_masker :  $q \geq 25$ 

Reads trimming &amp; Orphan reads exclusion

sickle :  $q \geq 25$ 

## Mapping

Mapping of paired reads against osca 1.2

bwa,  $q \geq 25$  M=10

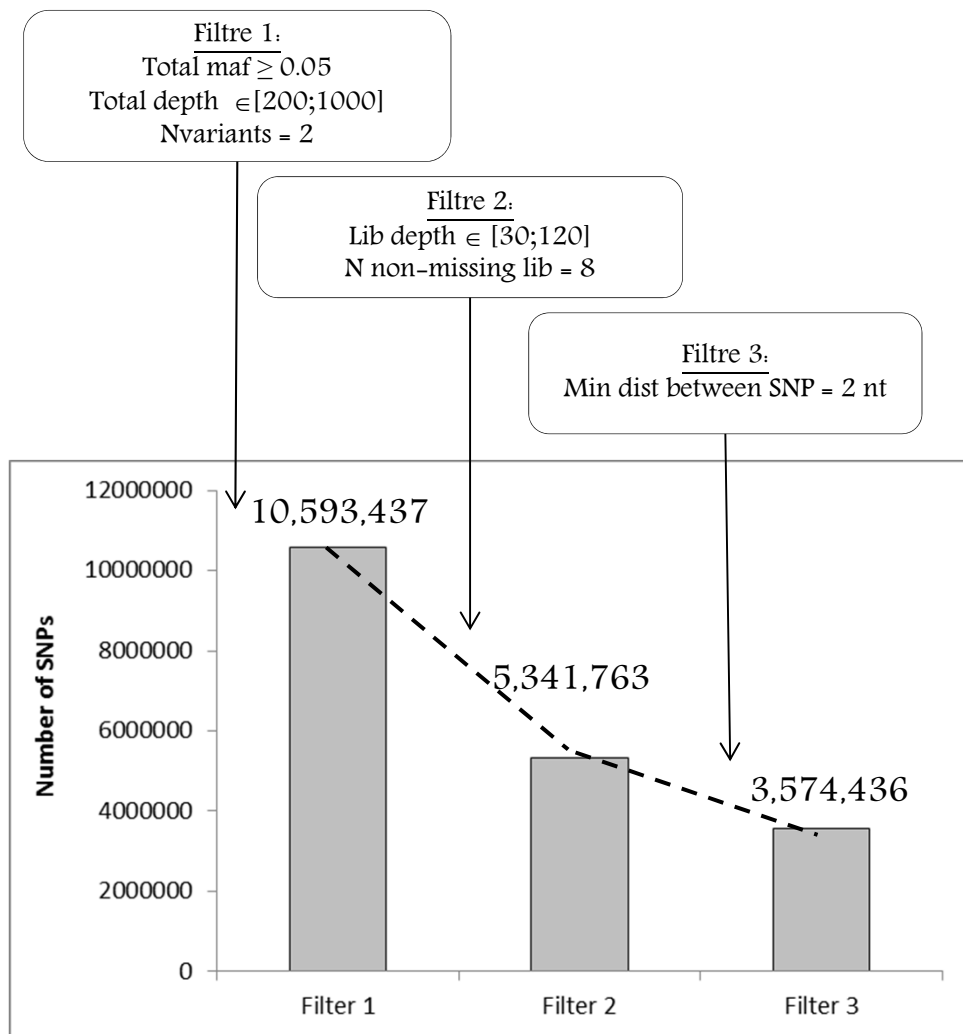
## SNP detection

SNP calling

Popoolation2 &amp; perl scripts

SNP filtering

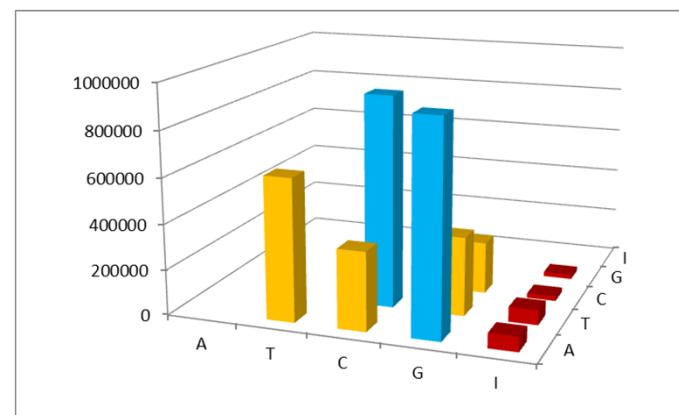
perl scripts



3 574 436 SNPs dans 26 454 scaffolds  
 (52.1% des scaffolds sont polymorphes)

Transitions : 51.92%  
 Transversions : 43.26%  
 Indels : 4.82%

SNP density:  
 1.23 / 100bp





Pipeline bioinformatique

## Reads cleaning

PhiX removal

bowtie2

Reads purity filtering

perl script

Individual bases masking

fastq\_masker :  $q \geq 25$ 

Reads trimming &amp; Orphan reads exclusion

sickle :  $q \geq 25$ 

## Mapping

Mapping of paired reads against osca 1.2

bwa,  $q \geq 25$  M=10

## SNP detection

SNP calling

Popoolation2 &amp; perl scripts

SNP filtering

perl scripts

## Selection tests

BayPass

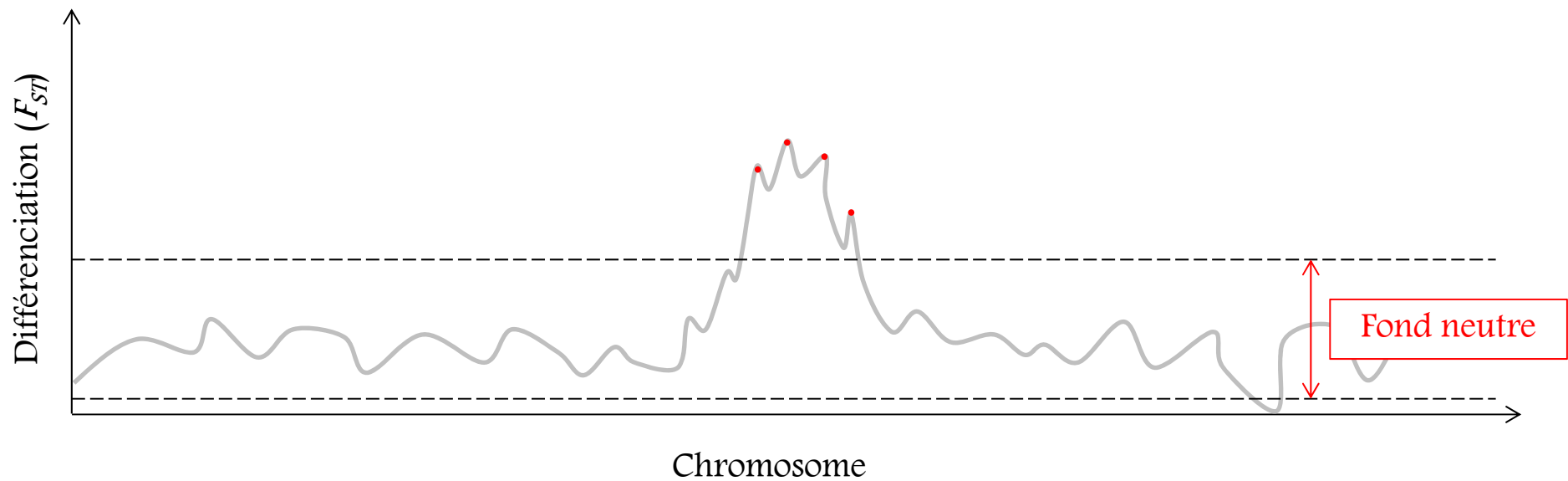
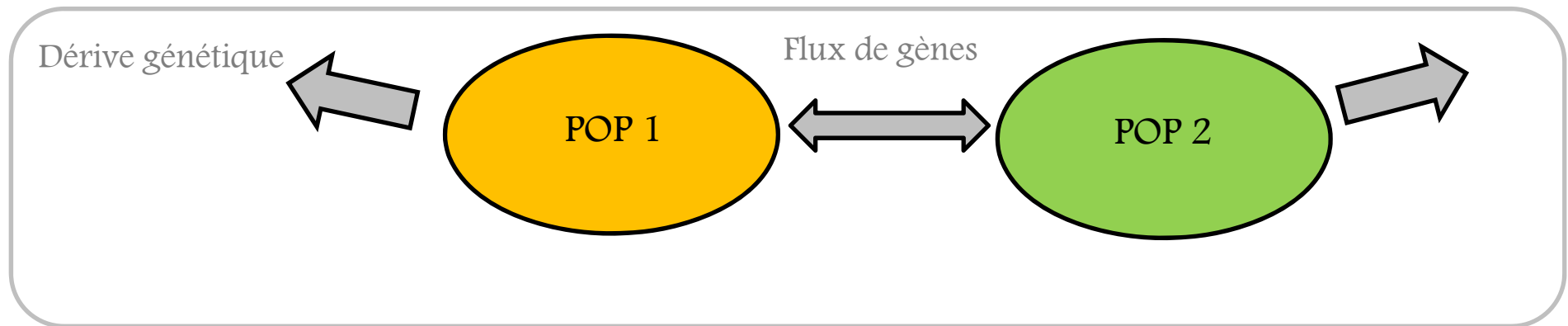
M. Gautier 2015

SelEstim

R. Vitalis et al. 2014

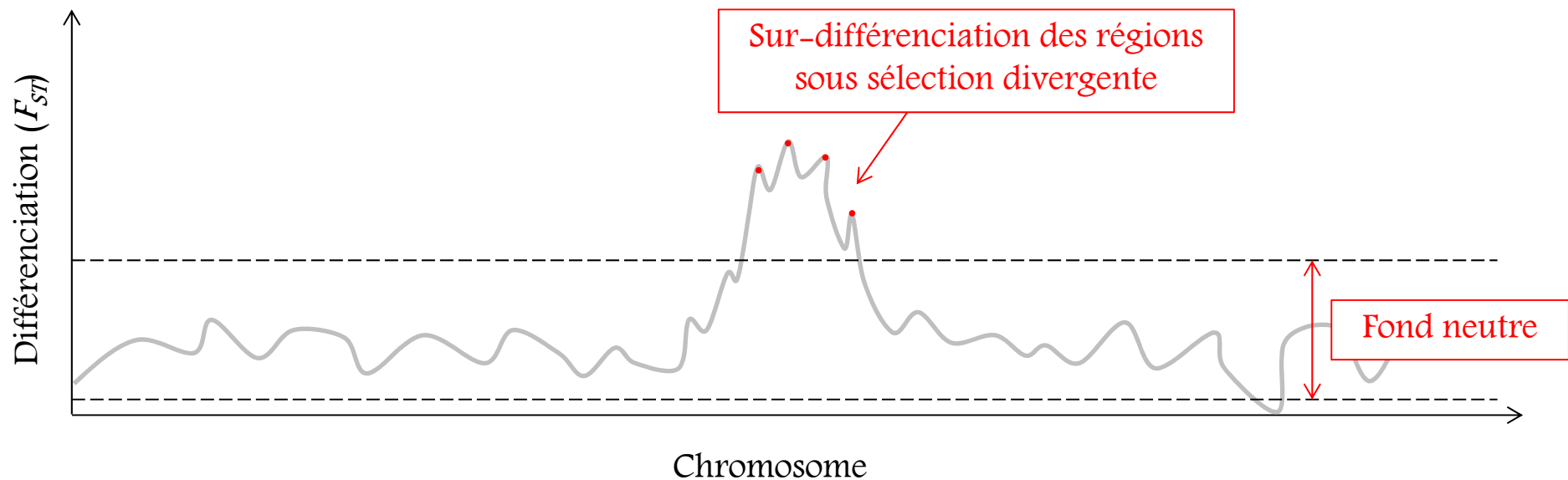
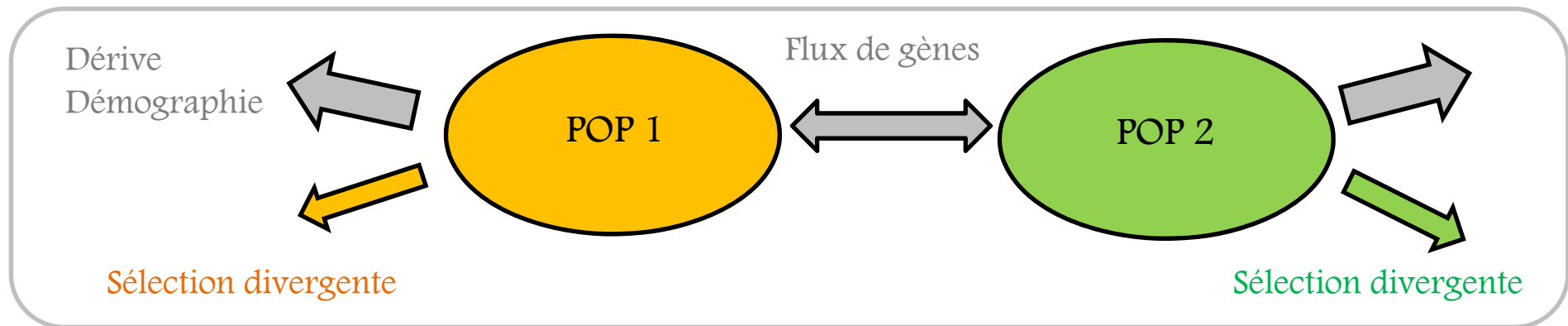
## Principe général

- Divergence neutre : « Génome-wide »



### Principe général

- ❑ Divergence neutre : Génome-wide
- ❑ Divergence adaptative : locus-spécifique



## BayPass (M. Gautier 2015, *Genetics*)

### **Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates**

**Mathieu Gautier<sup>1</sup>**

INRA, UMR CBGP (Centre de Biologie pour la Gestion des Populations), Campus International de Baillarguet, F-34988 Montpellier-sur-Lez, France, and IBC (Institut de Biologie Computationnelle), F-34095 Montpellier, France

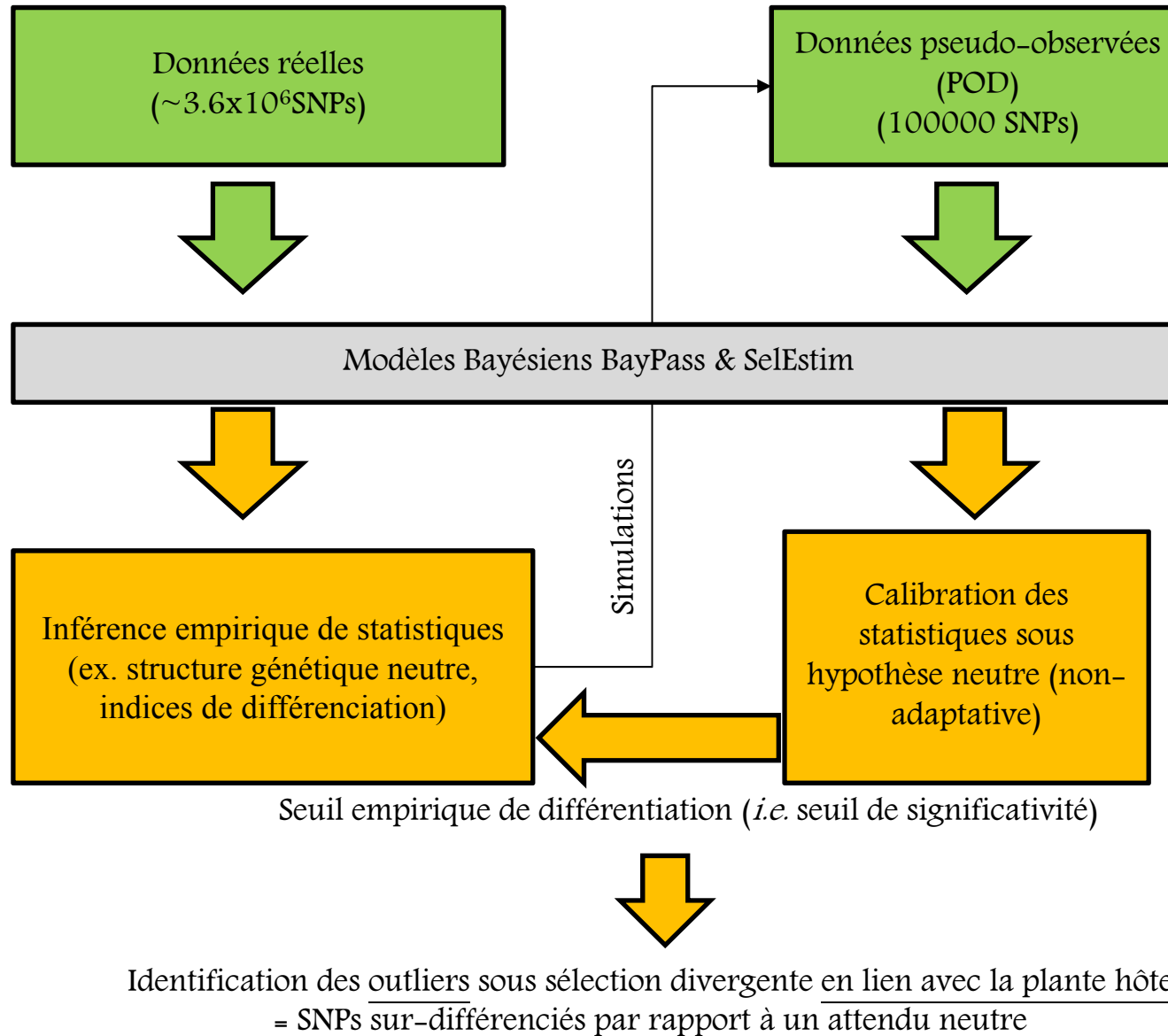
## SelEstim (R. Vitalis *et al.* 2015, *Genetics*)

### **Detecting and Measuring Selection from Gene Frequency Data**

**Renaud Vitalis,<sup>\*,1</sup> Mathieu Gautier,<sup>\*,1</sup> Kevin J. Dawson,<sup>2</sup> and Mark A. Beaumont<sup>3</sup>**

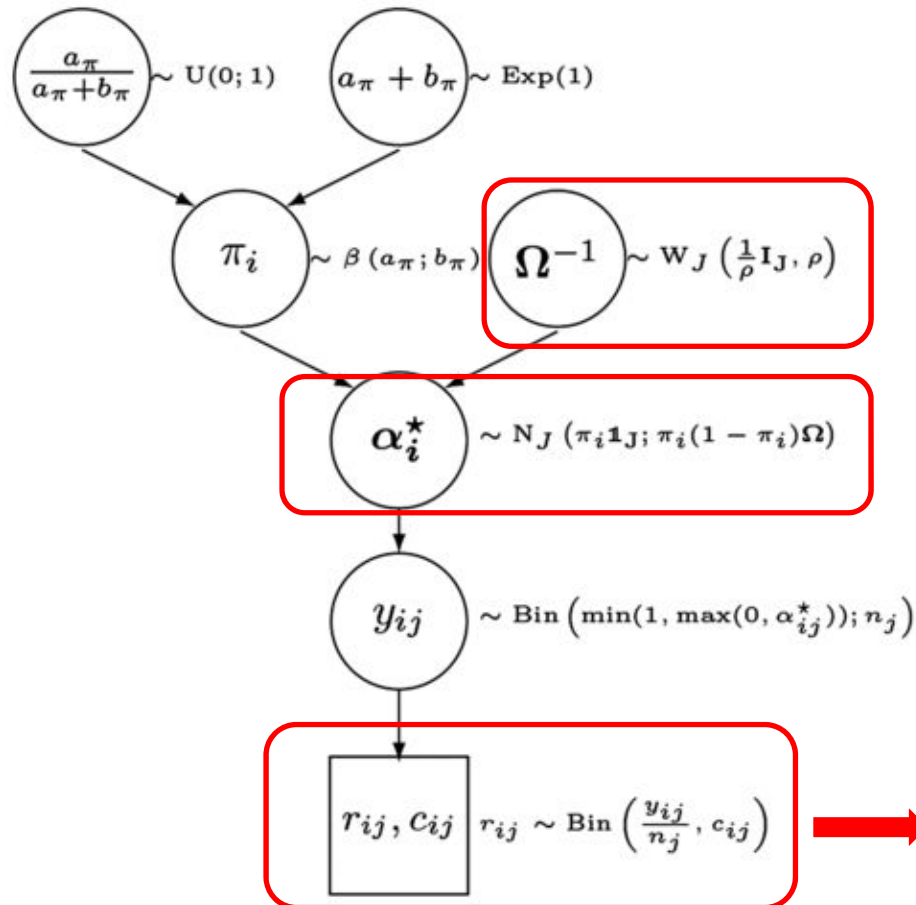
<sup>\*</sup>Institut National de la Recherche Agronomique, Unité Mixte de Recherche CBGP, (Inra, Ird, Cirad, Montpellier-SupAgro) 34988 Montpellier-sur-Lez Cedex, France, <sup>1</sup>Institut de Biologie Computationnelle, 34095 Montpellier Cedex, France, <sup>2</sup>Cancer Genome Project, The Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, United Kingdom, <sup>3</sup>Department of Mathematics and School of Biological Sciences, University of Bristol, Bristol BS8 1TW, United Kingdom

## Procédure



BayPass

A.2) Read count data (Pool-Seq)



$\Omega$  : Matrice de variance-covariance genome-wide (structure génétique neutre)

$\alpha_i^*$  : Fréquences alléliques standardisées, corrigées par matrice de var-covar

Comptages alléliques

BayPass

$XtX$ : Indice de différenciation locus-spécifique (parmi l'ensemble des pops)

$$\widehat{XtX}_i = \frac{1}{T} \sum_{t=1}^T \frac{\alpha_i^*(t) \Lambda(t) \alpha_i^*(t)}{\pi_i(t) (1 - \pi_i(t))}$$

Fonction de :

$\Lambda = \Omega^{-1}$  : matrice de précision (inverse de matrice de var-covar)

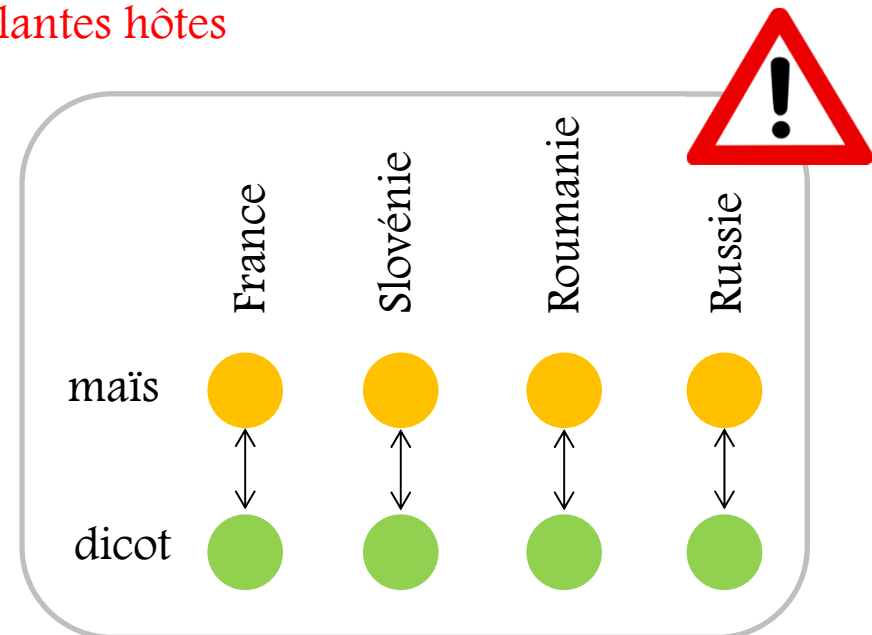
$\alpha_i^*$  : freq alléliques std

$I_{BayPass}$  : Indice de différentiation en lien avec les plantes hôtes

Localities :  $L = \{l1, l2, l3, l4\}$

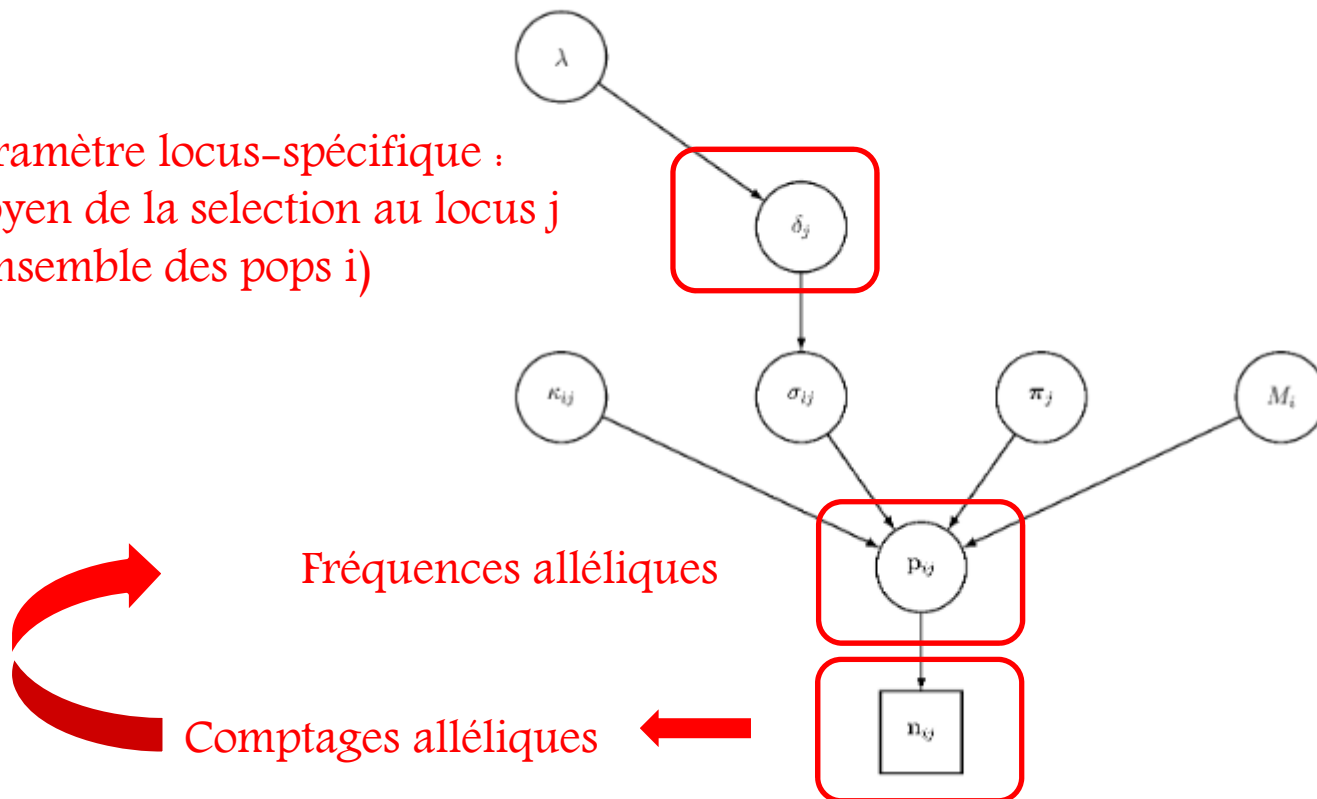
HostPlant :  $P = \{p1, p2\}$

$$I_{BayPass} = \left[ \sum_{l=1}^L (\alpha_{std_{l,p1}}) - \sum_{l=1}^L (\alpha_{std_{l,p2}}) \right]^2$$



SelEstim

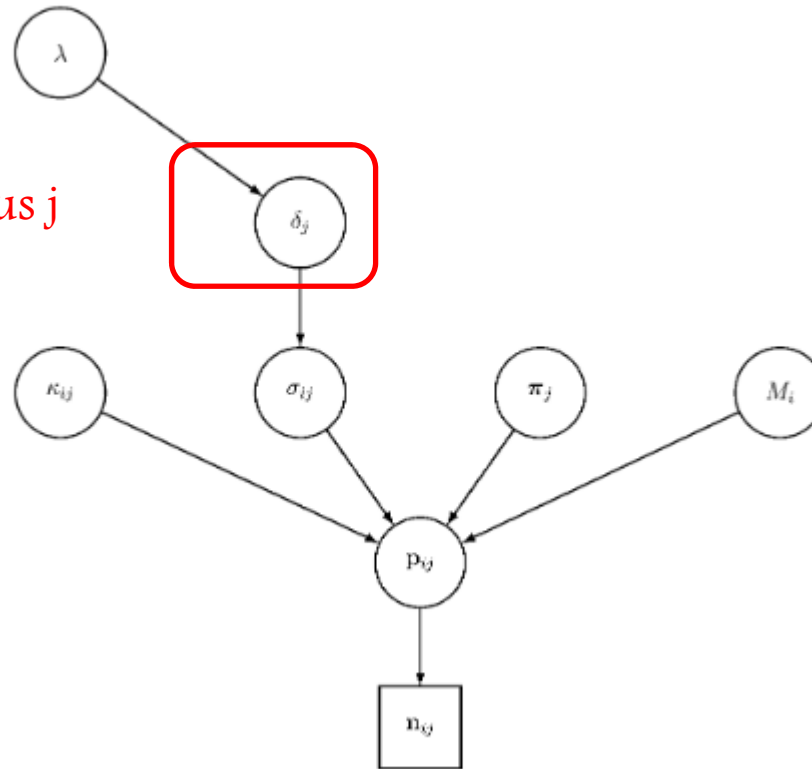
Hyper-paramètre locus-spécifique :  
~ effet moyen de la sélection au locus  $j$   
(parmi l'ensemble des pops  $i$ )





SelEstim

Hyper-paramètre locus-spécifique :  
~ effet moyen de la selection au locus j  
(parmi l'ensemble des pops i)

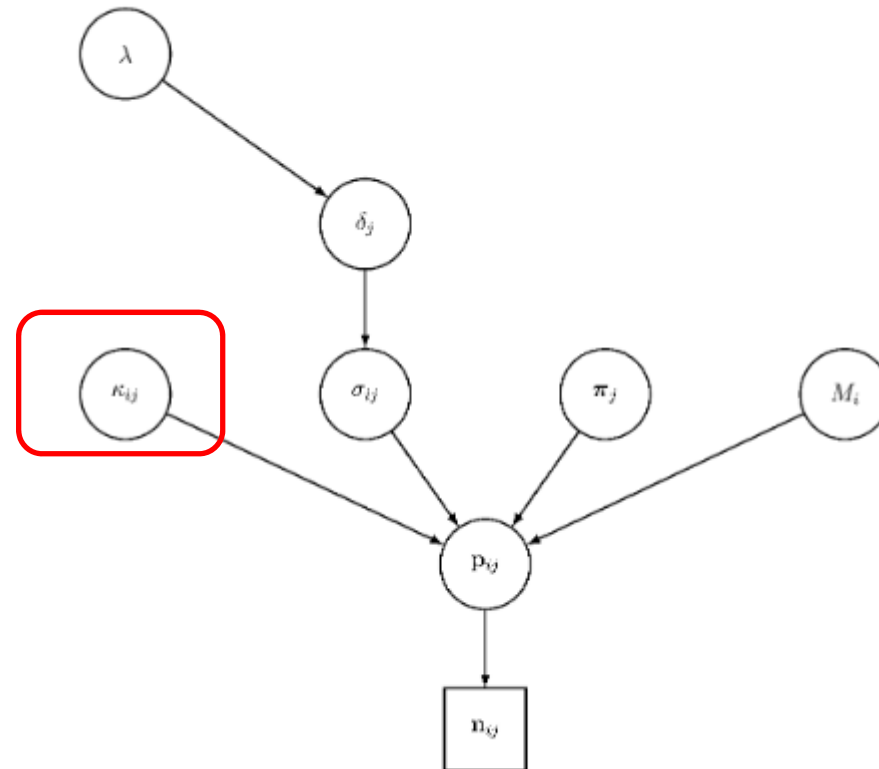


*KLD*: Indice de divergence de Kullback–Leibler

~ mesure de l'écart de  $\delta_j$  au locus j par rapport à la distribution de centrage (~ $\delta$  moyen aux différents loci j)

SelEstim

$K_{ij}$ : 'Indicator variable'  
locus-spécifique



SelEstim

$K_{ij}$ : 'Indicator variable'  
Locus-spécifique

- $K_{ij} = \sim 0.5$  sous hypothèse neutre

- $K_{ij} \rightarrow 1$  si allele de ref est sélectionné 'pour' dans la pop i au locus j
- $K_{ij} \rightarrow 0$  si allele de ref est sélectionné 'contre'

Locus neutre

	Dicot	Maïs
FR	$\sim 0.5$	$\sim 0.5$
SL	$\sim 0.5$	$\sim 0.5$
RO	$\sim 0.5$	$\sim 0.5$
RU	$\sim 0.5$	$\sim 0.5$

Locus sous sélection (sans lien avec la plante hôte)

	Dicot	Maïs
FR	$\sim 0$	$\sim 0.5$
SL	$\sim 0.5$	$\sim 0.5$
RO	$\sim 0.5$	$\sim 1$
RU	$\sim 1$	$\sim 0.5$

Locus sous sélection (en lien avec la plante hôte)

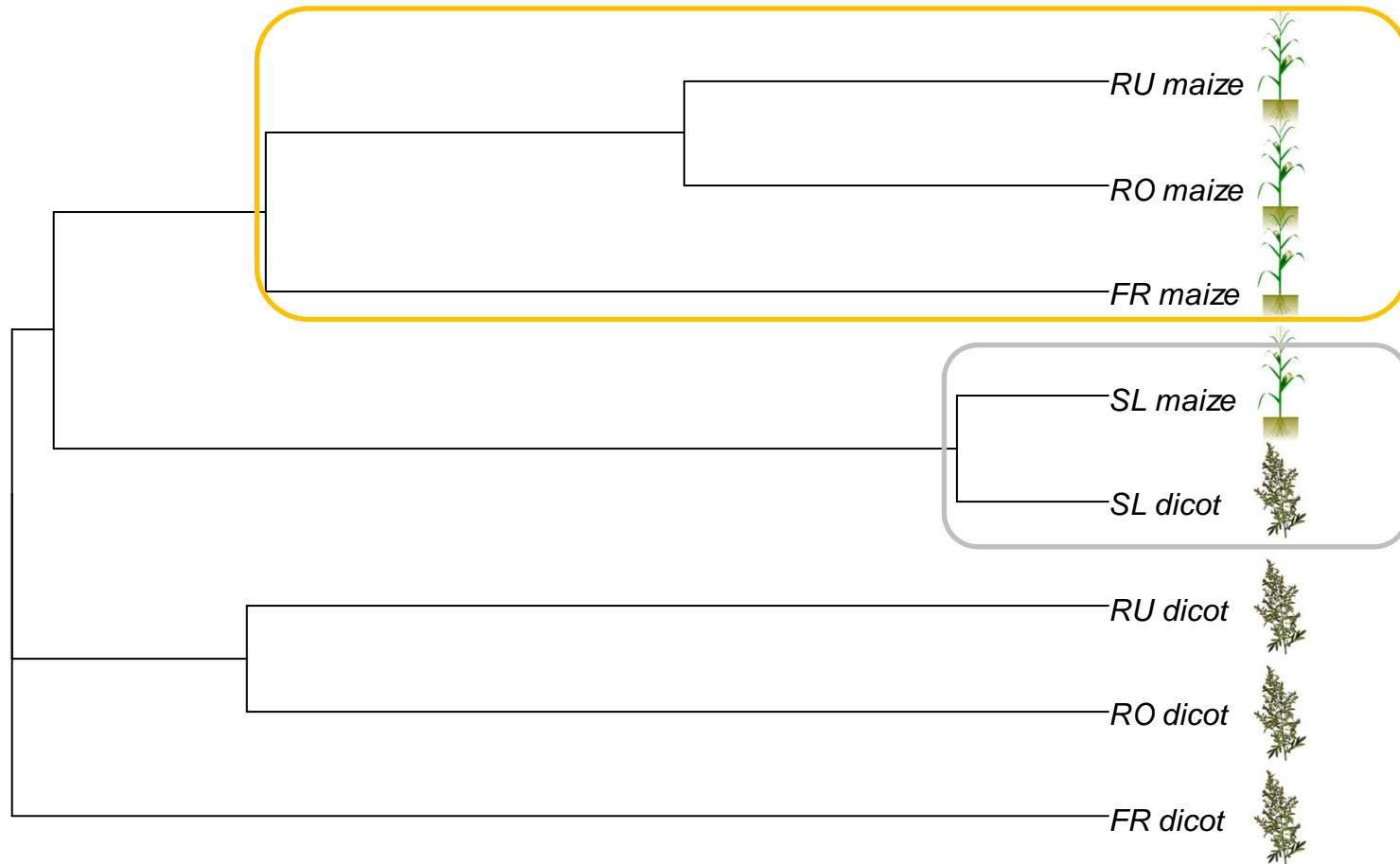
	Dicot	Maïs
FR	$\sim 0$	$\sim 1$
SL	$\sim 0$	$\sim 1$
RO	$\sim 0$	$\sim 1$
RU	$\sim 0$	$\sim 1$

	Dicot	Maïs
FR	$\sim 1$	$\sim 0$
SL	$\sim 1$	$\sim 0$
RO	$\sim 1$	$\sim 0$
RU	$\sim 1$	$\sim 0$

Localities :  $L = \{l1, l2, l3, l4\}$

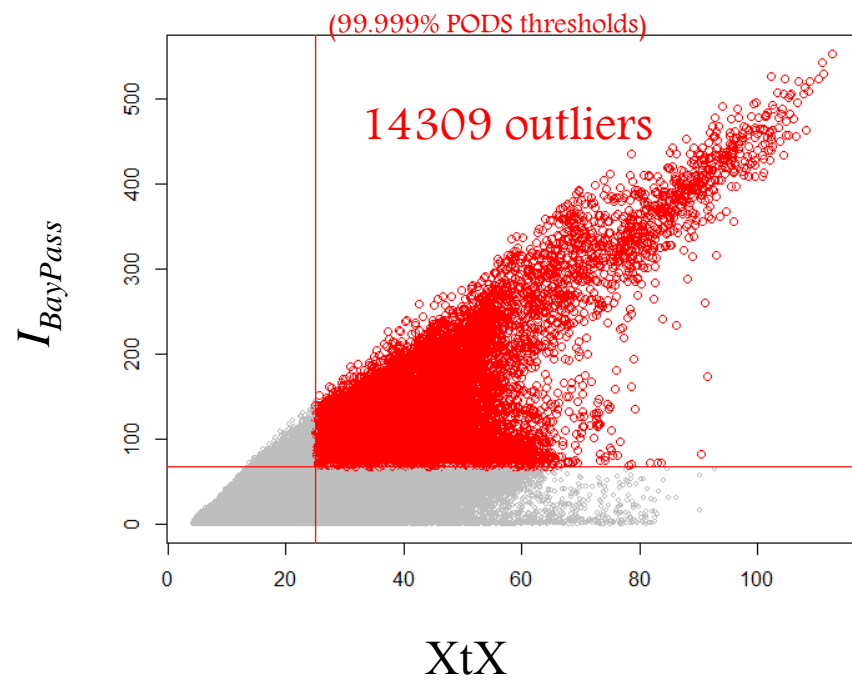
HostPlant :  $P = \{p1, p2\}$

$$I_{SelEstim} = \left[ \sum_{l=1}^L (\kappa_{l,p1}) - \sum_{l=1}^L (\kappa_{l,p2}) \right]^2$$

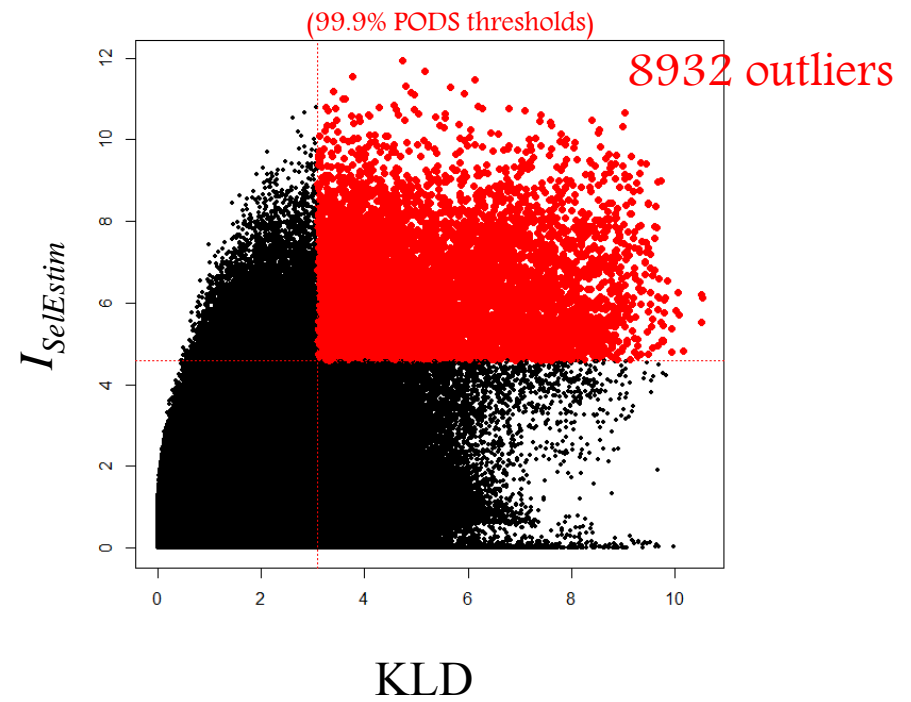
BayPass $\Omega$ 

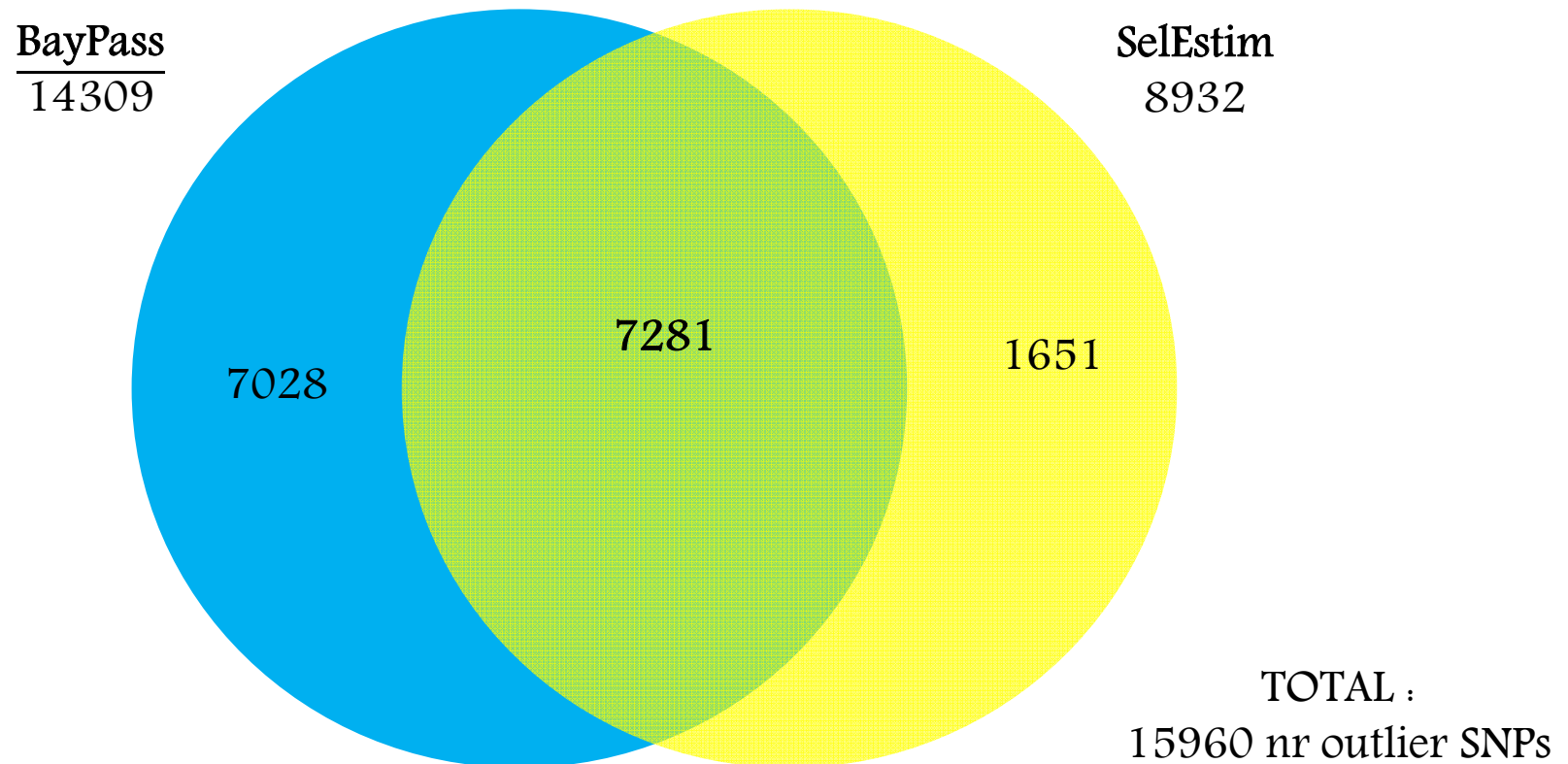
➔ Structure génétique en lien avec la plante hôte SAUF en Slovénie

## BayPass

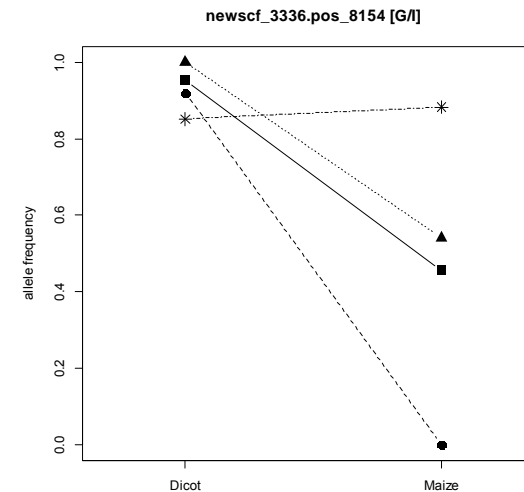
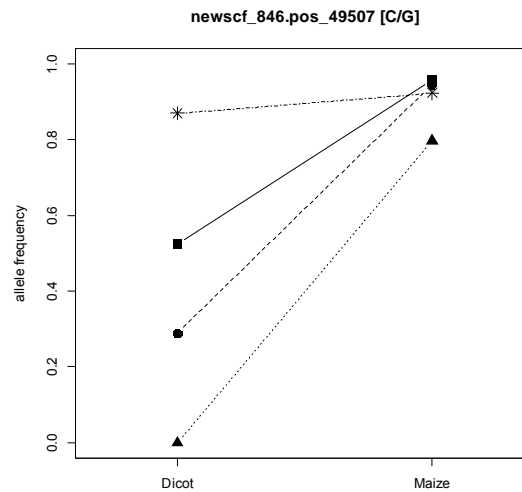
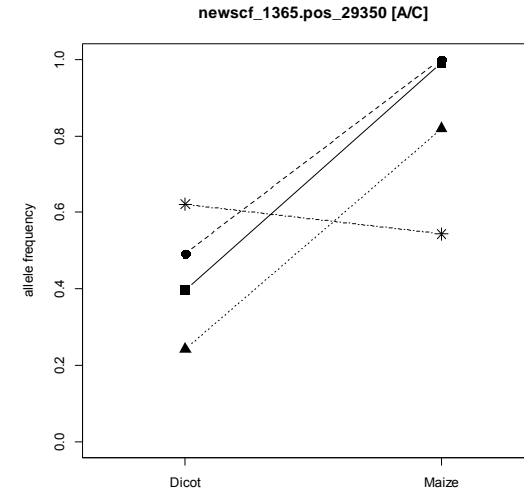
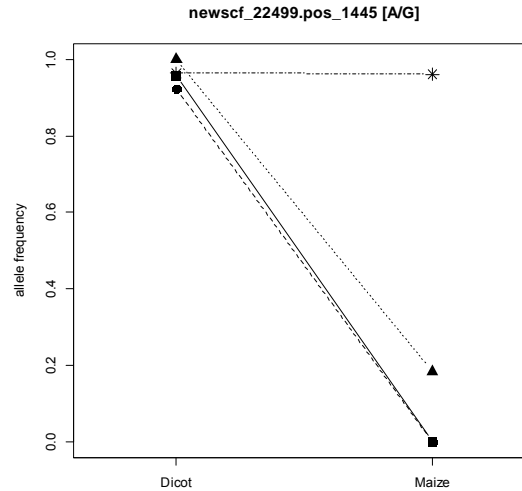
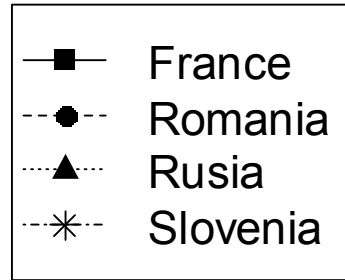


## SelEstim

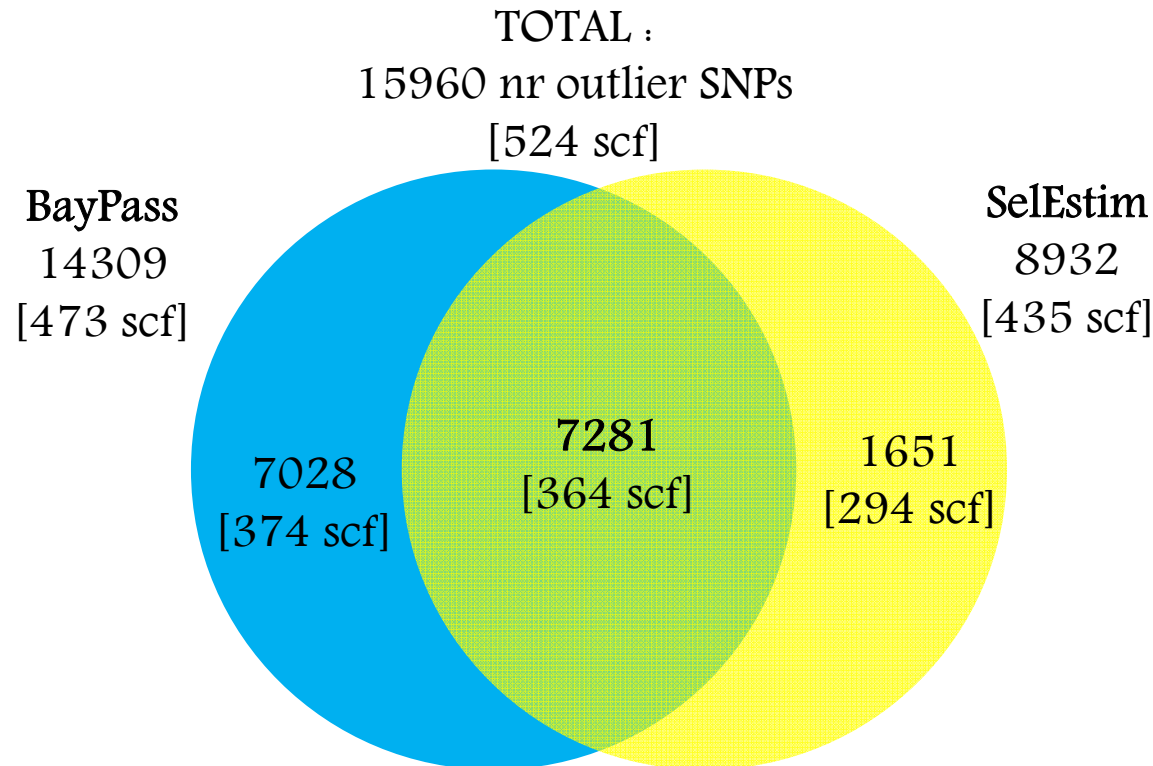




→ De nombreux outliers communs aux deux méthodes 😊



- ➔ Variations des fréquences alléliques entre plantes hôtes se répètent dans les différentes localités ('dans même sens') = sélection divergente en lien avec la plante hôte
- ➔ Vrai dans 3 localités sur 4 MAIS PAS en Slovénie :
  - isolation incomplète (divergence primaire) ???
  - contact secondaire (admixture) ???

Distribution des outliers dans les scaffolds :

→ Outliers groupés dans un nombre TRES réduit de scaffolds :

$364 \text{ scf} / 50738 = 0,7\%$  des scf alors que 50% des scf sont polymorphes (o.O)

- 'Îlots' de forte différenciation (fort DL du fait d'une divergence récente et/ou forte intensité de sélection) ????
- Résultats biaisés par un facteur non pris en compte ????



## Annotation de l'hétérochromosome Z (Gschloessl et al., en prep.)

Hétérochromosomes de la Pyrales (=Chromosomes sexuels) : ♂ (ZZ) / ♀ (ZW)

Importance des hétérochromosomes :

- Biologie : gènes d'intérêt (récepteurs olfactifs etc.)
- Méthodologie :

Taille efficace hétérosomes < Taille efficace autosomes

Taille efficace Z = $\frac{3}{4}$ (Taille efficace autosomes)
Taille efficace W = $\frac{1}{4}$ (Taille efficace autosomes)

- ➔ soumis à une plus forte dérive = se différencient plus vite  
= SNPs des hétérosomes sur-différenciés par rapport aux autosomes
- ➔ inclusion de SNPs des hétérosomes dans tests de sélection  
= risque de 'gonfler' artificiellement le fond neutre (génomique-wide)

Annotation du Z par séquençage Individuel :

- Séquençage de 4 bibliothèques (2 ♂ et 2 ♀) en 2x125 nt sur 4 lignes d'un séquenceur Illumina HiSeq
- Analyse du ratio des profondeurs entre bibliothèques mâles et femelles (méthode du AD-ratio: Bidon *et al.* 2015 *Genome Biology and Evolution*)

AD-ratio (♂ / ♀) = 2 pour le Z

AD-ratio (♂ / ♀) = 1 pour les autosomes

~~AD-ratio (♂ / ♀) = 0 pour le W (pas dans la référence génomique)~~

# Annotation de l'hétérochromosome Z (Gschloessl et al., en prep.)

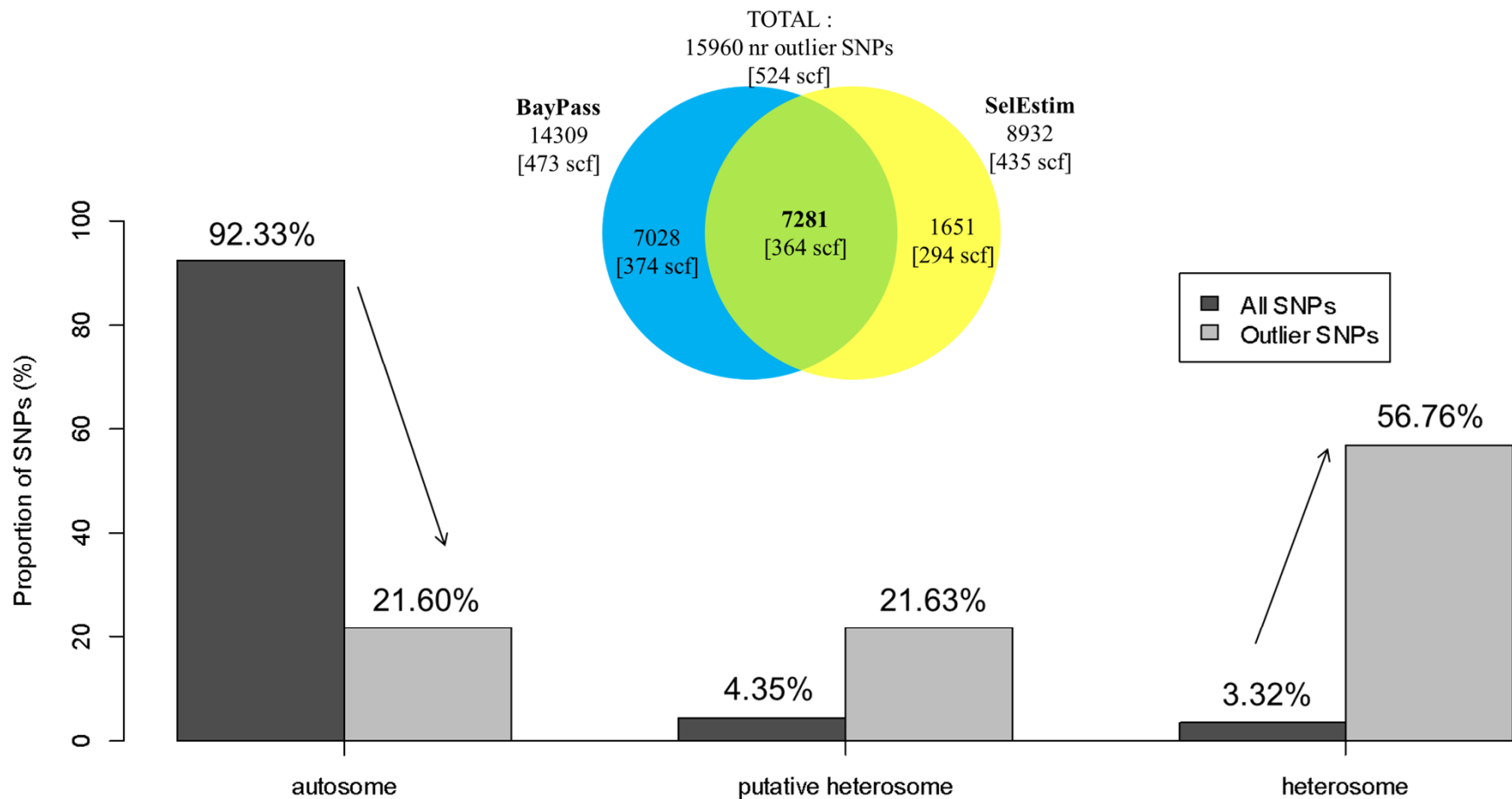
## Validation croisée avec les gènes connus:

Candidate_gene	Localisation dans la référence génomique (Blat)	Localisation autosomes vs hétérosome (connu)	Annotation AD-ratio
OR1	newscf_6616	Z	Heterosome Z
OR1	newscf_7058	Z	Heterosome Z
OR1	newscf_513	Z	Putative Heterosome Z
OR1	newscf_1365	Z	Putative Heterosome Z
OR2	newscf_7503	?	Autosome
OR2	newscf_998	?	Putative Heterosome Z
OR2	newscf_33824	?	Putative Heterosome Z
OR3	newscf_1365	Z	Putative Heterosome Z
OR3	newscf_7058	Z	Heterosome Z
OR3	newscf_513	Z	Heterosome Z
OR4	newscf_2328	Autosome	Autosome
OR4	newscf_25926	Autosome	Autosome
OR5	newscf_2328	Autosome	Autosome
OR6	newscf_8038	Z	Heterosome Z
OR6	newscf_16790	Z	Heterosome Z
OR6	newscf_1296	Z	Putative Heterosome Z
KET	newscf_106	Z	Heterosome Z
TPi	newscf_1555	Z	Heterosome Z

## Annotation de l'hétérochromosome Z (Gschloessl et al., en prep.)

Résultats :

	Autosomes	Putative Z	Hétérosome Z	NA
N scf	25760	4874	1273	18332
Size	342Mb	15Mb	29Mb	32Mb
N SNPs	3,272,724	154,109	117,780	



Outliers enrichis en SNPs situés sur l'hétérochromosome Z !!!

Erreurs de type I (faux-positifs parmi les SNPs du Z) et II (faux-négatifs parmi les SNPs des autosomes)

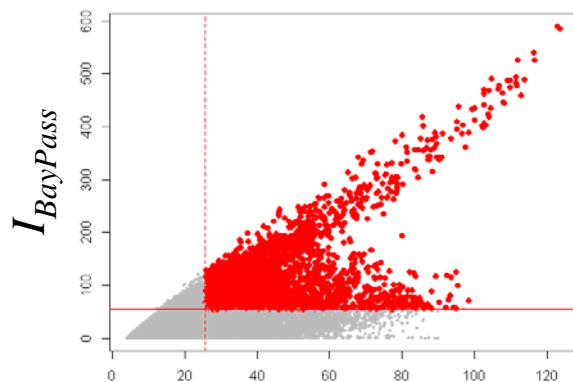
(1) Ré-Analyse des SNPs autosomes et hétérosome Z séparément

(2) Analyse des SNPs situés dans le génome mitochondrial

→ comparaison des proportions d'outliers entre autosomes, hétérosome Z et mitochondrie

BayPass

Autosomes



XtX

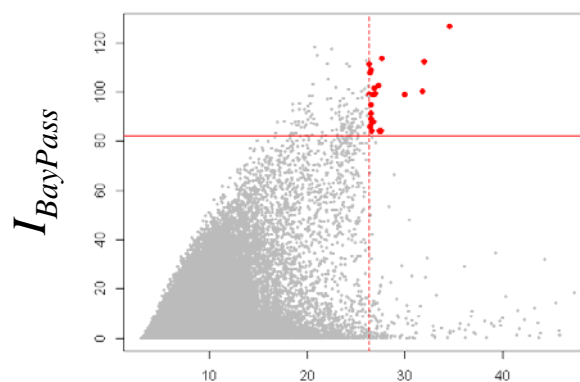
4138 outliers

3 272 724 SNPs



0.132%

Hétérosome Z



XtX

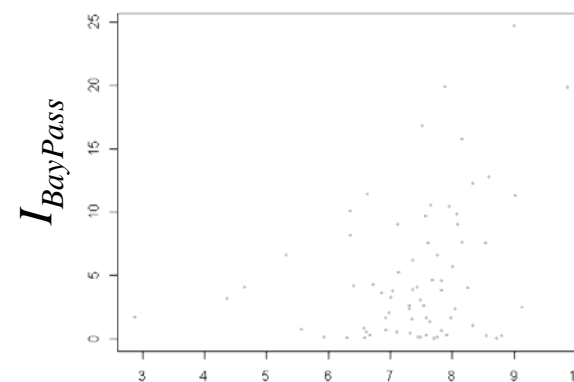
22 outliers

117 780 SNPs



0.018%

Mitochondrie



XtX

0 outliers

73 SNPs



0%

(3) Soumission de la référence génomique à Lepidodb et annotation fonctionnelle des régions génomiques avoisinant les SNPs d'intérêt (collaboration avec Bernhard Gschloessl)

→ id. voies métaboliques et fonctions biologiques des régions d'intérêt

(4) Génotypage massif de 500-1000 SNPs pour de nombreuses localités à travers l'Europe (~60 localités) :

- SNPs candidats à l'adaptation → validation de SNPs sous sélection
- SNPs neutres → reconstruction de l'histoire biogéographique du genre *Ostrinia* en Europe (tests scénarios démographiques et d'admixture)

→ meilleure compréhension du processus de divergence génomique et de spéciation (cas particulier du sud de l'Europe)

# Remerciements

➤ ANR « Adaptome » coordonné par Réjane Streiff



Sabine Nidelet



Bernhard Gschloessl



Philippe Audiot



Renaud Vitalis



Carole Smadja



Mathieu Gautier

