SNP discovery in pea: A powerful tool for academic research and breeding

Gilles Boutet [1]
Duarte J. [2]
Alves Carvalho S. [1,3]
Lavaud C. [1,]
Uricaru R. [1,3]
Peterlongo P. [3]
Pilet-Nayel M-L [1,]
Baranger A. [1,]
Rivière, N. [2]

[1]INRA, UMR1349 IGEPP, 35653 LE RHEU France
[2]BIOGEMMA, Upstream Genomics Team, 63720 CHAPPES France
[3]INRIA Rennes – Bretagne Atlantique/IRISA, EPI GenScale, RENNES, France

IFLRC VI & ICLGG VII
6th International Food Legumes Research Conference
7th International Conference on Legume Genetics and Genomics.

2014 / 07 / 07

Introduction

***Pisum sativum* :**

- **4.3 Gb** complex **genome**
- **Limited** genomic and sequencing **resources**
- Significant **challenges** against **biotic stress**
- breakthrough in **Marker Assisted Selection** needed by french breeders

Introduction

***Pisum sativum* :**

- **4.3 Gb** complex **genome**
- **Limited** genomic and sequencing **resources**
- Significant **challenges** against **biotic stress**
- breakthrough in **Marker Assisted Selection** needed by french breeders

Massive sequencing & SNP markers development
Two complementary NGS approaches :

Introduction

***Pisum sativum* :**

- **4.3 Gb** complex **genome**
- **Limited** genomic and sequencing **resources**
- Significant **challenges** against **biotic stress**
- breakthrough in **Marker Assisted Selection** needed by french breeders

Massive sequencing & SNP markers development
Two complementary NGS approaches :

1. **Sequencing** of **full length standardized cDNA** for 8 pea genotypes

Introduction

***Pisum sativum* :**

- **4.3 Gb** complex **genome**
- **Limited** genomic and sequencing **resources**
- Significant **challenges** against **biotic stress**
- breakthrough in **Marker Assisted Selection** needed by french breeders

Massive sequencing & SNP markers development
Two complementary NGS approaches :

1. **Sequencing** of **full length standardized cDNA** for 8 pea genotypes

2. **Genotyping by Sequencing** of **gDNA** for **48 RILs** segregating for a major biotic stress resistance

_01

454 Sequencing of eight pea cDNA normalized libraries

Boutet G., Duarte J., Alves Carvalho S., Lavaud C., Uricaru R., Peterlongo P., Pilet-Nayel M-L., Baranger A., Rivière, N.

Sequencing of eight pea cDNA normalized libraries

Full-length cDNA libraries

↓

Normalization

↓

454 sequencing

Duarte *et al.* 2014

Boutet G., Duarte J., Alves Carvalho S., Lavaud C., Uricaru R., Peterlongo P., Pilet-Nayel M-L., Baranger A., Rivière, N.

Sequencing of eight pea cDNA normalized libraries

Full-length cDNA libraries

↓

Normalization

↓

454 sequencing

3.8 M reads / 1.4 GB / 8 génotypes

Duarte *et al.* 2014

Sequencing of eight pea cDNA normalized libraries

Full-length cDNA libraries

↓

Normalization

↓

454 sequencing

3.8 M reads / 1.4 GB / 8 génotypes

No pea genome reference sequence available

Sequencing of eight pea cDNA normalized libraries

Full-length cDNA libraries

↓

Normalization

↓

454 sequencing

3.8 M reads / 1.4 GB / 8 génotypes

→

No pea genome reference sequence available

De novo assembling

Sequencing of eight pea cDNA normalized libraries

Full-length cDNA libraries

↓

Normalization

↓

454 sequencing

3.8 M reads / 1.4 GB / 8 génotypes

→

No pea genome reference sequence available

De novo assembling

80% reads assembled

(Contig longest: 4,1kb)

69 k contigs

(74% with hit blast on *Medicago truncatula*)

Sequencing of eight pea cDNA normalized libraries

Full-length cDNA libraries

↓

Normalization

↓

454 sequencing

3.8 M reads / 1.4 GB / 8 génotypes

→

No pea genome reference sequence available

De novo assembling

80% reads assembled

(Contig longest: 4,1kb)

69 k contigs

(74% with hit blast on *Medicago truncatula*)

↓

Over 74 k SNPs / **35 k SNPs** considered as robust spread over 10 k contigs

(98% with a hit blast on the genome of *Medicago truncatula*)

Sequencing of eight pea cDNA normalized libraries

Full-length cDNA libraries

↓

Normalization

↓

454 sequencing

3.8 M reads / 1.4 GB / 8 génotypes

→

No pea genome reference sequence available

De novo assembling

80% reads assembled

(Contig longest: 4,1kb)

69 k contigs

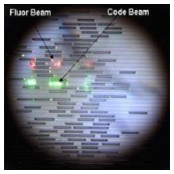
(74% with hit blast on *Medicago truncatula*)

↓

Over 74 k SNPs / **35 k SNPs** considered as robust spread over 10 k contigs

(98% with a hit blast on the genome of *Medicago truncatula*)

↓



1920 SNP genotyped on :

4 RILs population

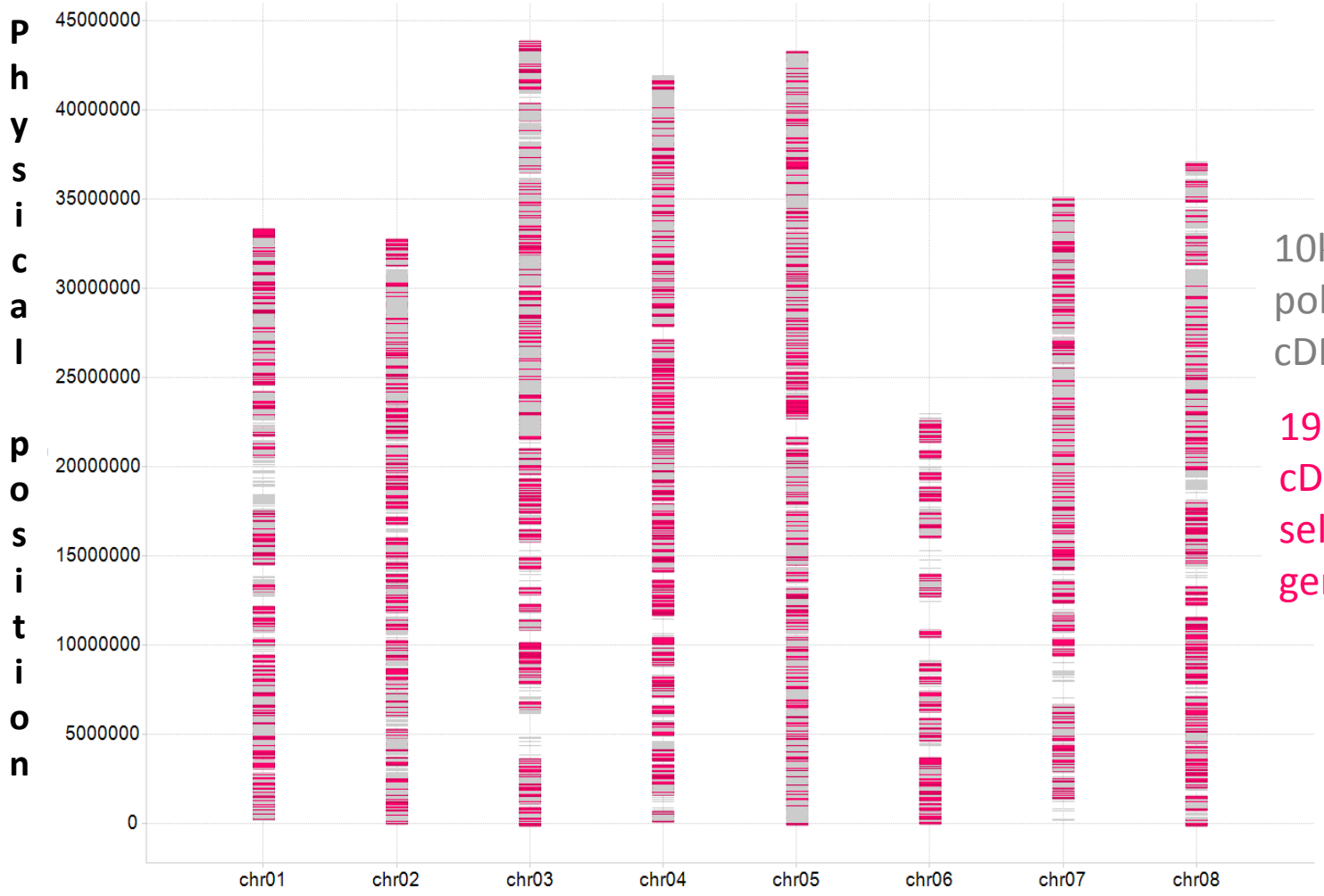
92 accession

biogemma

INRA SCIENCE & IMPACT

Duarte *et al.* 2014

Distribution of pea contigs along *Medicago truncatula* physical map

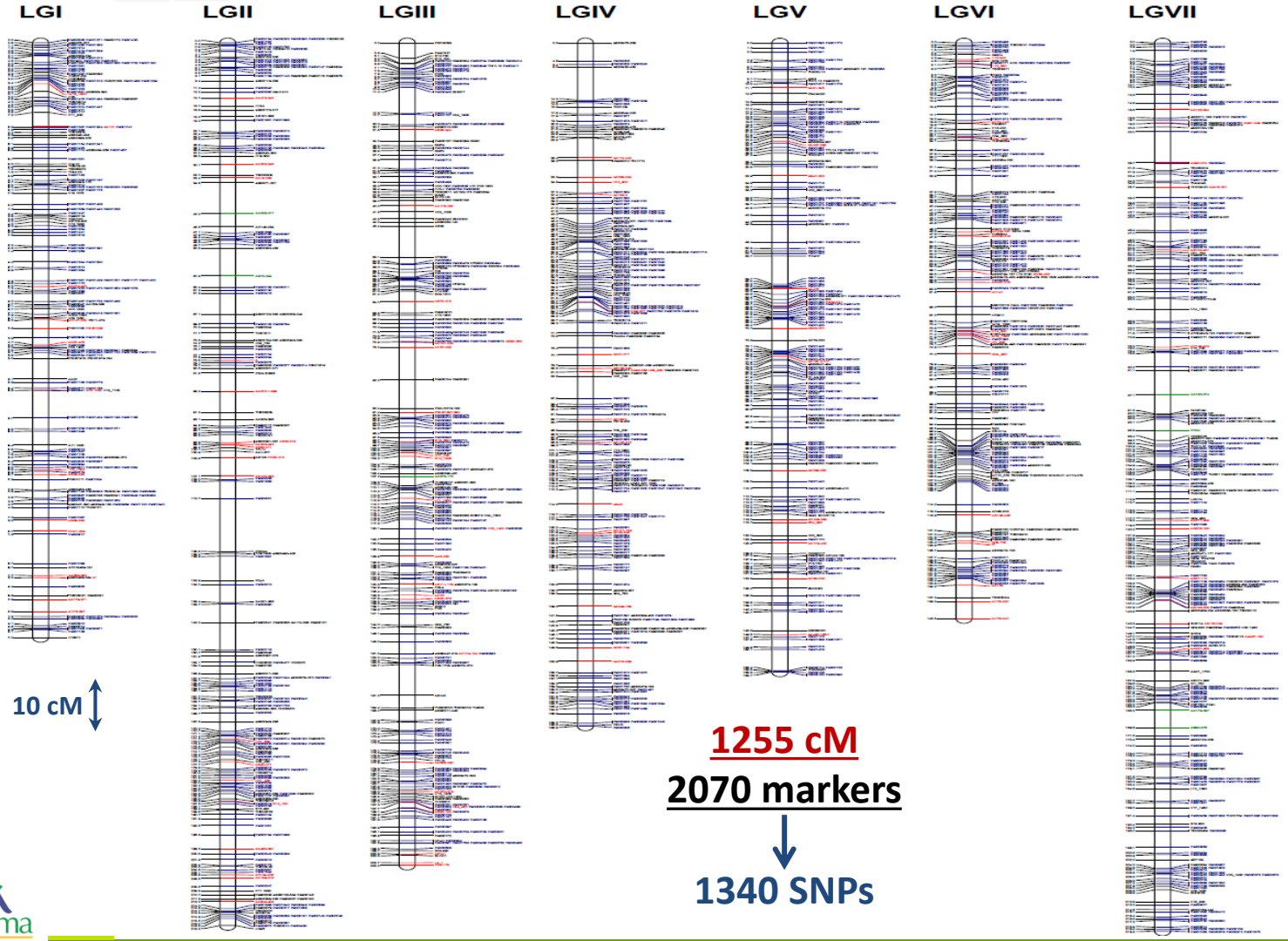


10k pea polymorphic cDNA contigs

1920 pea cDNA contigs selected for genotyping

Duarte *et al.* 2014

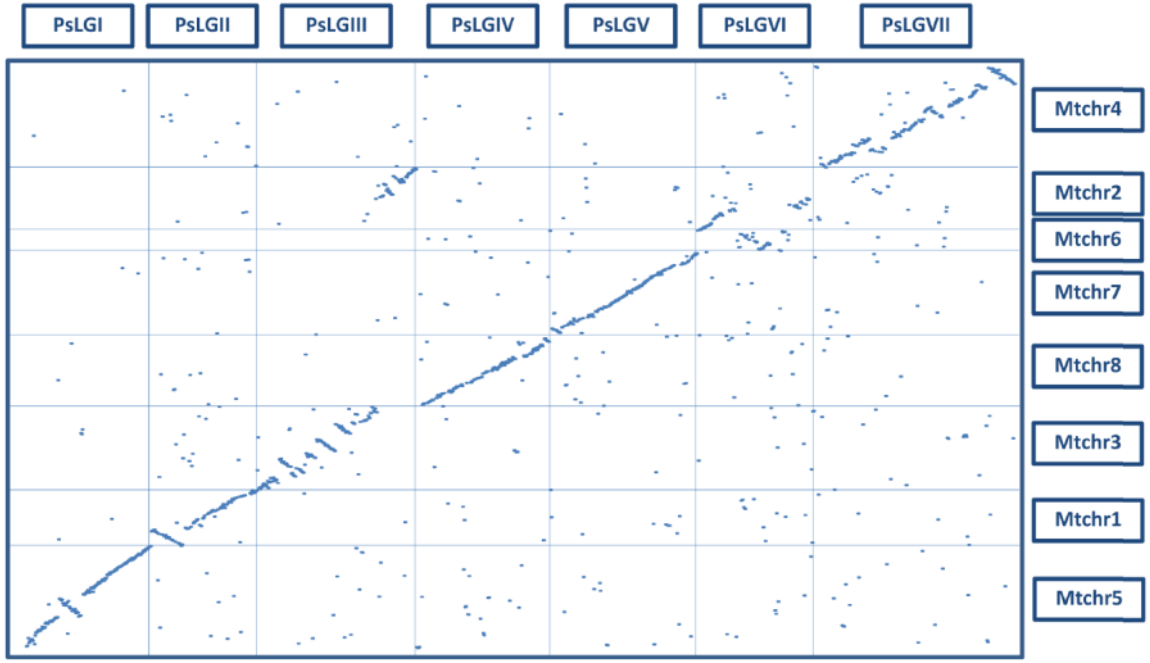
A high density reference composite genetic map anchored to the *M. truncatula* genome



Duarte *et al.* 2014

A high density reference composite genetic map anchored to the *M. truncatula* genome

1252 bridges between the two genomes



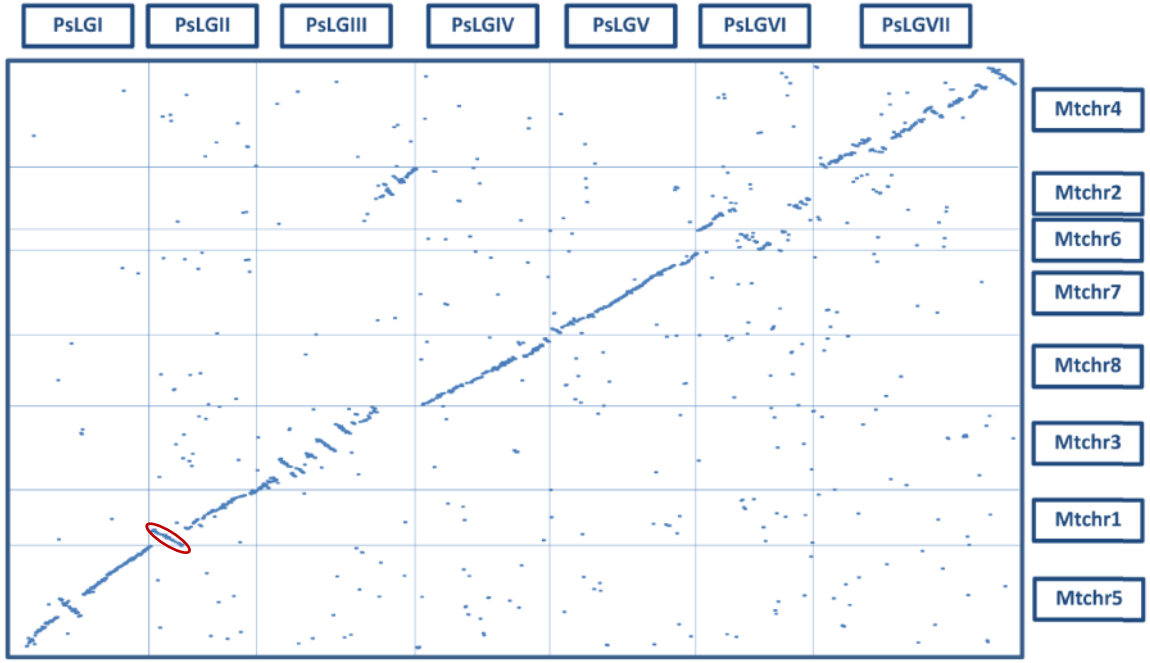
Syntenic relationships between the *P. sativum* LGs and the *M.truncatula* pseudo-chromosomes.

1340 SNPs

Duarte *et al.* 2014

A high density reference composite genetic map anchored to the *M. truncatula* genome

1252 bridges between the two genomes



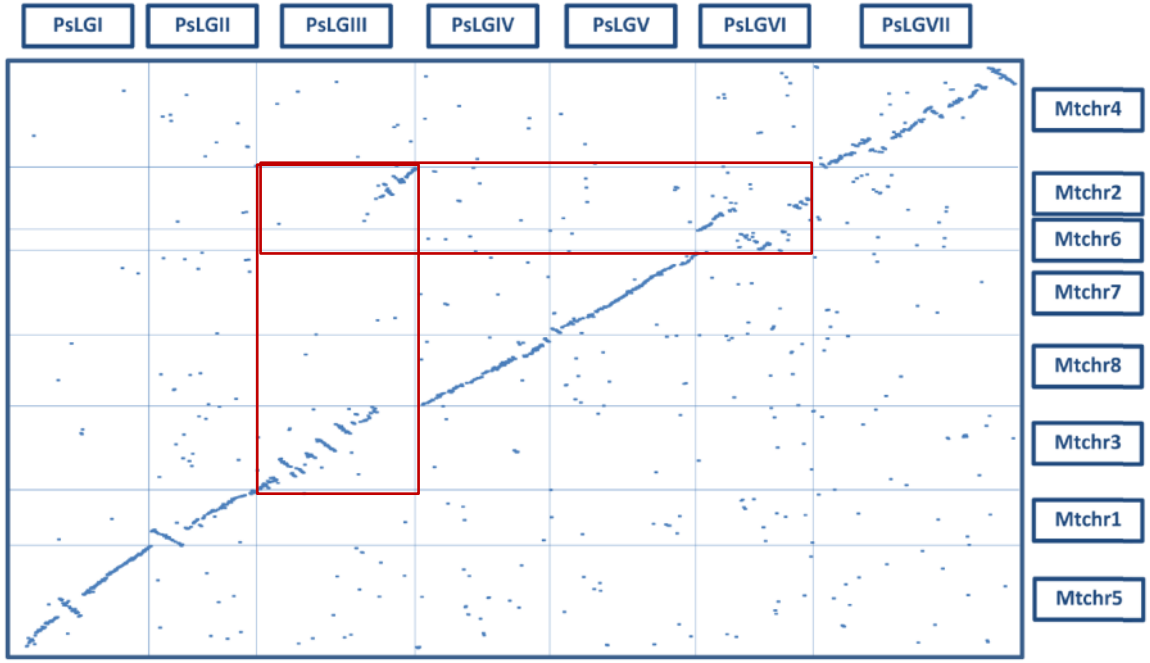
Syntenic relationships between the *P. sativum* LGs and the *M.truncatula* pseudo-chromosomes.

1340 SNPs

Duarte *et al.* 2014

A high density reference composite genetic map anchored to the *M. truncatula* genome

1252 bridges between the two genomes



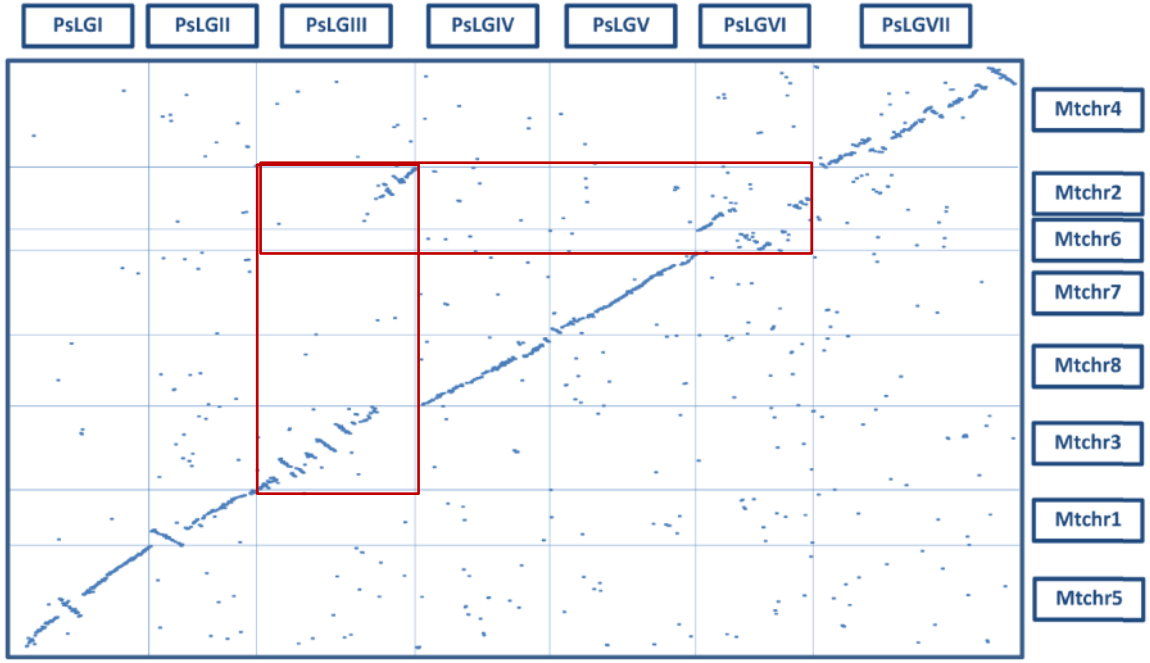
Syntenic relationships between the *P. sativum* LGs and the *M.truncatula* pseudo-chromosomes.

1340 SNPs

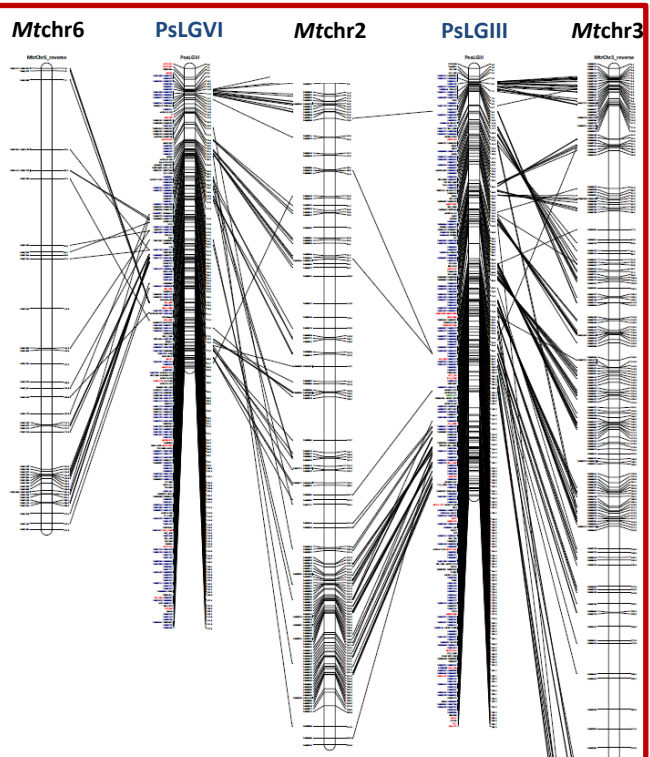
Duarte *et al.* 2014

A high density reference composite genetic map anchored to the *M. truncatula* genome

1252 bridges between the two genomes



Syntenic relationships between the *P. sativum* LGs and the *M.truncatula* pseudo-chromosomes.

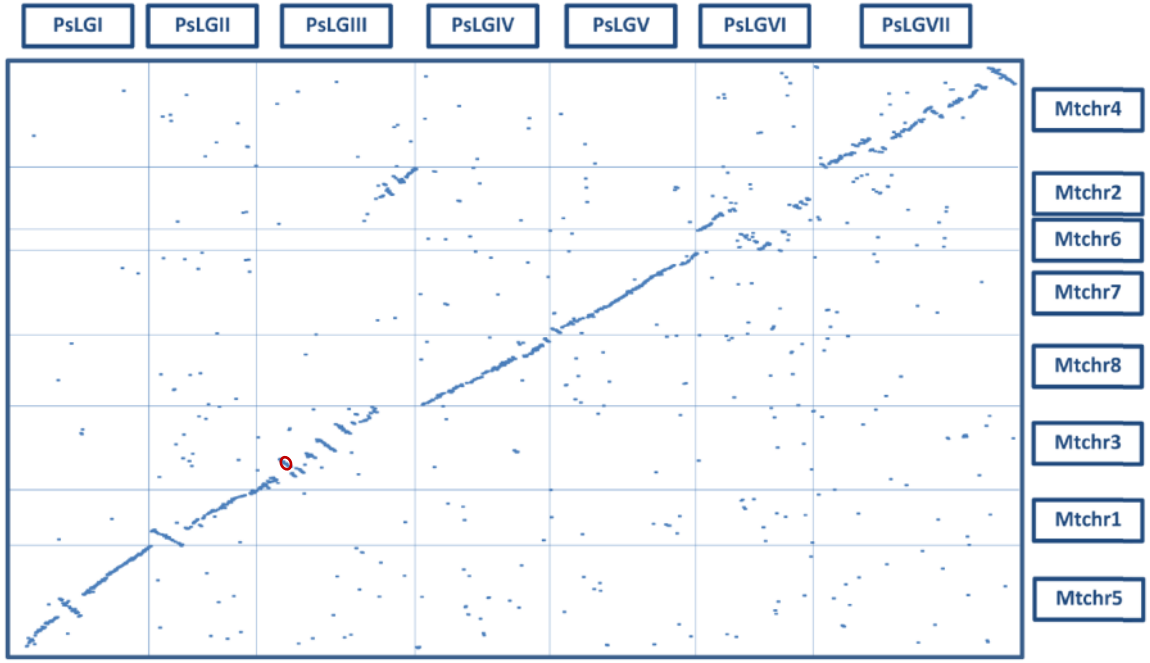


Collinear positions between the ***P.sativum*** composite genetic map and ***M.truncatula*** physical map

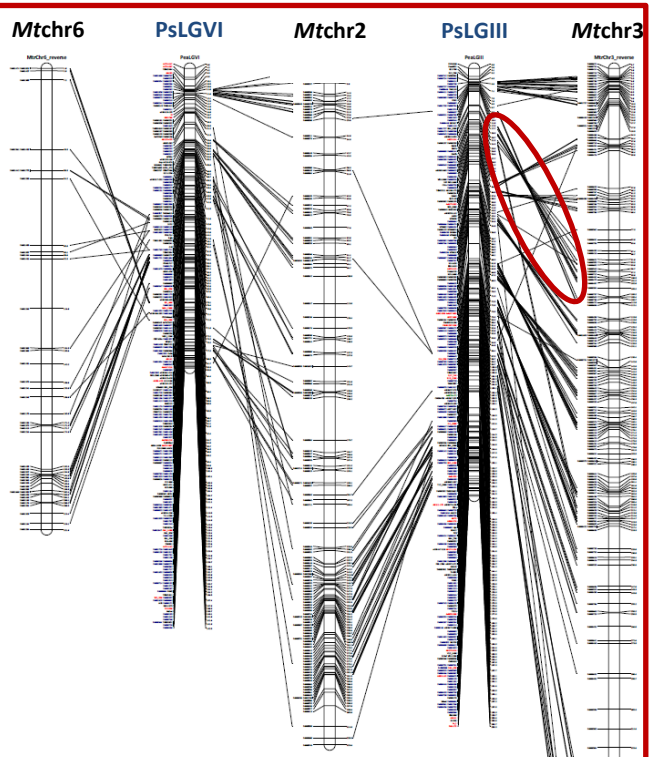
1340 SNPs

A high density reference composite genetic map anchored to the *M. truncatula* genome

1252 bridges between the two genomes



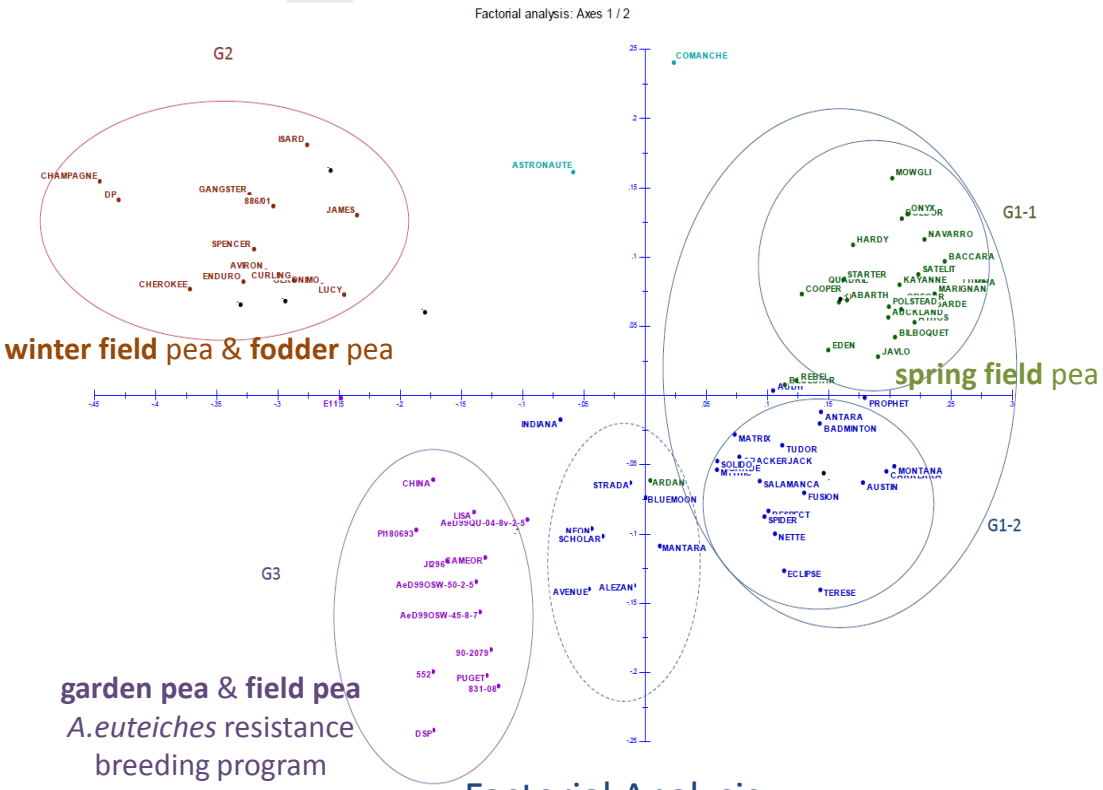
Syntenic relationships between the *P. sativum* LGs and the *M.truncatula* pseudo-chromosomes.



Collinear positions between the ***P.sativum*** composite genetic map and ***M.truncatula*** physical map

1340 SNPs

Diversity structuration of a collection of pea cultivars



winter field pea & **fodder** pea

spring field pea

garden pea & **field pea**
A.euteiches resistance
breeding program

Factorial Analysis

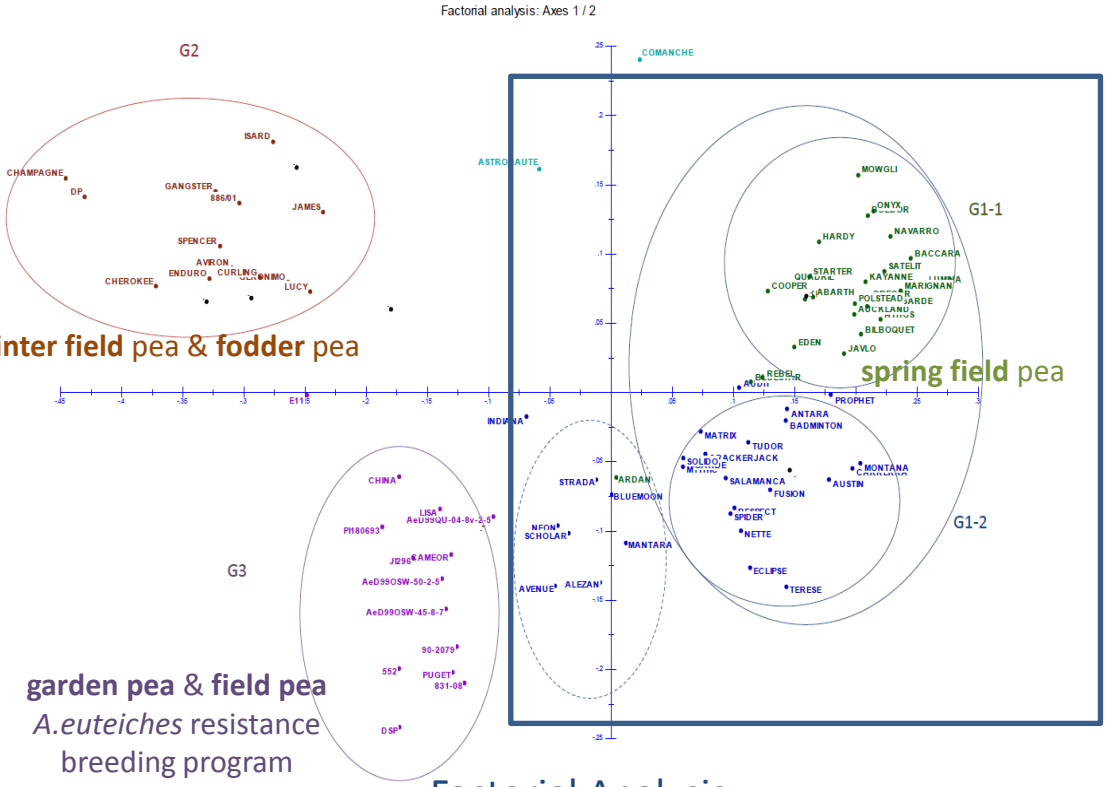
92 accessions and
cultivar diversity panel

1538 SNP markers

Duarte *et al.* 2014

Boutet G., Duarte J., Alves Carvalho S., Lavaud C., Uricaru R., Peterlongo P., Pilet-Nayel M-L., Baranger A., Rivière, N.

Diversity structuration of a collection of pea cultivars



winter field pea & **fodder** pea

spring field pea

garden pea & **field pea**
A.euteiches resistance
breeding program

Factorial Analysis

92 accessions and
cultivar diversity panel

1538 SNP markers

Duarte *et al.* 2014

Boutet G., Duarte J., Alves Carvalho S., Lavaud C., Uricaru R., Peterlongo P., Pilet-Nayel M-L., Baranger A., Rivière, N.

Diversity structuration of a collection of pea cultivars

winter field pea & **fodder** pea

spring field pea

garden pea & **field pea**
A.euteiches resistance
breeding program

Factorial Analysis

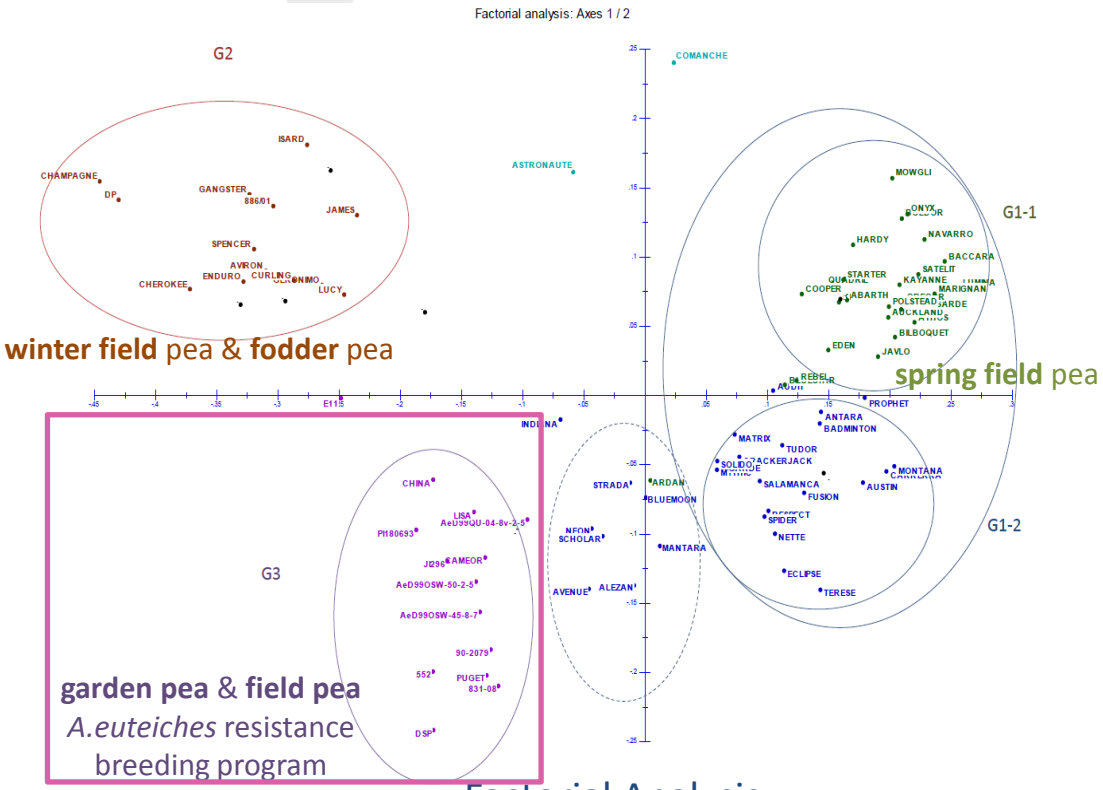
92 accessions and
cultivar diversity panel

1538 SNP markers

Duarte *et al.* 2014

Boutet G., Duarte J., Alves Carvalho S., Lavaud C., Uricaru R., Peterlongo P., Pilet-Nayel M-L., Baranger A., Rivière, N.

Diversity structuration of a collection of pea cultivars



winter field pea & **fodder** pea

spring field pea

garden pea & **field pea**
A.euteiches resistance
breeding program

Factorial Analysis

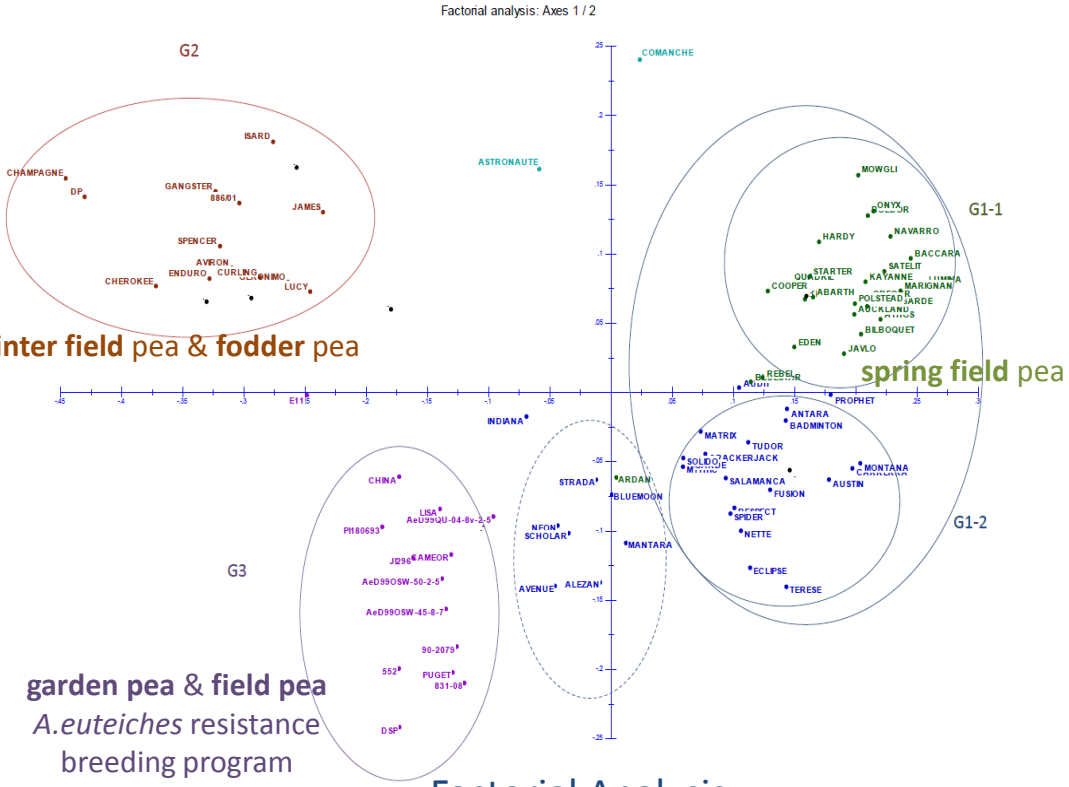
92 accessions and
cultivar diversity panel

1538 SNP markers

Duarte *et al.* 2014

Boutet G., Duarte J., Alves Carvalho S., Lavaud C., Uricaru R., Peterlongo P., Pilet-Nayel M-L., Baranger A., Rivière, N.

Diversity structuration of a collection of pea cultivars



winter field pea & **fodder** pea

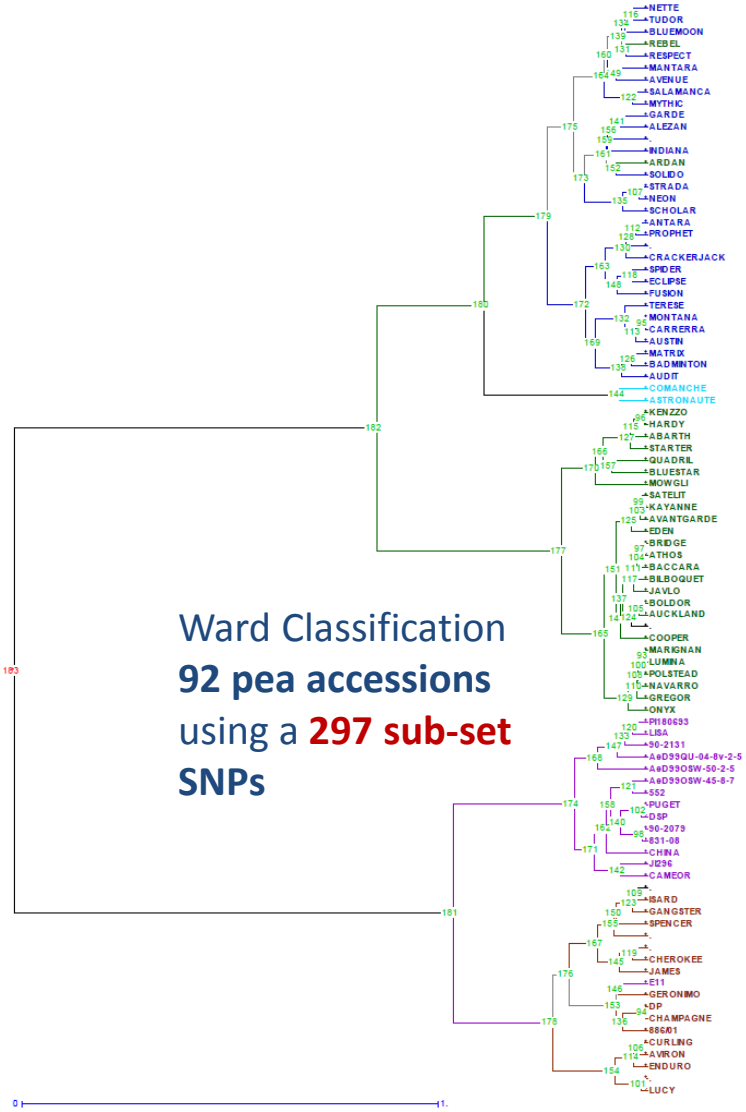
spring field pea

garden pea & **field pea**
A.euteiches resistance
breeding program

Factorial Analysis

92 accessions and
cultivar diversity panel

1538 SNP markers

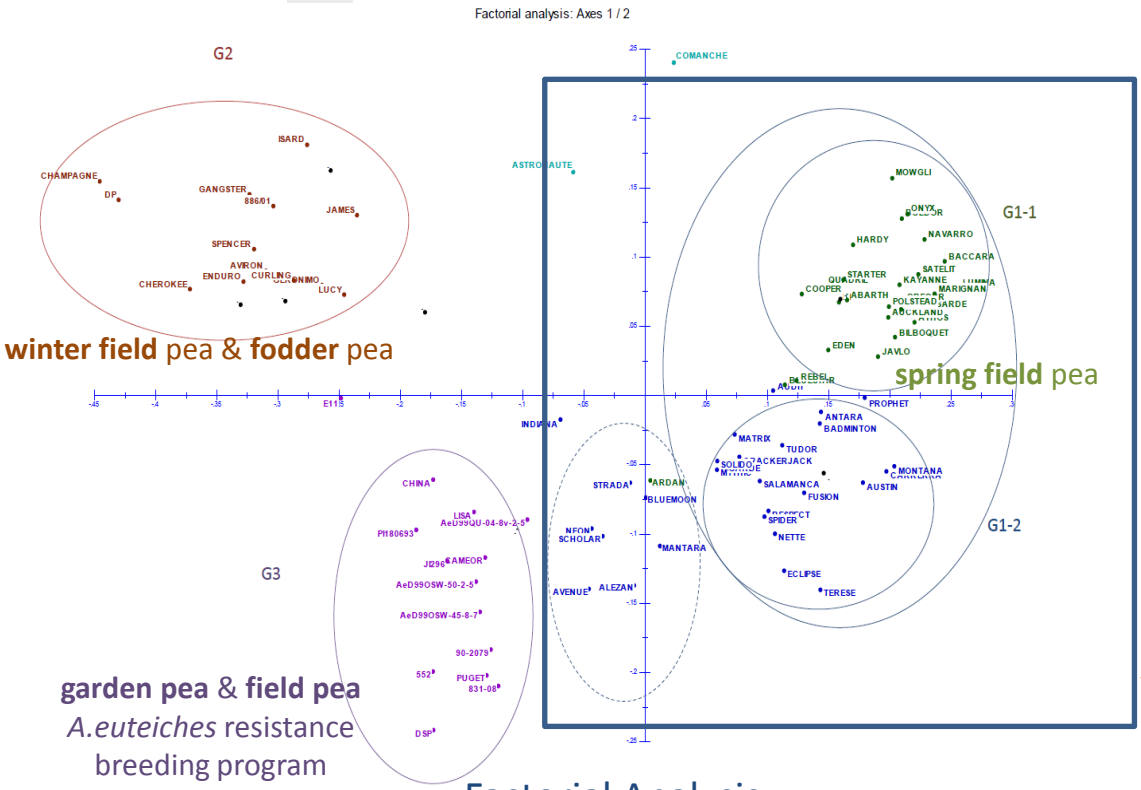


Ward Classification
92 pea accessions
using a **297 sub-set SNPs**

Duarte *et al.* 2014

Boutet G., Duarte J., Alves Carvalho S., Lavaud C., Uricaru R., Peterlongo P., Pilet-Nayel M-L., Baranger A., Rivière, N.

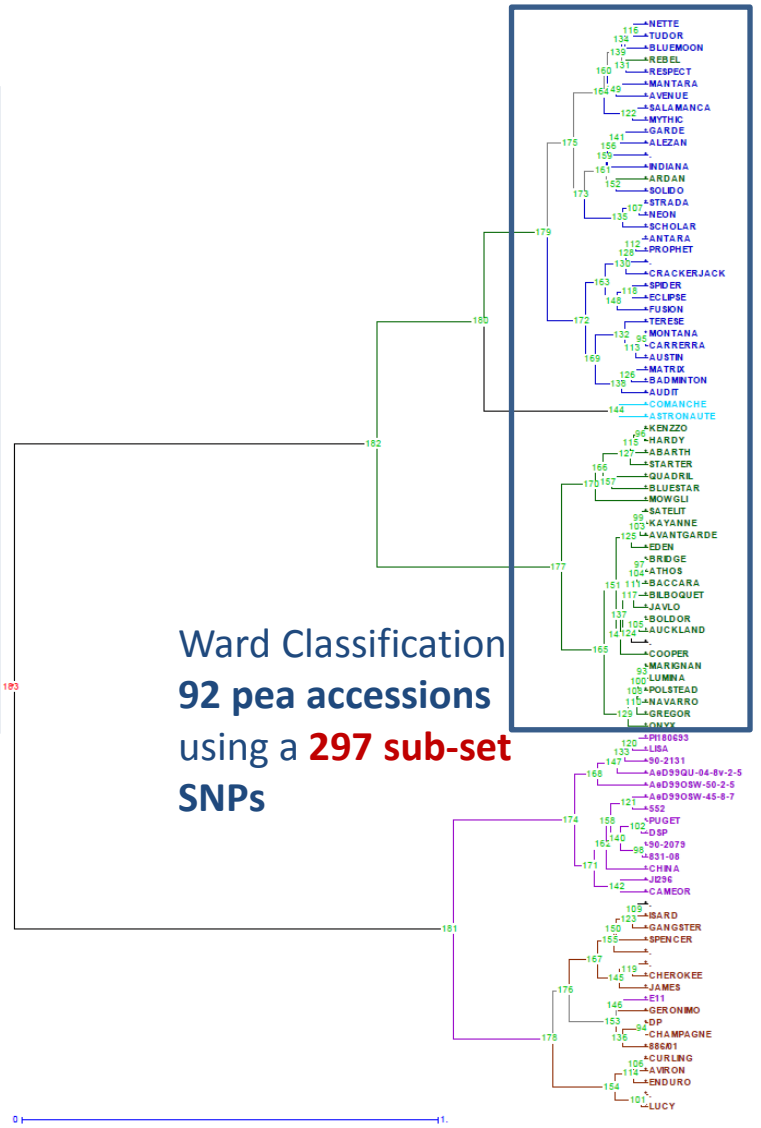
Diversity structuration of a collection of pea cultivars



Factorial Analysis

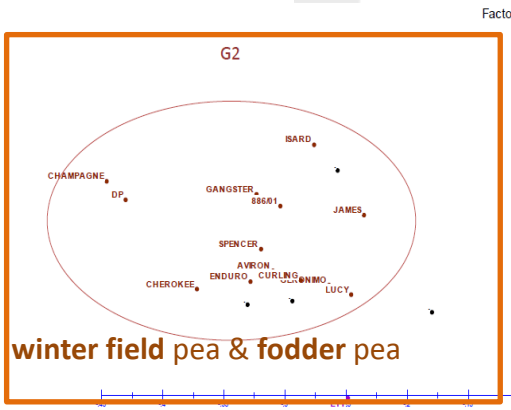
92 accessions and
cultivar diversity panel

1538 SNP markers

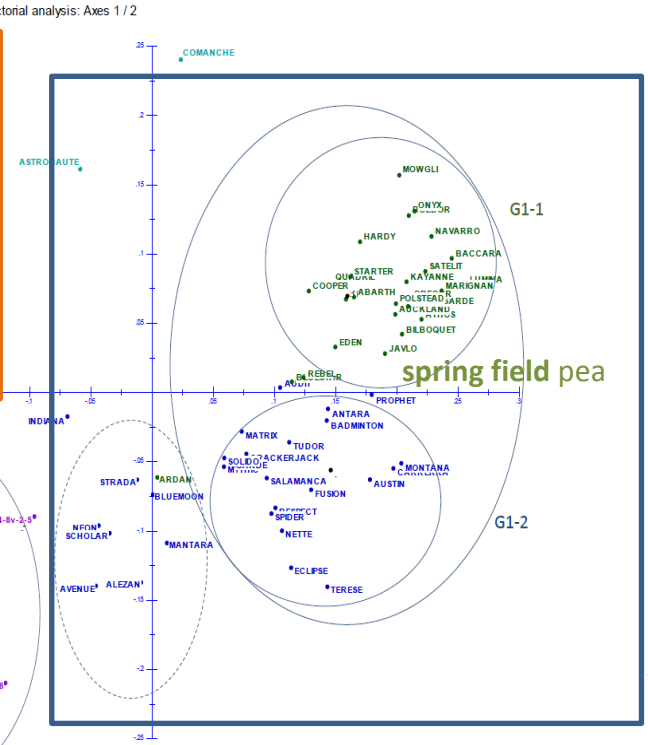


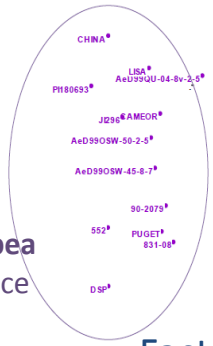
Ward Classification
92 pea accessions
using a **297 sub-set SNPs**

Diversity structuration of a collection of pea cultivars



winter field pea & **fodder** pea





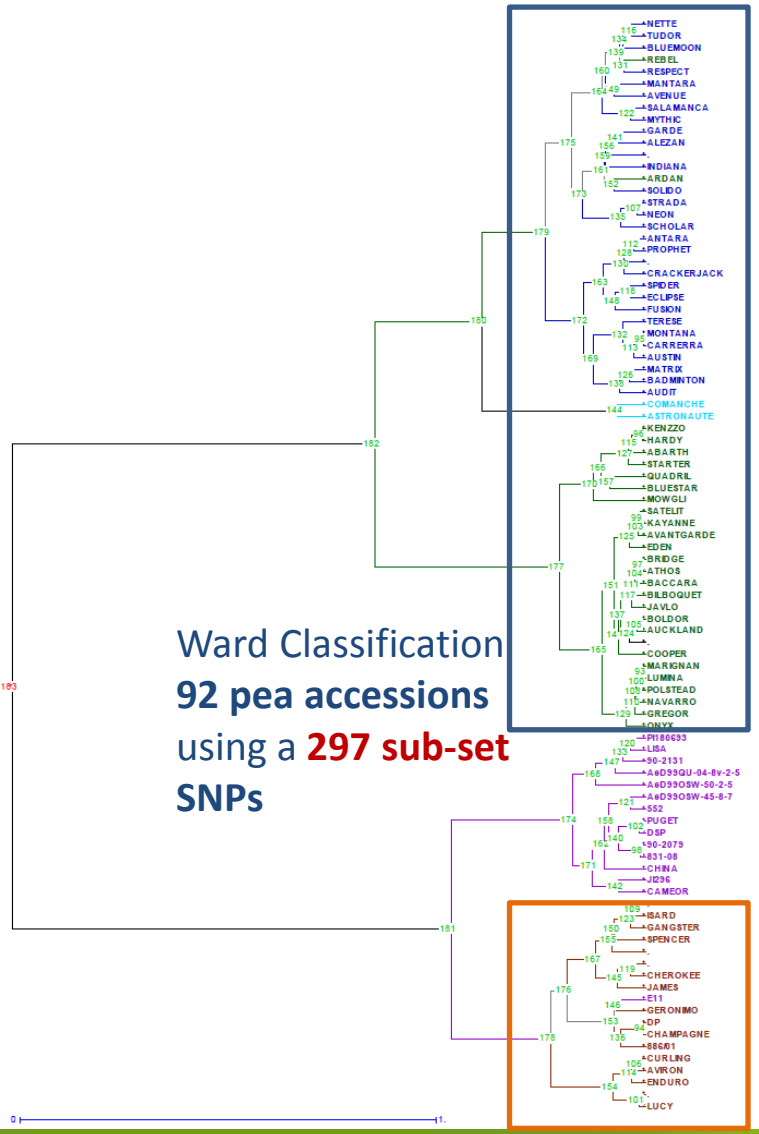
garden pea & **field pea**
A.euteiches resistance
breeding program

spring field pea

Factorial Analysis

92 accessions and
cultivar diversity panel

1538 SNP markers

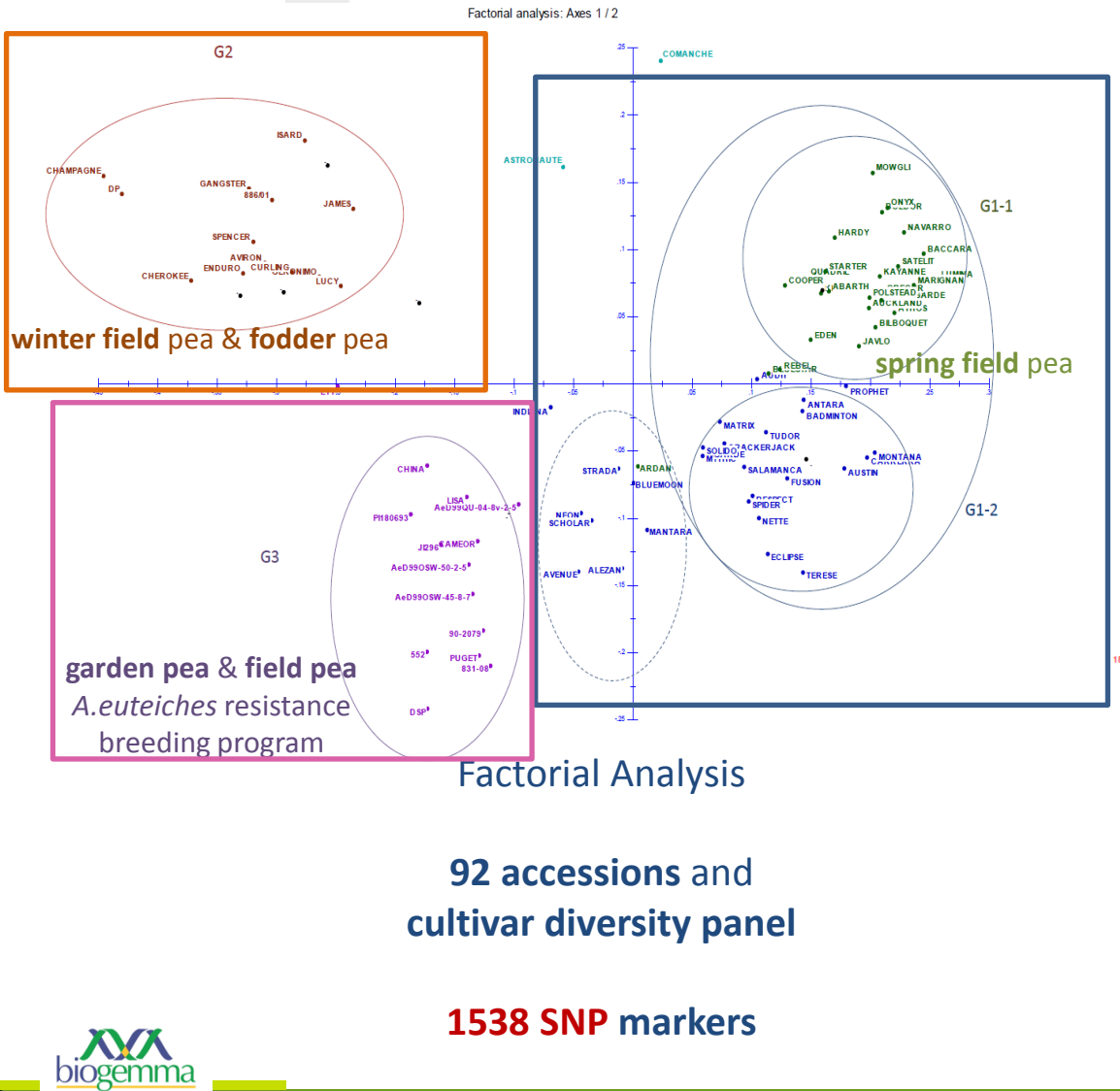


Ward Classification
92 pea accessions
using a **297 sub-set SNPs**

Duarte *et al.* 2014

Boutet G., Duarte J., Alves Carvalho S., Lavaud C., Uricaru R., Peterlongo P., Pilet-Nayel M-L., Baranger A., Rivière, N.

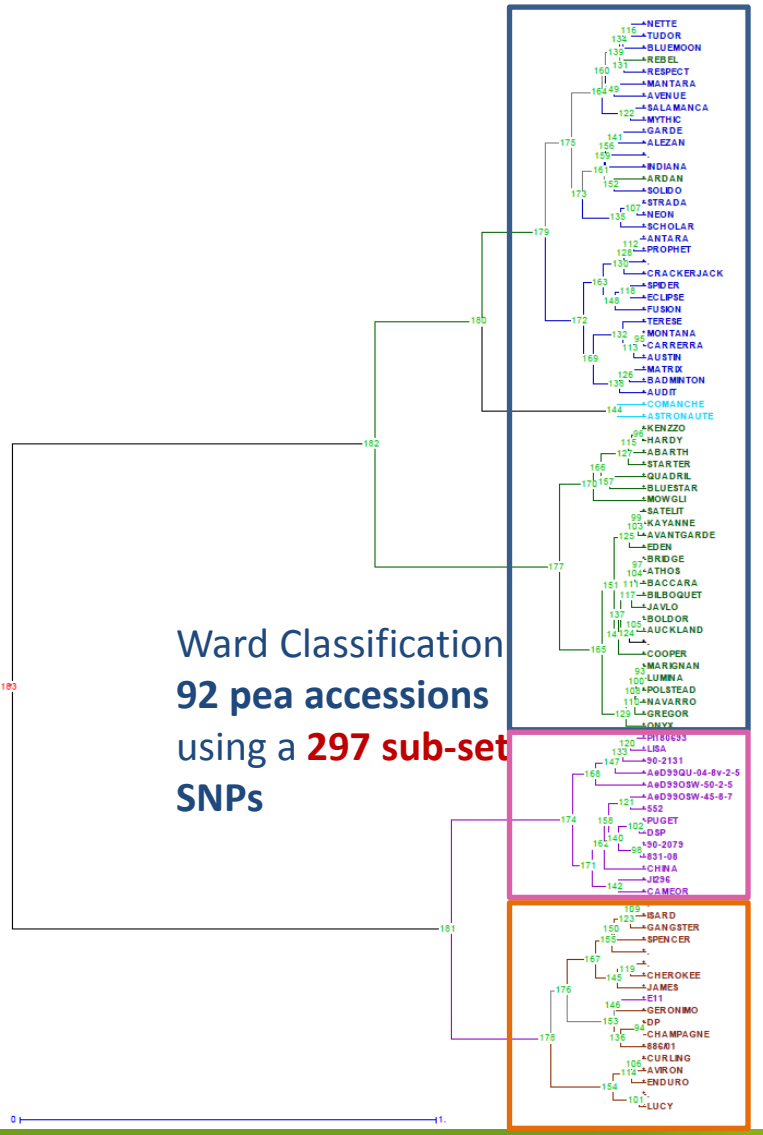
Diversity structuration of a collection of pea cultivars



Factorial Analysis

92 accessions and
cultivar diversity panel

1538 SNP markers



Ward Classification
92 pea accessions
using a **297 sub-set SNPs**

Duarte *et al.* 2014

Boutet G., Duarte J., Alves Carvalho S., Lavaud C., Uricaru R., Peterlongo P., Pilet-Nayel M-L., Baranger A., Rivière, N.

Aphanomyces euteiches disease

_02

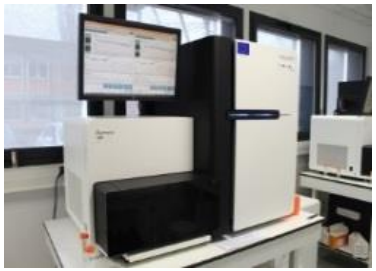
Genotyping By Sequencing on genomic DNA libraries from a 48 RILs mapping population segregating for resistance to *Aphanomyces euteiches*

Boutet G., Duarte J., Alves Carvalho S., Lavaud C., Uricaru R., Peterlongo P., Pilet-Nayel M-L., Baranger A., Rivière, N.

Genotyping By Sequencing on genomic DNA libraries from a 48 RILs mapping population

50 Genomic DNA libraries
(48 Rils + Baccara & PI180693)

INRA
Get-Plage
platform

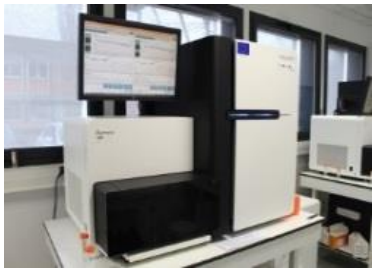


Illumina Hiseq2000 sequencing
(2 libraries/lane)

Genotyping By Sequencing on genomic DNA libraries from a 48 RILs mapping population

50 Genomic DNA libraries
(48 Rils + Baccara & PI180693)

INRA
Get-Plage
platform



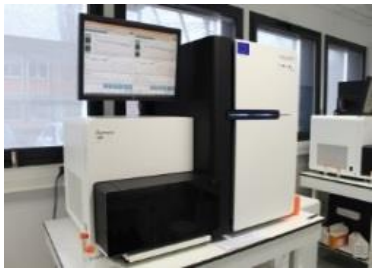
Illumina Hiseq2000 sequencing
(2 libraries/lane)

8.8 G reads / 877 GB

Genotyping By Sequencing on genomic DNA libraries from a 48 RILs mapping population

50 Genomic DNA libraries
(48 Rils + Baccara & PI180693)

INRA
Get-Plage
platform



Illumina Hiseq2000 sequencing
(2 libraries/lane)

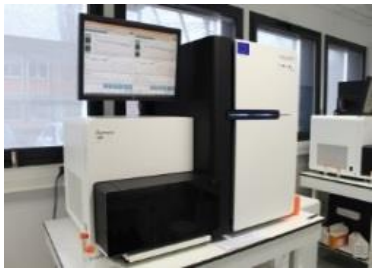
8.8 G reads / 877 GB

No pea genome reference sequence available

Genotyping By Sequencing on genomic DNA libraries from a 48 RILs mapping population

50 Genomic DNA libraries
(48 Rils + Baccara & PI180693)

INRA
Get-Plage
platform



Illumina Hiseq2000 sequencing
(2 libraries/lane)

8.8 G reads / 877 GB

No pea genome reference sequence available

discoSnp method

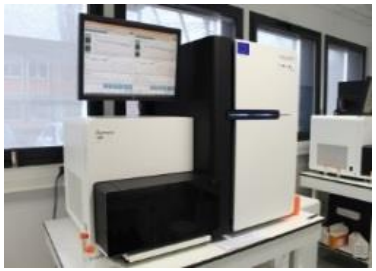
https://colibread.inria.fr/software/discosnp/

INRIA Genscale team
Uricaru *& al.* submitted

Genotyping By Sequencing on genomic DNA libraries from a 48 RILs mapping population

50 Genomic DNA libraries
(48 Rils + Baccara & PI180693)

INRA
Get-Plage
platform



Illumina Hiseq2000 sequencing
(2 libraries/lane)

8.8 G reads / 877 GB

No pea genome reference sequence available

discoSnp method

https://colibread.inria.fr/software/discosnp/

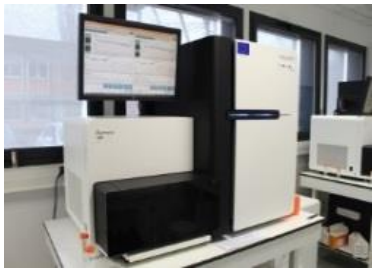
INRIA Genscale team
Uricaru *& al.* submitted

No reference genome needed

Genotyping By Sequencing on genomic DNA libraries from a 48 RILs mapping population

50 Genomic DNA libraries
(48 Rils + Baccara & PI180693)

INRA
Get-Plage
platform



Illumina Hiseq2000 sequencing
(2 libraries/lane)

8.8 G reads / 877 GB

No pea genome reference sequence available

discoSnp method

https://colibread.inria.fr/software/discosnp/

INRIA Genscale team
Uricaru *& al.* submitted

No reference genome needed

No Data assembly needed

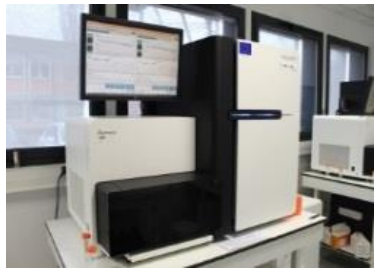
biogemma

INRA SCIENCE & IMPACT

Genotyping By Sequencing on genomic DNA libraries from a 48 RILs mapping population

50 Genomic DNA libraries
(48 Rils + Baccara & PI180693)

INRA
Get-Plage
platform



Illumina Hiseq2000 sequencing
(2 libraries/lane)

8.8 G reads / 877 GB

No pea genome reference sequence available

discoSnp method

https://colibread.inria.fr/software/discosnp/

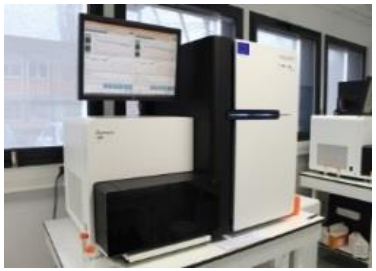
INRIA Genscale team
Uricaru *& al.* submitted

- No reference genome needed
- No Data assembly needed
- Quick & Low Memory

Genotyping By Sequencing on genomic DNA libraries from a 48 RILs mapping population

50 Genomic DNA libraries
(48 Rils + Baccara & PI180693)

INRA
Get-Plage
platform



Illumina Hiseq2000 sequencing
(2 libraries/lane)

8.8 G reads / 877 GB

No pea genome reference sequence available

discoSnp method

https://colibread.inria.fr/software/discosnp/

INRIA Genscale team
Uricaru *& al.* submitted

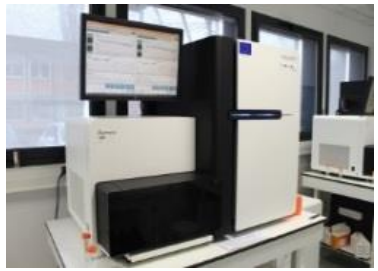
discoSnp module 1

Kissnp2: **SNPs detection** from sets of reads
(based on the **de Bruijn Graph**)

Genotyping By Sequencing on genomic DNA libraries from a 48 RILs mapping population

50 Genomic DNA libraries
(48 Rils + Baccara & PI180693)

INRA
Get-Plage
platform



Illumina Hiseq2000 sequencing
(2 libraries/lane)

8.8 G reads / 877 GB

No pea genome reference sequence available

discoSnp method

https://colibread.inria.fr/software/discosnp/

INRIA Genscale team
Uricaru *& al.* submitted

discoSnp module 1

Kissnp2: **SNPs detection** from sets of reads (based on the **de Bruijn Graph**)

discoSnp module 2

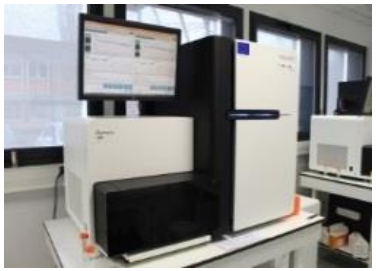
Kissreads: **kissnp2 results improving** / each read / each SNP:

1. **read coverage** calculating
2. **quality** of **reads** who generated SNP polymorphism

Genotyping By Sequencing on genomic DNA libraries from a 48 RILs mapping population

50 Genomic DNA libraries
(48 Rils + Baccara & PI180693)

INRA
Get-Plage
platform



Illumina Hiseq2000 sequencing
(2 libraries/lane)

8.8 G reads / 877 GB

No pea genome reference sequence available

discoSnp method

https://colibread.inria.fr/software/discosnp/

INRIA Genscale team
Uricaru *& al.* submitted

discoSnp module 1

Kissnp2: **SNPs detection** from sets of reads
(based on the **de Bruijn Graph**)

discoSnp module 2

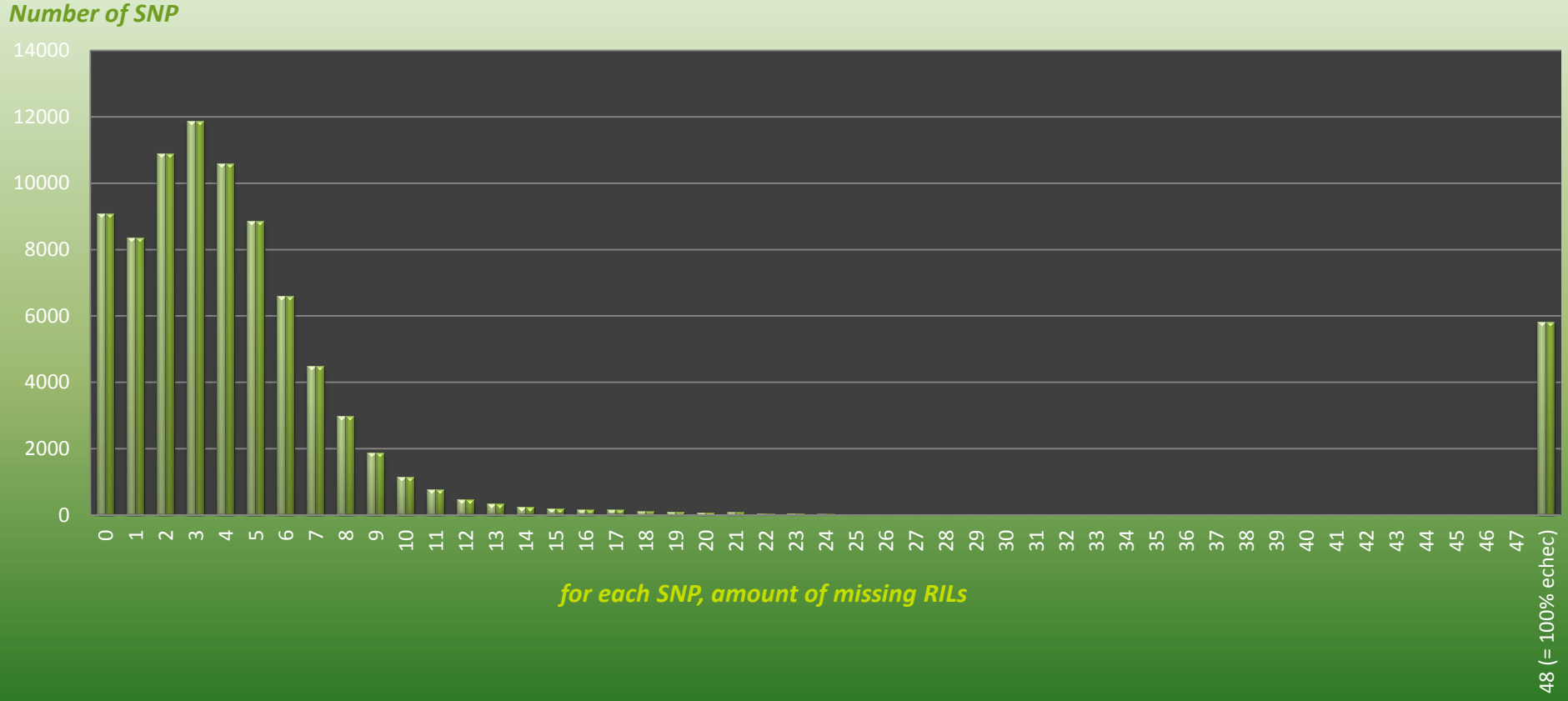
Kissreads: **kissnp2 results improving** / each read / each SNP:

1. **read coverage** calculating
2. **quality** of **reads** who generated SNP polymorphism

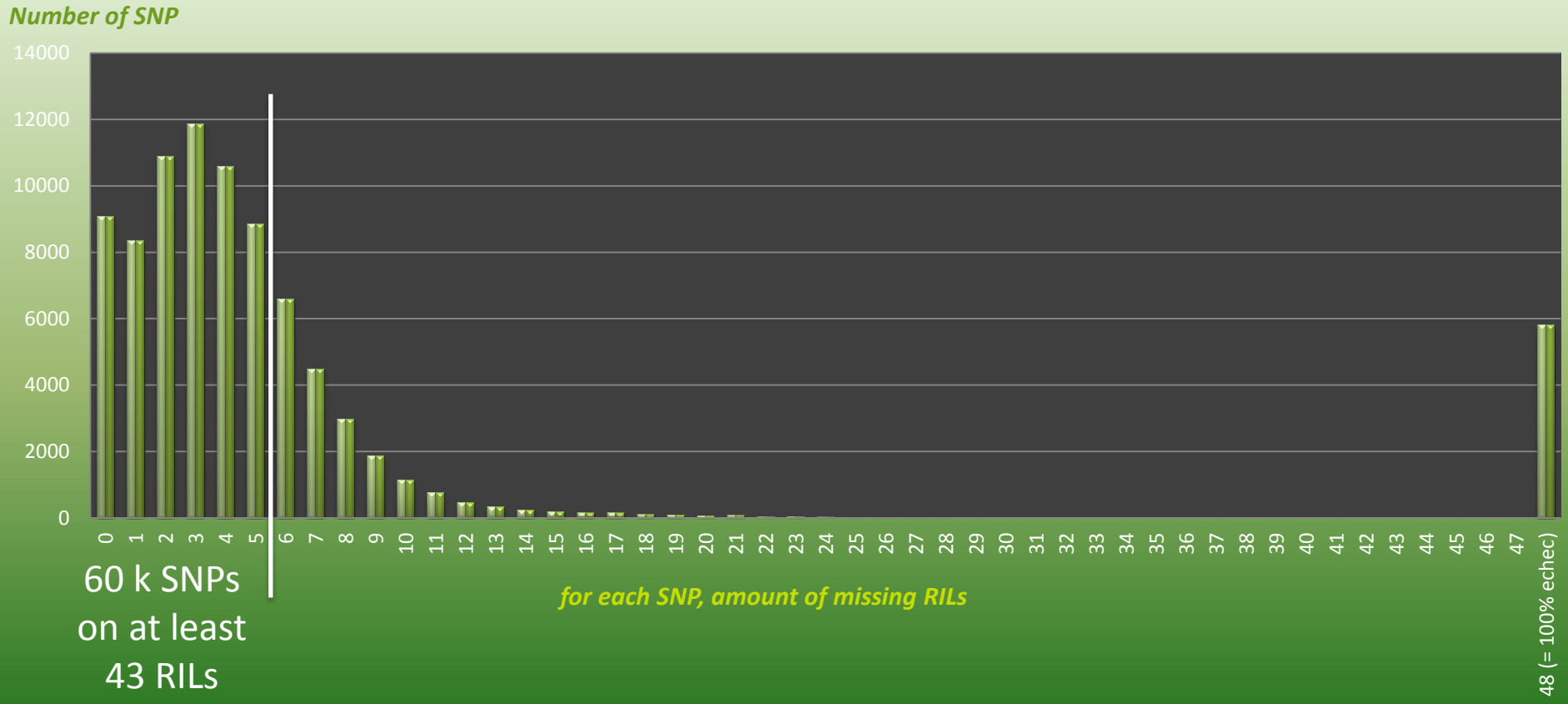
Post-discoSnp "project-specific" filters :

1. SNP **"putative false heterozygote"** coverage filter
2. **Minor allele coverage** filter

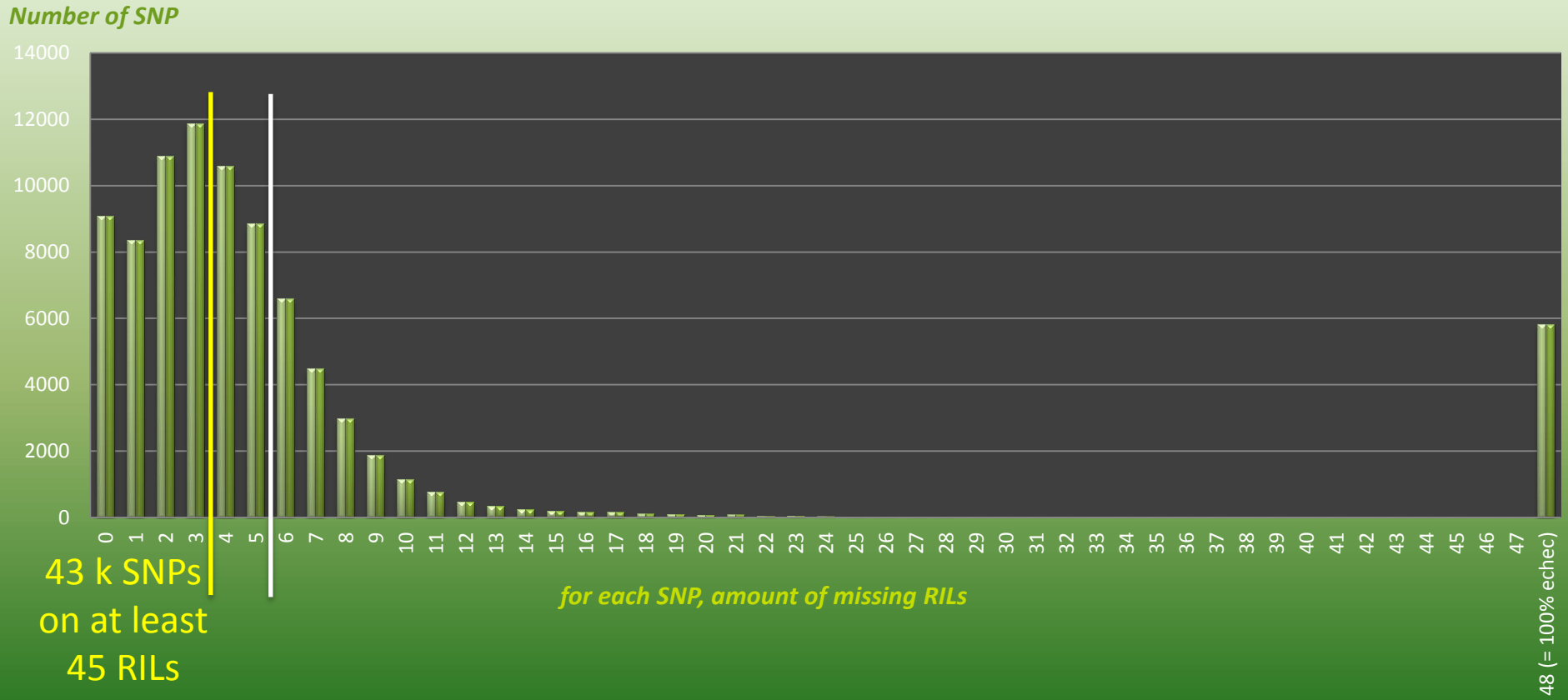
Over 75,000 genomic SNPs considered as robust



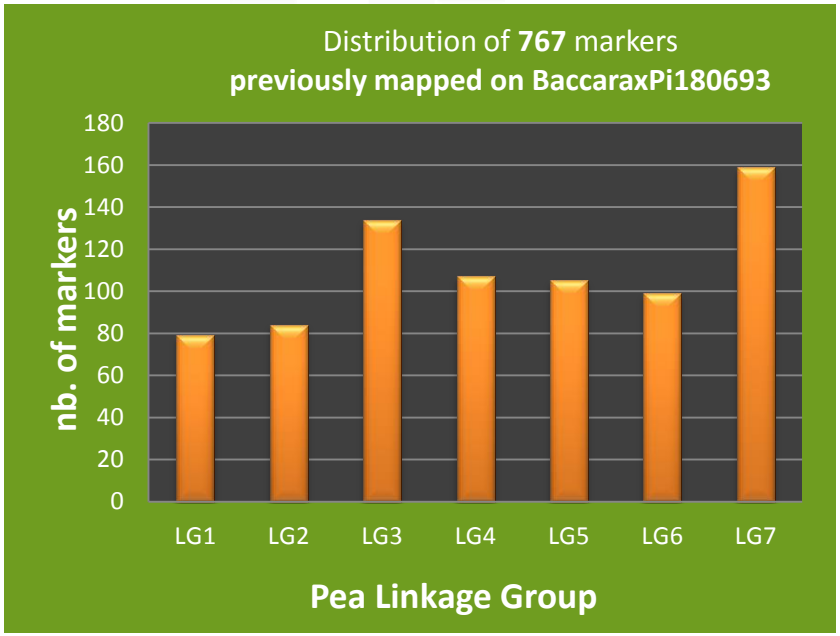
Over 75,000 genomic SNPs considered as robust



Over 75,000 genomic SNPs considered as robust

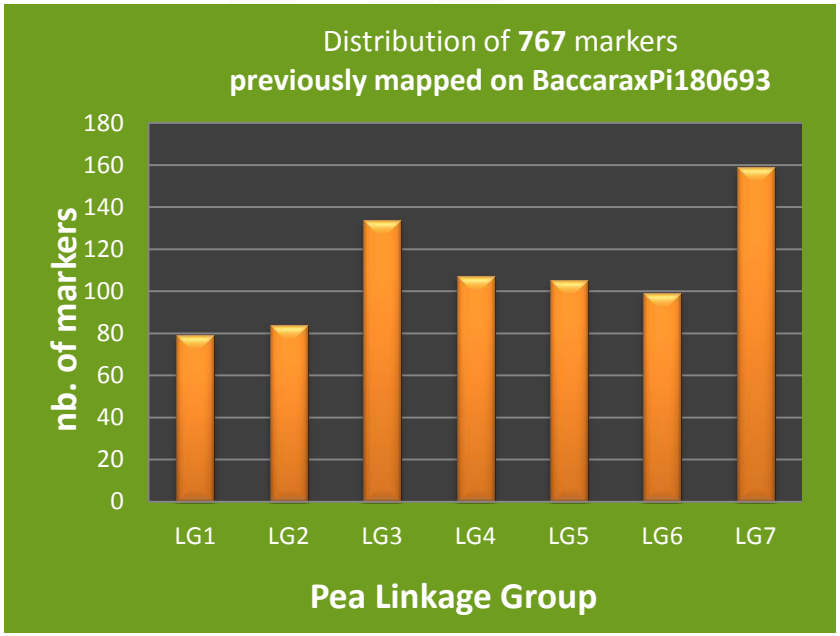


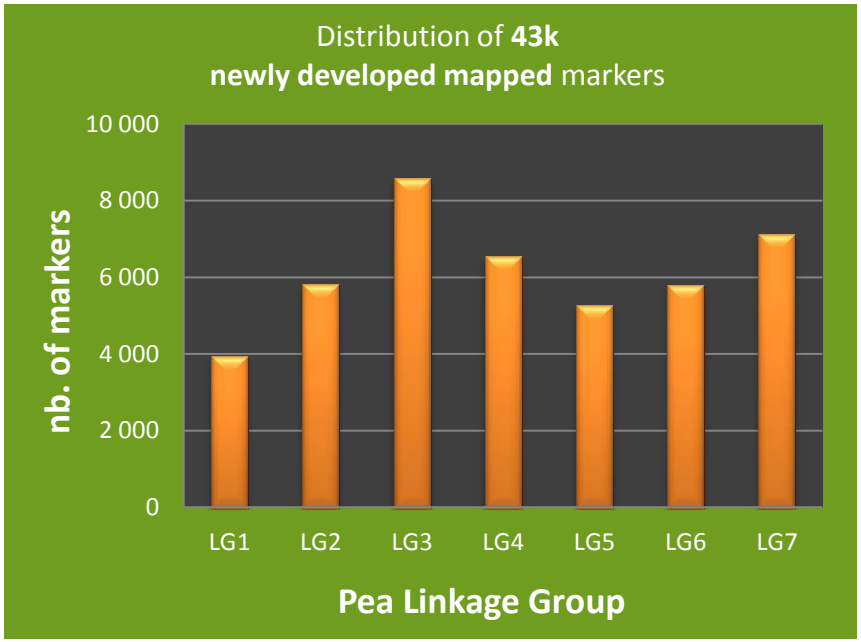
43,000 newly mapped SNPs



Average **100 markers/ LG**

43,000 newly mapped SNPs



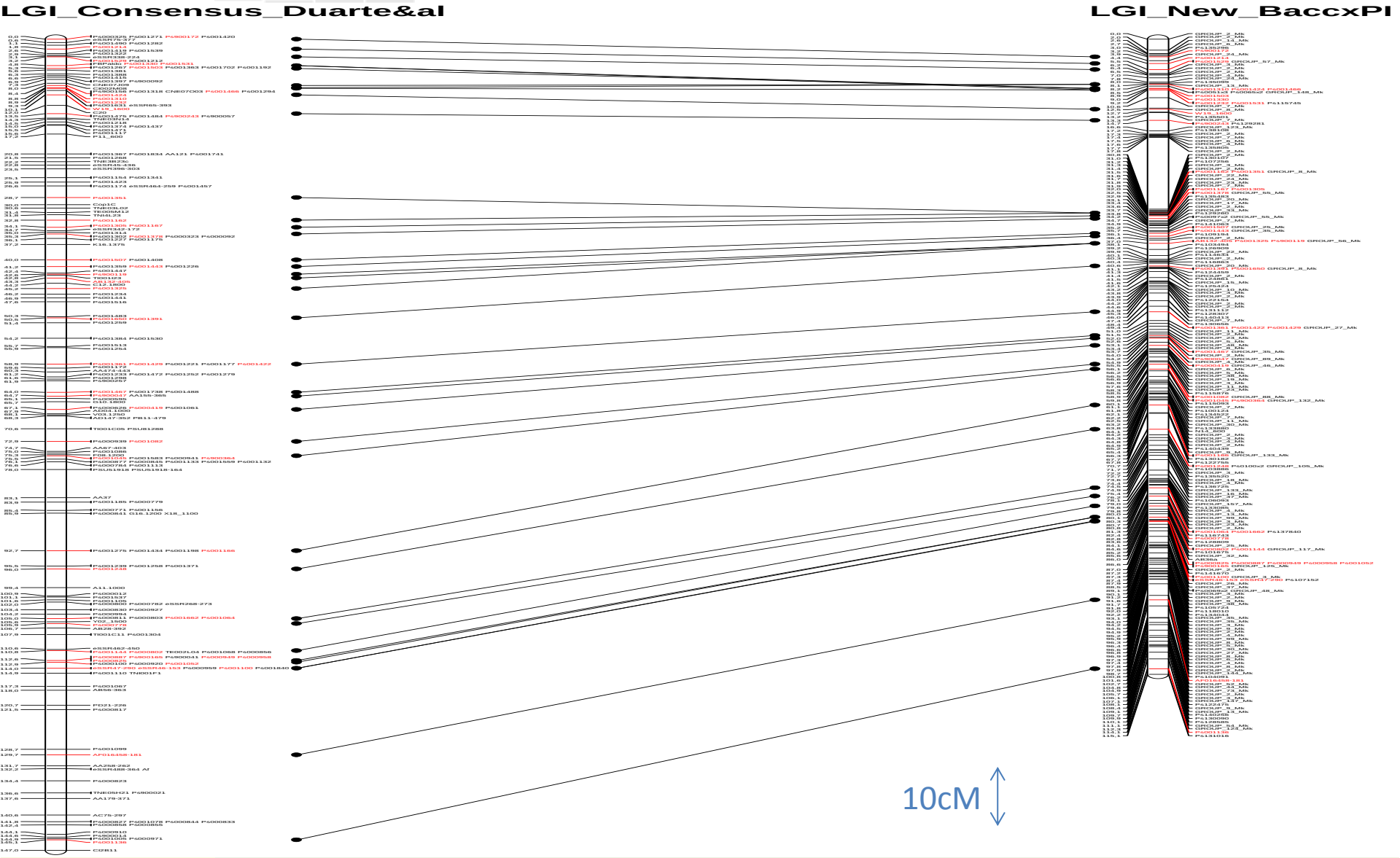


Average **100 markers/ LG**

Average **6000 markers/ LG**

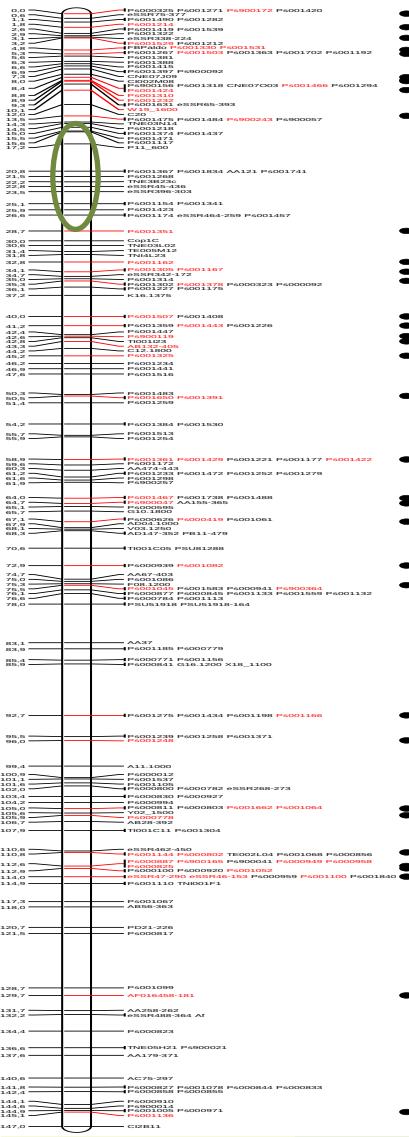
the same distribution pattern

Focus on *pisum sativum* LGI



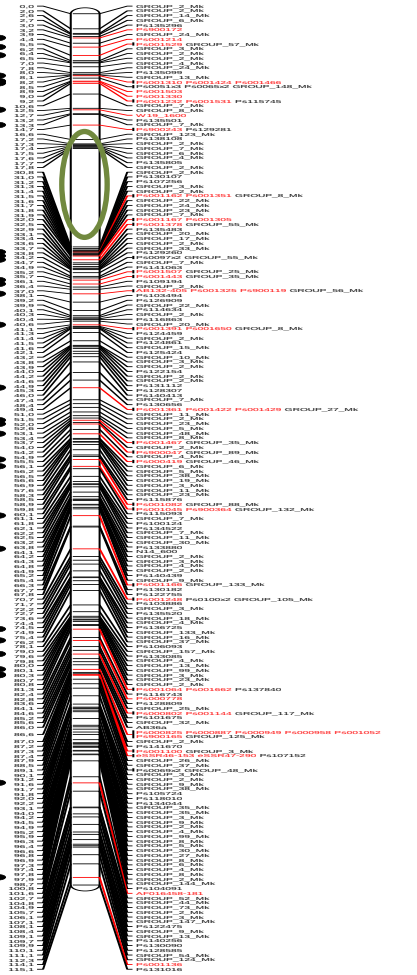
Focus on *pisum sativum* LGI

LGI_Consensus_Duarte&al



	LGI 4 RILs populations Pea Reference Consensus Map (Duarte *et al.*)	LGI Markers polymorphic BaccaraxPi on Pea Reference Consensus Map (Duarte *et al.*)	LGI GBS generated BaccaraxPI new map
Genetic Distance	147 cM	147 cM	**115** cM
Markers	235	105	**3878**
Nb positions	134	68	203 (grouping **1 to 147 Mk/locus**)
Map Coverage (position/cM)	0,9	0,5	**1,7**
Marker density (Marker/cM)	1,6	0,7	**34**
Gaps >10 cM	0	5	1

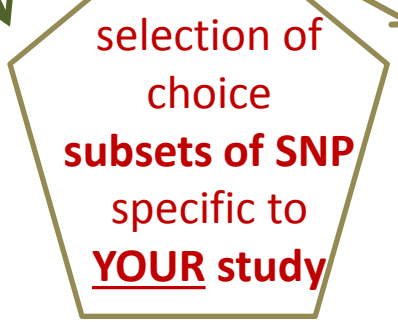
LGI_New_BaccxPI



10cM

Conclusion

Pea Genomic Resource
Pea Sequences
69 000 full length **cDNA contigs**
877 GB **genomic DNA sequences**
Pea Markers
43 000 **mapped genomic SNPs**
35 000 **expressed-SNPs**, *M. truncatula* genome anchored
1252 **bridges** between *M.truncatula* and **pisum sativum** maps
Polymorphism data and **structuration** of 92 **accessions** and **cultivar diversity panel** for 1,5 k **expressed SNP**



Academic research and breeding applications
pea genome **sequencing** projects
marker assisted breeding and cumulating alleles at **QTLs** for traits of interest
linkage-based and **genome wide association**
Synteny studies
Germplasm characterization for **directing creation of new ideotypes**

INRA UMR 1349 IGEPP (RENNES)

Susete Alves Carvalho (Genouest Platform)*

Raluca Uricaru (INRIA GenScale team)*

**Sofiproteol funding on PEAPOL project*

Clement Lavaud

Marie-laure Pilet-Nayel

Alain Baranger

Gilles Boutet

BIOGEMMA (CLERMONT-FERRAND)

Jorge Duarte

Nathalie Rivière

INRA GeT-PlaGe Platform (TOULOUSE)

Emeline Lhuillier

Olivier Bouchez

INRA UMR 320 GV (Gif sur Yvette)

Matthieu Falque

INRIA-IRISA (RENNES)

Olivier Collin (Genouest platform)

Pierre Peterlongo (Genscale team)

Thank You

The authors acknowledge the financial support of SOFIPROTEOL under the Project PEAPOL