

4th International Symposium on Broomrape, Bucharest Romania,  
July 2-4, 2018.

# GENETIC DIVERSITY OF SUNFLOWER BROOMRAPE POPULATIONS IN CHINA REVEALED BY GENOME RE-SEQUENCING

Luyang HU

PhD, INRA (France) and ZHEJIANG University (China)

Supervisor: Stéphane Muñoz (INRA)

Weijun Zhou (ZHEJIANG University)



# *Sunflower in China*

The most rapid development of sunflower had occurred in 1980s and 1990s. By the end of the last century, China had over 1,000,000 ha of sunflower.

Oil Sunflower : **40%** of the total sunflower production in China.

Confectionery Sunflower: more than **60%**

(Hua Zhang, et al., 16th International Sunflower Conference, Fargo, ND USA)



# Sunflower production in China

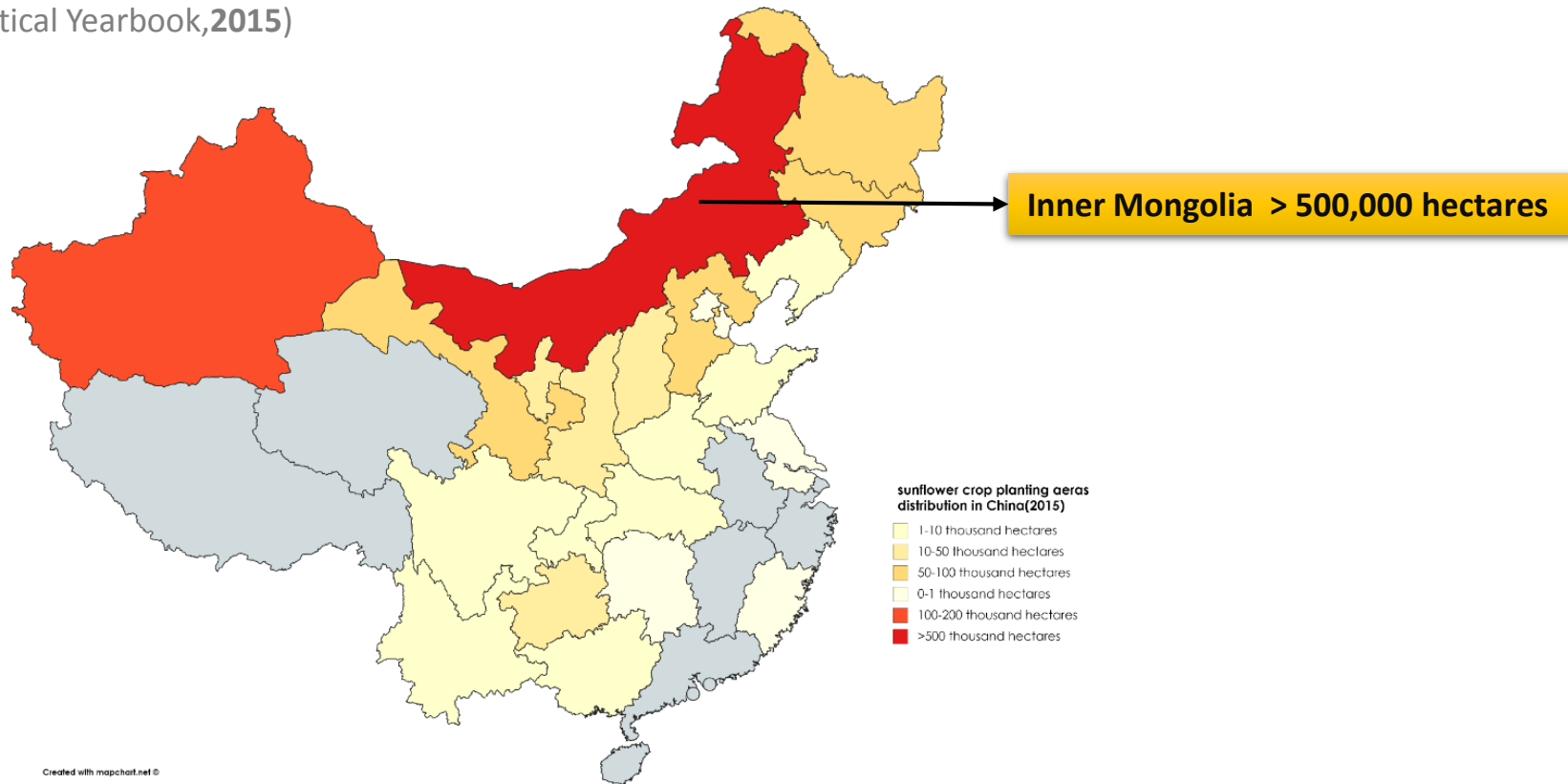
In year 2015 : China

1,036,000 hm<sup>2</sup> (hectares) sowing area

2,698,113 ton

0.62% of Total Sown Areas of Farm Crops

(China Statistical Yearbook, 2015)



Inner Mongolia is the **largest** area of sunflower production

# Race of *O. cumana* in China

First report of sunflower broomrape in 1979 in China, in recent 10 years, occurred in Inner Mongolia

2006: only a few found In Inner Mongolia

2014: 40% of the sunflower area with *O. cumana*, >70% of parasited sunflower

in China: Race A, D, E, F and G (Bixian shia, et al., 2017)



Expansion of *O. cumana*. But a still unclear history and characterization of *O. cumana* populations

# Objective : Genetic diversity of five populations from China (Inner Mongolia)

Seeds collected in Inner Mongolia

Manipulate in growth chamber → F1 (selfed seed)

Genomic DNA extraction Race characterization in growth chamber

Reference genome assembly → DNA seq (re-sequencing) ← Add 5 Spain populations

Mapping (BWA software)

SNP calling 12X depth (Varscan software)

Parse VCF file and filtering( VCF tools)

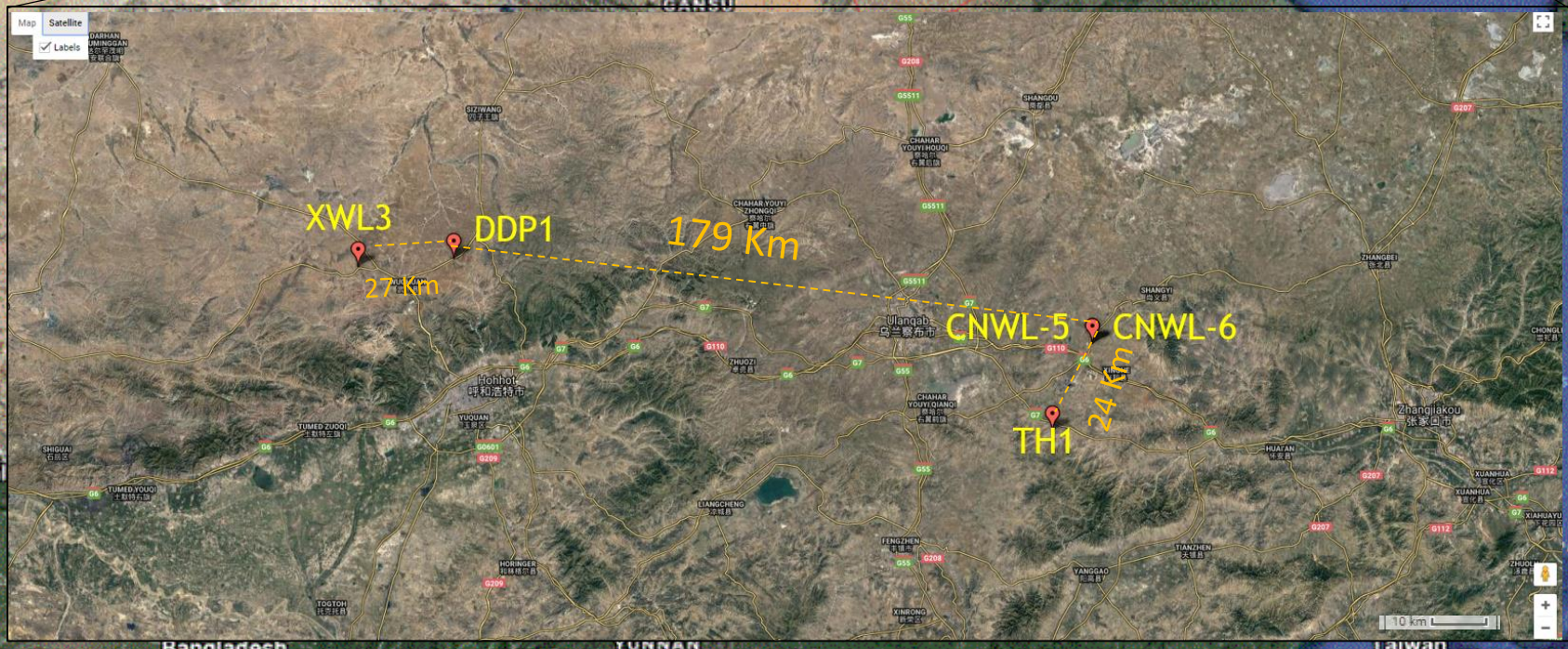
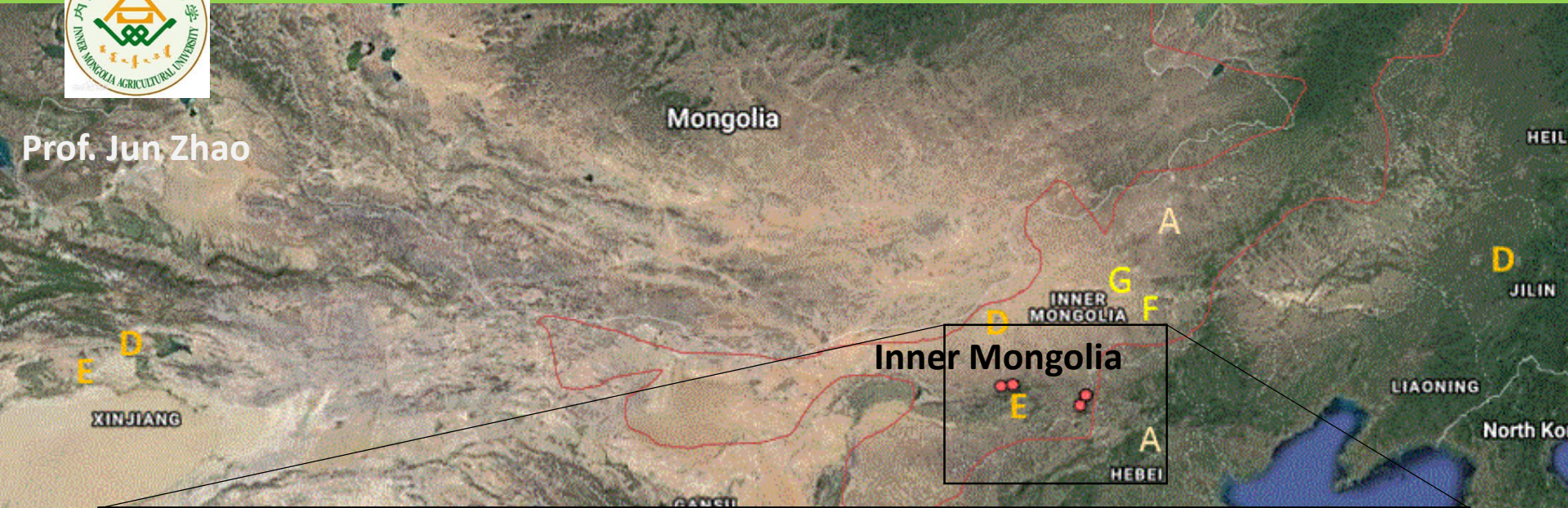
Merge data from Europe populations( add 12 populations)

Principle coordinate analysis ( 22 populations, R script)

# Collect *O. cumana* seeds in China

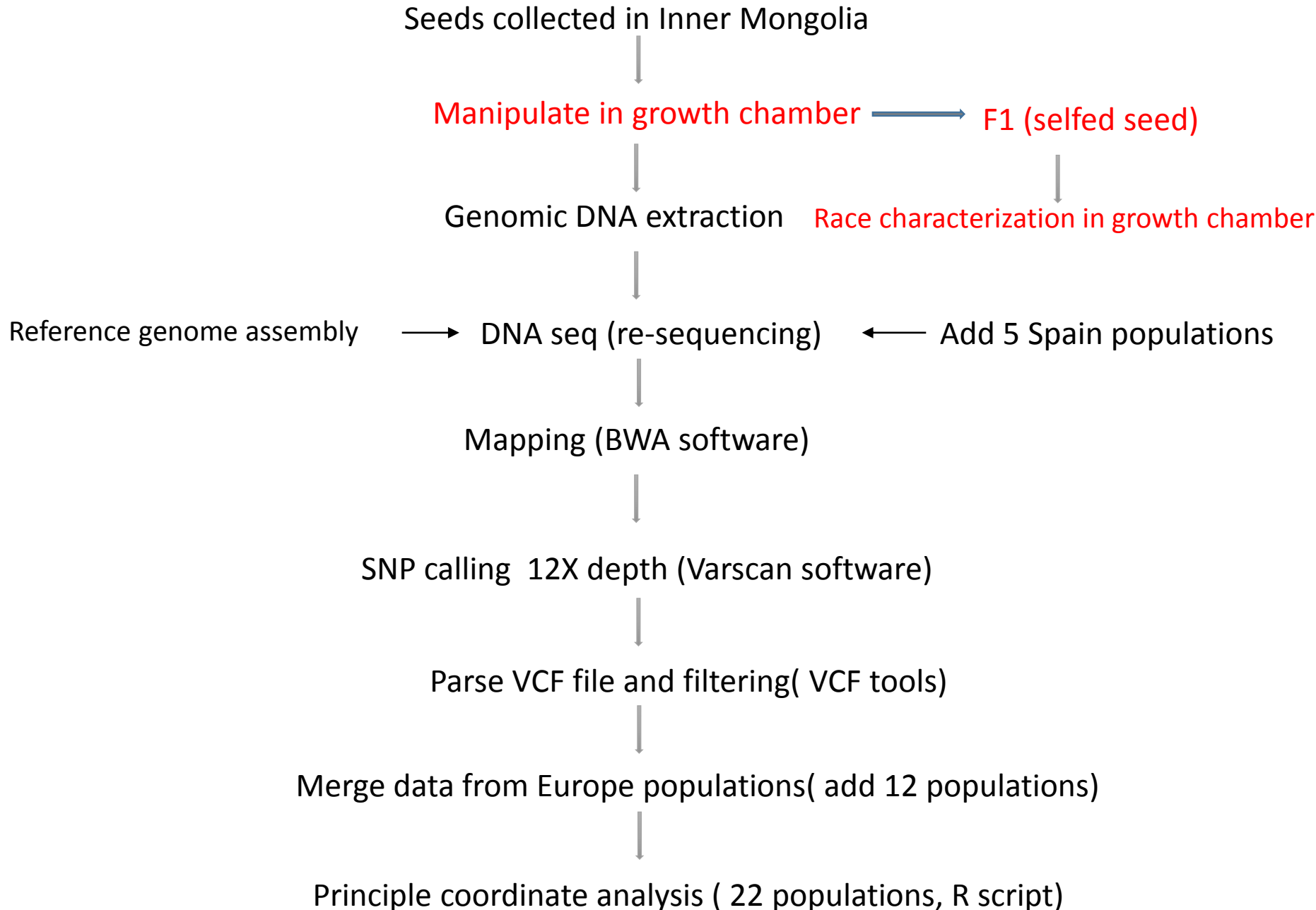


Prof. Jun Zhao

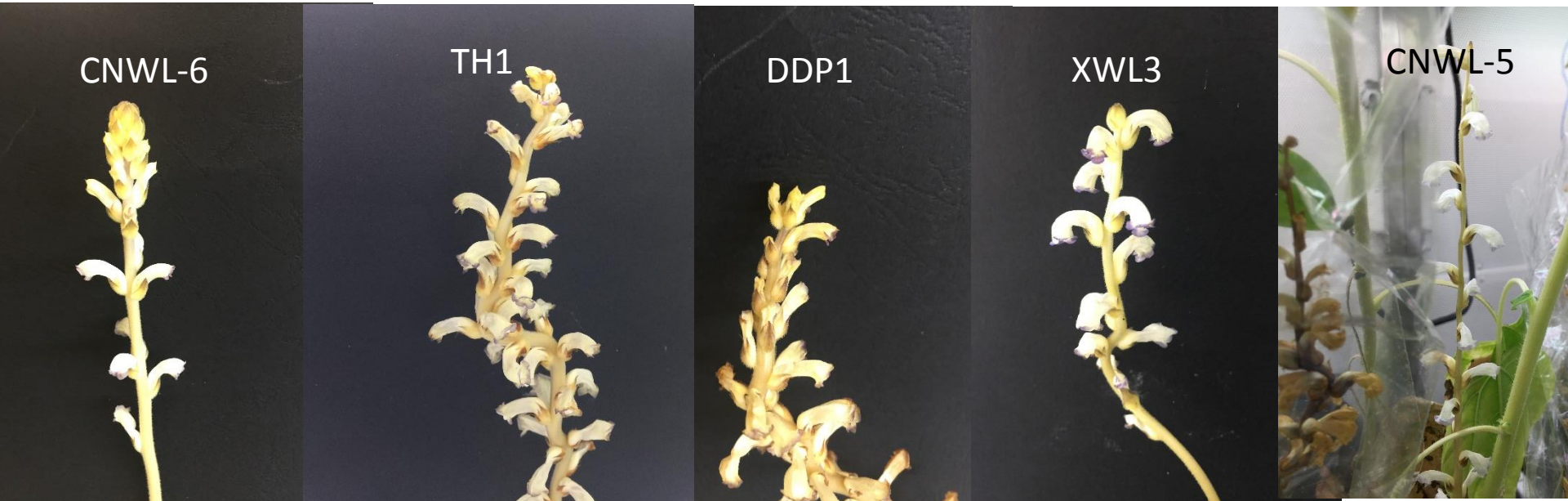


2017

# Objective : Characterization of five populations from China (Inner Mongolia)



# Flower phenotype of F1 *O. cumana*

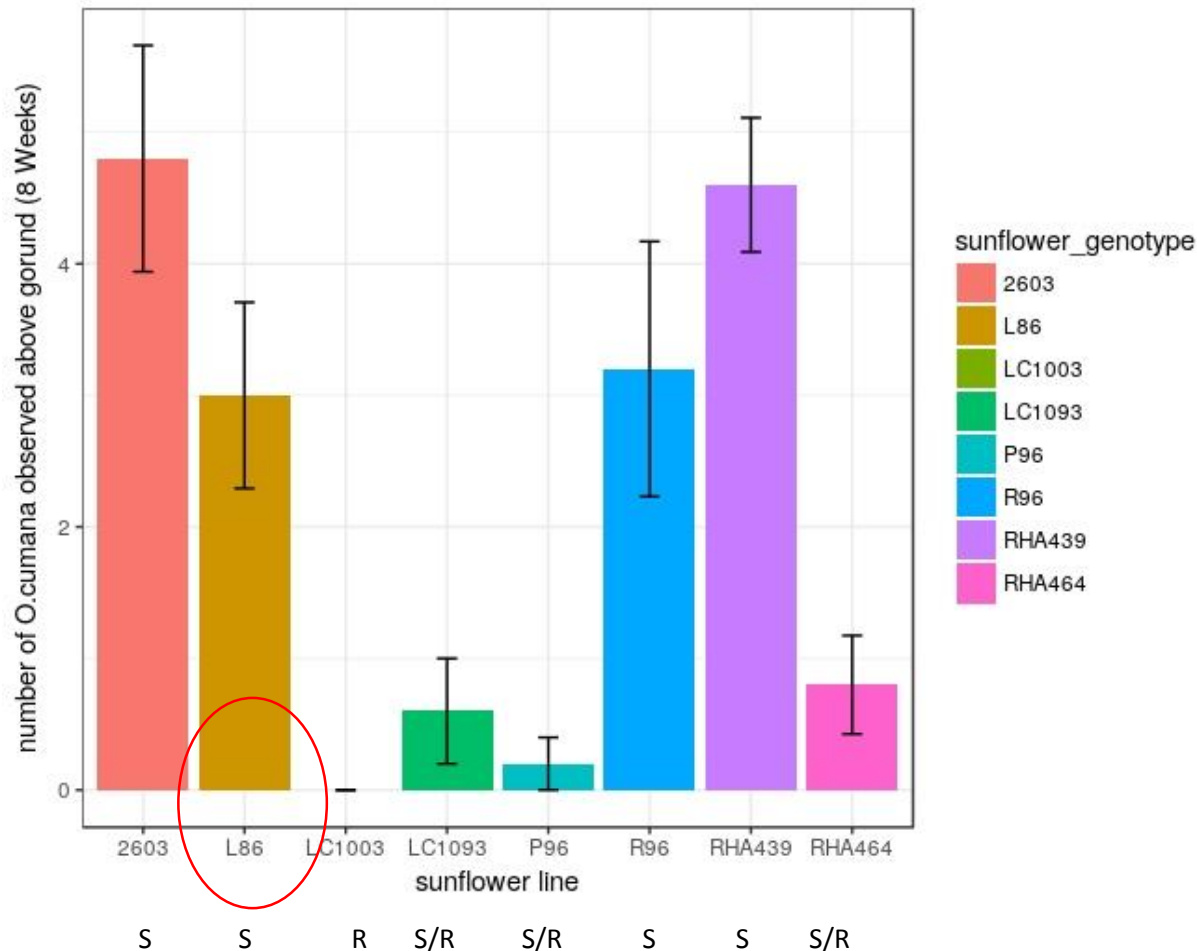


Picture took in October , 2017, Canon camera. In L3, LIPM,France



# Virulence (race) of population F1 CNWL-5

RACE STRUCTURE OF CNWL-5 POPULATION



Sunflower lines	Race E (Ecija, Spain)	Race F (Marchena, Spain)	Race G (Çeşmekolu, Turkey)
2603	S (18.89)	S (20.6)	S (8)
<b>L86</b>	R (0)	<b>R (0.2)</b>	S (15.3)
LC1093	R (0)	S (2)	S (3.5)
P96	R (0)	R (0)	S (3.3)
R96	R (0)	R (0.2)	S (2.2)

In plots (3L)  
5 plants/line  
Growth chamber  
One experiment

**Virulence of this population: between F and G**

# Objective : Characterization of five populations from China (Inner Mongolia)



N. Pouilly

(cut 1-2 bud /sample)

Reference genome assembly

Seeds collected in Inner Mongolia

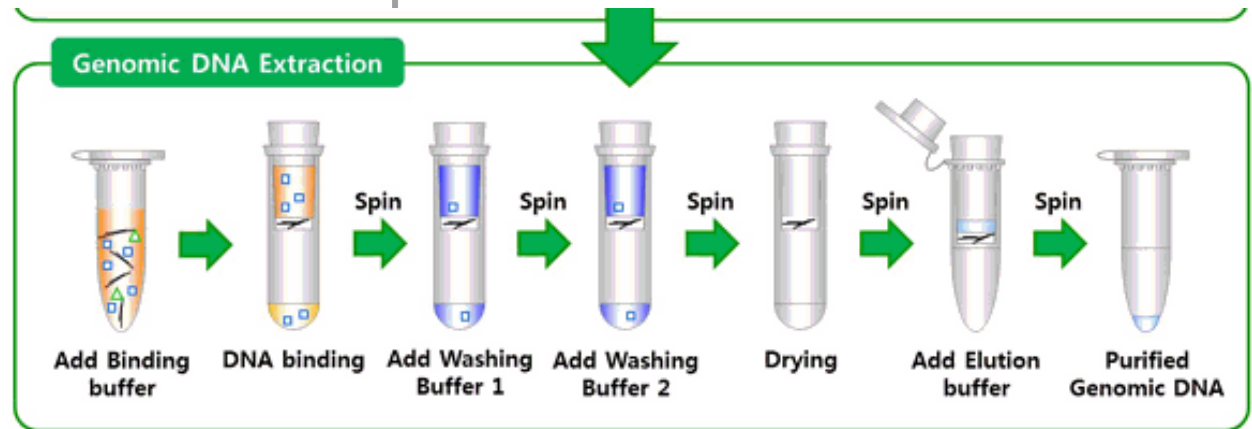
Manipulate in growth chamber → F1 (selfed seed)

Race characterization in growth chamber

**Genomic DNA extraction**

DNA seq (re-sequencing)

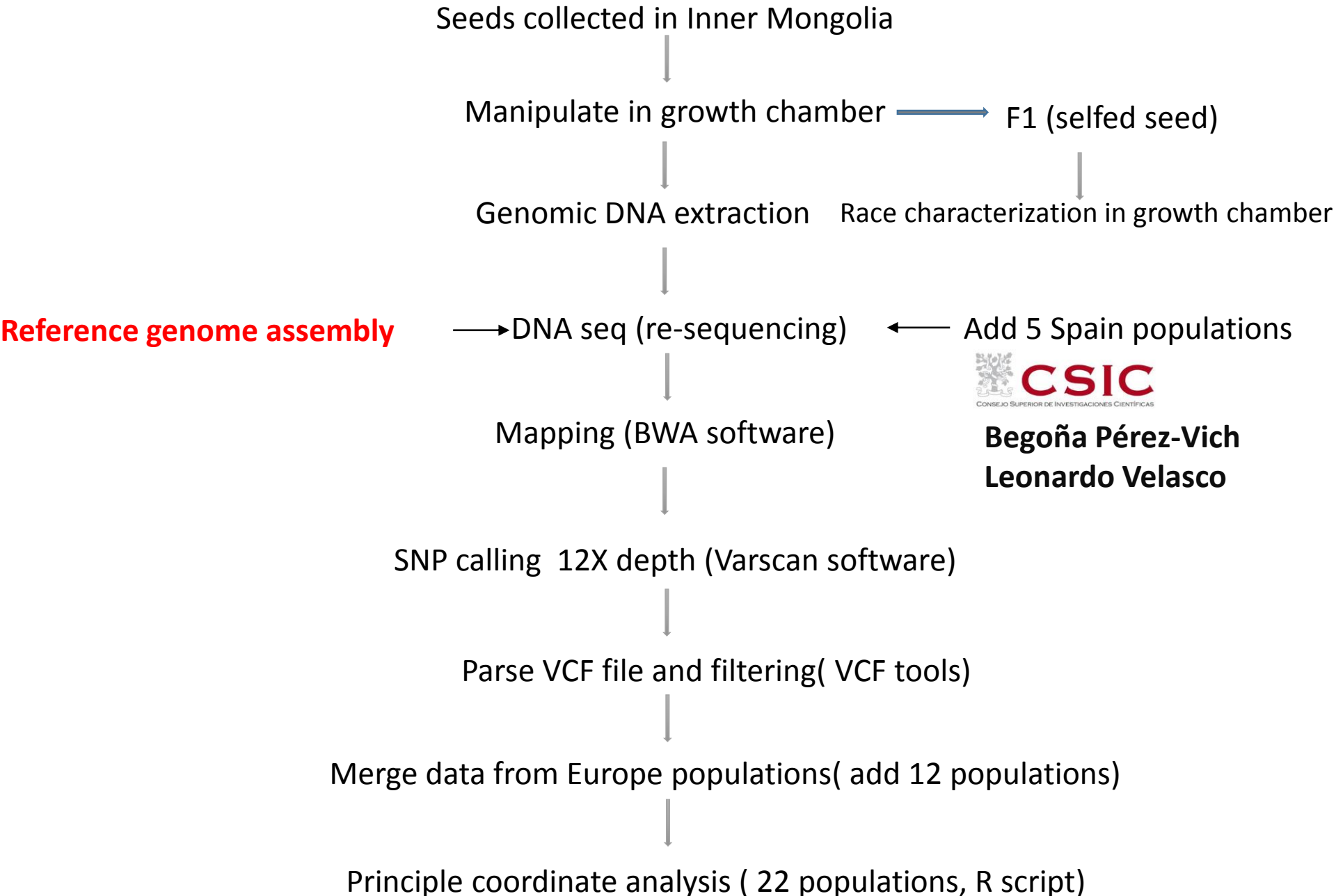
← Add 5 Spain populations



Merge data from Europe populations( add 12 populations)

Principle coordinate analysis ( 22 populations, R script)

# Objective : Characterization of five populations from China (Inner Mongolia)



# O. cumana reference genome assembly

HeliOr Project

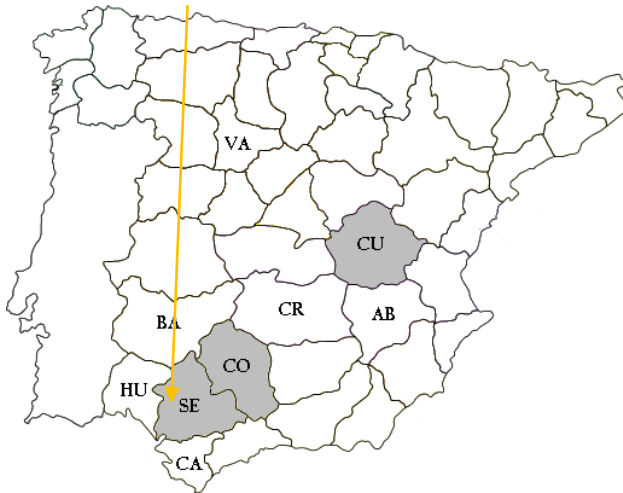


- 126 SMRT Cells (mean: 1.19 Gb/SMRT Cell)

- **Assembly: 1.4 Gb, 622 scaffolds, N50: 5.92Mb**

## IN23 from Spain

(selfed three times to increase homozygosity)



### INRA Sunflower Bioinformatics Resources



Lipm Bioinfo

Helianthus Wiki

### Genomes

<b>XRQ Genome Portal</b>  Assembly of the XRQ genome from PacBio data 2016	<b>HA412-HO Genome Portal (Bronze version)</b>  Assembly of the HA412-HO genome from Illumina and 454 Roche data (2014)	<b>HA412-HO Genome Portal (Newbler version)</b>  Assembly of the HA412-HO genome from Illumina and 454 Roche data (2012)	<b>O. cumana Genome Portal (OcH23-20170413)</b>  Assembly of O. cumana genome from PacBio data (2017)
--	---	--	---

### Transcriptomes

<b>XRQ Transcriptome portal HaT13I</b>  Assembly of the XRQ transcriptome	<b>HA412-HO Transcriptome portal Ha412T4I</b>  Assembly of the HA412-HO	<b>Sunflower and Downy Mildew Transcriptome portal H+P</b>  Assembly of the combined	<b>Helianthus Transcriptome portal Hu2007</b>  Assembly of the transcriptome from
---	---	--	---

# Objective : Characterization of five populations from China (Inner Mongolia)

Seeds collected in Inner Mongolia



Manipulate in growth chamber → F1 (selfed seed)



Genomic DNA extraction Race characterization in growth chamber



Reference genome assembly → **DNA seq (re-sequencing)** ← **Add 5 Spain populations**

**Mean depth of Re-sequencing (9 chinese individuals): 15.3 X**

2 individuals/population  
except for CNWL-6

SNI



re)



**Begoña Pérez-Vich**  
**Leonardo Velasco**

ols)

Merge c

2 populations)



Principle coordinate analysis ( 22 populations, R script)

# Objective : Characterization of five populations from China (Inner Mongolia)

Seeds collected in Inner Mongolia



Manipulate in growth chamber → F1 (selfed seed)



Genomic DNA extraction

Race characterization in growth chamber



Reference genome assembly → DNA seq (re-sequencing) ← Add 5 Spain populations

**Polymorphism (SNPs) identified**



**Begoña Pérez-Vich  
Leonardo Velasco**

Mapping (BWA software)



SNP calling 12X depth (Varscan software)



Parse VCF file and filtering( VCF tools)



Merge data from Europe populations( add 12 populations)



Principle coordinate analysis ( 22 populations, R script)

# Polymorphism (SNPs) identification

reference: OcIN23-20170413

Protocol

1. BWA mapping
2. SNP calling with VarScan 2.4.3 (min\_coverage: 7, min-reads2: 3, min-var-freq: 0.3)

Vcftools (Adam Auton and Anthony Marcketta 2009) for filtering and statistic analysis

Filter : Raw read depth as reported by SAMtools  $\geq 7$

3,551,971 SNPs

Stringent filtering of raw SNPs (3,551,971)

Variant allele frequency =100 define as **homozygous variant** genotype

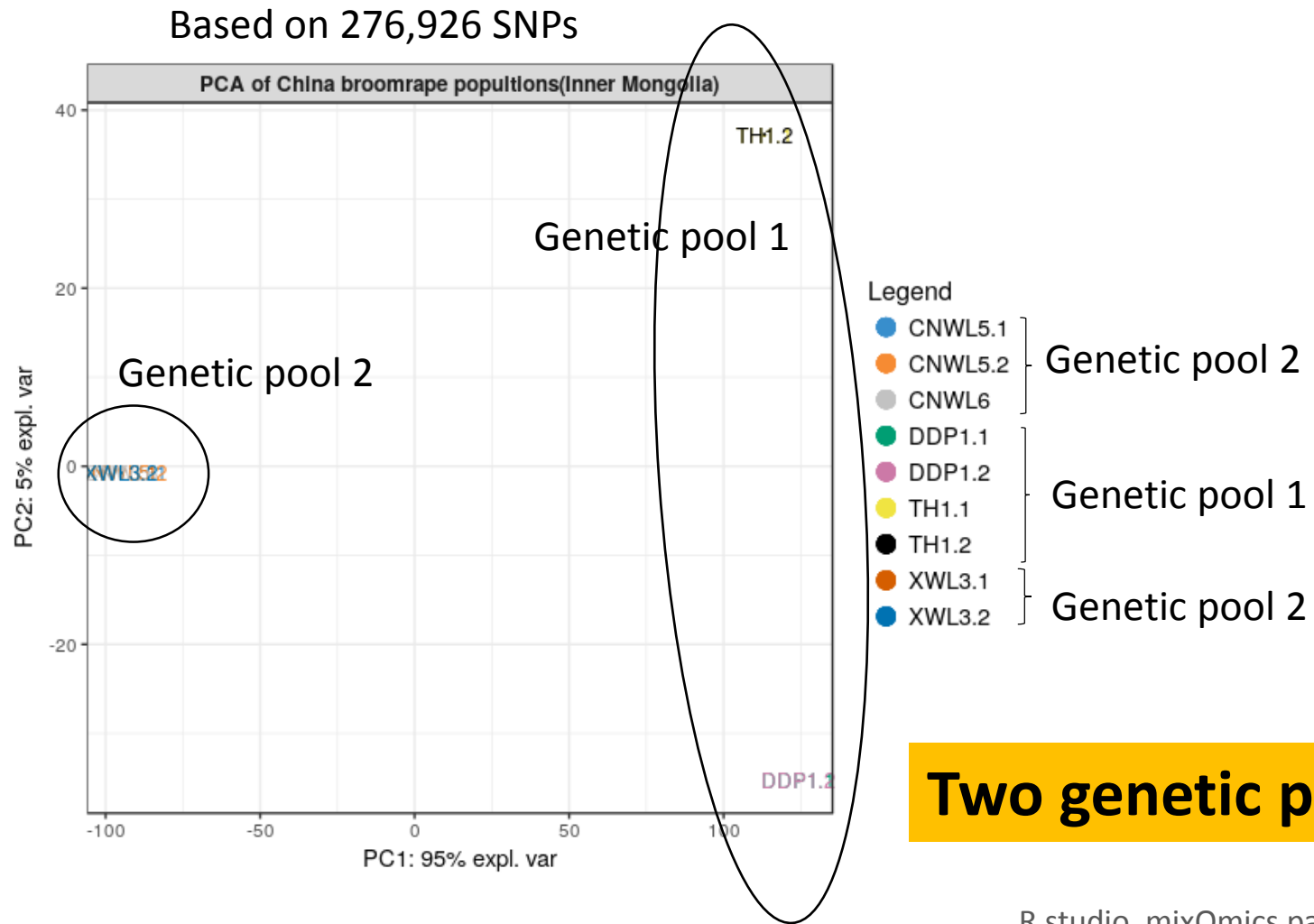
Variant allele frequency =0 define as homozygous **wild** genotype

Variant allele frequency [40,60] define as **heterozygous** genotype

Keep only SNPs that fits the criteria and without any missing data on all 10 populations (China + Spain):

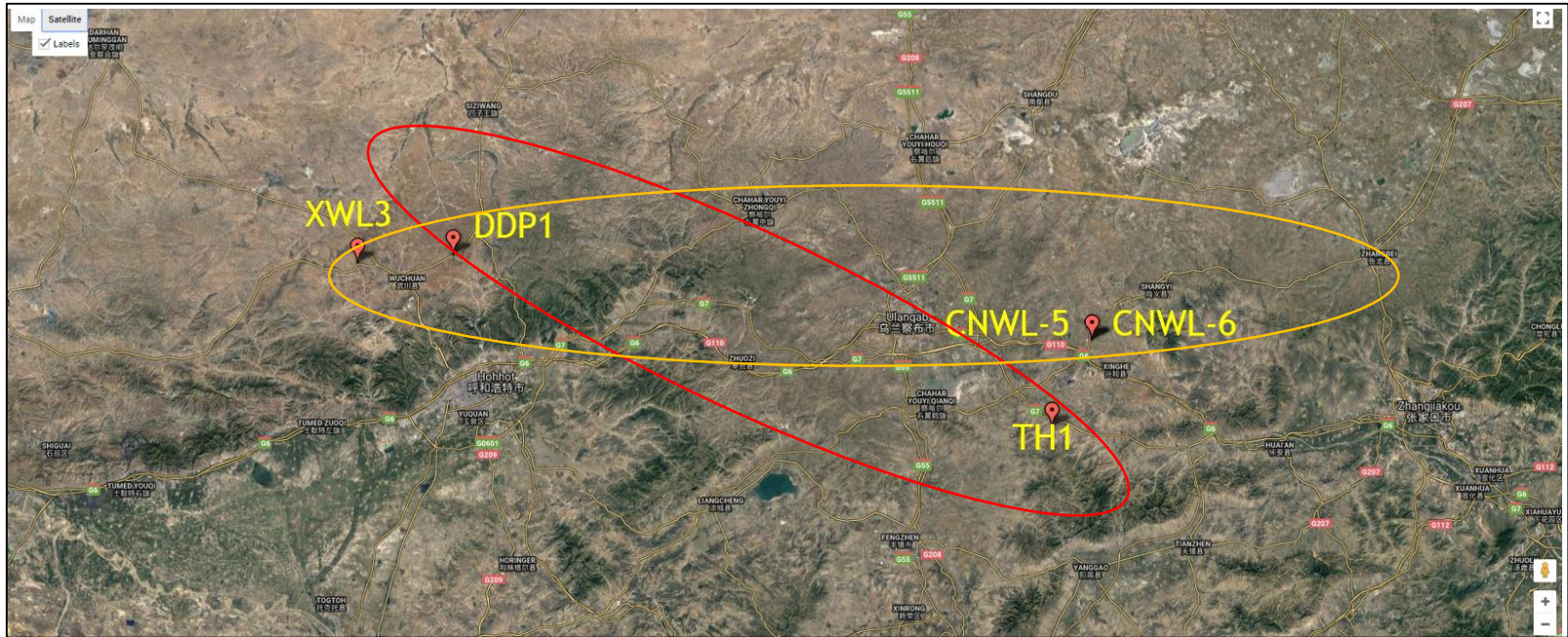
276,926 SNPs

# Genetic diversity of chinese populations

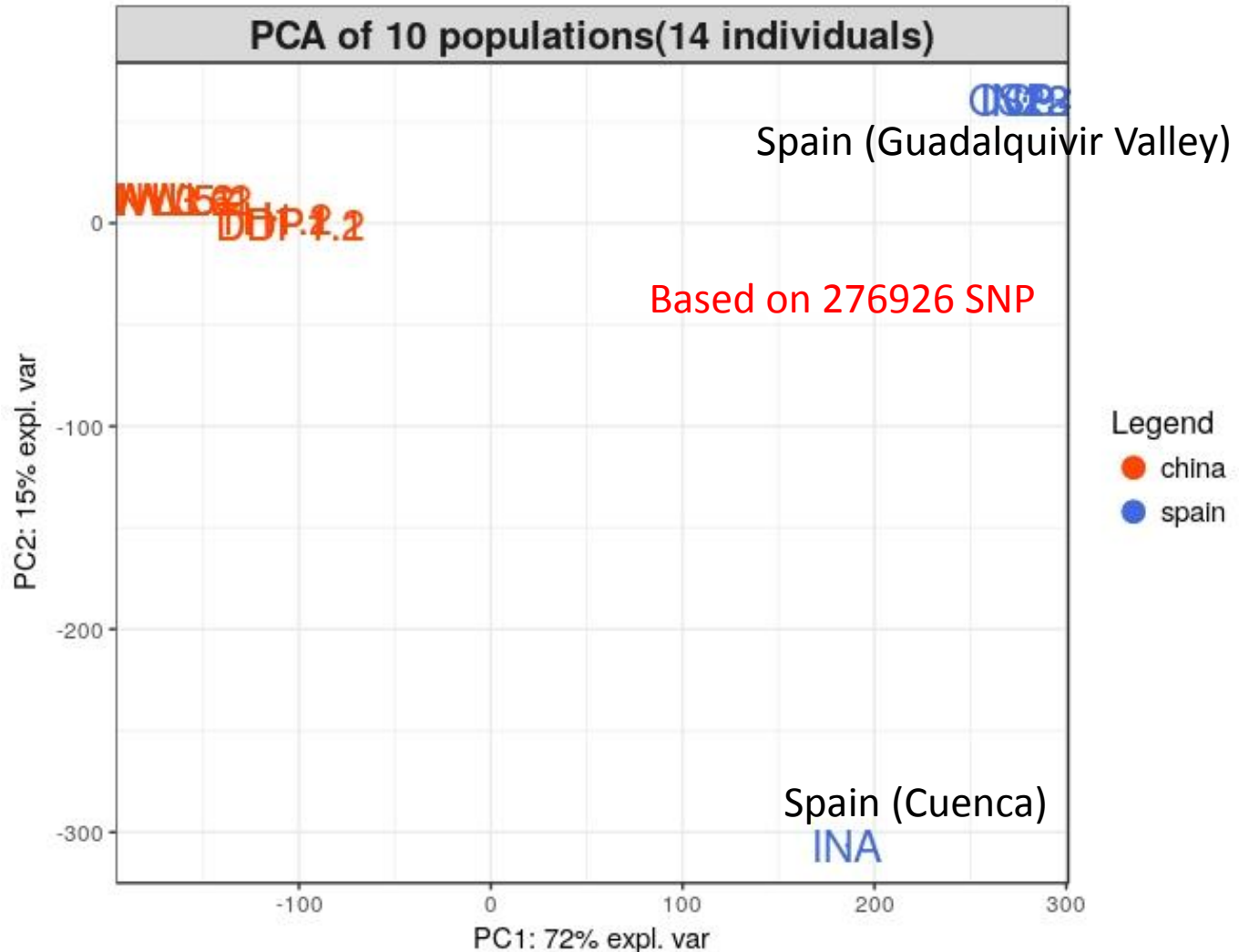




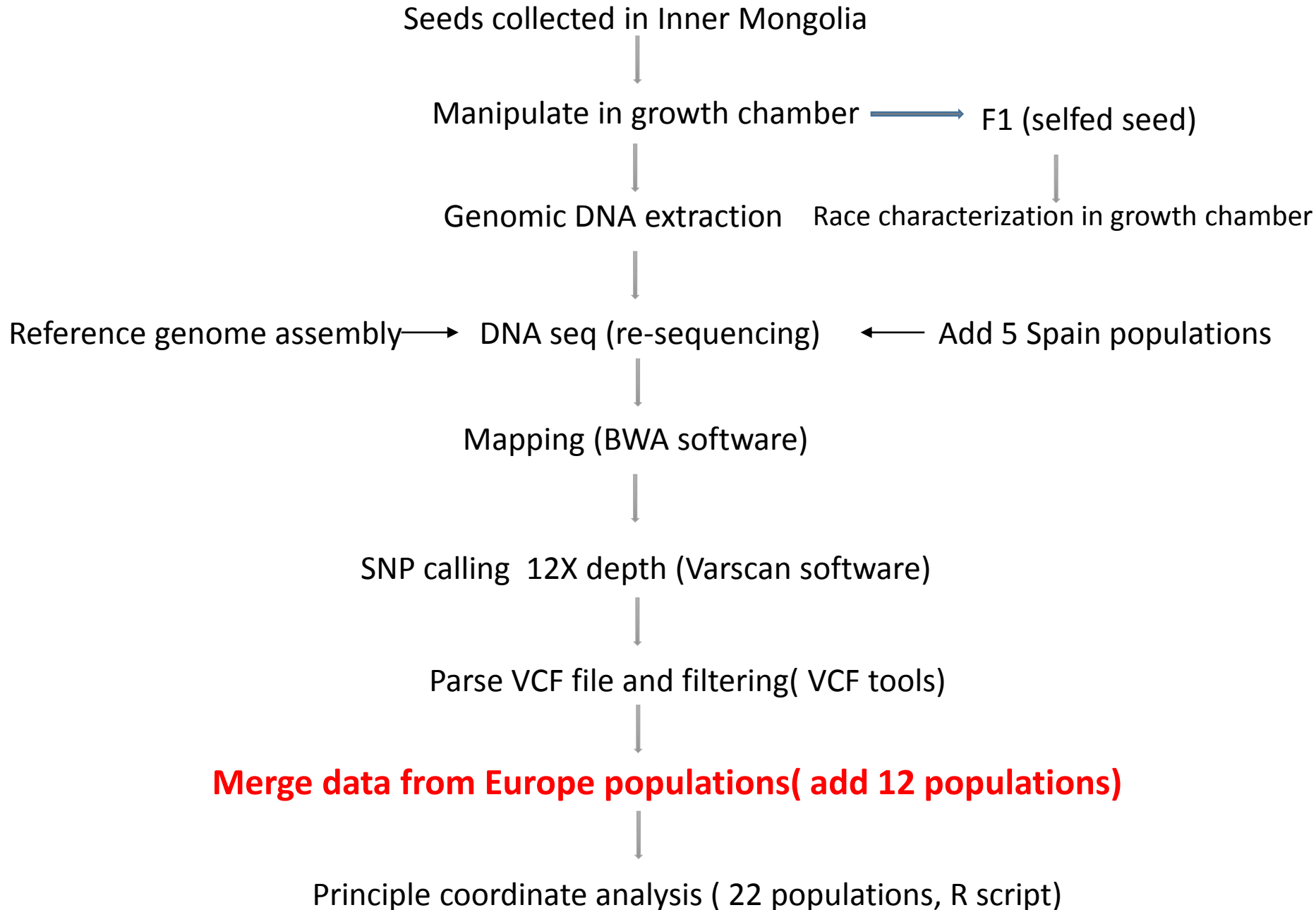
# No relatedness with geographical locations in China



# Comparasion of 5 chinese populations with 5 spanish populations



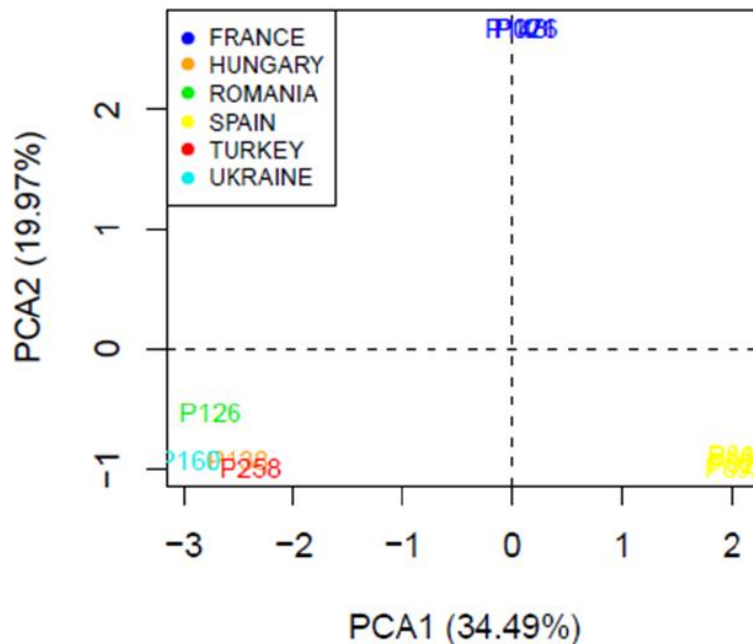
# Objective : Characterization of five populations from China (Inner Mongolia)





M. Coque

# Diversity analysis of 12 populations from Europe



	Total
FRANCE	3
HUNGARY	1
ROMANIA	1
SPAIN	5
UKRAINE	2
<b>Total</b>	<b>12</b>

Exome capture from the 12 populations : 362285 SNPs



1536 SNPs selected to maximize the diversity of the whole set

Fundings:  **SOFIPROTEOL**  
L'engagement durable

# Location of the 1536 SNPs on *O. cumana* genome

genotyping matrix of 12 *O. cumana* populations from Europe with the context sequences of the 1536 SNPs.



Obtain Fasta file



Blast on IN23 reference genome (1493 hits)



select the unic hits that fully aligned with more than 98% of identity (697 hits).



Obtain location of SNP on genome sequence



Too few with no missing data

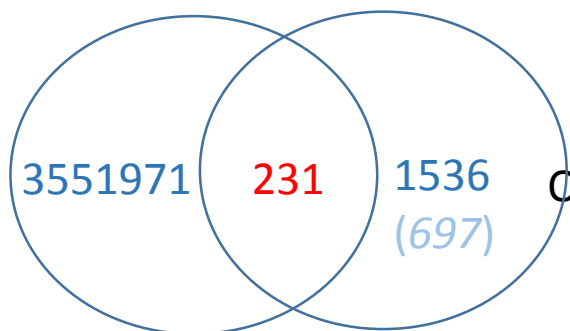
Find 231 SNPs same with that from re-sequencing genome (3,551,971 SNPs)



Merge data from Europe populations( add 12 populations)



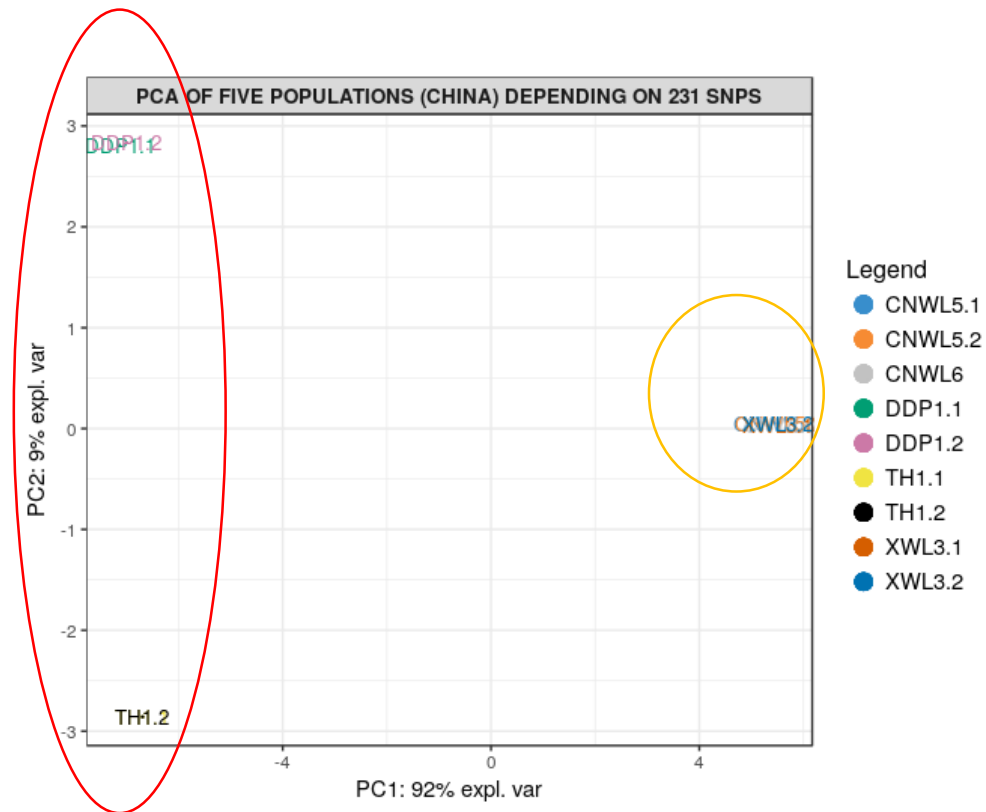
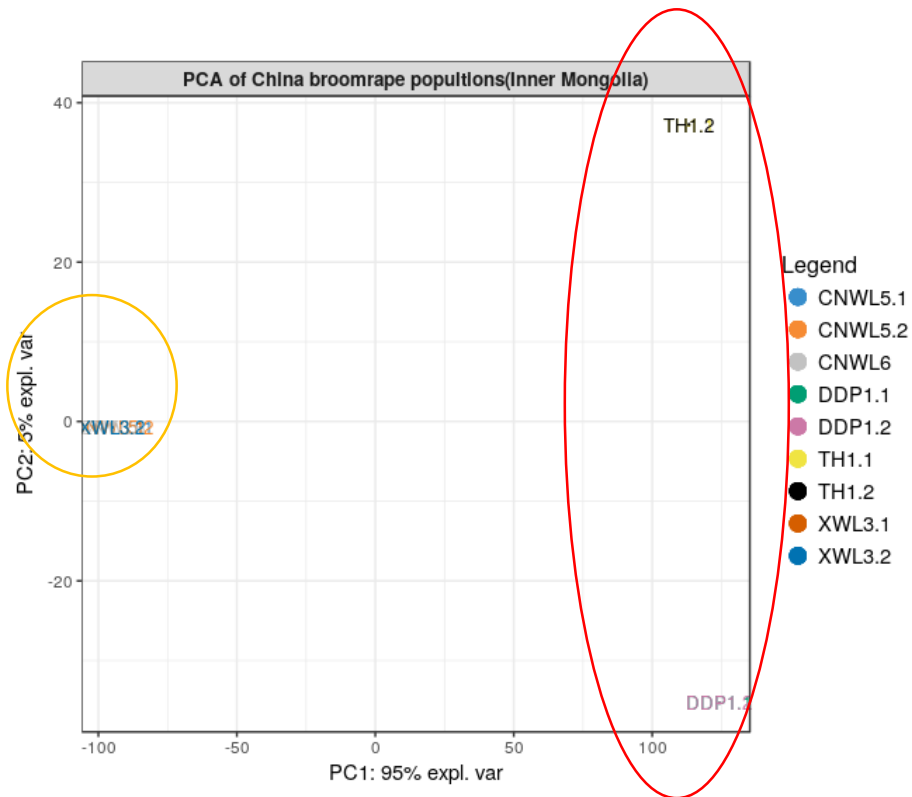
**Principle coordinate analysis ( 22 populations, R script)**



# PCA comparasion with subset of SNPs (231)

Based on 276,926 SNPs

Based on 231 SNPs

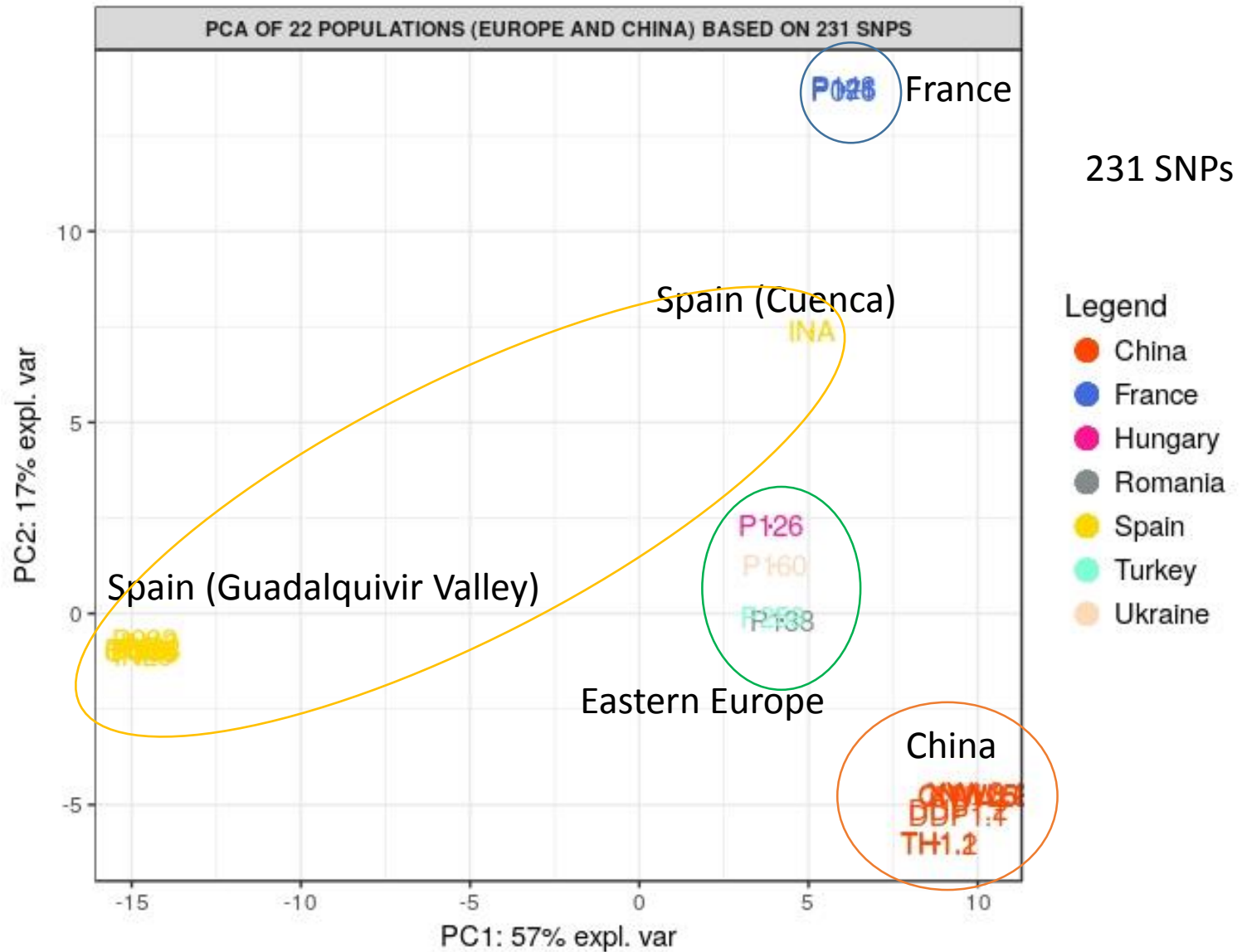


PC1 95%

PC1 92%

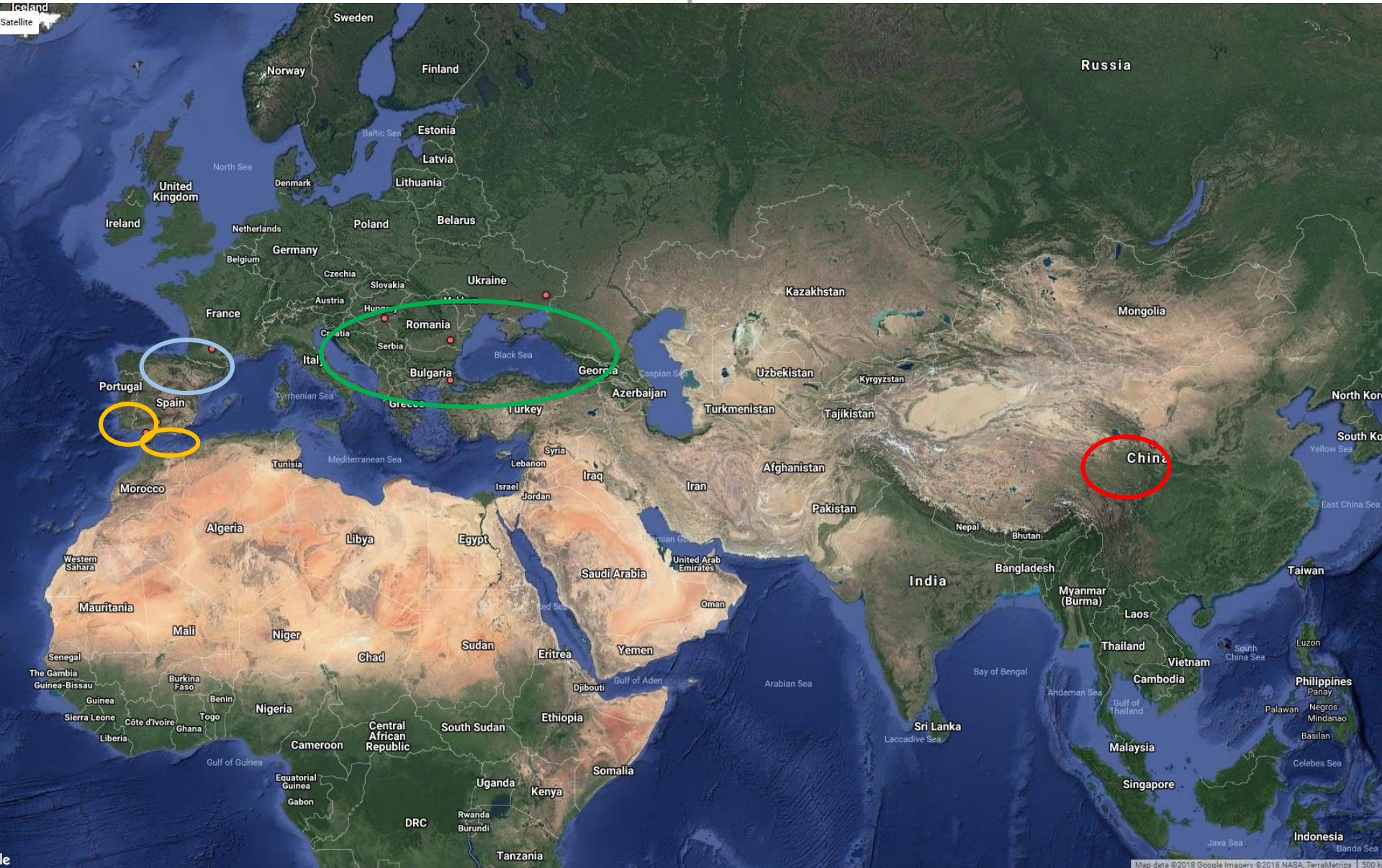
R studio, mixOmics package

# PCA of 22 populations (Europe and China) based on 231 SNPs



# Link with geographical locations at the world level

Seeds collected in Inner Mongolia





# Conclusions

- Find 231 SNPs shared among 22 populations from 7 countries.
- PCA showed that *O.cumana* in inner Mongolia(china) is different with that from other populations in Europe.
- Same situation as France (*O. cumana* since 2008 with different diversity profile).

Why?

Frequency-dependent selection?

# perspectives

- Race characterization for other four population from china
- Intra-population diversity analysis
- Analysis of genes content or non-synonymous polymorphisms in protein sequences

# THANKS TO TEAM AND COLLABORATORS

LIPM:  Laboratoire Interactions Plantes Micro-organismes

Stéphane Muñoz      Nicolas Langlade  
Alexandra Dühnen      Nicolas Pouilly  
Anne-Sophie Lubrano      Olivier Catrice  
Brigitte Mangin      Olivier Guillaume  
Camille Tappy      Pauline Duriez  
Charlotte Penouilh-      Prune Pegot-Espagnet  
Suzette

Florie Gosseau  
Harold Durufle  
Jérôme Gouzy  
Lolita Lorenzon  
Louise Gody  
Ludovic Cottret  
Ludovic Legrand

Marie-Claude  
Boniface  
Marion Laroque  
Mireille Chabaud

Nicolas Blanchet

Nicolas Langlade  
Nicolas Pouilly  
Olivier Catrice  
Olivier Guillaume  
Pauline Duriez  
Prune Pegot-Espagnet



Marie Coque

Inner Mongolia agriculture university  
Jun Zhao



ZHEJIANG UNIVERSITY:

Weijun Zhou  
Chong Yang  
Na Zhang  
Jiang Wang  
Islam Faysal



CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Begoña Pérez-Vich  
Leonardo Velasco



ISTIS



4th International Symposium on Broomrape  
Bucharest, Romania. July 2-4, 2018