Demographic history and dynamics of the genomic differentiation in the pea aphid

Marjolaine Rousselle-CBGP INRAE   10/05/2022

Demographic history and dynamics of the genomic differentiation in the pea aphid

Carole Smadja

Emmanuelle Jousselin

Jean-Christophe Simon & cie (RENNES)

Renaud Vitalis

Mathieu Gautier

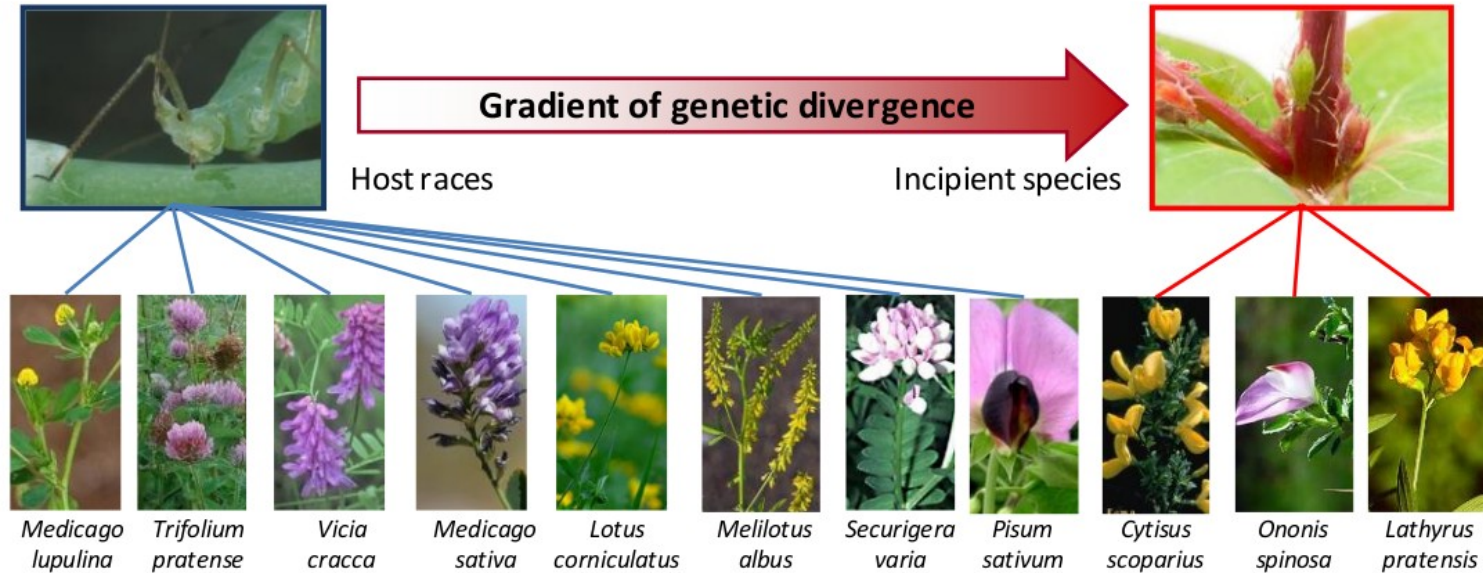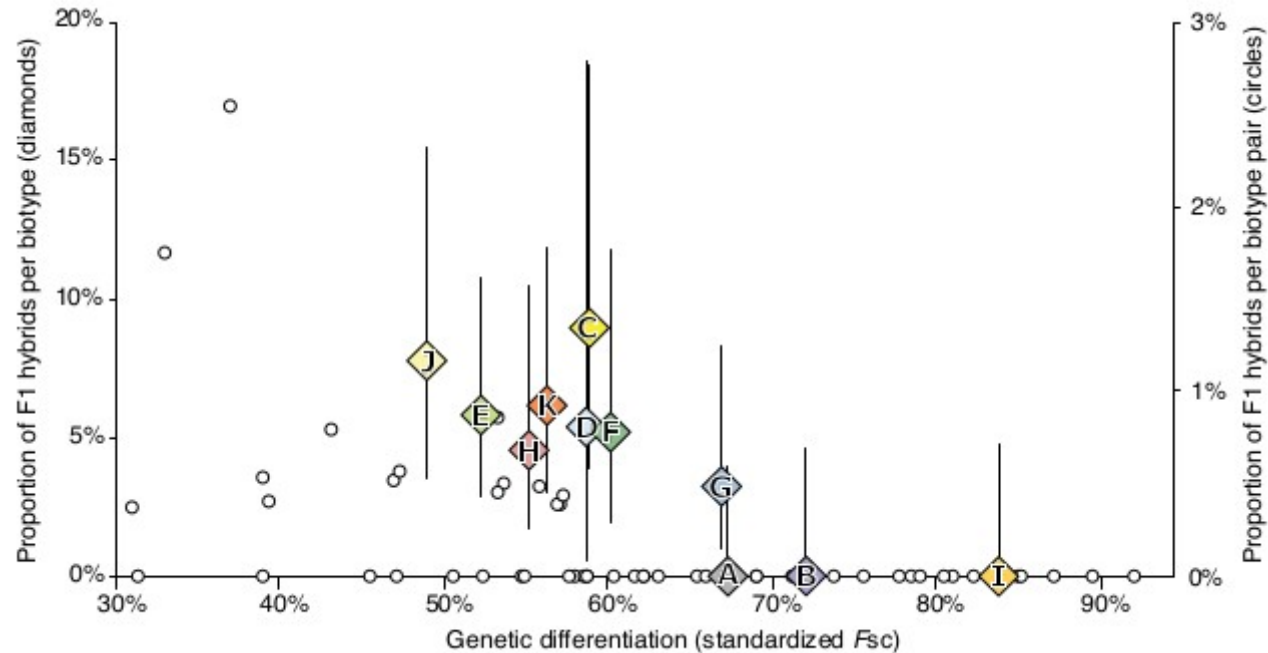Marjolaine Rousselle-CBGP INRAE   10/05/2022

Figure 1: The pea aphid, Acyrthosiphon pisum: a complex of biotypes forming a continuum of divergence from host races to incipient species. Only 11 biotypes adapted to distinct legumes are shown. (~16)

The story of an adaptive radiation

Context : A divergence continuum in the pea aphid in Europe

Peccoud et al. 2009

→ High specialization to the host plant
→Continuum in the degree of hybridization, which decreases with genetic distance → on the way to a sympatric speciation process ?

Context : A divergence continuum in the pea aphid in Europe

This divergence continuum raises two types of global questions :

→ **What are the genetic basis of the differentiation in Europe?**

-Differentiation characteristic of a **divergence in the presence of gene flow** scenario?
-**Spatial and temporal genomic dynamic of recruitment of barrier loci** as divergence follows ?
-**Functionnal nature of barrier loci**? Are they linked to **plant specialization**, especially in the first stages of the differentiation ?

→ **At a larger geographical scale, is there still a differentiation by the host plant ? Can we reconstruct the evolutionary history of biotype formation ?**

-**Single origin of host race formation** or repeated occurrence?
-**Timing of ecological specialisation = timing of divergence** or decoupled?
- **How old** is the radiation ? **Where** did it start from ?
- History of **gene flow** during divergence

Context : A divergence continuum in the pea aphid in Europe

This divergence continuum raises two types of global questions :

→ **What are the genetic basis of the differentiation in Europe?**

-Differentiation characteristic of a **divergence in the presence of gene flow** scenario?
-**Spatial and temporal genomic dynamic of recruitment of barrier loci** as divergence follows ?
-**Functionnal nature of barrier loci**? Are they linked to **plant specialization**, especially in the first stages of the differentiation ?

→ **Pool-seq data on 16 biotypes**

→ **At a larger geographical scale, is there still a differentiation by the host plant ? Can we reconstruct the evolutionary history of biotype formation ?**

-**Single origin of host race formation** or repeated occurrence?
-**Timing of ecological specialisation = timing of divergence** or decoupled?
- **How old** is the radiation ? **Where** did it start from ?
- History of **gene flow** during divergence

→ **Ind-seq data on ~250 individuals of worldwide origin**

Part 1 : The architecture of genomic differentiation towards speciation along the divergence continuum of the pea aphid in Europe

7

Step 1: Characterize the continuum
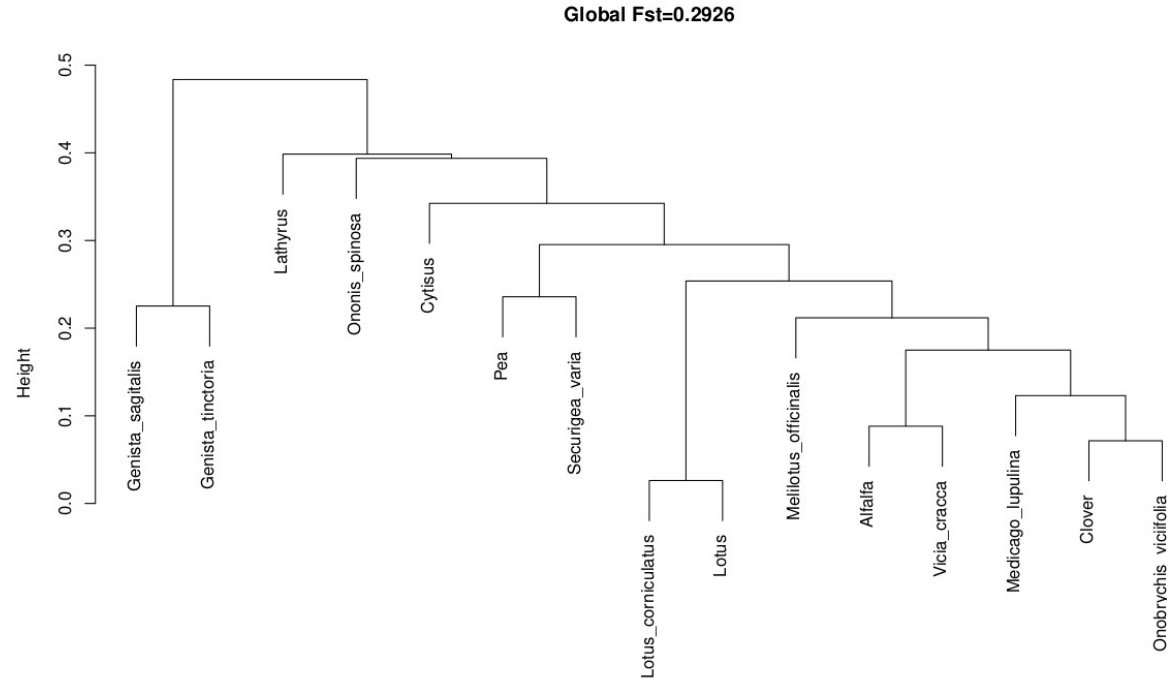
Global Fst=0.2926

Part 1 : The architecture of genomic differentiation towards speciation along the divergence continuum of the pea aphid in Europe
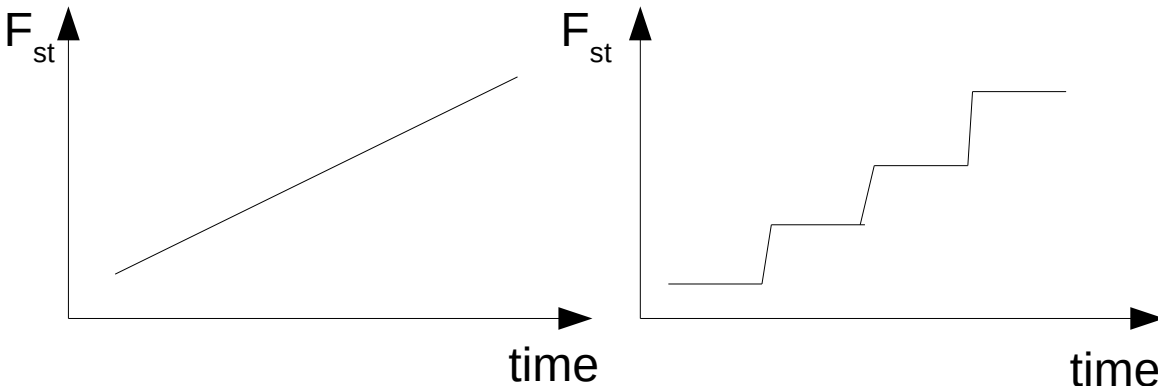
8

Step 2 : Caratacterize the global divergence and genetic differentiation for each pair along the continuum

-Measure F_{st} , d_{XY}
-Measure hte absolute **split time t** for each pair (ie each node of the tree)
-Measure proportion of gene flow via **D like statistics**

Part 1 : The architecture of genomic differentiation towards speciation along the divergence continuum of the pea aphid in Europe

9

Step 2 : Caratacterize the global divergence and genetic differentiation for each pair along the continuum

-Measure F_{st} , d_{XY}
-Measure hte absolute **split time t** for each pair (ie each node of the tree)
-Measure proportion of gene flow via **D like statistics**

Part 1 : The architecture of genomic differentiation towards speciation along the divergence continuum of the pea aphid in Europe

10

Step 2 : Caratacterize the global divergence and genetic differentiation for each pair along the continuum

-Measure F_{st} , d_{XY}
-Measure hte absolute **split time t** for each pair (ie each node of the tree)
-Measure proportion of gene flow via **D like statistics**

Stankowski et al. 2019
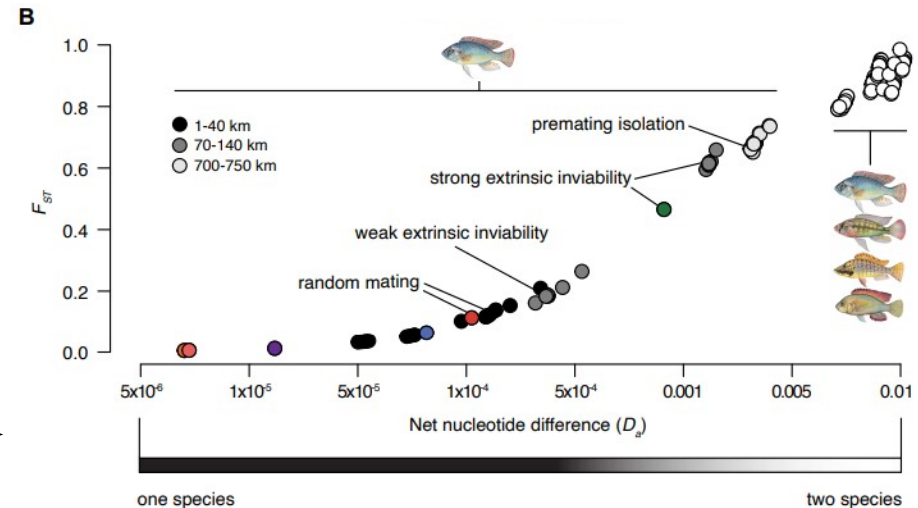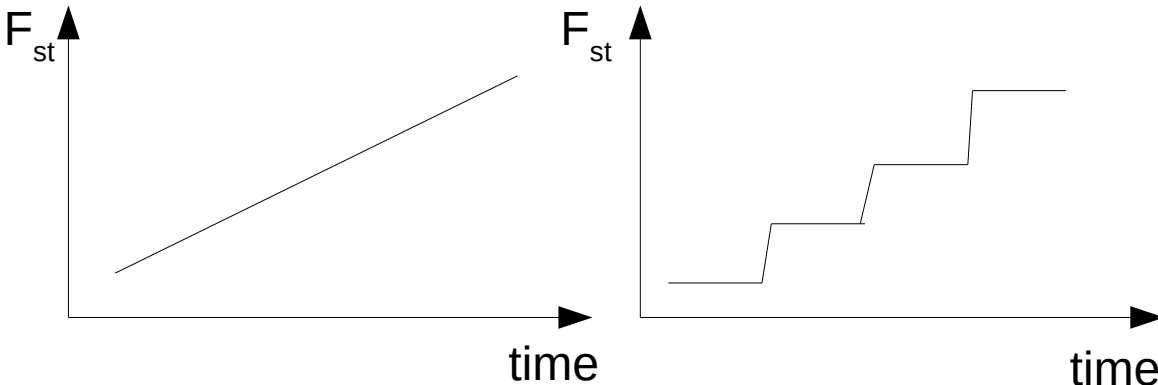
Part 1 : The architecture of genomic differentiation towards speciation along the divergence continuum of the pea aphid in Europe

11

Step 3 : Comparison of the differentiation along the genome for each pair and analysis & interpretation of differentiation outliers

-Measure the differentiation (F_{st}/d_{xy}) along the genome
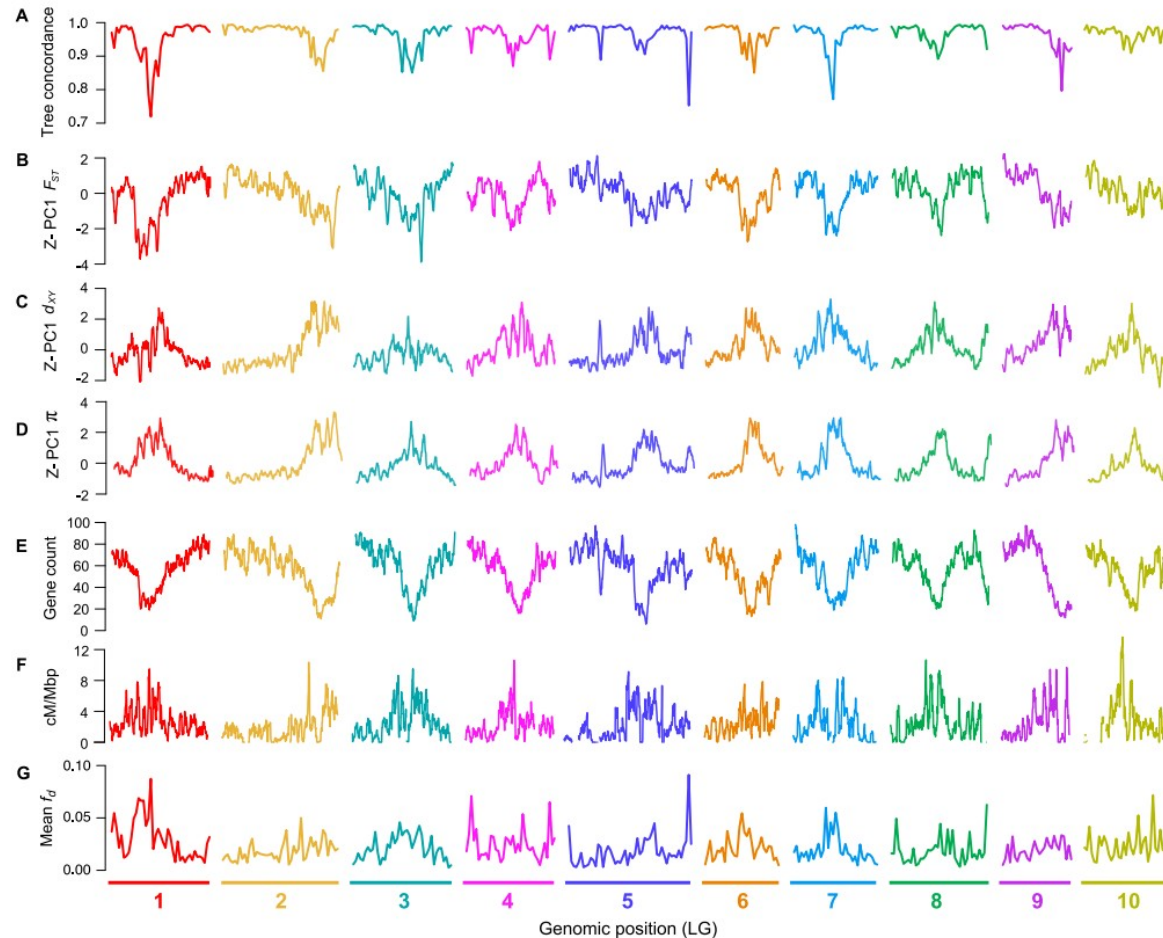
But **confounding effects** for F_{st} pics :

-Measure the recombination along the genome and compare landscape between biotypes
-Measure proxies of the strenght of selection along the genome (π, CDS density)
-Measure the strenght of introgression along the genome along the genome to identify barrier loci *vs*. porous loci via D-like statistics
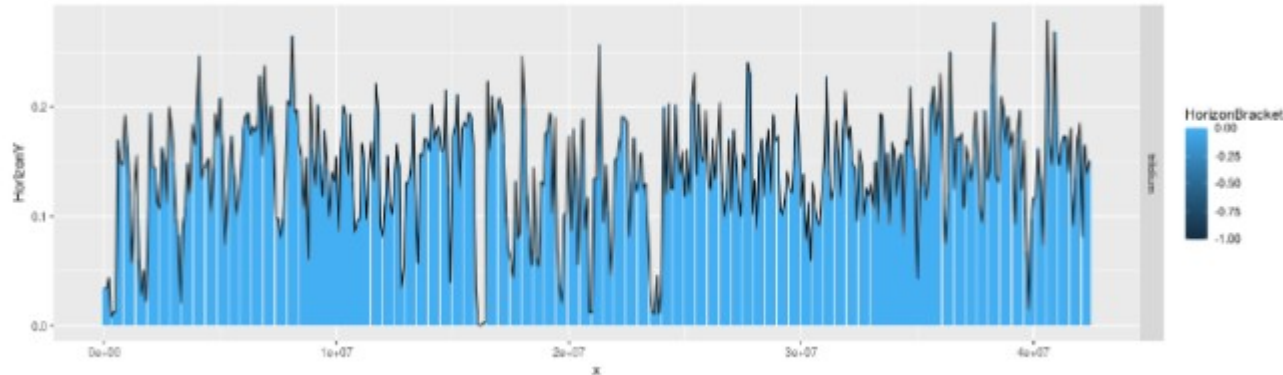
Stankowski et al. 2019

Step 3 : Comparison of the differentiation along the genome for each pair and analysis & interpretation of differentiation outliers

-How are new barrier loci **recruited** over time ?

-Are the **same barrier loci involved in each pair of the continuum** ?

-Are those barrier loci associated with SNPs **associated with host plant specialization** ?

-**Functionnal analysis of the genes** contained in barrier loci (especially **chemosensory genes**) : can we dsitinguish GO categories for the barrier loci evolving in the early divergence versus the latest stages of divergence ?
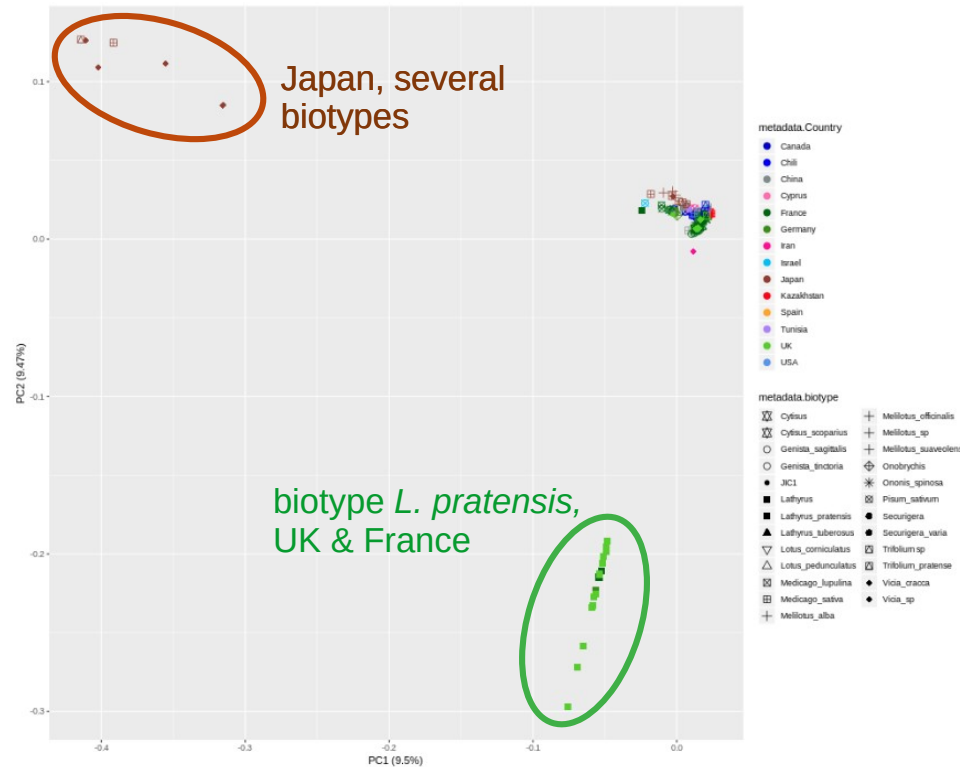
Part 2 : Reconstructing the evolutionary history of biotype formation

Step 1 : Analyze the genetic structurationof the worldwide sample (258 individuals, 14 countries and ~19 host plants)
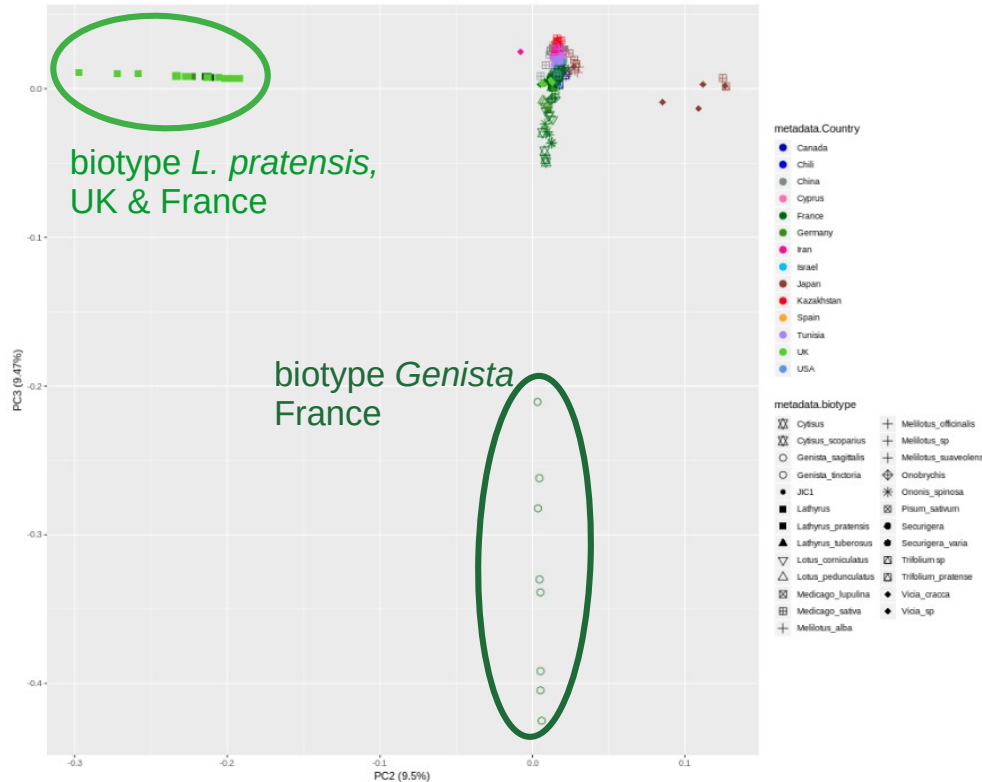
- How do individual cluster (NJ/ PCA) ? How many clusters (DAPC) ?

Part 2 : Reconstructing the evolutionary history of biotype formation

Step 1 : Analyze the genetic structurationof the worldwide sample

- How do individual cluster (NJ/ PCA) ? How many clusters (DAPC) ?

Step 1 : Analyze the genetic structurationof the worldwide sample

- How do individual cluster (NJ/ PCA) ? How many clusters (DAPC) ?

biotype *L. pratensis,* UK & France

biotype *Genista* France

Part 2 : Reconstructing the evolutionary history of biotype formation

Step 1 : Analyze the genetic structurationof the worldwide sample
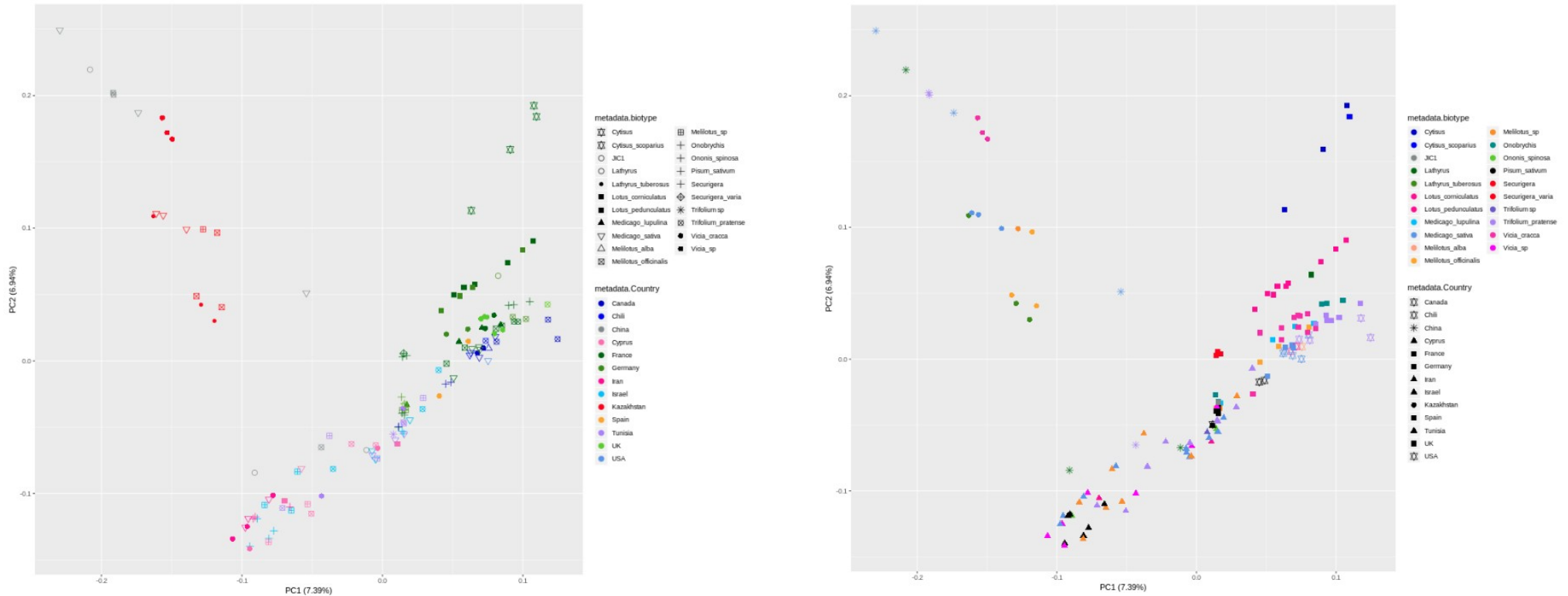
- How do individual cluster (NJ/ PCA) ? How many clusters (DAPC) ?

Part 2 : Reconstructing the evolutionary history of biotype formation

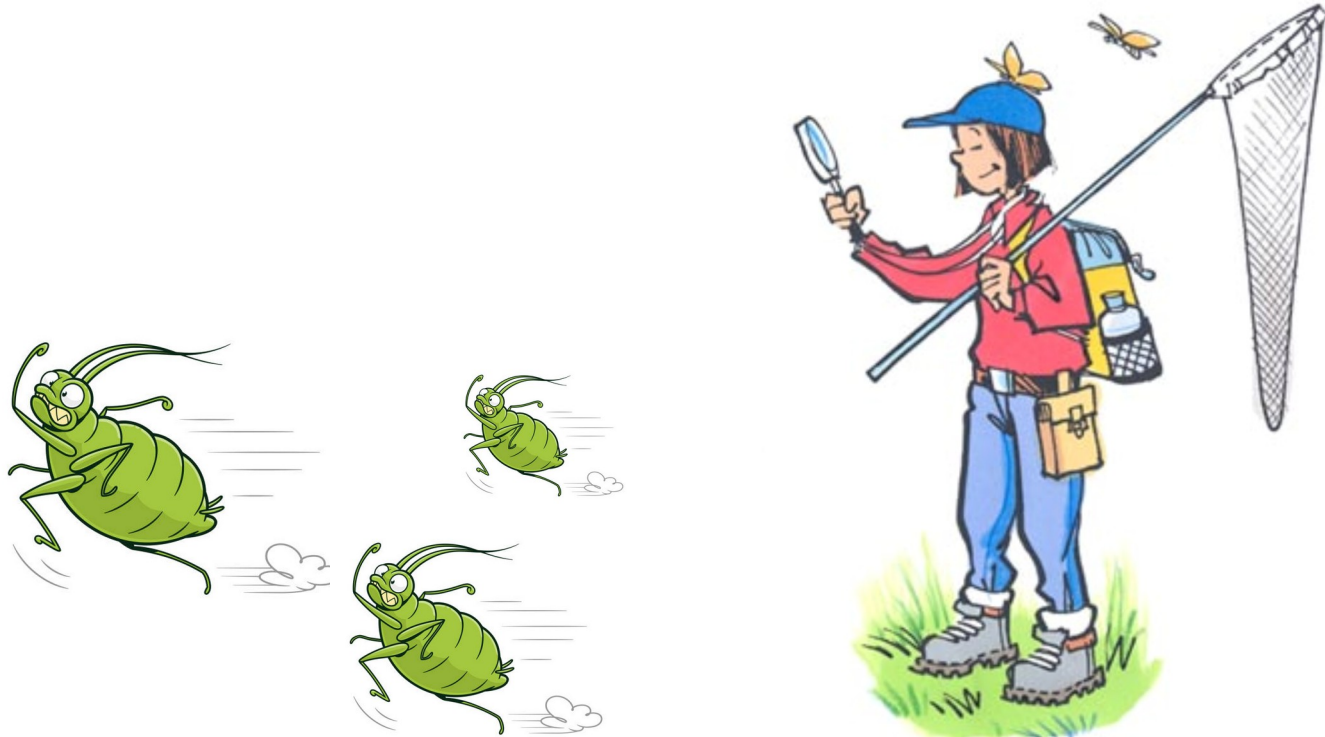**Step 2 : Datation and ancestral populations sizes**

-Coalescent Hidden Markov Models on subgroups of populations

Step 3 : History of gene flow

-Measure the past admixture on relevant trios of populations

Thank you for your attention

Field work in Kazakhstan next week!