

Multi-scale characterization of symbiont diversity in the pea aphid complex through metagenomic approaches



Cervin Guyomar^{1,2}, Fabrice Legeai^{1,2}, Christophe Mougel¹,
 Claire Lemaitre², Jean-Christophe Simon¹
 1 : INRA, UMR 1349 IGEPP, le Rheu, France
 2 : INRIA/IRISA GenScale, Campus de Beaulieu, Rennes, France

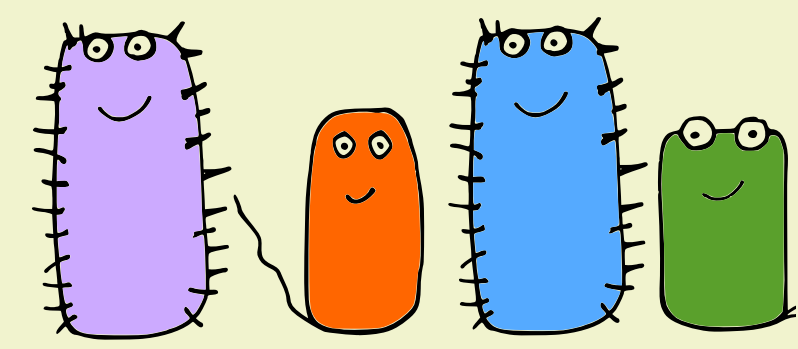


In a nutshell : exploit multi-sample metagenomic datasets to explore finely the pea aphid microbial community

The pea aphid complex



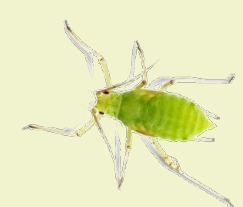
15 biotypes associated to host plant



A diverse symbiotic community
 - 1 obligatory (*Buchnera aphidicola*)
 - 8 documented secondary

Genomic material

Individual Sequencing



Low expected genomic diversity

Pool Sequencing



Higher coverage
 More polymorphism

Reference genomes

Pea aphid
 +
 7 symbionts
 Pea aphid symbiont when available,
 Closest strain when not
 +
 Several plasmids and a phage

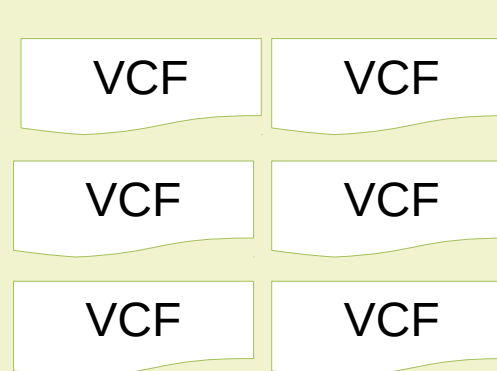
Workflow



Illumina readsets

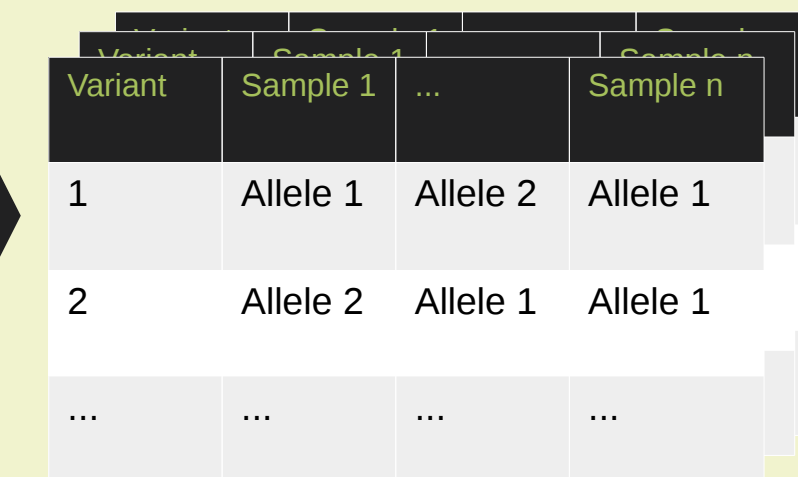


Mapping reads on reference genomes



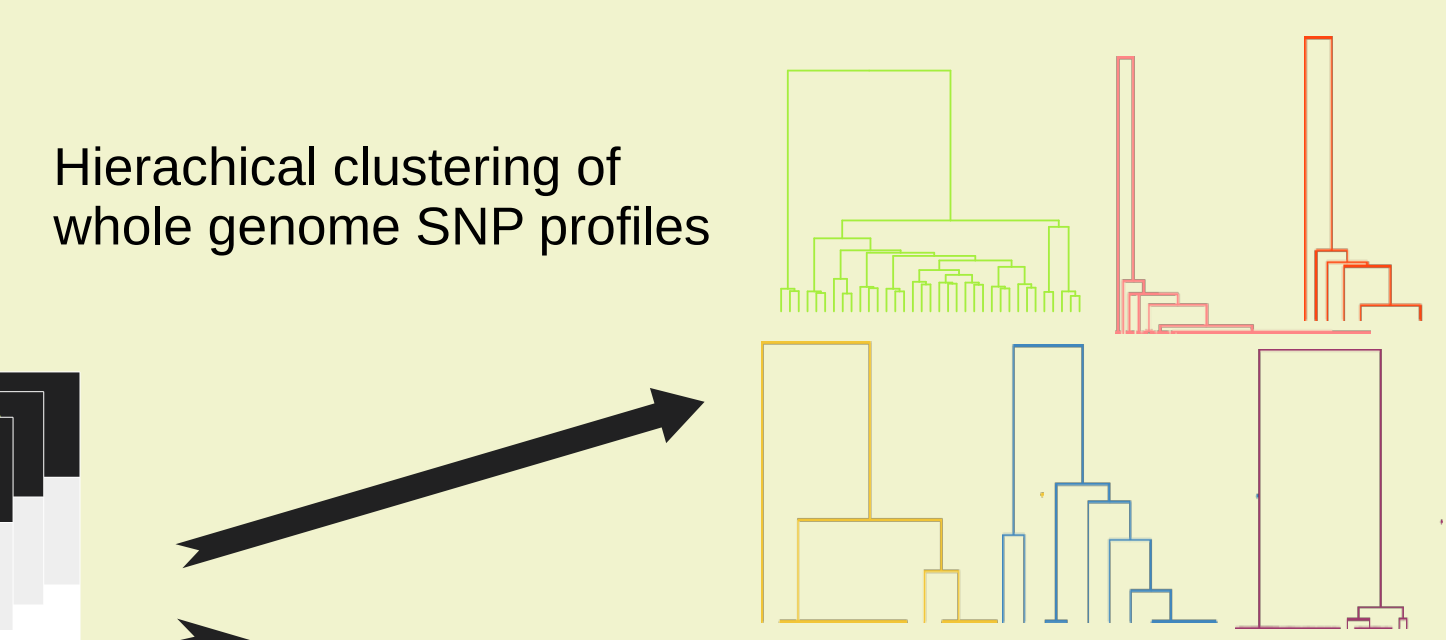
Careful SNP-calling and filtering

- Removing some regions :
 - Homologous between 2 reference genomes
 - Too covered
 - Uncovered
- Filtering rare variants



Building by-sample SNP profiles

- Computing main genotype in sample
- Discarding intra-sample variability



Phylogenetic trees

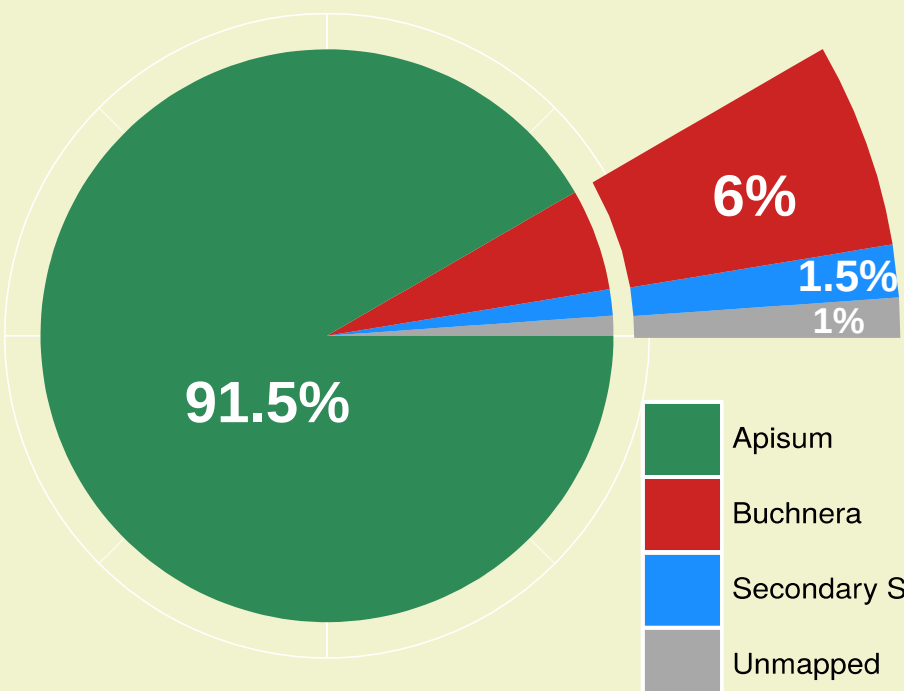
Inter-sample comparison

	Hierarchical clustering of SNP profiles	Phylogenomic analysis
+	Whole genome approach	Enables reconciliation approaches
-	No underlying evolutionary model	Restricted and biased by a list of genes

Question 1

Species level diversity in the pea aphid complex
 → Accurate taxonomic assignation of reads

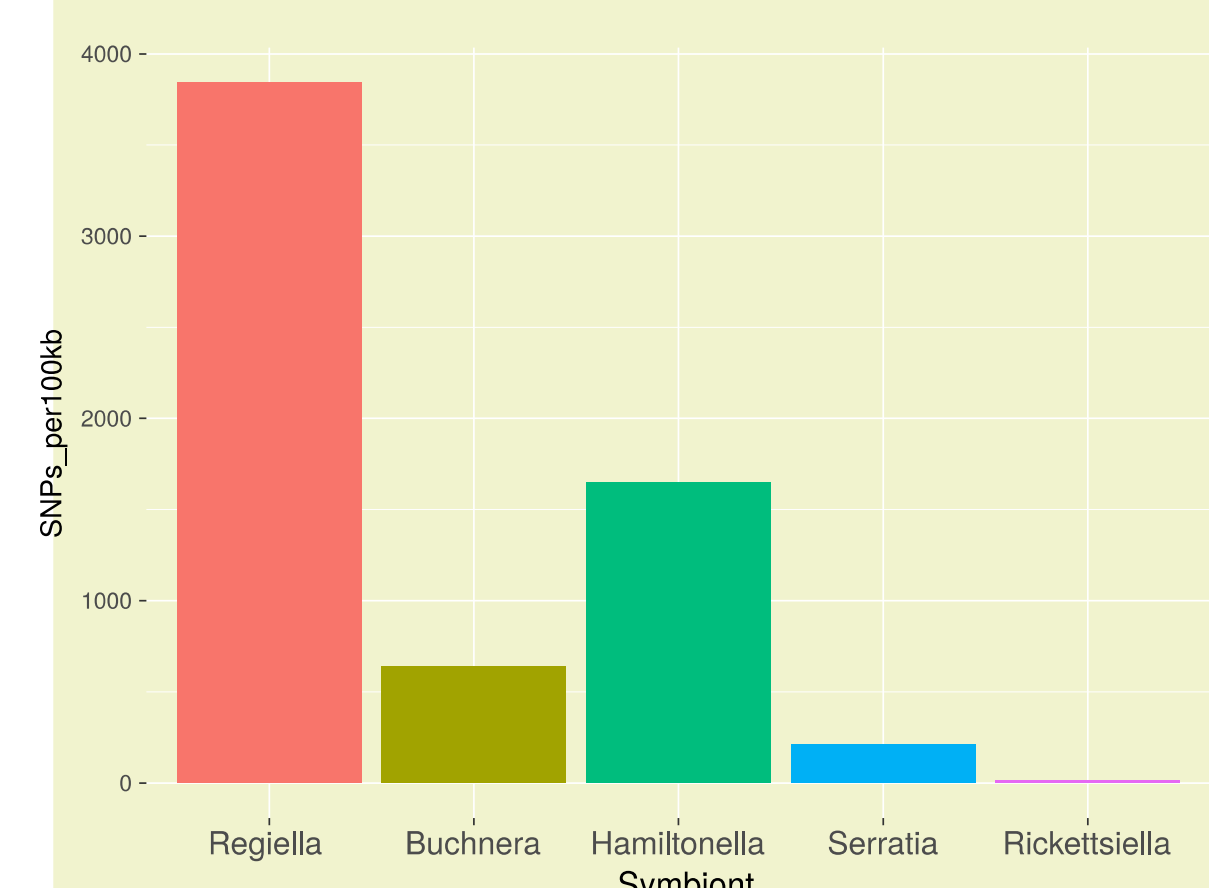
- Good enough reference set (~99% mapped reads)
- Abundance estimated by coverage, omitting homologous or chimeric regions
- More unmapped reads for remote reference sequences



Overall taxonomic assignation of reads

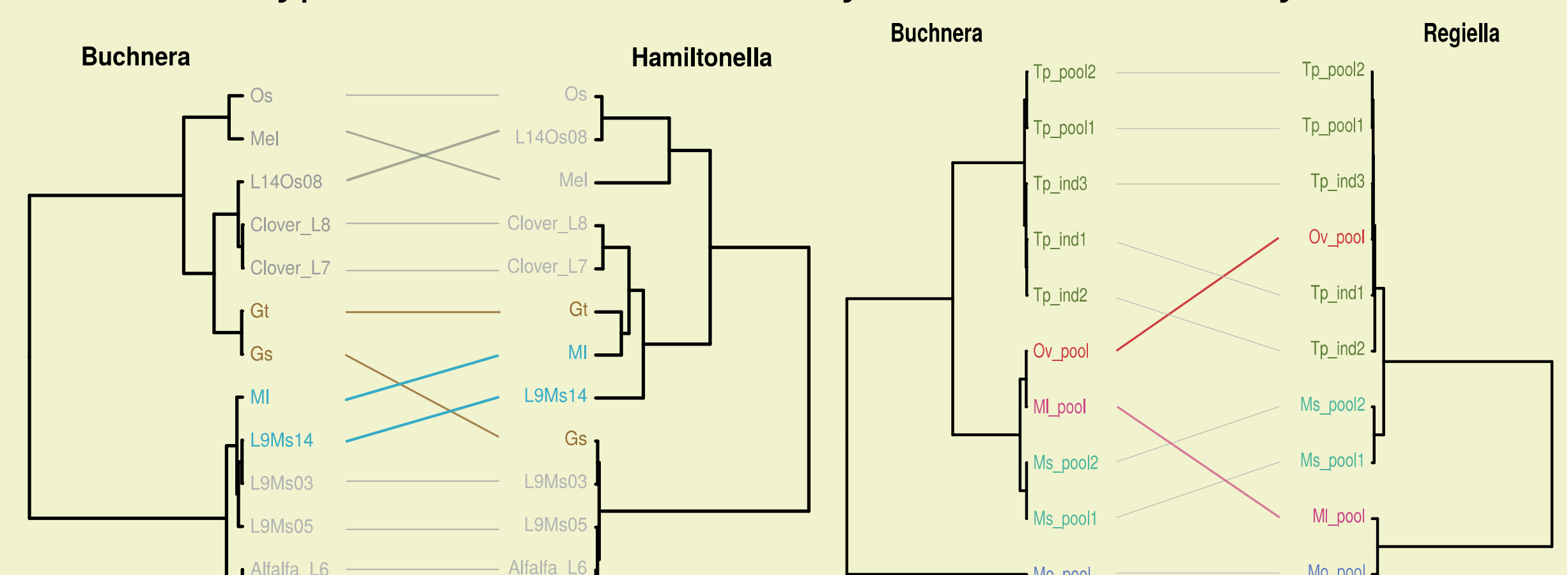
Question 2

Evolutionary dynamics of symbionts
 → SNP-level inter-sample comparison



Variable number of variants detected for the different symbionts
 → Different evolutionary stories

Buchnera aphidicola : vertically transmitted only
 "Null hypothesis" to test evolutionary scenarios for other symbionts



Hamiltonella defensa
 Horizontal transfers

Regiella insecticola
 2 events of acquisition

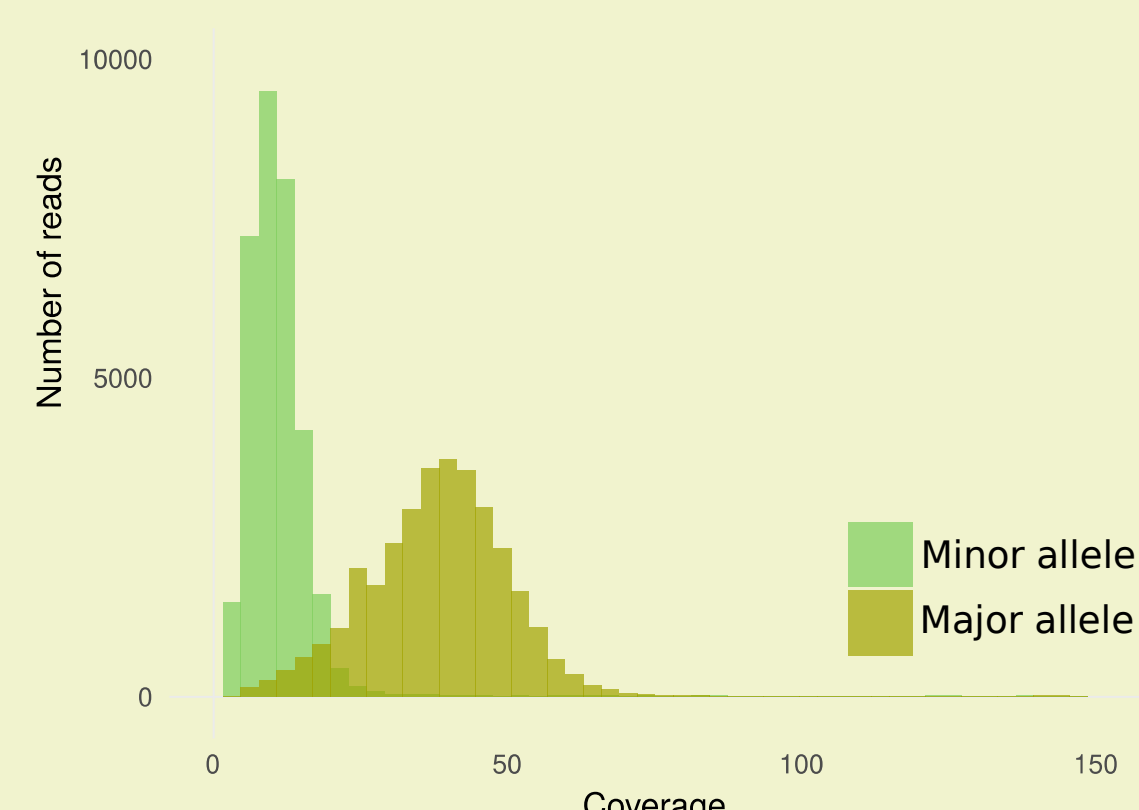
Question 3

Explore intra-sample genomic variability
 → Detection and characterization of several strains inside a single sample

Analyze minor genotypes in samples (discarded for Q.2)

Individual sequencing :

2 cases of intra-sample polymorphism
 2 strains of *Regiella* coexist
 with ~ 30 000 SNPs between them



Pool sequencing :

More than 2 strains may coexist
 Method unable to retrieve coexisting strains

Conclusions

Simple bacterial community finely explained by analysis of multi-sample metagenomic data

- Reference mapping able to capture the most of the diversity for this model
- SNP-calling to sketch evolutionary stories of secondary symbionts
- Able to exploit intra-sample polymorphism in some cases

Limits and todo-list

- Statistical testing of evolutionary scenarios from phylogenetic trees
- What about the 1% of unmapped? Large variant detection and reference free methods