Multi-scale characterization of symbiont diversity in the pea aphid complex through metagenomic approaches
Cervin Guyomar, Fabrice Legeai, Christophe Mougel, Claire Lemaitre, Jean-Christophe Simon

To cite this version:
Cervin Guyomar, Fabrice Legeai, Christophe Mougel, Claire Lemaitre, Jean-Christophe Simon. Multi-scale characterization of symbiont diversity in the pea aphid complex through metagenomic approaches. JOBIM 2017 - Journées Ouvertes en Biologie, Informatique et Mathématiques, Jul 2017, Lille, France. hal-01638884

HAL Id: hal-01638884
https://hal.science/hal-01638884v1
Submitted on 20 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Multi-scale characterization of symbiont diversity in the pea aphid complex through metagenomic approaches

Cervin Guyomar1,2, Fabrice Legeai1,2, Christophe Mougé1, Claire Lemaître2, 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
- Removing some regions:
  - Homologous between 2 reference genomes
  - Too covered
  - Uncovered
  - Filtering rare variants

**Pool Sequencing**
- Higher coverage
- More polymorphism
- Computing main genotype in sample
- Discarding intra-sample variability

**Workflow**
1. Illumina readsets
2. Mapping reads on reference genomes
3. Careful SNP-calling and filtering
   - Removing some regions:
     - Homologous between 2 reference genomes
     - Too covered
     - Uncovered
     - Filtering rare variants
4. Building by-sample SNP profiles
   - Computing main genotype in sample
   - Discarding intra-sample variability

**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

**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