



**Cervin Guyomar**



**Funding**  
MESR

**2015-2018**



**UMR IGEPP**

**Institute for Genetics,  
Environment and Plant  
Protection**

**INRA - Agrocampus  
Ouest - Université de  
Rennes 1**

**Team**

**Ecology and Genetic of  
insects**

**Direction**

Jean-Christophe Simon  
(IGEPP)

Claire Lemaitre (IRISA)

Christophe Mougel (IGEPP)

**Partners**

IRISA Rennes (Genscale  
team)



**Keywords**

Bioinformatics  
Metagenomics  
Aphids  
Symbionts



# Development and application of bioinformatics tools for the metagenomic analysis of insect related bacterial communities



## Social-economic context

This thesis, focused on microbial metagenomics, will help to better understand the links between insect hosts and their associated microorganisms. My work will be focused on the pea aphid, an important crop pest. The better understanding of its relation with its symbionts could lead to the identification of new beneficial symbionts of aphids, and therefore new possible strategies to control this pest.

## Scientific context

The pea aphid harbours a diversity of microorganisms, which can bring several benefits to their host. Therefore, the aphid phenotype results not only from its genotype but also from the genotypes of its associated microorganisms, according to the concept of extended phenotype. Next-Gen Sequencing and metagenomic approaches are able to access this genomic diversity and explain some of the host phenotypic traits expressed by the metagenome. This PhD project aims at the exploration of the aphid bacterial communities, by the use of new or existing bioinformatics tools for metagenomic assembly or variant calling.

## Objectives

The first objective of this project is to extend our knowledge of the microbial communities associated to the pea aphid. This can be done either with reference based techniques, by mapping sequencing data to reference sequences of known organisms, or with reference free *de novo* assembly techniques for metagenomes. This last part will likely require the development of new bioinformatics tools, based on the Minia assembler developed in the Genscale team.

The next step of the project will be to study the microbial diversity of these communities at the inter-specific or even intra-specific level. This will be possible thanks to the development of new variant calling methods for metagenomes, and the study of the genomic diversity across the available metagenomic datasets.

## Results

The firsts months of this thesis (started in October 2015) have been devoted to the constitution of a new reference metagenome for the pea aphid. This has been done by comparing available resequencing data and the current databases, and by closely examining the reads which did not map to any reference. It was then possible to estimate the abundances of the several known symbionts across datasets, and to select a few sample for which a better sequencing using PacBio technology may lead to new reference genomes for some symbionts.

## Perspectives

The first months of this project brought a better knowledge of the metagenomic content of the available datasets. The upcoming PacBio sequencing of several promising samples will make possible the assembly of high quality reference genomes for several species. In the meantime, the next months will be focused on the intra and inter specific diversity for some of the best represented symbionts.

