SHORT TERM SCIENTIFIC MISSION (STSM)
SCIENTIFIC REPORT

This report is submitted for approval by the STSM applicant to the STSM coordinator

Action number: CA16107
STSM title: **Genetics of bacterial fitness of *X. campestris in vitro* and *in planta***

STSM start and end date: 20/01/2018 to 26/01/2018
Grantee name: Alice Boulanger

PURPOSE OF THE STSM:

(max.200 words)
In order to identify the pathways important for bacterial fitness at different steps of plant infection and in different *in vitro* conditions, we have initiated a study using the TnSeq method (described in Wetmore et *al*., 2015), a high-throughput screen of barcoded transposon (Tn) mutants of *Xcc*. This project is developed in collaboration of both Jennifer Lewis and Adam Deutschbauer's lab from the University of California Berkeley (CA, USA). Jennifer Lewis's group is interested in *P. syringae* virulence factors at the Plant Gene Expression Center (PGEC) and Department of Plant & Microbial Biology (PMB) at UC Berkeley. Adam Deutschbauer's group is expert in the construction and validation of barcoded libraries at Lawrence Berkeley National Labs and UC Berkeley.
The objective of this STSM was (i) to acquire data on *Xcc* library from Deutschbauer's group as well as the Bioinformatic and statistical tools, (ii) to determine conditions for future TnSeq experiments especially for *in planta* TnSeq assays, (iii) to develop a common strategy for *Pseudomonas* and *Xanthomonas* fitness analysis *in planta*.

DESCRIPTION OF WORK CARRIED OUT DURING THE STSMS

(max.500 words)
A meeting was organized between J. Lewis, A. Deutschbauer and Morgan Price. M. Price is in charge of the script development for TnSeq biostatistical analysis and the website presenting the database regrouping fitness results obtained by their team and collaborators.
TnSeq data obtained with *Xcc* 8004 before and after growth in nine different *in vitro* conditions, minimal medium with different sugar or amino acid contents, have been analyzed. The same type of preliminary experiments were performed on *Pseudomonas*. Samples from these *in vitro* assays performed on both strains were already sent for BarSeq analysis in the Deutschbauer's lab. Different issued about *in vitro fitness* experiments in *Xcc* were discussed: (i) There is a poor correlation between the two biological repeats of each experience; (ii) the level of fitness defect for genes expected to be involved in tested conditions is low.

To develop a strategy for *Xcc* and *Pseudomonas* fitness comparisons and acquire expertise on

Pseudomonas pathosystem, I have spent 4 days with members of Jennifer Lewis' group.

This time have been dedicated to the comparison of TnSeq library properties and to discuss first *in vitro* and *in planta* experiments results obtained on *Xcc* and *Pseudomonas*. Using these preliminary results we have determined bioinformatic and statistical analysis requirements that will drive the conditions for future TnSeq experiments especially for *in planta* TnSeq assays (number of technical and biological repeats, conditions of plant inoculations, …).

This mission allowed me to present a seminar on our project at PGEC on Thursday 01/25/2018 and to present the EuroXanth cost action. This seminar was followed by meetings with PhD students and young researchers as well as PGEC's principal investigators.

It also gave me the opportunity to meet researchers who are heavily involved in the field of phytopathology at UC Berkeley, including Rodrigo Almeida (one of our COST Observer) and Steven Lindow on Tuesday 01/23/2018.

.

DESCRIPTION OF THE MAIN RESULTS OBTAINED

- **TnSeq mapping of *Xcc* library:**

A. Deutschbauer's group was in charge of redefining the *Xcc* library pool of mutants correctly mapped on the genome.

Mapping quality requires a threshold of at least 10 reads to map one barcode to one position with enough confidence.

Because of the *Xcc* library complexity, three runs of TnSeq (three independent libraries sequenced by NGS on the Illumina sequencer) were required to obtain good quality mapping for a large enough pool of mutant strains. These TnSeq analyses allowed Deutschbauer's group to correctly map around 240000 barcodes representing our new library pool of Tn strains. This is consistent with previous TnSeq library analyses performed by the group on *Phaeobacter inhibens*, or *Shewanella oneidensis* for which only 40 to 60 % of the library was correctly mapped. This pool is large enough to perform *in vitro* and *in planta* analyses.

- **Biostatistical and bioinformatic tools:**

We discussed the importance of sharing the raw data for TnSeq and fitness analyses between Deutschbauer's group and our group. It was decided to establish a specific site where it would be possible to download all the sequencing raw data acquired.

Concerning fitness analyses, Morgan Price has developed a script to produce fitness results and biostatistical analyses from BarSeq results. The script is available for us to use. Furthermore, each time M. Price runs the script to analyse our BarSeq data, it will automatically generate a website with a description of all results and normalizations.

Finally, our team has re-annotated the *Xcc* 8004 genome using RNAseq data, we will send the file with the new annotation to M. Price.

- **Analyses of the first high-throughput genetic screens of *Xcc* library performed *in vitro*:**

M. Price will perform a new analysis using the newly defined library pool. This should decrease the noise in the dataset. It was also decided that more repetitions with many passage on each medium will be done to increase the correlation level and to identify genes involved in growth fitness in rich and poor media.

However, the fact that we are still able to identify genes expected to be affected with only two generations of growth is a strong evidence of a good quality library ready to be used for more fitness assays.

- **Strategy developed for future *in planta* TnSeq assays:**

Lewis' group has performed *in planta* fitness assay with two different strains of *Pseudomonas* by infiltrating their library into *Arabidopsis thaliana* leaves. They have tested different designs for this experiment. The results suggest that there is a strong bottleneck for library entry into plant tissue. Furthermore, Steven Lindow 's lab has performed the same kind of assay on *Phaseolus* leaves and had to treat a thousand plants to cover their library (300000 mapped barcoded Tn mutants).

This is also supported by data published recently by Dangl's group were five thousand plants were required to get proper results.

Altogether, these data suggest that we will have to scale-up analyses by pooling hundreds of samples to get results of sufficient quality.

FUTURE COLLABORATIONS (if applicable)

This STSM allowed us to determine the work coordination to apply in the collaboration between the three groups.
The overall strategy determined during these work discussions is to first perform *in planta* assay on different compartments on as many plants as possible, pool each samples together per repetition, prepare DNA from these pools, and send DNA for sequencing and analysis to Deutschbauer's group.

The foliar pathogens *X. campestris* and *P. syringae* are relevant microorganisms for this study, as they represent the main bacterial populations found on the leaf surface and present in a wide range of crops. With J. Lewis' group, we have designed a common strategy to perform *in planta* fitness analysis allowing direct comparison of these two bacteria.
Lewis' group will engineer the library entry into plant tissue when we will engineer DNA collection after the passage on different plant compartments.