

Deciphering genomic markers to predict plant pathogenicity in *Erwinia* genus

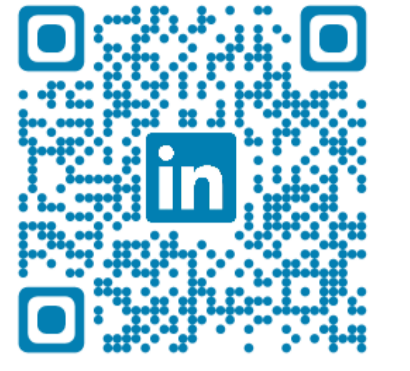
Felipe LIRA¹, Gilles HUNAUULT², Martial BRIAND¹, Perrine PORTIER¹, Claudine LANDES¹, and Marion FISCHER-LE SAUX¹



¹IRHS, INRA, Université d'Angers, Agrocampus-Ouest, SFR 4207 QuaSaV, 49071, Beaucauzé, France.

²Hémodynamique, Interaction Fibrose et Invasivité Tumorales Hépatiques Laboratory, Unité Propre de Recherche de l'Enseignement Supérieur 3859, Structure Fédérative de Recherche 4208, Bretagne Loire University, Angers, France.

Corresponding Author: felipelira3@gmail.com



BACKGROUND

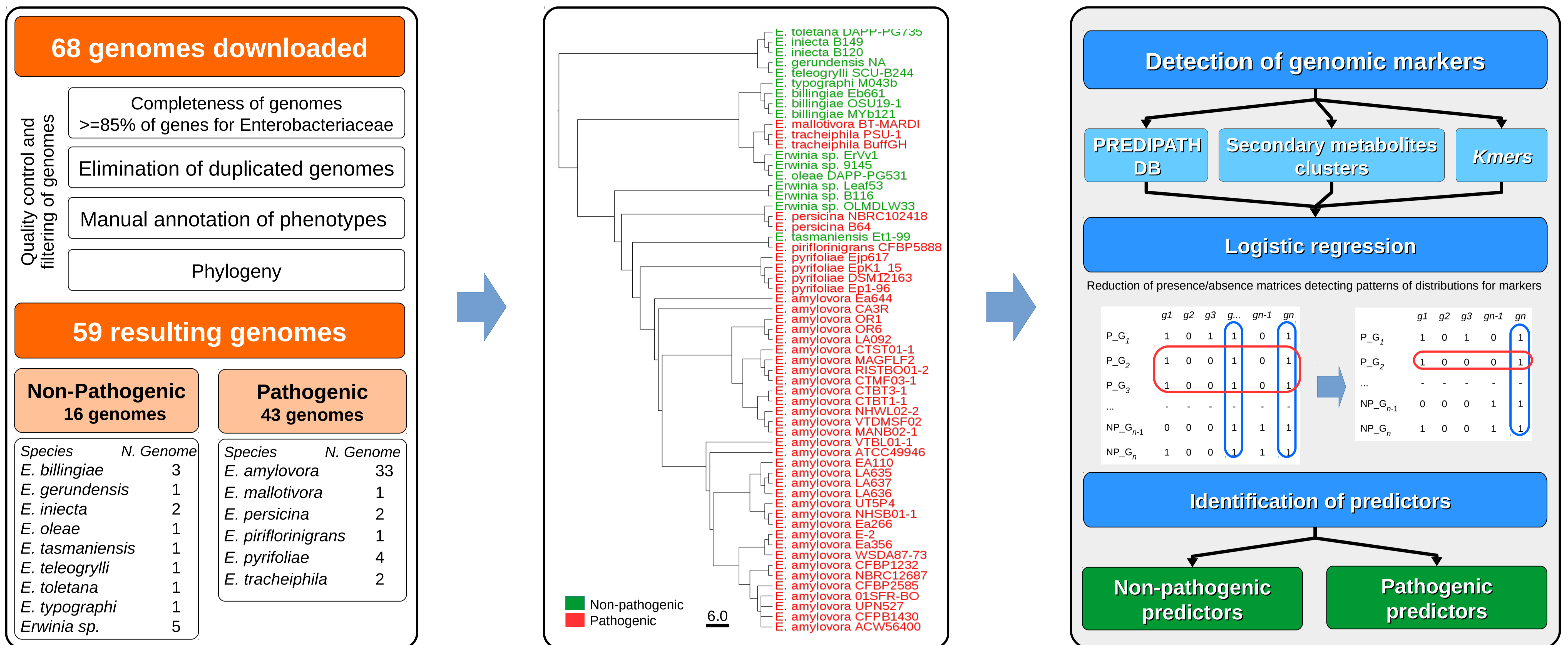
Erwinia genus is formed by species which can be classified as **non-plant-pathogenic** and **plant-pathogenic**. However, non-plant-pathogenic organisms may be found associated with plants, as commensals, or ass. Comparative genome-wide characterization has been proving its efficiency e to predict phenotypes of interest differentiating between predetermined classes formed by isolates with known phenotypes. It is known that extensive intra-species variations can be recognize based on the presence or absence of whole genes, genes clusters, and *kmers* to infer their genotype and phenotype associations.

OBJECTIVE

Our aim is to perform a comparative genomics analysis of closely related strains and species from the genus *Erwinia* to describe their genetic variability and to identify genomic elements which enable select predictors to distinguish between pathogenic from non-pathogenic individuals.

METHODS

Bacterial genomes of genus *Erwinia* were downloaded and submitted to the **PREDIPATH** workflow to discover potential predictors of phenotypes.



RESULTS

Phylogeny calculated using complete genomes, bacterial core genes, and *k*-mers composition, didn't demonstrated a separation between classes **NP** and **P**. Otherwise, PREDIPATH DB, biosynthetic gene clusters (BGCs), and *k*-mers composition, permitted to predict **NP** and **P** organisms.

From 228 genes detected using PREDIPATH DB, eight of them permitted to predict the potential pathogenicity of plant-associated species. *Erwinia tasmaniensis* revealed their close relationship with pathogenic bacteria, presenting genes commonly attributed to pathogenic bacteria.

Seven BGCs highlighted the potential to predict plant-pathogenic organisms in genus *Erwinia*. Biosynthetic - **thiopeptide**, **nrps**, **siderophore**, **terpene**, **butyrolactone**, **arylpolyene t1pks**, and the **homoserine lactone cluster**, was present in 100% of genomes from NP.

Genome-Wide Association Study indicated the presence of 512 overlapping *k*-mers with significant distribution through the two groups. Class **NP** presented the greater richness of exclusive *k*-mers to be used as predictors for *Erwinia* spp.

| Genes | NP | | P | |
|-------------|-----|-------|-----|-------|
| | sum | % | sum | % |
| <i>hrpF</i> | 1 | 6.25 | 41 | 95.35 |
| <i>hrcN</i> | 1 | 6.25 | 40 | 93.02 |
| <i>hrpX</i> | 1 | 6.25 | 39 | 90.70 |
| <i>invB</i> | 0 | 0.00 | 38 | 88.37 |
| <i>hrpV</i> | 1 | 6.25 | 38 | 88.37 |
| <i>spaR</i> | 0 | 0.00 | 37 | 86.5 |
| <i>spaQ</i> | 1 | 6.25 | 37 | 86.05 |
| <i>acrD</i> | 14 | 87.50 | 3 | 6.98 |

| <i>kmers</i> | NP | | P | |
|--------------|-----|--------|-----|-------|
| | sum | % | sum | % |
| 774086 | 16 | 100.00 | 0 | 0.00 |
| 1126723 | 16 | 100.00 | 1 | 2.33 |
| 1137966 | 16 | 100.00 | 5 | 11.63 |
| 1005477 | 16 | 100.00 | 6 | 13.95 |
| 577828 | 15 | 93.75 | 0 | 0.00 |
| 1122208 | 15 | 93.75 | 1 | 2.33 |
| 833362 | 15 | 93.75 | 2 | 4.65 |
| 926871 | 14 | 87.50 | 2 | 4.65 |
| 606155 | 14 | 87.50 | 5 | 11.63 |
| 135477 | 14 | 87.50 | 5 | 11.63 |
| 454779 | 14 | 87.50 | 6 | 13.95 |

CONCLUSIONS

- Phylogenetic distribution was not able to distinguish Pathogenic and Non-pathogenic organisms in genus *Erwinia*.
- BGCs involved in quorum sensing communication and formation of biofilms, such as homoserine lactone, offer insight about the presence QS in the class of non-plant-associated *Erwinia* spp.
- Our approach enables the compilation of a complete datasets formed by genomic sequences, such as **genes related to infective pattern**, **biosynthetic gene clusters** and **kmers** to predict the potential bacterial pathogenicity on plants.
- PREDIPATH** workflow allowed the creation of exclusive datasets of predictors associated *Erwinia* spp. according their phenotypes offering clues to understanding the evolution of these pathogens and the mechanisms involved at their pathogenicity.

- Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM, BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics, *Molecular Biology and Evolution*, Volume 35, Issue 3, March 2018, Pages 543–548
- Blin K, Wolf T, Chevrette MG, et al. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* 2017;45(W1):W36-W41.
- Jaillard M, Lima L, Tournoud M, Mahé P, et al. (2018) A fast and agnostic method for bacterial genome-wide association studies: Bridging the gap between kmers and genetic events. *PLoS Genetics* 14(11).
- Chen LH, Yang J, Yu J, Yao ZJ, Sun LL, Shen Y and Jin Q, 2005. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 36 (Database issue):D539-D542.
- Jia B, Raphenya AR, Alcock B, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 2016;45(D1):D566-D573.
- Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E., Larsson, DGJ. (2014) BacMet: antibacterial biocide and metal resistance genes database, *Nucleic Acids Res.*, 42, D737-D743
- Na, S.-I. I., Kim, Y. O., Yoon, S.-H. H., Ha, S. min, Baek, I., & Chun, J. (2018). UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *Journal of Microbiology*, 56(4), 281–285.