

Comparative genomic approaches to predict ecological niche of *Ralstonia* spp.: PREDIPATH

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

¹ IRHS, INRA, Université d'Angers, Agrocampus-Ouest, SFR4207 QuaSaV, 49071, Beaucouzé, France

² Hémodynamique, 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
felipelira3@gmail.com

Abstract. The genus *Ralstonia* can be divided in plant-associated (PA) and not plant-associated (NPA) species. PA bacteria are responsible for devastating diseases on many crops while NPA species are opportunistic pathogens in immunocompromised patients. The specificity of living habits of both groups suggests the potential to detect specific elements to predict the potential pathogenicity on plants. Determination of pathogenicity **to for** plants relies on microbiological methods which are time consuming and require specialists. The specificity of living habits of each group suggests the possibility to detect specific genetics elements in their genome that could help to predict their potential pathogenicity. This study proposes the PREDIPATH workflow to detect genomic markers associated to bacterial phenotypes. It uses a customized **data-base** composed by protein sequences from public repositories and secondary metabolites **clusters** and *k*-mer detection. This workflow has allowed to create exclusive datasets of markers that could be used as predictors of the potential pathogenicity on plant of *Ralstonia* spp.. It has also revealed common traits that could be linked to ecological adaptation.

Keywords: phenotype-genotype association, *k*-mers, antimicrobial resistance genes, secondary metabolite biosynthesis gene clusters, virulence genes

1 Introduction

Pathogenic bacteria need specific mechanisms for surviving in adverse environment and invading their host. These common adaptive strategies suggest the presence of specific and/or exclusive genes or alleles which could enable us to predict and characterize the bacterial phenotypes from genome content. In this manner, the PREDIPATH workflow has been developed to detect these specific markers for a given taxonomic group representing contrasting ecological features.

2 Creation of PREDIPATH-DB and detection of markers

Compilation and clustering of sequences from public databases resulted in a **database**, named PREDIPATH-DB (<https://github.com/felipelira/PrediPath>), of 14.248 protein sequences related to antimicrobial and heavy metal resistance, and to virulence [1]. Complementary approaches to detect class-specific markers, focused on the detection of secondary metabolites **clusters** [2] and specific *k*-mers using SkIf (<https://iris.angers.inra.fr/galaxypub-cfbp/>) [3].

Our approach was tested using 147 *Ralstonia* spp. genomes, 108 classified as PA and 39 as NPA. Phylogenetic analysis was performed with using UBCG software v.3.0 [4]. Calculation of overall genome relatedness using KI-S [5] was used for species delineation.

3 Results

Phylogenomics allowed us to rename misidentified genomes. Among the 126 genes detected using the PREDIPATH-DB, a class D β -lactamase-encoding gene was detected exclusively in all genomes from species responsible for nosocomial infections (class NPA). Plant pathogenic species presented 13 exclusive genes from type II and type III secretion systems. Analysis of clusters responsible for secondary metabolite synthesis revealed that all *Ralstonia* strains presented the potential to produce terpenes and that only genomes in class PA presented non-ribosomal peptide synthetase and homoserine lactone clusters. This finding suggests that *Ralstonia* spp. may have the ability to use terpenes for interkingdom communication, considering their environmental origin [6]. Class-specific *k*-mers were detected. In total, 4487 fragments were distributed along 1395 CDS for PA, and 708 throughout 506 CDS of NPA. In total, 1408 gene variants and two secondary metabolite clusters were exclusive of genomes in class PA and 507 gene variants were exclusive of genomes in class NPA. These two sets of markers, one for each class, will help to predict the potential phenotype of unknown *Ralstonia* spp. It is interesting that the PREDIPATH workflow has evidenced traits that make sense like resistance to carbapenems in all clinical species and microorganism communication via quorum sensing in plant pathogens. Further studies will help to decipher their importance as niche adaptive traits.

References

1. PREDIPATH Homepage, <https://github.com/felipelira/PrediPath>, last access: 2019/08/02
2. Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., Weber, T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline, *Nucleic Acids Research*, Volume 47, Issue W1, Pages W81–W87, (2019). doi: 10.1093/nar/gkz310
3. Na, S.I., Kim, Y.O., Yoon, S.H., Ha, S.M., Baek, I. & Chun, J.. UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *Vol. 56, No. 4*, pp. 280–285 (2018). doi: 10.1007/s12275-018-8014-6
4. Identification of genetic relationships and subspecies signatures in *Xylella fastidiosa*; Nicolas Denancé, Martial Briand, Romain Gaborieau, Sylvain Gaillard & Marie-Agnès Jacques; *BMC Genomics*; volume 20, Article number: 239 (2019)., doi: 10.1186/s12864-019-5565-9
5. Briand, M., Bouzid, M., Hunault, G., Legeay, M., Fischer-Le Saux, M., Barret, M. A rapid and simple method for assessing and representing genome sequences relatedness. *BioRxiv.*, (2019). doi: 10.1101/569640
6. Schmidt, R., Jager, V., Zühlke, D., Wolff, C., Bernhardt, J., Cankar, K., Beekwilder, J., Ijcken, W., Sleutels, F., Boer, W., Riedel, K., Garbeva, P. Fungal volatile compounds induce production of the secondary metabolite Sodorifen in *Serratia plymuthica* PRI-2C. *Scientific reports*, 7(1), 862. (2017). doi:10.1038/s41598-017-00893-3