

Comparative genomics of *Xanthomonas* spp. focusing on CAZymes associated with host-pathogen specificity

Gabriela Persinoti¹, Mario Tyago Murakami¹,

1 CTBE/CNPq

Abstract

Xanthomonas is a genus of Gram-negative Gammaproteobacteria that cause infections in leaves and fruits of plant hosts. Around 400 plants may be infected by *Xanthomonas* species, among them, many are economically important ones, such as citrus, rice, tomato and banana. A high degree of host specificity is observed among *Xanthomonas* pathogenic species and pathovars. For instance, *X. citri* pv. *citri* exclusively infects citrus while, other pathovars such as *X. citri* pv. *mangiferaeindicae* infects mango and *X. vesicatoria* may infect tomato and pepper. The goal here was to perform a large scale comparative genomics analysis of *Xanthomonas* focusing on the relationship of CAZyme-genome content and host-bacterium specificity. For the comparative genome analysis, 51 complete genomes of *Xanthomonas* species and pathovars were used. To investigate the phylogenetic relationship of *Xanthomonas*, 699 single copy genes with members in 51 species were identified. Single-gene alignments longer than 100 residues after excluding low-scoring alignment sites were concatenated into a supermatrix. The resulting supermatrix is composed of 111,390 distinct alignment patterns. Maximum likelihood trees were estimated using either FastTree using WAG+CAT model and RAxML using a distinct model for each of the 582 partitions. Both trees presented strong support values (Bootstrap > 95%) and the same topology, which was assessed by Robinson-Foulds distance implemented in phangorn R package. Two distinct groups were clearly formed among *Xanthomonas* species. One group is composed by *X. sacchari*, *X. albilineans*, *X. translucens*, and *X. hyacinthi*, while another larger group is composed by *X. campestris*, *X. arboricola*, *X. gardneri*, *X. hortorum*, *X. fragariae*, *X. cassavae*, *X. bromi*, *X. oryzae*, *X. vasicola*, *X. fuscans*, *X. citri*, *X. axonopodis*, *X. euvesicatoria* and *X. perforans* including several pathovars. Regarding the CAZyme-genome content, *X. fuscans* and *X. hyacinthi* present, respectively, the fewer and greater number of CAZymes identified. For instance, *X. citri* pv. *citri*, a mesophyllic pathogen, which infects the intercellular spaces of the mesophyll tissue causing citrus canker, presents 239 CAZymes, being many organized as PUL (Polysaccharide Utilization Loci). In the other hand, *Xylella fastidiosa*, also a citrus plant pathogen, but a vascular pathogen, which infects the xylem elements of the vascular system, presents only 82 CAZymes, and none of them are organized as PULs. This suggests that the CAZyme may play a role in host-pathogen interactions in *Xanthomonas*.

Funding: FAPESP and CNPq.