Comparative Genomic Analysis of *Erwinia carotovora* subsp. *atroseptica*: Evidence For Extensive Horizontal Gene Transfer with Plant Associated Bacteria

**Leighton Pritchard**, Paul R.J. Birch, Ian K. Toth

Plant Pathology,SCRI, Invergowrie, Dundee, Scotland, DD2 5DA

**Introduction**

The family Enterobacteriaceae includes many devastating animal pathogens, including *Escherichia coli*, and species of *Yersinia* and *Shigella*, yet it also contains important plant pathogens, such as *Erwinia carotovora* subsp. *atroseptica* (*Eca*). Horizontal Gene Transfer (HGT) of genes encoding determinants of bacteria-plant interaction has been proposed as a mechanism to explain the acquired plant-associated lifestyle and phytopathogenicity of *Eca*. We present results from comparative genomics of *Eca* with thirteen animal- or plant-pathogenic enterobacteria (APE) and fourteen plant- or animal-associated bacteria (PAB) demonstrating that nearly 500 coding sequences (CDS) from *Eca* (over 10% of the genome) bear greater average percentage sequence identity to sequences from PAB than to sequences from APE. These sequences are distributed throughout the genome, but are frequently found in clusters, and often cluster at, or adjacent to 17 putative horizontally-acquired islands (HAIs).

By way of comparison, a similar comparative genomics analysis of *Salmonella enterica* subsp. *enterica* serovar Typhi CT18 (STy) reveals very few CDS with greater percentage sequence identity to PAB than to APE. These results are consistent with the proposal that *Eca* may have acquired determinants of its plant-associated lifestyle from plant-associated bacteria that were already present in that ecological niche.

CDS with greater percentage identity to PAB than APE sequences were found to be preferentially distributed in HAIs than in the chromosomal backbone of *Eca*, and this tendency was most pronounced for functional classes such as transport, binding, and regulation.

**Comparisons**

The *Eca* genome carries 17 major putative HAIs, and several additional minor HAIs, members of both carry genes involved in pathogenesis and/or ecological adaptation. The *Eca* genome is highly similar to the APE genomes, as seen in figure 1: three quarters of all CDS in *Eca* have RBHs in APE genomes. This is consistent with *Eca*’s immediate evolutionary history as a member of the Enterobacteriaceae. The overall pattern of distribution of genes is one in which core functions in *Eca* (energy metabolism, motility, etc.) are shared with both APE and PAB genomes, and are located on the chromosomal ‘backbone’, but CDS that encode for proteins associated with cell surface, pathogenicity, plant-associated functions, transport and regulation are proportionally over-represented in the major HAIs, and also show a greater than expected proportion of RBHs to PAB.

The doughnuts each represent a different functional grouping: (a) all CDS in *Eca*; (b) CDS associated with pathogenicity; (c) CDS associated with transport and binding; (d) CDS associated with regulatory functions. In all four functional groups, the proportion of CDS with RBH to enterobacteria is lower in the putative HAIs than in the chromosomal backbone, also, the proportion of CDS making hits only to PAB is greater in the HAIs than in the chromosomal backbone. For functional classes associated with transport and binding, and regulatory functions, the majority of CDS in HAIs make no RBH to enterobacteria. In all cases the proportion of CDS that make hits to PAB and APE is much reduced in the putative HAIs.

**Discussion**

The doughnuts each represent a different functional grouping: (a) all CDS in *Eca*; (b) CDS associated with pathogenicity; (c) CDS associated with transport and binding; (d) CDS associated with regulatory functions. In all four functional groups, the proportion of CDS with RBH to enterobacteria is lower in the putative HAIs than in the chromosomal backbone, also, the proportion of CDS making hits only to PAB is greater in the HAIs than in the chromosomal backbone. For functional classes associated with transport and binding, and regulatory functions, the majority of CDS in HAIs make no RBH to enterobacteria. In all cases the proportion of CDS that make hits to PAB and APE is much reduced in the putative HAIs.

**References**


Toth, I.K., Pritchard, L. and Birch, P.R.J. (in press) Annual Reviews in Phytopathology

**Materials and Methods**

Reciprocal best hit (RBH) analyses were carried out using FASTA 3.2.42XS with the protocol described in Bell et al. (2004). RBHs were counted if sequences showed greater than 30% sequence identity over 80% of the shortest sequence length. Functional classifications were taken from the Sanger Institute functional annotation of *Eca*. Circular diagrams were constructed using the GenomeDiagram package described in Pritchard et al. (2006). Genome sequences for comparison were downloaded from ftp.ncbi.nih.gov/genera/Bacteria.