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

Scottish Crop Research Institute

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-pathogenic enterobacteria (APE) and fourteen plant-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 (HAI).

By way of comparison, a similar comparative genomics analysis of *Salmonella enterica* subsp. *enterica* serovar Typhi CT18 (*S*Ty) 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

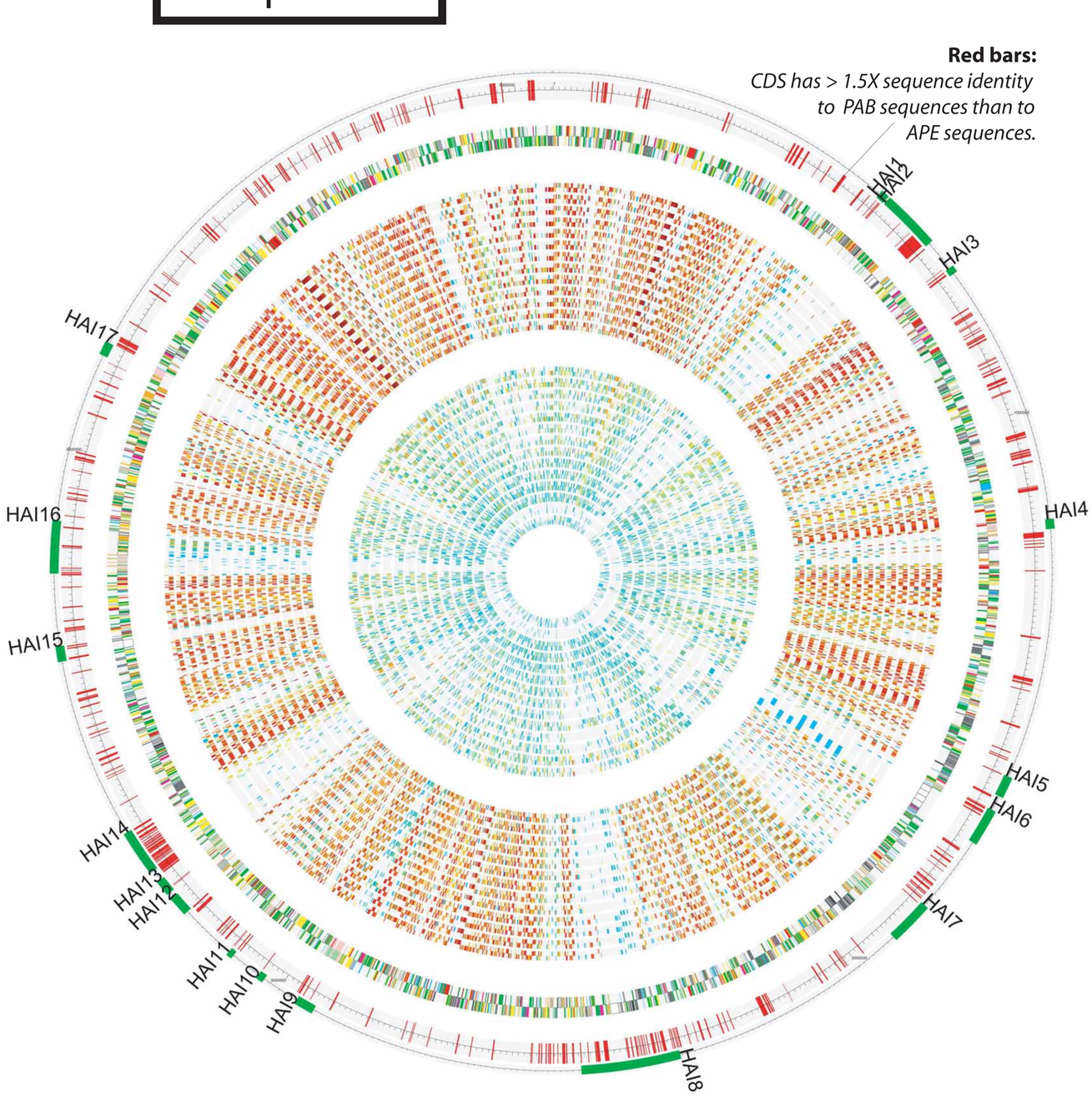

Figure 1 (left): Circular representation of the *Eca* chromosome indicating coding sequences (CDS), reciprocal best hits (RBHs) for each CDS to other bacterial genomes, and the ratio PAB:APE of percentage identity to best hits in PAB and APE.
From the outer edge, successive rings represent: locations of putative horizontally-acquired islands (HAIs); CDS for which the PAB:APE ratio > 1.5; CDS from *Eca* in forward and reverse directions, coloured by functional class; RBHs to CDS from other bacteria, in order of decreasing mean percentage identity per CDS. The comparison sequences are split at the division between the thirteen APE (outer group) and fourteen PAB. (inner group).
The inner rings representing RBHs to each CDS are coloured individually on a scale from 30% (cyan) to 100% (brick red) amino acid identity.
A larger, linear version of this diagram is also available nearby for more detailed viewing.

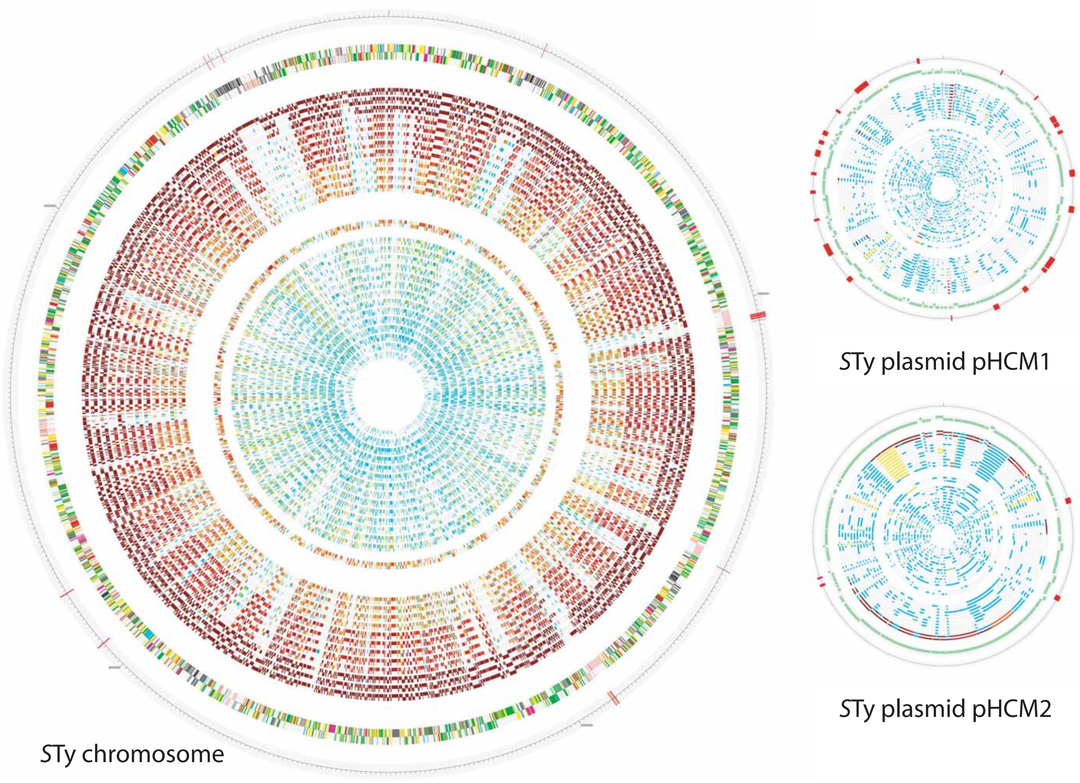

Figure 2 (above): Circular representations of the *S*Ty chromosome and plasmids pHCM1 and pHCM2, and comparisons with thirteen APE and fourteen PAB. From the outer edge, successive rings represent: CDS for which the PAB:APE ratio > 1.5; CDS from *S*Ty in forward and reverse directions, coloured by functional class; RBHs to CDS from other bacteria, in order of decreasing mean percentage identity per CDS. The comparison sequences are split at the division between twelve APE (outer group) and *Eca*, and between *Eca* and the fourteen PAB. (inner group).

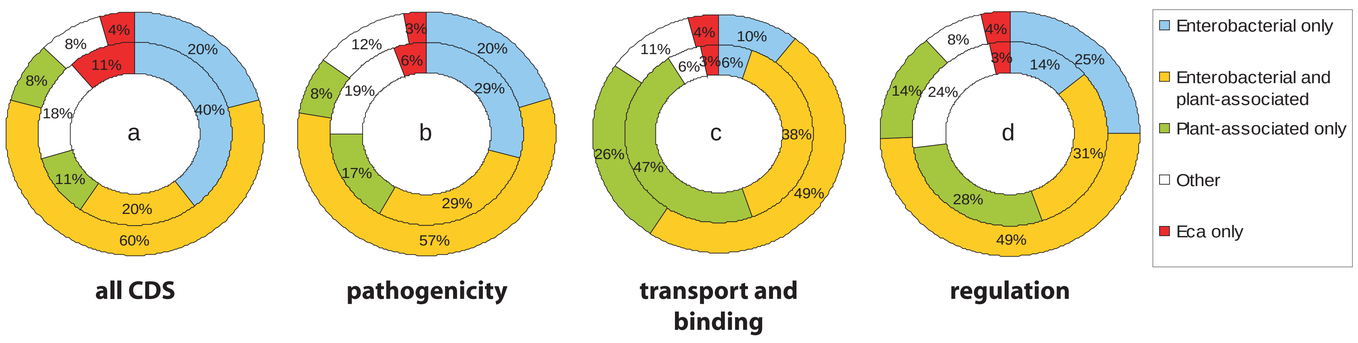

Figure 3 (left): Doughnut charts indicating the percentages of CDS in the *Eca* chromosome found in the 17 putative major HAI (inner rings) and the chromosome backbone (outer rings). CDS are classified according to whether they have RBH only with APE (blue), APE and PAB (yellow), PAB only (green), only to bacteria outwith the PAB and APE (white), or have no RBH in a comparison with 229 other bacteria (red). 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 *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.
From figures 1 and 2 it can be seen that the majority of CDS from the *Eca* and *S*Ty chromosomes are, as expected, more similar to their counterparts from APE than PAB. CDS with greater sequence identity to PAB than to APE are associated with sites of putative HGT in both organisms. In *S*Ty, 6/7 chromosomal locations where this is the case are associated with phage and transposon-related sequences. In *Eca*, 14/17 major HAIs include clusters of PAB-associated sequences, and further minor clusters of CDS outwith the HAI are found in regions where there is little similarity to APE.
The overall extent of CDS with greater percentage sequence similarity to PAB than APE is substantially greater for *Eca* than for *S*Ty, recapitulating the view that HGT is influenced more by physical proximity than by phylogenetic relatedness. The great extent of putative HGT between *Eca* and PAB is consistent with the role of a `hub' of horizontal gene transfer proposed by Kunin *et al.* (2005), and that so many putative HGT events appear to be associated with plant-associated bacteria is consistent with the *Eca*'s adaptation to a plant-associated lifestyle (Toth *et al*. (in press)).

References

Kunin, V., Goldovsky, K., Darzentas, N. and Ouzounis, C.A. (2005) Genome Research **15** 954-959

Pritchard, L., White, J.A., Birch, P.R.J. and Toth, I.K. (2006) *Bioinformatics* **22** 616-617

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 3.24t25 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/genomes/Bacteria.