# New discoveries with *Erwinia* genomics

I.K. Toth, L. Pritchard, M.C. Holeva, L.J. Hyman, K.S. Bell, S.C. Whisson, A.O. Avrova & P.R.J. Birch

*Erwinia carotovora* subsp. *atroseptica* (*Eca*) is an economically important pathogen of potato, causing blackleg of plants in the field and soft rot of tubers post-harvest. Its pathogenicity is primarily dependant on the tightly regulated production of large amounts of extracellular enzymes that degrade plant cell walls, with other factors such as iron acquisition and mechanisms to defend against plant attack also playing a role. In recent years, however, it has become clear that soft rot pathogenesis is more complex than previously thought and the relationship between *Eca* and potato / non-host plants is still far from understood. As a new approach to gene discovery in *Eca,* the complete genome sequence and annotation of *Eca* was determined in collaboration with the Sanger Institute, Cambridge, UK and SCRI through SEERAD funding. The genome is ca 5 Mb with 4, 491 coding sequences.

Analysis of the genome, and comparison with 60 other bacterial genomes using bioinformatics has revealed a wealth of new information, including putative pathogenicity factors previously unknown in this organism. For example, we have discovered i) a number of putative toxin genes, including those possibly involved in the formation of the polyketide-based coronafacic acid (part of the plant toxin coronatine produced by *Pseudomonas syringae* during infection); ii) a cluster of genes similar to a type IV secretion system that, in the plant pathogen *Agrobacterium tumefaciens,* plays a major role in the disease process.

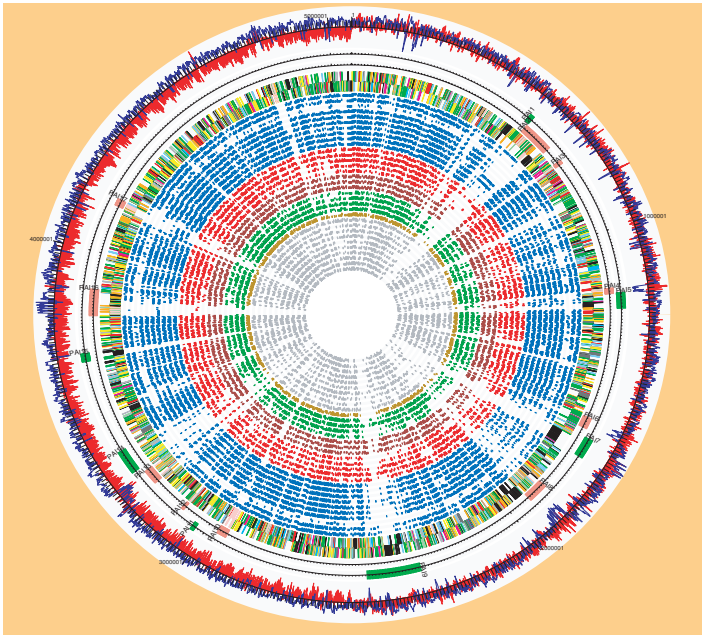

**Figure 1** Comparison of the *Eca* genome sequence with other bacterial genomes: Inner to outer tracks: the locations of reciprocal best hits found by reciprocal FASTA of *Eca* CDSs against those from 32 bacterial genomes: Gram+ (grey); *Shewanella oneidensis* (ochre); non-enteric animal pathogens (green); plant-associated bacteria (brown); non-enteric plant pathogens (red); enterobacteria (blue). The locations of CDSs on the *Eca* genome, coloured by functional class. Two tracks indicating islands listed in Table 1: islands with evidence of recent acquisition (red bars), possible islands based on reciprocal FASTA analysis (green bars). A plot of G+C skew (red) and %GC content (blue).

We have also found that the number of pathogenicity genes acquired from more distantly-related bacteria, possibly *via* horizontal gene transfer, was greater than expected. Many of these distantly-related bacteria are plant pathogenic or plant associated, suggesting that *Eca* may have developed its plant pathogenic lifestyle through gain of important genes, following exchange of DNA with bacteria relevant to a plant associated lifestyle.

In collaboration with the *Phytophthora infestans* group at SCRI, we have developed a 'transposon mutation grid', allowing pooled libraries of transposon mutants to be searched rapidly for mutations in any given gene in the genome. We also have potato plants, obtained as miniplants from a commercial source, available for disease testing throughout the year. Using this dual approach over the last 6 months, we have isolated over 20 *Eca* mutants and determined the role of some important novel genes in pathogenicity, including those associated with both the coronafacic acid and type IV secretion system.

Finally, a number of other functional genomics programmes are being developed i) at SCRI, including microarrays containing the complete set of *Eca* coding sequences, to study the genome at the gene expression level both *in vitro* and *in planta*; ii) in collaboration with other institutions, such as Cambridge University and Moredun Research Institute, including proteomics to study the genome at the protein level.