Erwinia Genomics: A new era in the battle against potato disease

Toth I.K., K.S. Bell, M.C. Holeva, G.J. Bryan, A.O. Avrova, L.J. Hyman and P.R.J. Birch

***Erwinia* research at SCRI** For over 30 years SCRI has been a world-leader in research on the potato pathogen *Erwinia carotovora* subsp. *atroseptica* (*Eca*) (Fig. 1), the causal agent of blackleg disease of field

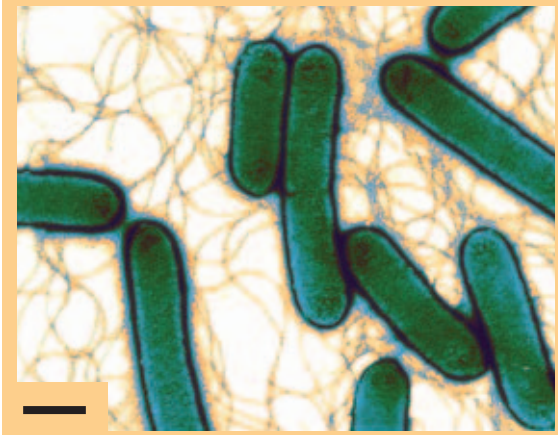

Figure 1 Electron micrograph of *Erwinia carotovora* subsp. *atroseptica* following colour enhancement. Bar = 1.0 µm.

crops and soft rot disease of stored tubers (Fig. 2). This research has focussed on the development of diagnostics, both as predictive tools for growers and for use in epidemiological studies, allowing pathogen movement to be followed through the growing season and between crop generations. The research has had a major impact on the way growers manage their crops and, as a consequence, on decreasing disease incidence. However, it remains a major problem in many potato-growing regions and more research is thus needed to continue the fight against this disease.

Figure 2 Potato plant with symptoms of blackleg disease.

The growth of genomics and microbial genome sequencing Since the structure of DNA was first described in the 1950s, molecular biology has revolutionised plant pathology. Less than 50 years on, in 1995, the genome of *Haemophilus influenza* was the first free-living organism to be completely sequenced (Fleischmann *et al.* 1995). Seven years later there are over 70 complete genome sequences, and almost 200 in progress, providing information relevant to many aspects of microbial life. These sequences are being used successfully to find genes involved in disease, new targets for drugs, and a plethora of other basic physiology, cell biology and evolutionary functions (Doolittle 2002).

Microbial genome sequencing has led to a number of exciting new discoveries, *e.g.* how microbes have evolved and adapted to different environments and ways of life, and how genomes lose and acquire genetic information. Currently, around a quarter of the genome sequences are from medically-related pathogens. These genomes have allowed comparisons between virulent and non-virulent strains of the same organism to pin-point the genetic bases for these differences in virulence (Perna *et al.* 2001). Sequence comparisons have also shown clearly how organisms from a common ancestor can adapt to different niches during divergent evolution (Cole *et al.* 2001). Common mechanisms between pathogenic bacteria have been discovered, such as the Type III secretion system used to inject proteins into plant and animal cells or pathogenicity islands, which are large regions of DNA carrying clusters of genes correlated with virulence. Microbial genome sequences are the basis on which new drugs, vaccines and diagnostic tools are now being developed (Doolittle 2002), and some genomes may prove useful in bioremediation and biotechnology.

The first plant pathogen genome to be fully sequenced was *Xylella fastidiosa* (Simpson *et al.* 2000), and there are now a handful of others, *i.e.*, *Agrobacterium tumefaciens, Ralstonia solanacearum,*

Xanthomonas campestris pv. *campestris* and *X. axonopodis* pv. *citri*, with more on the way. New insights into the life of these pathogens are emerging. For example, the genome sequence of *X. fastidiosa* appears to lack avirulence genes and the Type III secretion system needed for their injection into host cells. *R. solanacearum,* however, does have a Type III system and over 40 candidate proteins for secretion *via* this mechanism, many of which had not been identified prior to genome sequencing (Salanoubat *et al.* 2002).

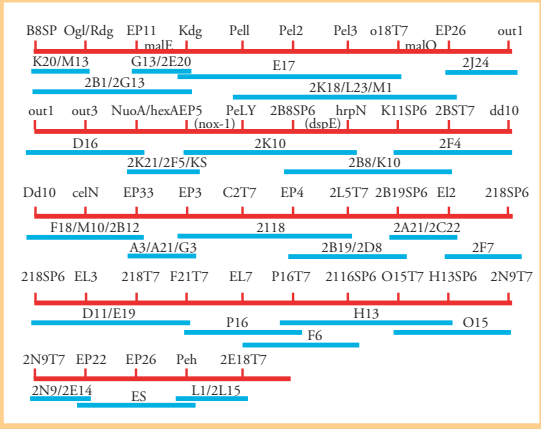

Figure 3 Physical map of part of the *Erwinia carotovora* subsp. *atroseptica* genome. Blue bars represent individual BAC clones. Red line represents map with a number of known genes and other sequences marked.

Molecular research on *Erwinia* at SCRI In 1998, through collaboration with the plant genomics group at SCRI, physical maps of selected regions of the genomes of *Eca* and *Phytophthora infestans* were made using bacterial artificial chromosome (BAC) libraries (Bell *et al.* 2002; see article by Whisson *et al.*). The *Eca* library, consisting of DNA fragments of approximately 100 kb (from a total of around 5 Mb), was ordered using AFLP fingerprinting, and many published *Erwinia* genes and genes from our research were placed on the physical map through hybridisation (Fig. 3). From this mapping exercise, a Type III secretion system (*hrp* cluster) was identified within the genome of *Eca* and is now being assessed for its role in pathogenicity and host range.

The genomic region spanning the *hrp* cluster was examined in detail by sample sequencing of two overlapping BAC clones (*ca* 200 kb) (Bell *et al.* 2002). These sequences revealed the presence of 28 *hrp* genes previously found in the fruit pathogen *Erwinia amylovora* (*Ea*) but also a number of unknown potential pathogenicity genes. Targeted sequencing was used to close gaps between these gene sequences and produce the entire hrp cluster sequence, which was analysed using bioinformatic software, including ARTEMIS and BioEdit to predict the positions, and structural and functional properties of individual genes within the cluster (Fig.4). A number of structural genes involved in the formation (*hrcC* and *hrcV*) and genes involved in the regulation (*hrpL*) of the Type III system were mutated, along with others thought to be exported *via* this system (*dspE*, *hrpN* and *hrpW*). These mutants are currently being tested on both tobacco (non-host) and potato (host) to determine the role of the hrp cluster in host range and pathogenicity. Potential pathogenicity genes that were previously unknown in *Eca* included *dsp*EF (essential for pathogenicity in *Ea*), adhesin-, haemagglutinin- and haemolysin-like genes, (involved in attachment and pathogenicity in animal pathogens and also present in the genome sequence of *Xylella fastidiosa*), and sequences similar to opine catabolism genes (synthesised *in planta* during infection by *Agrobacterium tumefaciens* to provide a specialized nutrient source).

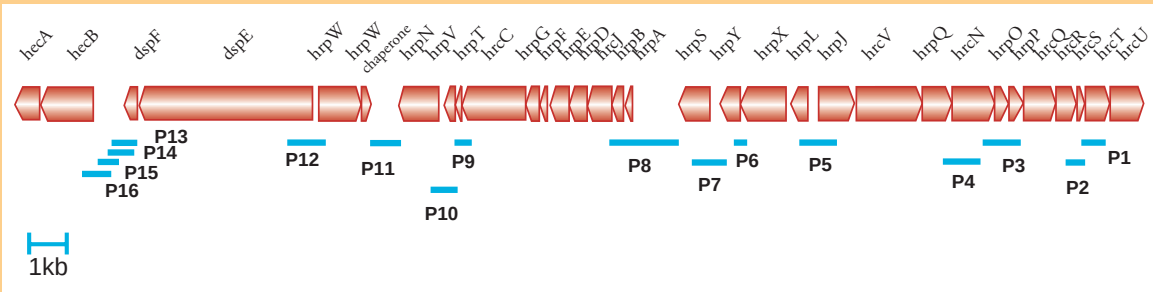

Figure 4 Size and position of genes within the hrp cluster in *Erwinia carotovora* subsp. *atroseptica* together with the position of amplification products P1-P16 used to close gaps in the sequence.

Erwinia genome sequencing Collaboration with the Wellcome Trust Sanger Institute in Cambridge (http://www.sanger.ac.uk/Projects/E_carotovora/) was initiated in 2002, to sequence the complete genome of *Eca*. This project also involves high throughput gene functional analyses based on the complete sequence. Shotgun sequencing of the genome to 8-fold coverage is now complete and almost all sequence gaps in the genome have been closed. Once complete, the sequence will be annotated using the software ARTEMIS to predict structural and functional properties of genes within the sequence. Computer-based predictions relating to evolution, regulation, secretion, plant interaction, protein folding and many other factors will be examined in collaboration with both BioSS and the University of Dundee.

The *E. chrysanthemi* (*Ech*) genome is also being sequenced in the USA. This project, led by Dr. Nicole Perna from the University of Wisconsin, is being undertaken by an international consortium, of which SCRI is a member (http://www.ahabs.wisc.edu:16080/~pernalab/index.html). Once completed, the *Eca* and *Ech* genomes will be compared with each other to identify variations in genomic content that may account for differences in host range or disease symptoms. They will also be compared with genome sequences from the closely related animal pathogens *E. coli*, *Salmonella* sp. and *Yersinia pestis,* and a variety of plant pathogens to identify genetic similarities and differences that may be responsible for common or unique modes of attack.

Future research From studying the *Eca* genome sequence, we already have a number of new lines of research. These include proteomic analyses, currently being undertaken in collaboration with the Moredun Research Institute (MRI) to identify, for example, proteins secreted *via* the Type III system that may interact directly with the plant to cause disease (effectors) or trigger a resistance response (elicitors). We have also made a mutation grid of 5,000 transposon mutants, with plans to expand to 40,000, to allow rapid selection and screening of gene mutations in plant tests (assessing changes in disease symptoms and non-host reactions). Gene expression during the early *Eca*-potato interaction, currently being studied within the group using cDNA-AFLP (Dellagi *et al.* 2000), will continue to be an important part of our research, and will include techniques such as suppression subtractive hybridisation (SSH) and microarray analyses.

The potential of genomic approaches for blackleg control Understanding the interaction between *Eca* and potato is paramount to controlling blackleg disease and we have been doing this on a field-scale for many years. The new research aims to complement this by looking at the potato-*Eca* interaction at a molecular level. Genes from *Eca* proteins found to interact directly with the plant will be used to investigate resistance responses in potato. For example, expression of secreted pathogen proteins in potato species within the Commonwealth Potato Collection (CPC) may identify novel sources of resistance, and rapidly facilitate introgression of such resistance into *Solanum tuberosum.* Type III secretion systems are found in many pathogens, both plant and animal, and have potentially important biotechnology applications. For example, homologous genes in *Ea* have been patented, as their proteins (under the name "Messenger"), when used as crop sprays, both elicit resistance responses in the crop and enhance plant growth(http://www.edenbio.com/tk/tkmain_whitepaper.html).Potential elicitors in *Eca* similar to "Messenger", and other as yet unknown targets, will be investigated for commercial use as well as a role in disease reduction. Ways are also being sought to prevent *Eca* attacking the plant, *e.g.* by identifying an "achilles heel" within the pathogen that may respond to small increases in levels of natural plant anti-microbials or oxidative stress (Lopez-Solanilla *et*

al., 1998; Reverchon *et al.,* 2002) or disrupting the disease process by altering natural levels of chemicals/proteins, such as homoserine lactone (HSL), within the pathogen (our work and Mae *et al.* 2001).

In summary, genomic research offers a new and very powerful approach to investigating both the pathogen and its host, offering the potential to provide biotechnological solutions to blackleg control. The research is also expected to lead to the discovery of targets for commercialisation, as well as providing high quality science, forming the core of future *Erwinia* research at SCRI.

For further information visit the *Eca* sequencing web site at http://www.scri.sari.ac.uk/TiPP/Erwinia.htm

References

Bell K. S., *et al.,* (2002). *Microbiol.* **148,** 1367-1378.

Cole *et al.*, (2001). *Nature* **409,** 1007-1011.

Dellagi A., *et al.,* (2000). *Mol Plant Microbe Interact.* **13,** 1092-1101.

Doolittle R.F. (2002). *Nature* **416,** 697-700.

Fleischmann RD, *et al.* (1995). *Science* **28,** 496-512.

Lopez-Solanilla, E., *et al.*,(1998). *Plant Cell,* **10,** 917-924.

Mae A, *et al* (2001). *Molecular Plant-Microbe Interactions* **14,** 1035-1042.

Perna N.T. *et al.* (2001). *Nature* **409,** 529-533.

Reverchon, S., *et al.,* (2002). *J. Bacteriol.* **184,** 654-665.

Salanoubat *et al.* (2002). *Nature* **415,** 497-502.

Simpson A.J.G. *et al.* (2000). *Nature* **406,** 151-159.

Wood, D.W., *et al.* (2001). *Science* **294,** 2317-2323.