## **Bioinformatics in Plant Pathology**

L. Pritchard, J.A. White, S.C. Whisson, I.K. Toth & P.R.J. Birch

The way biology is carried out has been changed **I** forever by the recent explosion in genome sequences, and modern post-genomic and highthroughput experimental techniques which have the capacity to bring an unprecedented quantity of information to practically any biological field. Mathematical approaches to biology have acquired great value to biologists, and otherwise unattainable insights now flow regularly from computational techniques. These approaches require a blend of biological and computational expertise, flexible access to appropriate computing hardware and facilities, and the ability to innovate freely at the cutting edge of the emerging discipline of bioinformatics.

The Plant-Pathogen Interactions (PPI) bioinformatics team continues to contribute to the BioPython project and to develop tools useful to the wider community, such as the *PyZerg* wrapper to the *Zerg* BLAST parser and the *GenomeDiagram* programming library, which was integral to the comparative genomics analyses of *Erwinia carotovora* subsp. *atroseptica (Eca). GenomeDiagram* (Fig. 1) is a free drawing package that allows the user to generate publication-quality graphics of their genome or biological sequence data, which has been downloaded worldwide and is finding use in the ongoing *Clavibacter michiganensis* sequencing study at Colorado State University.



Figure 1 GenomeDiagram output and graphical interface screenshots from OS X and Windows XP.

In PPI, the *Erwinia* and *Phytophthora* groups in particular have generated large amounts of sequence data, typical of biology in this genomic era, and the *Eca* bacterial genome<sup>1</sup> has been sequenced by SCRI in collaboration with the Sanger Institute. Comparative genomics analyses of *Eca* and all other available bacterial genomes suggest that many of the disease-causing characteristics of *Eca* 'jumped' into this pathogen from other bacteria<sup>1</sup>. For *Pi*, DNA sequence databases comprising tens of thousands of ESTs and partial

genome sequences are available, and with these data the PPI bioinformatics team has located and investigated regions of similarity between the oomycetes *Pi* and *Hyaloperonospora parasitica* (in collaboration with Warwick HRI). This revealed, surprisingly, that gene synteny is conserved at a key region in each of these organisms that is responsible for determining whether infection occurs

on host plants. A pipeline for the identification of Short Interspersed Elements (SINEs) in genomic sequences was also developed and applied to *Pi*, further characterising the pathogen and its evolution.

PPI have designed and are currently using custom Agilent microarrays that contain probes for the complete set of *Eca* coding sequences and selected potato sequences, and the groups also have access to results from the *Pi* Affymetrix microarrays *via* the Syngenta *Phytophthora* Consortium. These arrays, and the bioinformatic approaches described above, have great potential to improve our understanding of what goes on at a fundamental level in the pathogens themselves, and in their interactions with host plants as disease progresses.

## References

<sup>1</sup>Bell, K.S., Sebaihia, M., Pritchard, L., Holden, M.T.G., Hyman, L.J., Holeva, M.C., Thomson, N.R., Bentley, S.D., Churcher, L.J.C., Mungall, K., Atkin, R., Bason, N., Brooks, K., Chillingworth, T., Clark, K., Doggett, J., Fraser, A., Hance, Z., Hauser, H., Jagels, K., Moule, S., Norbertczak, H., Ormond, D., Price, C., Quail, M.A., Sanders, M., Walker, D., Whitehead, S., Salmond, G.P.C., Birch, P.R.J., Parkhill J. & Toth, I.K. (2004). *Proceedings of the National Academy of Sciences of the USA* **101**, 11105-11110.