Aphids join the genomics age

Brian Fenton, Gaynor Malloch, Agnese Winfield, Alex Wilson[1], John Ramsey[2] & Georg Jander[2]

[1]Department of Biology, University of Miami, Coral Gables, Florida 33146, USA; [2]Boyce Thompson Institute for Plant Research, Tower Road, Ithaca, NY 14853, USA

In recognition of their importance as plant pests, a group of international scientists, including a representative from SCRI, met in Paris, France, 2003 under the banner of starting AGIN (aphid genomic international network). These scientists shared an interest in the analysis of aphid genomes and were gathered to decide which aphid species should be the subject of a genome sequencing project. The vote was close, but the pea aphid (*Acyrthosiphon pisum*) won, with the peach potato aphid or green peach aphid (*M. persicae*), SCRI's choice, a close second. *M. persicae* is considered the main vector which spreads viruses within the Scottish seed potato crop. The sequencing of pea aphid genome is now complete and annotation is underway. While the pea aphid exhibits genome differences these are minor, and many genes have high identity and a presumed similarity in function to *M. persicae* genes. The Paris meeting also forged new alliances, and from this developed cooperation between SCRI and Cornell University in the production of expressed sequence tags (ESTs) for *M. persicae*. This cooperation has led to a publication (Ramsey *et al*. 2007) and production of a *M. persicae* microarray. The output from aphid EST analysis has already identified challenges for the immediate future. The joint *M. persicae* EST analysis produced ~27,000 ESTs, which is an awful lot of information, and analysis has required the use of bioinformatics, a combination of computing power and software expertise. Using bioinformatics it is possible to search databases for similarities to genes from other insects and these similarities allocate an EST to a functional class, based on their known involvement in biological processes or molecular functions (GO – gene ontongeny). Another insect, *Drosophila melanogaster*, the genetically well known fruitfly, is used as the comparator. Figures 1 and 2 are diagrams using data generated from an analysis of the collection of *M. persicae* ESTs using the GO method. While many of the sequences could be matched to a *Drosophila* protein with a known function, it was clear that almost half had no known match and were labelled as 'orphans'. An EST analysis of the pea aphid also discovered that 59% of the sequences identified in their study had no known equivalent (Sabater-Muñoz *et al*. 2006). Only half the ESTs cross correlated between the two studies, and this means that there are

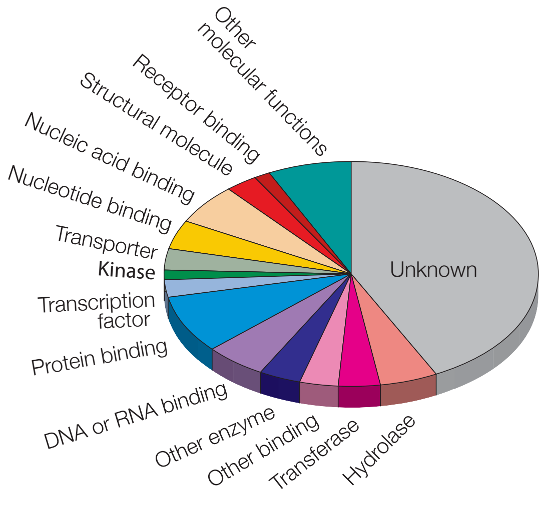

Figure 1 Distribution of Gene Ontology biological process annotations, compared between non-normalized libraries.

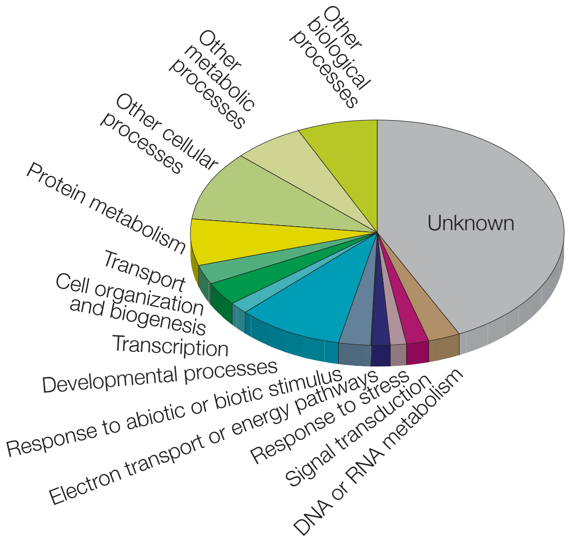

Figure 2 Distribution of Gene Ontology molecular function annotations, compared between non-normalized libraries.

probably many more unique aphid genes that will only be identified on the completion of the aphid genome project. It is possible that the figures represent slight overestimates of the true number of unknown genes, but it is already clear that many do correspond to coding sequences where the protein products have an as yet unknown function. These ‘orphan’ sequences represent completely unknown proteins involved in as yet uncharacterised biochemical or biological processes. Aphids manipulate plants in a variety of ways and the potential for discovering genes with potential use in biotechnological applications is high.

References

Ramsey, J.S., Wilson, A.C.C.,de Vos, M., Sun, Q., Tamborindeguy, C., Winfield, A., Malloch, G., Smith, D.M. Fenton, B., Gray, S.M., and Jander, G. 2007. Genomic resources for *Myzus persicae*: EST sequencing, SNP identification, and microarray design. *BMC Genomics,* **8**, 423.

Sabater-Munoz, B., Legeai, F., Rispe, C., Bonhomme, J., Dearden, P. K., Dossat, C., Duclert, A., Gauthier, J-P., Ducray, D.G., Hunter, W., Dang, P., Kambhampati, S., Martinez-Torres, D., Cortes, T.,Moya, A., Nakabachi, A., Philippe, C., Prunier-Leterme, N., Rahbe, Y., Simon, J-C., Stern, D.L., Wincker, P., & Tagu, D. 2006. Large-scale gene discovery in the pea aphid *Acyrthosiphon pisum* (Hemiptera). *Genome Biology*, **7**, R21