# REQUEST FOR CONSTRUCTION OF A BAC LIBRARY OF THE PEA APHID (*ACYRTHOSIPHON PISUM*)

David L. Stern, Princeton University
Nancy Moran, University of Arizona

## 1 – Importance of the pea aphid to biological and biomedical research

Aphids display a diversity of biological problems that are not easily studied in other genetic model systems. Aphids are the premier model for the study of bacterial endosymbiosis and excellent models for the study of viral vectoring, insect-plant interactions, sympatric speciation, phenotypic plasticity and the evolution of parthenogenesis. Aphid genomics will provide important phylogenetic information because aphids are hemimetabolous insects and can serve as an outgroup to the many genomes being sequenced in holometabolous insects (flies, beetles, bees, moths). Finally, the facultatively parthenogenetic lifecycle of aphids enables genetic experiments not possible in other animal systems.

Aphid biology is relevant to human health in several ways. First, aphids cause crop damage on the order of hundreds of millions of dollars in lost production each year[1,2]. Second, pest aphid populations are controlled primarily by application of pesticides. These pesticides may persist on harvested crops and in the environment, providing a potential detriment to human health and the environment. Third, the biology of bacterial symbiosis in aphids may serve as a general model for understanding the process of bacterial infection[3-5]. Aphids host a variety of opportunistic bacterial symbionts that are closely related to *Yersinia*, *Salmonella* and other important mammalian pathogens [6,7]. Genomes of aphid symbionts contain homologs of pathogenicity factors, such as that including shiga-toxin, Type III Secretion Systems and others [8], implying that the molecular processes underlying aphid-symbiont relationships will provide a model for invasive and chronic bacterial infections in humans. These symbionts have been shown to have a variety of subtle effects on host responses to other environmental stresses, such as heat and parasites (Montllor et al. 2002, Oliver et al. 2003), providing a system for examining how chronic infections mediate host environmental tolerances. Fourth, aphids transmit some viruses in ways that resemble insect-borne human viruses[9,10]. Aphids may therefore serve as models for studying insect-vector viral disease. Finally, aphids display dramatic phenotypic plasticity[11,12]. There is growing awareness that phenotypic plasticity, for example differential response to drugs, is an important component of human welfare[13]. Aphids provide a model system for studying the interaction of environmental and genetic factors in determining the phenotype.

The pea aphid is a species of the Aphididae, the largest family of aphids that contains most of the important crop pests. The Aphididae radiated within the past ~150 million years and different species display 5-10% sequence divergence in

protein-coding genes. We have found that genomic information can be applied readily between species of this family. We are promoting the pea aphid as a genomic model system for aphids for several reasons. First, the pea aphid has served as the focus of laboratory investigation for many years[14]. The pea aphid can be raised in Petri dishes, simplifying genetic analysis[15]. The pea aphid is the only aphid for which genetic analysis is routine. Second, the international community has begun investing in developing pea aphid genetics and genomics. Significant EST sequencing is in progress (22,000 ESTs to date, another 80,000 in progress). Microarrays based on this EST sequencing are in production in our labs. Genetic mapping is underway in the Stern lab. Third, the genome of the primary endosymbiont *Buchnera aphidicola* has already been sequenced[16] and the genome of one host plant for the pea aphid, *Medicago truncula*ta, is being sequenced (medicago.org/genome). Genomic information from the pea aphid will therefore allow the genomic analysis of tri-trophic interactions.

**2 – Uses to which the BAC library will be put, in addition to genomic sequencing**

The labs of Nancy Moran and David Stern have an immediate need for a BAC library. Moran has funding from NSF to utilize a BAC library for studies of thermal tolerance in pea aphids. An aim of this research is to determine the role of stress response systems in determining the limits to geographic range and seasonal distributions. Availability of the BAC library will enable isolation of genes responding to stress and of their cis-regulatory regions, enabling identification of the relevant genetic differences between clones with different levels of heat tolerance. Additionally, Moran would make use of the BAC libraries to isolate aphid genes underlying innate immune responses, in connection with studies on the infection of aphids by symbiotic and other bacteria.

Stern has an immediate need for the BAC library for two projects. First, the Stern lab is cloning developmental genes of interest to study the comparative embryology of sexual and parthenogenetic development [17]. A BAC library would simplify the analysis of *cis*-regulatory regions surrounding developmental genes of interest. Second, the Stern lab is mapping genes that control dramatic polymorphisms and polyphenisms in the aphid[17,18]. The ultimate goal is to identify these genes by positional cloning. This work would be greatly accelerated by the availability of a BAC-based physical map and, ultimately, a complete genome sequence. We are actively engaged in developing additional markers to develop a high-density map of the region surrounding one gene of interest and such targeted marker development is greatly hampered by the absence of a physical map or genome sequence.

Additional labs in the US and Europe have expressed interest in the availability of BAC library and derivative resources. For example, several groups are interested in cloning families of genes involved in pesticide resistance and odor perception. Please see associated letters of support.

**3 - The size of the research community that could potentially use the BAC library and the community's interest in and support for having a BAC library**

We estimate that there are at least 35 labs with a potential interest in a pea aphid BAC library. There are fourteen labs that coordinate the activities of the International Aphid Genomics Consortium (IAGC) and who have written the genome sequencing white paper. There are an additional 21 labs with which we are in regular communication, in part through the Aphid Genomics listserver (www.eco.princeton.edu/mailman/listinfo/aphidgenomics).

**4 - Whether the organism will be, or has been, proposed to NHGRI or another publicly funded agency for BAC-based genomic sequencing and the status of that request**

The pea aphid was proposed by the International Aphid Genomics Consortium (IAGC) to NHGRI during their last round in Spring 2004 (White paper available at (www.princeton.edu/~dstern/AphidResources.htm). NHGRI reported that the sequencing of a hemipteran genome would be important, but they felt that the hemipteran of stronger medical relevance, *Rhodnius prolixus*, the vector of Chagas disease, should be sequenced first. The IAGC believes that the pea aphid genome would provide important benefits on its own merits and will also assist annotation of the *Rhodnius* genome. The IAGC will, therefore, re-submit an updated proposal during the next call for proposals.

**5 - Other genomic resources that are available that will complement this resource**

As mentioned above, there are over 22,000 pea aphid ESTs in dbEST. There are an additional 4,300 ESTs from related aphid species in dbEST. Another approximately 80,000 pea aphid ESTs are in production in France from libraries constructed from a variety of different organs (head, antenna, salivary gland, bacteriocytes, etc.)

Our French colleagues, led by Denis Tagu, have begun microarray production and the labs of Moran, Stern and Wayne Hunter are collaborating to produce a microarray based on clones sequenced at the Hunter lab. These arrays should be available by November 2004.

A low-resolution genetic map of the pea aphid was published in 2001. The Stern lab is building on this by fine-scale mapping a gene on the X chromosome that controls a male wing polymorphism[15].

**6 - The strain of the organism proposed and rationale for its selection**

One consideration when choosing a strain for genomic analysis is to minimize polymorphism to simplify genome assembly. Wild-caught clones of pea aphids maintained in the Stern lab display low levels of nucleotide diversity (0.007: 7 polymorphisms in 1000 bp from two randomly chosen alleles) and it is therefore not of crucial significance which clone is chosen. However, it is most useful for future genome sequencing to use a strain with the lowest levels of heterozygosity.

We have, therefore begun inbreeding a wild strain of pea aphid (LSR1: collected from alfalfa in Tompkins County, NY) that was one of the parental strains used in the mapping crosses performed in the Stern lab. The first generation of inbreeding is now hatching and we should have the second generation available by the time of BAC library construction. Two generations of inbreeding will greatly reduce heterozygosity (by 75% on average), although this reduction will be patchy in the genome.

**7 - The size of the genome**

*A. pisum* has a haploid genome size of approximately 300Mb[19] (www.genomesize.com) on four holocentric chromosomes. This genome size estimate was confirmed by measurements performed by J. Spencer Johnston at Texas A&M. Chromosome *in situ* hybridization has been developed[20], which will allow assignment of BAC clones, and therefore physical maps and the complete genome, to chromosomes.

**8 - The need for an additional BAC library if one or more already exists**

A BAC library of the pea aphid was constructed by Jeff Tompkins at the Clemson University Genomics Institute in 2002 (available from www.genome.clemson.edu/orders). There are 18,432 clones of average insert size is 120 kb. We had originally estimated that this would provide 7X coverage of the 300Mb genome. We have recently discovered, however, that the library is heavily contaminated with clones containing DNA from the primary endosymbiotic bacterium *Buchnera aphidicola*. We estimate that approximately 50% of the clones contain *Buchnera*. The current library therefore provides approximately 3X coverage of the genome, which is insufficient for most purposes.

The level of *Buchnera* contamination observed in this BAC library is much higher than we had expected from previous experiments. In total DNA preps, *Buchnera* DNA usually accounts for approximately 10% of total DNA. We have made a small insert plasmid library, average size 2kb, and found that ten percent of clones contained *Buchnera* DNA. It is clear, therefore, that the *Buchnera* DNA is heavily overrepresented in this BAC library and it is possible that *Buchnera* DNA is cloned preferentially into BACs.

The primary challenge for construction of a high coverage BAC library is, therefore, the minimization of *Buchnera* clones. We believe there are several ways to address this problem, which we detail below. In addition, we believe this is likely to be a problem for many organisms. For example, ten percent of insect species contain obligate endosymbiotic microorganisms. Construction of an aphid BAC library with low levels of bacterial contamination will, therefore, provide a test case for developing BAC libraries from other insect species that contain endosymbionts.

*Proposed methods to minimize bacterial contamination*

A – Reduce level of *Buchnera* in aphid through heat treatment or antibiotics: *Buchnera* is heat-sensitive, and heat-treated aphids show major reductions in the relative frequency of symbiont gene copies to aphid gene copies. Quantitative PCR in the Moran lab shows that *Buchnera* genome copies decrease to about 4% of the original levels, relative to host genome copies, after 5 d at 30$^{o}$C. Antibiotics can also be used to eliminate symbionts, though aphids have difficulty growing without *Buchnera*. We are now performing experiments to optimize heat and/or antibiotic treatment to grow adult aphids with minimal bacterial levels.

B – Purify aphid specific DNA using Pulsed-Field Gel Electrophoresis: The Stern lab is collaborating with an expert on DNA separation, Edward Cox at Princeton University, in an attempt to separate aphid genomic DNA from bacterial DNA on pulsed-field gels. Cox routinely performs similar separations in his laboratory and experiments on pea aphid DNA should be available by the end of 2004.

C – Generate an aphid cell line: An ideal method for eliminating bacterial contamination entirely is to isolate gDNA from a cell line. Previous attempts to generate aphid cell lines have utilized tissues from the parthenogenetic stages (adults and parthenogenetic embryos). These attempts have succeeded in keeping cells alive, but these cells were non-proliferative and did not generate a stable cell line[21]. One alternative is to generate a cell line from true embryonic tissues, such as those present in the sexual eggs. The Stern lab routinely generates large quantities of sexual eggs for genetic crosses and studies of development. We are now attempting to generate a stable cell line from these embryonic tissues. Results from this work should be available by the end of 2004.

D – Screen BAC library for removal of *Buchnera* clones: If none of the aforementioned approaches reduce bacterial contamination sufficiently an alternative approach involves selective elimination of clones containing *Buchnera*. Since this is not a typical practice, the precise method used for removing these clones would need to be discussed with the center constructing the library. One simple approach is to screen an arrayed library with a collection of *Buchnera* genes evenly spaced throughout the *Buchnera* 600kb genome. The

clones that do not carry *Buchnera* could then be rearrayed into a new library.

### 9 - The availability of a source of DNA for construction of the BAC library

The source strain for DNA will be provided by the lab of David Stern. We will use a strain free of secondary symbionts and, as discussed above, we will prepare aphids, cells or PFGE extracted DNA for libraries.

### 10 - Specifications for the library (e.g., library depth, BAC insert size) and supporting scientific rationale for these specifications.

Given the previous difficulties with bacterial contamination we propose that enough clones be spotted to provide 10X coverage of the aphid genome. BAC insert sizes should be as large as possible, preferably greater than 120kb on average. Two major proposed uses of the BAC library are to construct a physical map and to provide large clones to aid genome sequencing. 10X coverage of the 300Mb genome with clones of 120kb requires spotting 25,000 clones.

### 11 - The time frame in which the library is needed

The library is required as soon as possible for experiments that are already planned and funded in the labs of David Stern and Nancy Moran. Both of these labs had planned to use the BAC library at CUGI, but this library has provided insufficient coverage due to excessive bacterial contamination.

### 12 - Other support that is available or has been requested for the construction of the desired library

We have not requested other funds for construction of the BAC library. David Stern has recently submitted a proposal to NIH for research on the pea aphid that includes a request for funds to generate a physical map from the existing BAC library. If both this request and the NIH grant are successful, the money would be applied to generating a physical map from the library generated from this effort.

### References


1. Oerke, E.-C. in *Crop production and crop protection: estimated losses in major food and cash crops* (eds. Oerke, E.-C., Dehne, H.-W., Schonbeck, F. & Weber, A.) 179-296 (Elsevier, Amsterdam, 1994).
2. Morrison, W. P. & Peairs, F. B. in *Response model for an introduced pest--the Russian wheat aphid* (eds. Quisenberry, S. S. & Peairs, F. B.) (Entomological Society of America, Lanham, MD, 1998).
3. Baumann, P., Moran, N. A. & Baumann, L. in *The Prokaryotes [Online]* (ed. Dworkin, M.) (Springer, New York, 2000).

4. Moran, N. A. & Wernegreen, J. J. Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends in Ecology and Evolution* **15**, 321-326 (2000).
5. Dale, C., Plague, G. R., Wang, B., Ochman, H. & Moran, N. A. Type III secretion systems and the evolution of mutualistic endosymbiosis. *PNAS* **99**, 12397-12402 (2002).
6. Russell, J. A., Latorre, A., Sabater-Munoz, B., Moya, A. & Moran, N. A. Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. *Mol Ecol* **12**, 1061-75 (2003).
7. Sandström, J. P., Russell, J. A., White, J. P. & Moran, N. A. Independent origins and horizontal transfer of bacterial symbionts of aphids. *Molecular Ecology* **10**, 217-228 (2001).
8. van der Wilk, F., Dullemans, A. M., Verbeek, M. & van den Heuvel, J. F. Isolation and characterization of APSE-1, a bacteriophage infecting the secondary endosymbiont of *Acyrthosiphon pisum*. *Virology* **262** (1999).
9. Blackman, R. L. & Eastop, V. F. *Aphids on the world's crops: an identification and information guide* (John Wiley & Sons Ltd., Chichester, 2000).
10. Nault, L. R. Arthropod transmission of plant viruses: A new synthesis. *Ann Ent Soc Am* **90**, 521-541 (1997).
11. Dixon, A. F. G. *Aphid Ecology* (Chapman & Hall, London, 1998).
12. Moran, N. A. The evolution of aphid life cycles. *Annual Review of Entomology* **37**, 321-348 (1992).
13. Bateson, P. et al. Developmental plasticity and human health. *Nature* **430**, 419-21 (2004).
14. Harper, A. M., Miska, J. P., Manglitz, G. R., Irwin, B. J. & Armbrust, E. J. The literature of arthropods associated associated with alfalfa. III. A bibliography of *Acrythosiphon pisum*. *Special Publ., Publ. Agric. Exp. Stn, Coll. Agric., University of Illinois at Urbana Champaign.* **50**, 89 pp. (1978).
15. Braendle, C., Caillaud, M. C. & Stern, D. L. Genetic mapping of *aphicarus* - a sex-linked locus controlling a wing polymorphism in the pea aphid (*Acyrthosiphon pisum*). *Heredity* **in press** (2004).
16. Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. & Ishikawa, H. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407**, 81-86 (2000).
17. Miura, T. et al. A comparison of parthenogenetic and sexual embryogenesis of the pea aphid Acyrthosiphon pisum (Hemiptera: Aphidoidea). *J Exp Zoolog Part B Mol Dev Evol* **295**, 59-81 (2003).
18. Braendle, C. et al. Developmental Origin and Evolution of Bacteriocytes in the Aphid-Buchnera Symbiosis. *PLoS Biol* **1**, E21 (2003).
19. Finston, T. L., Hebert, D. N. & Foottit, R. B. Genome size variation in aphids. *Insect Biochem. Molec. Biol.* **25**, 189-196 (1995).
20. Bizzaro, D., Mandrioli, M., Zanotti, M., Giusti, M. & Manicardi, G. C. Chromosome analysis and molecular characterization of highly repeated DNAs in the aphid Acyrthosiphon pisum (Aphididae, Hemiptera). *Genetica* **108**, 197-202 (2000).

21. Mitsuhashi, J. Development of highly nutritive culture media. *In Vitro Cell Dev Biol Anim* **37**, 330-7 (2001).