**Proposal To Sequence the Genome of the Pea Aphid (*Acyrthosiphon pisum*)**

**The International Aphid Genomics Consortium (IAGC) Steering Committee (in alphabetical order):** Marina Caillaud[a], Owain Edwards[b], Linda Field[c], Danièle Giblot-Ducray[d], Stewart Gray[e], David Hawthorne[f], Wayne Hunter[g], Georg Jander[h], Nancy Moran[i], Andres Moya[j], Atsushi Nakabachi[k], Hugh Robertson[l], Kevin Shufran[m], Jean-Christophe Simon[d], David Stern[n], Denis Tagu[d]

Contact: D. Stern; Ph. 609-258-0759; FAX 609-258-7892; dstern@princeton.edu

**Abstract**

We propose sequencing of the 300Mb nuclear genome of the pea aphid, *Acyrthosiphon pisum*. Aphids display a diversity of biological problems that are not easily studied in other genetic model systems. First, because they are the premier model for the study of bacterial endosymbiosis and because they vector many well-studied plant viruses, aphids are an excellent model for studying animal interactions with microbes. Second, because their normal life cycle displays extreme developmental plasticity as well as both clonal and sexual reproduction, aphids provide the opportunity to understand the basis of phenotypic plasticity as well as the genomic consequences of sexual versus asexual reproduction. Their alternative reproductive modes can also be exploited in genetic experiments, because clones can be maintained indefinitely in the laboratory with sexual generations induced at will[1,2]. Third, aphids provide some of the best studied instances of adaptation, in the form of both insecticide resistance, which has evolved through several molecular mechanisms, and host plant adaptation, which has repeatedly generated novel aphid lineages specialized to particular crop plant cultivars and which is presumably the basis for the radiation of aphids onto many specialized host plants during their long evolutionary history. Finally, an aphid genome will provide important phylogenetic information, serving as an outgroup to the many genomes being sequenced in holometabolous insects (flies, beetles, bees, moths) and providing a valuable resource for annotation of the genomes of other hemimetabolous insects, most of which have much larger genomes than the aphid.

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[3,4]. Second, pest aphid populations are controlled primarily by pesticides. These pesticides may persist on harvested crops and in the environment, to the detriment of human health and environmental quality. Third, bacterial symbiosis in aphids may serve as a general model for understanding processes of bacterial infection[5-7]. Fourth, aphids transmit some viruses in ways that resemble insect-borne human viruses and thus provide models for studying insect-vector viral disease[8,9]. Finally, because aphids display dramatic phenotypic plasticity[10,11], they provide a model system for how environmental and genetic factors interact in determining the phenotype. There is growing awareness that phenotypic plasticity, for example differential response to drugs, is an important component of human welfare[12].

As a group, aphids are among the most intensively studied insects and are the focus of a large community of researchers. Several labs have developed genomic tools for studying aphids, including genetic maps, BAC libraries, EST sequences and microarrays.

An aphid genome sequence would provide immediate benefits, including the bioinformatic discovery of many new potential targets for biological control. In addition, the pea aphid, with its relatively small genome compared to most hemimetabolous insects, will provide a valuable bioinformatics resource for annotating the genome of other hemimetabolous insects, including important vectors of human disease.

**Summary of major benefits of a pea aphid genome**

1. **Model for interactions with micro-organisms**
    a. **Best-studied model of bacterial-animal endosymbiosis**
    b. **Excellent model of insect-vectored virus transmission**
2. **Model for developmental plasticity**
    a. **Displays many cases of extreme developmental plasticity**
    b. **Excellent model of origins of asexuality in an animal**
3. **Model of adaptation**
    a. **Insecticide resistance (leads to heavier insecticide use on crops)**
    b. **Host-plant adaptation – Excellent model of insect-plant interactions, with genomic resources available for plant**
4. **Comparative genomics**
    a. **Hemimetabolous genome provides an outgroup to all other holometabolous insects**
    b. **Small aphid genome provides an important resource for annotation of other hemimetabolous insects that often have much larger genomes. Other hemimetabolous insects include many other agricultural pests and human disease vectors.**

**Rationale for choice of Pea Aphid for genome sequencing**

Approximately 4,000 species of aphids have been described, but only a few species have been adopted for laboratory study. The IAGC believes that the genome should be sequenced for one species, the pea aphid (*Acyrthosiphon pisum)*, which is the primary species used in laboratory and genetic studies.

*Closely related species are major agricultural pests*

*A. pisum* is a member of the Aphidinae, which includes almost all aphid pest species, and specifically of the largest aphid tribe, the Macrosiphini, which includes both most important aphid agricultural pests, including the peach-potato aphid (*Myzus persicae*) and the Russian wheat aphid (*Diuraphis noxia*), and most species used in laboratory studies. Nucleotide divergences for sequenced open reading frames of orthologous genes range from 5% to 10% in comparisons of *A. pisum* to other Macrosiphini and up to 15% for comparisons to other Aphidinae[13,14]. These values indicate less divergence than for human-mouse (78% identity of orthologs) and far less than for mosquito-Drosophila (56% identity of orthologs). Assuming that rates of nucleotide divergence and genomic rearrangements are similarly correlated in aphids and other animals, these values are a strong indicator that *A. pisum* will show substantial synteny of gene order and orientation with other Aphidinae. For example, over 90% of human and mouse genes are estimated to fall within chromosome blocks that are syntenic

for these two species, and even mosquito-Drosophila genomes show considerable regions of synteny. Given the much lower levels of sequence divergence among Aphidinae, the prospects are excellent for being able to extend genomic information from *A. pisum* to other aphid species. Furthermore, no genome sequencing efforts have been initiated for any member of the larger insect clade containing aphids, the Hemiptera, which contains many additional major agricultural pests (whiteflies, scale insects, planthoppers, and leafhoppers) and vectors of human disease (*Rhodnius*). The *A. pisum* sequence would undoubtedly be useful to researchers working on those insects.

*Major benefits of choosing the pea aphid,* Acyrthosiphon pisum

*A. pisum* is the primary aphid used in laboratory studies due to its relatively large size and the simplicity of rearing. Unlike many aphid species, *A. pisum* can be reared through the entire life cycle on a single host plant, and methods for rearing this aphid in Petri dishes on excised leaves or on artificial diets have been developed[15-17]. Genetic maps are available[2], and efforts are underway to perform transgenesis and RNA interference. There are extensive data on the genomics and physiology of the endosymbionts[18,19]. Many races of *A. pisum* can be found in the wild that differ in their host plant preferences, and some of these races represent incipient speciation events[1,2]. *A. pisum* can be raised in the lab on many host plants, including the genomic model system *Medicago truncatula*.

*Genome size and other relevant genetic data*

*A. pisum* has a haploid genome size of approximately 300Mb[20] (www.genomesize.com) on four holocentric chromosomes. This genome size estimate has been confirmed by J. Spencer Johnston at Texas A&M. Chromosome *in situ* hybridization has been developed[21], which will allow assignment of BAC clones, and therefore physical maps and the complete genome, to chromosomes.

**Major research areas that will be influenced by a pea aphid genome sequence**

Aphids are used to investigate many biological questions in fields as diverse as genetics, physiology, agriculture, microbiology, virology, ecology, evolution, development and behavior. Here we review the major fields of enquiry that we expect will be influenced by the sequencing of the pea aphid genome and indicate some of the immediate uses of an aphid genome.

***1 – Interactions with micro-organisms***

*1.a – Bacterial Symbiosis*

Aphids provide the best-studied model for maternally transmitted symbionts, and aphid-bacterial symbioses are a prime illustration of the progress possible using genomic approaches. Aphids display highly coevolved, ancient, mutualistic intracellular symbiosis as well as more recent, conditionally beneficial and facultative associations and pathogenesis.

*Obligate primary symbiosis in aphids:* The primary symbiont of aphids, *Buchnera aphidicola*, is maternally transmitted, inhabits specialized cells (bacteriocytes), and is required for host development, growth and reproduction. The *Buchnera* genome has undergone massive reduction, as is typical of endosymbiotic lineages. Three *Buchnera*

genomes have been sequenced[18,22,23] and they contain only 0.61-0.65 Mb, encoding just 510-570 proteins; another *Buchnera*, now being sequenced by Moya et al in Spain, is the smallest known bacterial genome at 0.45 Mb. The close relationship of *Buchnera* to *E. coli* and to other fully sequenced and well-studied bacteria has enabled the inference of function for about 90% of *Buchnera* genes. The gene inventories have confirmed and extended experimental data showing that *Buchnera* is able to provision its hosts with nine essential amino acids that are limiting in the phloem sap diet[24].

*Buchnera* also affects aspects of aphid biology other than nutrition. For example, the sensitivity of aphids to heat, a major factor determining their geographic and seasonal distributions, can be ascribed in large part to heat sensitivity of *Buchnera*[23,25]. Heat shock genes are expressed at very high levels in *Buchnera* even in the absence of thermal stress, possibly to mask effects of mutations that affect protein stability. One of these, the chaperonin GroEL, is present in the aphid hemocoel, where it has been shown to provide protection to plant-pathogenic viruses that are vectored by the aphids[26].

Although the genomic studies of *Buchnera* provide some answers to how this intimate symbiosis functions, they also raise several questions for which answers cannot be obtained without information from the host genome. *Buchnera* lacks many genes that would appear to be essential to bacterial function, including most genes responsible for phospholipid biosynthesis, implying that *Buchnera* cannot synthesize its own cell membrane. Most genes encoding transcriptional regulators are also missing from the *Buchnera* genomes, making it unclear how *Buchnera* regulates its activities with the host.

*Facultative symbiosis in aphids:* Aphids can also contain any of several maternally inherited symbiotic lineages that coexist with *Buchnera*[27-31]. For example, naturally occurring strains of *A. pisum* can contain any of three gamma-proteobacterial symbiont lineages (or none), as well as *Rickettsia* and *Spiroplasma* species; these organisms are also found in other aphid species and other insect groups, where they are known to disrupt sex ratios.

The more recent, facultative symbionts of aphids provide examples of the transition between pathogenic and harmless or mutualistic interactions[7]. For example, the gamma-proteobacterial symbionts have been shown experimentally to confer hosts with increased tolerance to heat stress and with the ability to overcome and kill internally developing parasitoid wasps[32,33]. They are closely related to human pathogens such as *Yersinia pestis* (the agent causing plague), *Escherichia coli*, and *Salmonella* (causing typhus), and they contain some of the same genes implicated in pathogenesis, such as a homolog of the Shiga toxin (van der Wilk et al 1999). Therefore, these symbionts serve as excellent models for studying bacterial infection and the early evolutionary stages of symbiosis.

*1.b - Virus infection*

A number of viruses, including both RNA and DNA viruses, have been characterized in aphids, and specifically in *A. pisum*. Several of these genomes have been sequenced and biological effects have been evaluated. These viruses vary in host range, mode of transmission, and effects on hosts, which range from innocuous to extreme pathogenesis[34-37]. These provide potential viral models for studies of the mechanisms and dynamics of infectious disease.

*1.c – Plant virus vectoring*

Viruses are obligate parasites that are unable to survive outside their hosts. In contrast to animal and insect viruses, plant viruses must be vectored from one immobile host to another and aphids vector hundreds of plant viruses[9,38].

Virus transmission is directly related to aphid feeding. Aphids have piercing-sucking mouthparts and their feeding behavior consists of probing plant epidermal tissues and ingesting phloem sap. Virus particles can be transmitted in either a circulative or non-circulative manner. In both cases, virus transmission relies on specific interactions between virus and aphid molecules.

The majority of plant viruses are transmitted in a non-circulative (non-persistent) manner. Acquisition from an infected plant requires an association of virus particles with aphid mouthparts and the anterior part of the alimentary tract[39]. Virus particles are released in subsequent probing by the aphid and can thus be transferred to uninfected plants. Non-circulative transmission relies on both virus and aphid genetic properties[40], and specific epitopes on viral coat proteins are required for attachment to aphid mouthparts[41,42].

During circulative (persistent) transmission, virus particles cross aphid cell membranes, apparently through a receptor-assisted endocytosis-exocytosis mechanism[40,43,44]. Virus particles pass through the aphid gut and hemocoel to the accessory salivary gland to facilitate transmission to a plant via the saliva[9]. Aphid proteins that are linked to transmission efficiency have been found; some of these proteins are able to specifically bind transmissible virus particles and may therefore represent putative virus receptors[45]. Furthermore, genetic crosses have demonstrated that the genotype of the aphid facilitates the transmission of virus[46].

The similarities in modes of virus transmission in different species suggest that genomic study of aphid-vectored viruses will provide important insight into transmission of viruses by whiteflies, which represents an increasing threat to agriculture, and also as a valuable model for insect-transmitted animal viruses.

2 – *Model for developmental plasticity*

2.a – *Polyphenisms*

Aphids can produce multiple alternative phenotypes, called polyphenisms, in response to specific environmental changes. For example, aphids can switch between sexual and asexual reproduction and they can switch between a winged form that colonizes new host plants and an unwinged form that is specialized to reproduce at a high rate. Polyphenisms are common in insects, and there is some understanding of the physiological mechanisms underlying polyphenisms in some species[47]. For example, some hormones are implicated in control of some aphid polyphenisms[48,49], but the physiological control of most polyphenisms in aphids is unknown. In addition, almost nothing is known of the genetic systems controlling polyphenisms in any insect. Aphids provide a convenient system for exploring the genetic control of polyphenisms, since genetic clones can be induced to produce alternative phenotypes.

2.b – *Clonality in animals*

One particularly exceptional example of aphid polyphenism is the switch between

asexual and sexual reproduction. Aphids typically reproduce by cyclical parthenogenesis - that is they have many parthenogenetic (asexual, clonal) generations - followed by a sexual generation that produces over-wintering eggs.

Aphids also provide hundreds of natural examples of obligately parthenogenetic lineages. A large proportion of species show a mix of obligate clones and cyclically sexual clones and many distinct species lack any known sexual stages[11,50,51]. These lineages arise readily because ancestral aphid life cycles incorporate clonal reproduction, so permanent asexual, apomictic lines are easily derived through elimination of the annual, inducible sexual stage. In most other animal lineages, asexual reproduction is abnormal, and the loss of sex is accompanied by other complications, such as obligate selfing each generation. Thus aphids provide an unusual opportunity to examine the role of sex and recombination in long-term evolution, one of the prime issues in evolutionary biology. Karyotype studies show that aphids that have lost the sexual stage undergo atypical genome evolution, including rapid evolution of chromosome structure and localization of the ribosomal RNA copies on only one of the two X chromosomes. The loss of recombination may cause other loci to be inactivated on one or even both chromosomes. Many of these obligately asexual aphid lineages are major pests, for example *D. noxia*, *M. persicae*, *S. graminum*, *R. padi*, *Aphis craccivora*, *Aphis gossypii* and others are exclusively or largely asexual in many regions in which they are crop pests.

## 3 – *Models of adaptation*

### *3.a – Insecticide resistance*

A major challenge for biological control of agricultural pests is the evolution of insecticide resistance in pest populations. On many crops, insecticides provide a simple solution for aphid control. However, wide scale application of insecticides is becoming increasingly unacceptable, and alternative means of controlling pest populations are desirable. In addition, aphids rapidly develop resistance to the current generation of insecticides, with *Myzus persicae* showing resistance to more insecticides than any other insect[52]. For example, in France farmers have increased the application of insecticides from four to ten applications per season between 1994 and 2001. Over twenty aphid species are known to be resistant to current insecticides.

The mechanisms of insecticide resistance in aphids are diverse, making it difficult to predict responses to insecticide treatment. In most cases, insecticide resistance arises either by an increase in detoxification enzymes (esterases, hydrolases, glutathione transferases or P450s) or by mutation of the genes encoding the target-site proteins (acetylcholinesterase, sodium channels and the GABA receptor)[53-56]. The best-studied example of insecticide resistance in an aphid is in *M. persicae* where elevated esterases and insensitive acetylcholinesterase and sodium channels have been found[57]. In the case of altered acetylcholinesterase the situation has been complicated by the fact that some aphids, unlike *Drosophila* and the house fly, have at least two loci encoding acetylcholinesterase.

Application of insecticides has both direct and indirect consequences on human welfare and health. The direct costs include the purchase and application of insecticides, the indirect costs include the inadvertent killing of beneficial organisms, contamination of ground water, as well as the direct poisoning of humans, household pets, and wild and

agricultural animals. The direct costs for all pesticides (insecticides, herbicides, etc) in the US alone have been estimated at $4 billion and a conservative estimate of the indirect costs is roughly double the direct costs, or $8 billion[58].

*3.b - Host plant adaptation*

Most aphids display striking specialization to feed on a particular plant family or even a few plant species within a family. Such specialization plays an important role in population divergence and speciation[1,2]. *A. pisum* has served as a model system for studies of host plant specialization revealing a possible sympatric speciation event in this species[2].

The study of plant-insect interactions would be greatly enhanced by the development of genomic tools for both an insect herbivore and its host plant. *A. pisum* feeds on the genomic model legume *Medicago truncatula* (barrel medic), and microarrays with 16,000 predicted *M. truncatula* unigenes are available (www.medicago.org). Sequencing of genomic DNA has been initiated (NSF Plant Genome Project 9872664) and genomic and metabolomic approaches are being used at the Noble Foundation (www.noble.org) and at CSIRO in Australia (www.csiro.au) to identify *M. truncatula* responses to aphid feeding. Researchers at CSIRO Entomology have already mapped and are now cloning a major dominant gene in *M. truncatula* conferring resistance to an aphid in the same genus as *A. pisum* (*A. kondoi*). Therefore, the genome sequence of the pea aphid would provide a powerful pair of genomic model systems (plant and aphid) for exploring host plant adaptation and the molecular interactions that mediate host plant use.

4 - *Comparative genomics*

Insects are divided into the basal hemimetabolous lineages, in which immature stages resemble adults (e.g. cockroaches, termites, grasshoppers, aphids), and the derived holometabolous clade, in which immature stages undergo a complete metamorphosis yielding adults that look dramatically different (e.g. flies, butterflies, beetles, bees). To date, the genome sequences of several holometabolous insects have been determined. To our knowledge, no genome sequencing projects have been approved for any hemimetabolous insects. One drawback for genome sequencing of hemimetabolous insects is that most species have large genomes. However, more basal insect groups may provide key insights into understanding the fundamental adaptations that allowed the insects to colonize land and to become such an important component of terrestrial ecosystems. An aphid genome would provide the first hemimetabolous genome sequence to allow comparison with the multiple existing holometabolous genome sequences. In addition, a hemimetabolous genome would provide an outgroup for the several holometabolous insect genomes, allowing determination of the ancestral states for genomic events observed in the holometabolous lineages. Of equal importance, the relatively small genome size of the aphid will provide a valuable resource for annotation of the larger genomes of most other hemimetabolous insects.

The aphid community has embarked on an EST project to enable genomic approaches to studying aphid biology. We have analyzed the first 25,300 ESTs produced from tissue-specific cDNA libraries constructed from whole animals, heads, antennae, guts, and bacteriocytes. (We have funding to produce additional ESTs from a variety of

tissue-specific libraries. See **Current Genomic Resources** section below.) Additional ESTs have been sequenced from *Toxoptera citricida*[59] (>4,000 ESTs) and *Rhopalosiphum padi*[60] (1000 ESTs). All of these sequences have been deposited in dbEST at the NCBI, and have been annotated through collaboration with the French Genoplante-Info Laboratory. The clusters, contigs and annotations are freely available at http://urgi.infobiogen.fr/data/gpi_seq/run.php.

The ESTs cluster into 6,541 contigs, including 4,094 singletons and 2,447 contigs of at least two ESTs. Approximately half of these 6,541 contigs (51%) are orphan genes, having no significant hits ($p<10^{-5}$) to genes in BLAST searches of protein databases (Swissprot and TRembl). Many of these transcripts may represent genes that are unique to the Hemipterans, which diverged from holometabolous insects approximately 300-400MYA[61,62]. For the 49% of contigs that have significant similarity to proteins in the databases, more hits are found for human proteins than for Drosophila proteins (Table 1). This suggests that at least some aphid genes are shared with humans but not with Drosophila, reinforcing the importance of genomic sequencing from insect model systems more basal than the holometabolous insects.

Table 1. Number of significant BLAST hits for all pea aphid contigs to the Drosophila and human genomes. The left column is the total number of significant BLAST hits and the right column is the total number of significant hits that were more similar to either Drosophila or human sequences.

| Species | # BLAST hits | # Best BLAST hits |
|---|---|---|
| *D. melanogaster* | 1271 | 601 |
| *H. sapiens* | 2579 | 804 |

We performed automated annotation of the pea aphid contigs for gene ontology using the Amigo Browser for the three main branches of the tree ontology: "Biological process", "Cellular component" and "Molecular function". The Gene Ontology classification indicates that the pea aphid ESTs represent genes involved in a wide diversity of biological processes (Table 2) and molecular functions (Table 3). This EST sequencing has provided a diverse assemblage of pea aphid genes that the community has already begun to exploit for a variety of purposes, including the construction of cDNA microarrays. Additional EST sequencing is focusing on other organ-specific libraries to increase the diversity of sequenced genes and to address the large array of biological questions being pursued by different laboratories. This EST data will thus provide valuable tools for annotation of an aphid genome as well as for annotation of genomes of related hemimetabolous insects.

**Table 2**: Gene Ontology classification of pea aphid contigs by Biological Process and Cellular Component

| **Biological Process** | | | **Cellular Component** | |
|---|---|---|---|---|
| GO Title | # Contigs | | GO Title | # Contigs |
| Behavior | 21 | | Cell | 2045 |
| Unknown | 37 | | Unknown | 59 |
| Cellular process | 207 | | Extracellular | 49 |
| Development | 210 | | Immunoglobulin complex | 0 |
| Obsolete | 6 | | Obsolete | 0 |
| Physiological processes | 2187 | | Unlocalized | 14 |
| Viral life cycle | 1 | | Virion | 0 |

**Table 3**: Gene Ontology classification of pea aphid contigs by Molecular function

| **Molecular Function** | |
|---|---|
| GO Title | # Contigs |
| Anticoagulant activity | 0 |
| Antifreeze activity | 1 |
| Antioxidant activity | 1 |
| Apoptosis regulator activity | 0 |

| | |
|---|---|
| Binding | 587 |
| Catalytic activity | 737 |
| Cell adhesion molecule activity | 0 |
| Chaperone activity | 45 |
| Chaperone regulator activity | 0 |
| Cytoskeletal regulator activity | 1 |
| Defense/immunity protein activity | 0 |
| Enzyme regulator activity | 49 |
| Ice nucleation activity | 0 |
| Unknown | 61 |
| Motor activity | 9 |
| Nutrient reservoir activity | 0 |

| | |
|---|---|
| Obsolete | 27 |
| Protein stabilization activity | 2 |
| Protein tagging activity | 1 |
| Signal transducer activity | 70 |
| Structural molecule activity | 139 |
| Surfactant activity | 0 |
| Toxin activity | 0 |
| Transcription regulator activity | 74 |
| Translation regulator activity | 47 |
| Transporter activity | 213 |
| Triplet codon-amino acid adaptator activity | 0 |

## Current Genomic Resources

*1 – Stocks*: Three hundred F2 clones of *A. pisum* used for genetic mapping are available and more are being generated. Over 600 natural isolates of *A. pisum* are available that vary for the following traits: resistance to parasitoid wasps or fungi, host plant preference and performance, presence or absence of secondary symbionts, heat tolerance, life cycle including ability to produce sexual stages, and polyphenism induction.

2 - *ESTs*: ESTs are an important resource for genome annotation. The community has submitted 26,945 aphid ESTs to dbEST (22,183 from *A. pisum,* 4,304 from *Toxoptera citricida* [59] and 1000 from *Rhopalosiphum padi* [60]) and we have briefly reviewed the annotation of these ESTs earlier in this white paper. Table 4 lists the current and planned EST sequencing projects for *A. pisum*.

**Table 4**: Current and planned EST sequencing projects

| Biological question | Source of library | Number of EST planned | Number of EST done | Planned date of completion |
|---|---|---|---|---|
| General expressed genes | Full body, mixed stage | 18,000 | 18,000 | Finished |
| Polyphenism control | Head | 13,000 | 3,000 | Nov-04 |
| Olfaction | Antennae | 6,000 | 1,000 | Nov-04 |
| Digestion, viral transmission | Gut | 11,000 | 1,000 | Jan-05 |
| Feeding, viral transmission | Salivary glands | 15,000 | 0 | March-05 |
| Endosymbiosis | Bacteriocytes | 10,000 | 2,300 | **?** |
| Reproduction,maternal transcripts | Nurse cells, oocytes | 5,000 | 0 | March-05 |
| Sexual embryogenesis | 0-12h after egg laying | 5,000 | 0 | **?** |
| Sexual embryogenesis | 12-24h after egg laying | 5,000 | 0 | **?** |
| Parthenogenetic embryogenesis | Early parthenogenetic embryos | 5,000 | 0 | Nov-04 |
| **TOTAL** | | **103,000** | **25,300** | |

*3 – Databases:* A crucially important element of any genome project is the availability of resources and expertise to perform ongoing annotation of the assembled genome. The IAGC is demonstrating this capability by establishing Aphidbase, a genomic database for aphids. This effort is being led by our French partners at INRA, Rennes, with participation from several other members in the US and Japan. We are installing the GMOD architecture, an open-access database structure developed by Flybase that is being adopted by many other genomic communities. This database will allow archiving and browsing of all genomic data and provides a structure for continued annotation of the genomic data by the scientific community.

*4 – Microarrays:* Microarrays containing clones selected from the EST projects are being

developed by several labs.

*5 - BAC Library*: An arrayed BAC Library of *A. pisum* genomic DNA containing 18,432 clones with an average insert size of 120kb (~7X coverage of the 300Mb genome) is available from the Clemson University Genomics Institute (www.genome.clemson.edu).

6 - *Linkage maps*: A linkage map of 150 markers is available for *A. pisum*[2].

7 - *Symbiont genomics*: The entire sequence of the primary endosymbiont *Buchnera aphidicola* has been determined from three divergent aphid species[18,22,23], including *A. pisum*, and one more is in progress. Microarrays for the entire *B. aphidicola* genome have been constructed and used to study expression of the symbiont genes[63]. In addition, the genomes of two facultative secondary symbionts are being sequenced: one symbiont of *A. pisum* in the gamma-Proteobacteria (collaboration between N. Moran and J. Eisen at TIGR) and one symbiont from the aphid *Cinara tujafilina* (collaboration between M. Hattori and A. Nakabachi of RIKEN, Japan and A. Latorre and A. Moya of the University of Valencia, Spain).

**Material to be sequenced**

The IAGC proposes that one strain of *A. pisum* should be sequenced. The strain selected (LSR1) was collected from alfalfa in Tompkins County, NY and has been used for genetic mapping experiments and the construction of a BAC library. This strain is maintained in at least three laboratories to ensure its survival. It is preferable to perform genome sequencing with a strain showing low levels of polymorphism to simplify assembly. Strains found in the New World are derived from recent introductions of European populations. Studies of mitochondrial diversity have found extremely low levels of genetic polymorphism (about 0.2%) in both European[64] and American[64] populations. These results suggest that the European (and by extrapolation, the American) population underwent a bottleneck less than 100,000 years ago[64]. Unpublished data from the Hawthorne lab found levels of nucleotide diversity of about 0.007 for two randomly chosen genomic clones.

We have not yet made contact with a genome-sequencing center and will seek the advice of the NHGRI in picking an appropriate center. The precise sequencing strategy will be determined in consultation with the genome-sequencing center.

**Size and activity of research community**

We estimate that there are at least 35 labs with a potential interest in a pea aphid genome sequence. 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 (68 individuals) with which we are in regular communication, in part through the AphidGenomics listserver (www.eco.princeton.edu/mailman/listinfo/aphidgenomics).

In addition, there is a large community of biologists that use aphids in some aspect of their research. For example, in a search of Science Citation Index we found that 8359 papers published since 1945 included the word "aphid"; 1976 of these papers were published in the last 5 years and 449 were published in the last year. Not all of these labs currently use genetic approaches, but we expect that the availability of a genome and associated resources would catalyze a renaissance in the study of aphid biology.

## References


1. Via, S. Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution* **53**, 1446-1457 (1999).
2. Hawthorne, D. J. & Via, S. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature* **412**, 904-907 (2001).
3. 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).
4. 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).
5. Baumann, P., Moran, N. A. & Baumann, L. in *The Prokaryotes [Online]* (ed. Dworkin, M.) (Springer, New York, 2000).
6. Moran, N. A. & Wernegreen, J. J. Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends in Ecology and Evolution* **15**, 321-326 (2000).
7. 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).
8. Blackman, R. L. & Eastop, V. F. *Aphids on the world's crops: an identification and information guide* (John Wiley & Sons Ltd., Chichester, 2000).
9. Nault, L. R. Arthropod transmission of plant viruses: A new synthesis. *Ann Ent Soc Am* **90**, 521-541 (1997).
10. Dixon, A. F. G. *Aphid Ecology* (Chapman & Hall, London, 1998).
11. Moran, N. A. The evolution of aphid life cycles. *Annual Review of Entomology* **37**, 321-348 (1992).
12. Bateson, P. et al. Developmental plasticity and human health. *Nature* **430**, 419-21 (2004).
13. Von Dohlen, C. D. & Teulon, D. A. J. Phylogeny and historical biogeography of New Zealand indigenous Aphidini aphids (Hemiptera, Aphididae): An hypothesis. *Annals of the Entomological Society of America* **96**, 107-116 (2003).
14. Moran, N. A., Kaplan, M. E., Gelsey, M. J., Murphy, T. G. & Scholes, E. A. Phylogenetics and evolution of the aphid genus *Uroleucon* based on mitochondrial and nuclear DNA sequences. *Systematic Entomology* **24**, 85-93 (1999).
15. Sasaki, T., Hayashi, H. & Ishikawa, H. Growth and reproduction of the symbiotic and aposymbiotic pea aphids, *Acyrthosiphon pisum,* maintained on artifical diets. *Journal of Insect Physiology* **37**, 749-756 (1991).
16. Febvay, G., Delobel, B. & Rahbé, Y. Influence of the amino acid balance on the improvement of an artificial diet for a biotype of *Acyrthosiphon pisum* (Homoptera: Aphididae). *Canadian Journal of Zoology* **66**, 2449-2453 (1988).
17. Shingleton, A., Sisk, G. & Stern, D. L. Diapause in the pea aphid (*Acyrthosiphon pisum*) is a slowing but not a cessation of development. *BMC Developmental Biology* **3**, 1-12 (2003).
18. 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).
19. Douglas, A. E. Nutritional interactions in insect-microbial symbioses: Aphids and Their Symbiotic Bacteria Buchnera. *Annu. Rev. Entomol.* **43**, 17-37 (1998).
20. Finston, T. L., Hebert, D. N. & Foottit, R. B. Genome size variation in aphids. *Insect Biochem. Molec. Biol.* **25**, 189-196 (1995).
21. 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).
22. Tamas, I. et al. 50 million years of genomic stasis in endosymbiotic bacteria. *Science* **296**, 2376-2379 (2002).

23. van Ham, R. C. H. J. et al. Reductive genome evolution in *Buchnera aphidicola*. *PNAS*, 0235981100 (2003).
24. Douglas, A. E., Minto, L. B. & Wilkinson, T. L. Quantifying nutrient production by the microbial symbionts in an aphid. *J Exp Biol* **204**, 349-58 (2001).
25. Lambert, J. D. & Moran, N. A. Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences, USA* **95**, 4458–4462 (1998).
26. van den Heuvel, J., Verbeek, M. & van der Wilk, F. Endosymbiotic bacteria associated with circulative transmission of potato virus by *Myzus persicae*. *Journal of General Virology* **75**, 2559-2565 (1994).
27. Chen, D. Q. & Purcell, A. H. Occurrence and transmission of facultative endosymbionts in aphids. *Curr Microbiol* **34**, 220-5 (1997).
28. 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).
29. Sandström, J. High variation in host adaption among clones of the pea aphid, Acyrthosiphon pisum on peas, Pisum sativum. *Entomologia Experimentalis et Applicata* **71**, 245-256 (1994).
30. Fukatsu, T. Secondary intracellular symbiotic bacteria in aphids of the genus Yamatocallis (Homoptera: Aphididae: Drepanosiphinae). *Appl Environ Microbiol* **67**, 5315-20 (2001).
31. Simon, J. C. et al. Host-based divergence in populations of the pea aphid: insights from nuclear markers and the prevalence of facultative symbionts. *Proc R Soc Lond B Biol Sci* **270**, 1703-12 (2003).
32. Oliver, K. M., Russell, J. A., Moran, N. A. & Hunter, M. S. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *PNAS* **100**, 1803-1807 (2003).
33. Montllor, C. B., Maxmen, A. & Purcell, A. H. Facultative bacterial endosymbionts benefit pea aphids, Acyrthosiphon pisum under heat stress. *Ecological Entomology* **27**, 189-195 (2002).
34. 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**, 104-13 (1999).
35. Van Munster, M. et al. Sequence analysis and genomic organization of Aphid lethal paralysis virus: a new member of the family Dicistroviridae. *J Gen Virol* **83**, 3131-8 (2002).
36. van Munster, M. et al. Characterization of a new densovirus infecting the green peach aphid Myzus persicae. *J Invertebr Pathol* **84**, 6-14 (2003).
37. van Munster, M. et al. A new virus infecting Myzus persicae has a genome organization similar to the species of the genus Densovirus. *J Gen Virol* **84**, 165-72 (2003).
38. Blackman, R. L. & Eastop, V. F. *Aphids on the World's Crops* (Wiley, Chichester, 2000).
39. Pirone, T. P. & S., B. Helper-dependent vector transmission of plant viruses. *Annu Rev Phytopathol* **34**, 227-247 (1996).
40. Gildow, F. E. in *The Luteoviridae* (ed. Barker, H.) 88-113 (CAB International, 1999).
41. Liu, S., He, X., Park, G., Josefsson, C. & Perry, K. L. A conserved capsid protein surface domain of Cucumber mosaic virus is essential for efficient aphid vector transmission. *J Virol* **76**, 9756-62 (2002).
42. Bauwe, H. & Kolukisaoglu, U. Genetic manipulation of glycine decarboxylation. *J Exp Bot* **54**, 1523-35 (2003).
43. Garret, A., Kerlan, C. & Thomas, D. Ultrastructural study of acquisition and retention of potato leafroll luteovirus in the alimentary canal of its aphid vector, *Myzus persicae* Sulz. *Arch Vir* **141**, 1279-1292 (1996).
44. Peiffer, M. L., Gildow, F. E. & Gray, S. M. Two distinct mechanisms regulate luteovirus transmission efficiency and specificity at the aphid salivary gland. *J Gen Virol* **78 ( Pt 3)**, 495-503 (1997).
45. Li, C., Cox-Foster, D., Gray, S. M. & Gildow, F. Vector specificity of barley yellow dwarf virus (BYDV) transmission: Identification of potential cellular receptors binding BYDV-MAV

in the aphid, *Sitobion avenae*. *Virology* **286**, 125-133 (2001).
46. Gray, S. M. & Gildow, F. Luteovirus-Aphid Interactions. *Annu Rev Phytopathol*, in press. (2003).
47. Nijhout, H. F. Control mechanisms of polyphenic development. *Bioscience* **49**, 181-192 (1999).
48. Hardie, J. Juvenile hormone mimics the photoperiodic apterization of the alate gynopara of aphid, *Aphis fabae*. *Nature* **286**, 602-604 (1980).
49. Hardie, J. in *Photoperiodic Regulation of Insect and Molluscan Hormones* 240-253 (Pitman, 1984).
50. Simon, J.-C., Delmotte, F., Rispe, C. & Crease, T. Phylogenetic relationships between parthenogens and their sexual relatives: the possible routes to parthenogenesis in animals. *Biological Journal of the Linnean Society* **79**, 151-163 (2003).
51. Simon, J.-C., Rispe, C. & Sunnucks, P. Ecology and evolution of sex in aphids. *Trends in Ecology and Evolution* **17**, 34-39 (2002).
52. Vasquez, B. L. Resistant to Most Insecticides. http://wufbir.ifas.ufl.edu/chap15.htm (1995).
53. Loxdale, H. D., Brookes, C. P., Wynne, I. R. & Clark, S. J. Genetic variability within and between English populations of the damson-hop aphid, *Phorodon humuli* (Hemiptear:Aphididae), with special reference to esterases associated with insecticide resistance. *Bulletin of Entomological Research* **88**, 513-526 (1998).
54. Barber, M. D., Moores, G. D., Tatchell, G. M., Vice, W. E. & Denholm, I. Insecticide resistance in the current-lettuce aphid, Nasonovia ribisnigri (Hemiptera:Aphididae) in the UK. *Bulletin of Entomological Research* **89**, 17-23 (1999).
55. Ono, M., Swanson, J. J., Field, L. M., Devonshire, A. L. & Siegfried, B. D. Amplification and methylation of an esterase gene associated with insecticide-resistance in greenbugs, Schizaphis graminum (Rondani) (Homoptera : Aphididae). *Insect Biochemistry and Molecular Biology* **29**, 1065-1073 (1999).
56. Wang, K. Y., Liu, T. X., Yu, C. H., Jiang, X. Y. & Yi, M. Q. Resistance of *Aphis gosypii* (Homoptera:Aphididae) to fenvalerate and imidacloprid and activities of detoxification enzymes on cotton and cucumber. *Journal of Economic Entomology* **95**, 407-413 (2002).
57. Field, L. M. & Foster, S. P. Amplified esterase genes and their relationship with other insecticide resistance mechanisms in English field populations of the aphis, *Myzus persica* (sulzer). *Pest Management Science* **58**, 889-894 (2002).
58. Pimentel, D. et al. Environmental and economic costs of pesticide use. *BioEssays* **42**, 750-760 (1992).
59. Hunter, W. B. et al. Aphid biology: Expressed genes from alate Toxoptera citricida, the brown citrus aphid. *Journal of Insect Science* **3**, 7 (2003).
60. Tagu, D. et al. Annotated expressed sequence tags for studies of the regulation of reproductive modes in aphids. *Insect Biochem Mol Biol* **34**, 809-22 (2004).
61. Burmester, T., Massey, H. C., Jr., Zakharkin, S. O. & Benes, H. The evolution of hexamerins and the phylogeny of insects. *J Mol Evol* **47**, 93-108 (1998).
62. Gaunt, M. W. & Miles, M. A. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Mol Biol Evol* **19**, 748-61 (2002).
63. Wilcox, J. L., Dunbar, H. E., Wolfinger, R. D. & Moran, N. A. Consequences of reductive evolution for gene expression in an obligate endosymbiont. *Mol Microbiol* **48**, 1491-500 (2003).
64. Birkle, L. M. & Douglas, A. E. Low genetic diversity among pea aphid (*Acyrthosiphon pisum*) biotypes of different plant affiliation. *Heredity* **82**, 605-612 (1999).

Institutional affiliations of Steering Committee members:
[a] Ithaca College, Ithaca, NY, USA;
[b] CSIRO, Wembley, Australia
[c] Rothamsted Resarch Station, BBSRC, UK
[d] INRA, Rennes, France
[e] Cornell University, Ithaca, NY, USA
[f] University of Maryland, College Park, MD, USA
[g] USDA, Fort Pierce, FL, USA
[h] Boyce Thompson Research Institute, Ithaca, NY, USA
[i] University of Arizona, Tucson, AZ, USA
[j] University of Valencia, Valencia, Spain
[k] RIKEN, Wako, Japan
[l] University of Illinois, Urbana, IL, USA
[m] USDA, Stillwater, OK, USA
[n] Princeton University, Princeton, NJ, USA