

Aphid Genomics White Paper II: Proposal to complete development of the aphid model

The International Aphid Genomics Consortium¹

Prepared by Denis Tagu, INRA Rennes, France, Alex Wilson, University of Miami, USA and Owain Edwards, CSIRO Australia.

With input from Federica Calevro, James Carolan, Chung-che Chang, Hubert Charles, Michele Cilia, Stefano Colella, Linda Field, Astrid Forneck, Toni Gabaldón, Nicole Gerardo, Wayne Hunter, Stéphanie Jaubert-Possamai, Isgouhi Kaloshian, Amparo Latorre, Carlos Llorens, Nancy Moran, Andres Moya, Navdeep Mutti, Atshushi Nakabachi, Miguel Pignatelli, Yvan Rahbé, Stephen Richards, Claude Rispe, Julio Rozas, Jean-Christophe Simon, Guy Smagghe, Thomas Wilkinson, and Jing-Jiang Zhou.

Contact: D. Tagu, Denis.Tagu@rennes.inra.fr; A. Wilson, acwilson@bio.miami.edu; O. Edwards, Owain.Edwards@csiro.au.

Abstract

Following publication of the first aphid genome in February 2010, aphid biology is entering a new era focused on deciphering many of the specialized biological adaptations underlying the role of aphids as plant pests. Meeting this challenge will be facilitated by the accessibility of NextGen Sequencing strategies, functional genome annotation and characterization of the aphid proteome.

Introduction

Publication of the draft 6.2X genome of the pea aphid, *Acyrtosiphon pisum* (1), marks completion of the first goal of the International Aphid Genomics Consortium (IAGC) in achieving its aim of developing the aphid model to the same level of molecular, cellular and developmental biological understanding as other model insects. This document, prepared in consultation with members of the IAGC, defines the near-term goals of the aphid community, with a particular focus on four aims:

Aim 1: Improvement of the current pea aphid genome annotation

Aim 2: The sequencing of new aphid genomes

Aim 3: Development of an AphidAtlas: a gene atlas for the aphid that establishes a roadmap for all Hemiptera and includes both development of a controlled vocabulary for anatomical structures and deep characterization of transcripts and proteins.

Aim 4: Bioinformatics and development of tools and services for the IAGC community

The unusual biology of aphids makes them ideal models for the study of biological phenomena that are not readily studied in other genetic model systems.

Aphids are the premier model for studies of bacterial endosymbiosis and because they vector many agriculturally important plant viruses, aphids also are a model for studying animal-microbe interactions. Aphid life cycles display extremes in developmental plasticity and many

¹ A list of IAGC members current April 23, 2010 can be found in Appendix 1.

aphids can reproduce both sexually and asexually, so they are excellent subjects for investigating the basis of phenotypic plasticity as well as the ways that alternate reproductive modes shape genomic architecture. Aphids also display well-studied examples of adaptation, exemplified by both insecticide resistance, which has evolved multiple times through several molecular mechanisms, and host plant adaptation, which is implicated in diversification and speciation. Unexpected and remarkable features of the pea aphid genome, such as its high incidence of gene duplications make aphids an ideal model for investigating the role of genome architecture in the evolution of complex life cycles. Finally, aphids are the nexus of an ecological and evolutionary network that is increasingly well-studied at a genomic level; a network that includes the plants upon which aphids feed, the endosymbiotic bacteria they harbor, the plant viruses they transmit, and the natural enemies they fight (Figure 1).

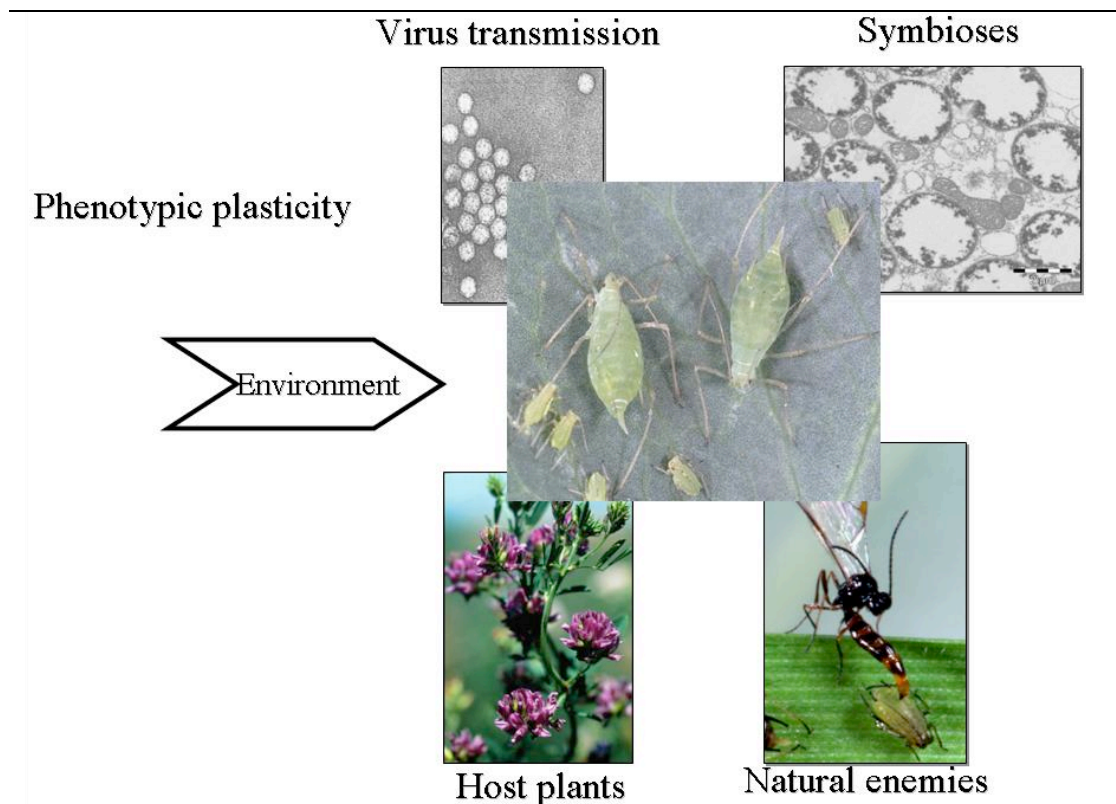


Figure 1. Aphids are the nexus of an ecological and evolutionary network that is well-studied at a genomic level (2).

Aphid biology is relevant to human health and economic wellbeing in several important ways:

1. Aphids and their transmission of plant viruses cause hundreds of millions of dollars in lost agricultural production annually.
2. Pest aphids are largely controlled by pesticides, which are known to persist in the environment and on harvested food crops to the detriment of human health and the environment.
3. Bacterial symbioses in aphids provide the opportunity to investigate the process of bacterial infection and immune system function.
4. Aphids transmit some plant viruses in ways that resemble insect-borne human viruses and thus provide models for studying insect-borne viral disease.

Sequencing of the pea aphid genome resulted in a number of unexpected results and is facilitating characterization of genes associated with aphid-symbiont and aphid-host interactions.

In February 2010, the draft genome sequence of the pea aphid (*Acyrtosiphon pisum*) was published (1). The project, funded by the National Human Genome Research Institute (USA) and completed as a collaboration between the Baylor College of Medicine (Houston, USA) and the IAGC, represents both the first genome of a plant sap-sucking insect, and the first genome of a hemipteran species. Major findings include²:

- Extensive gene duplications that may have facilitated the evolution of extensive polyphenism and cyclical parthenogenesis in aphids.
- A larger number of coding genes than any other sequenced insect genome. This includes expansion of thousands of gene families in aphids relative to other insects and the presence of orphan genes that represent 20% of the pea aphid gene count.
- Sharing of some metabolic genes and gene products between the aphid and its bacterial symbiont, *Buchnera aphidicola*. Affected pathways include those associated with purine and amino acid metabolism.
- Functional gene transfer from non-*Buchnera* bacteria and fungi to the aphid genome.
- An absence of functional transfers of *Buchnera* genes to the aphid genome.
- Loss of conserved pathways including the urea cycle, selenocysteine biosynthesis and the IMD-immune pathway.

The benefits to agriculture of having an aphid genome sequence are being immediately realized. For example, the genome sequence has facilitated analysis of the salivary proteome of the pea aphid (3) and has enabled the identification of olfactory and gustatory receptors and binding proteins that will facilitate elucidation of the molecular basis of host plant localization and acceptance.

Aim 1: Improvement of the current pea aphid genome annotation

The pea aphid genome is set to become the reference for aphid comparative and functional genomics. Ongoing efforts must focus on improving annotation, a genetic map and an increased number of manually-curated gene models. These efforts will improve our understanding of aphid global genome architecture and facilitate identification of signatures associated with whole genome duplication events. Meeting these objectives requires improvements in sequence coverage, the generation of more extensive genetic maps, additional full-length cDNA and RNA_seq data and development of more markers for mapping purposes.

1.1 Generation of an improved annotation of the pea aphid genome

Following completion of assembly Acyr 1.0 the Baylor College of Medicine (BCM) generated 6 Gb of genomic sequence data from the sequenced pea aphid strain LSR1.AC.GC, using 454 next generation sequencing technology. Assembly of Acyr 2.0 has been completed by BCM and has been submitted to NCBI for public release in July 2010. Overall, the second assembly is a tremendous improvement on Acyr 1.0 (Table 1). Of note, the scaffold N50 increased from 86.9 kb to 518.5 kb. Half the genome is now found in only 280 scaffolds and the size of the assembly now exceeds the flow-cytometry estimated 517 Mb size of the LSR1.AC.G1 genome (1).

² A full list of publications associated with publication of the pea aphid genome sequence can be found in Appendix 2.

Table 1. Assembly statistics for Acyr 1.0 and Acyr 2.0

	Acyr 1.0	Acyr 2.0
No. Contigs	72,844	60,598
Contig N ₅₀ (kb)	10.8	27.0
No. Scaffolds	22,801	23,929
Scaffold N ₅₀ (kb)	86.9	518.5
Assembly size (Mb)	464.3	541.7

The predicted pea aphid gene complement of Acyr 1.0 is composed of two sets: (i) the NCBI *A. pisum* RefSeq catalog of 10,248 predictions with very strong biological evidence from ESTs and/or high homologies with other insect genes; and (ii) the *A. pisum* Glean catalog, that covers the remaining 24,355 predicted genes. The Glean set represents a consensus set of gene models generated by different *ab initio* predictors. An immediate annotation goal of the IAGC is to increase the size of the RefSeq gene set from the new Acyr 2.0 genome. This goal will be supported by generation of the following genomic resources by IAGC members:

- Generation of additional ESTs in the form of RNA-Seq data from next-generation sequencing platforms.
- Generation of new full-length cDNAs. In the order of 50,000 full-length cDNAs are now available for the pea aphid, however many more are required to cover the complete gene set.
- Sequencing of small non-coding RNAs, as evidence for non-protein coding genes. High-throughput sequencing of small non-coding RNAs is underway in different IAGC labs (4).
- Identification and characterization of proteins. High-throughput whole organism and tissue specific proteomic studies are currently being conducted by various members of the IAGC.

A new RefSeq gene set will be generated based on information from newly generated EST, full-length cDNA, sncRNA and protein data, and community annotation of assembly Acyr 1.0. In addition, repeats and transposable elements in the newly assembled genome will be scanned with the newest version of REPET pipeline (Quesneville Lab, INRA, Versailles, France) to (i) improve identification of the different repeats and transposable elements and (ii) better annotate the transposable element families (see section 3.3.3).

1.2 Generation of genetic maps

The IAGC is taking four approaches to generating genetic maps of the pea aphid genome with the ultimate goal of a final assembly of four scaffolds corresponding to the four chromosomes of the pea aphid. The first approach to mapping scaffolds and improving the assembly was to conduct fragmentation mapping of the pea aphid genome via Happy Mapping (5). Progress on Happy Mapping of the pea aphid genome has been slow and problematic, thus current efforts have shifted from fragmentation mapping to the second approach, recombination mapping. A low-density genetic map for the pea aphid that defines four linkage groups was published in 2001 however at the time of publication the markers were not made publicly available and remain unavailable (6). A second genetic map, that employs a different mapping panel and a different set of microsatellite markers is under construction in the lab of Jean-Christophe Simon (INRA, Rennes, France) and should be available by the end of 2010. Significant improvements of this and subsequent genetic maps requires an increase in the number of markers available. Currently, the most significant increases in marker density will be obtained via the identification of single nucleotide polymorphism (SNPs) from large next-generation genomic sequence data

sets from additional divergent pea aphid lines (see below Section 2.2.1). A third approach being taken involves restriction site associated (RAD) mapping of genomic aliquots. The RAD mapping approach generates a high density of markers based on sequence polymorphisms detected through Illumina sequencing and is planned for the pea aphid genome by Justin Pachebat and Stephen Richards. The fourth mapping approach, using the mapping population described in (7) and also based on sequence polymorphisms detected through Illumina sequencing, is in progress by Jennifer Brisson (University of Nebraska-Lincoln) in collaboration with David Stern and Peter Andolfatto (Princeton University).

1.3 Manual curation of predicted genes

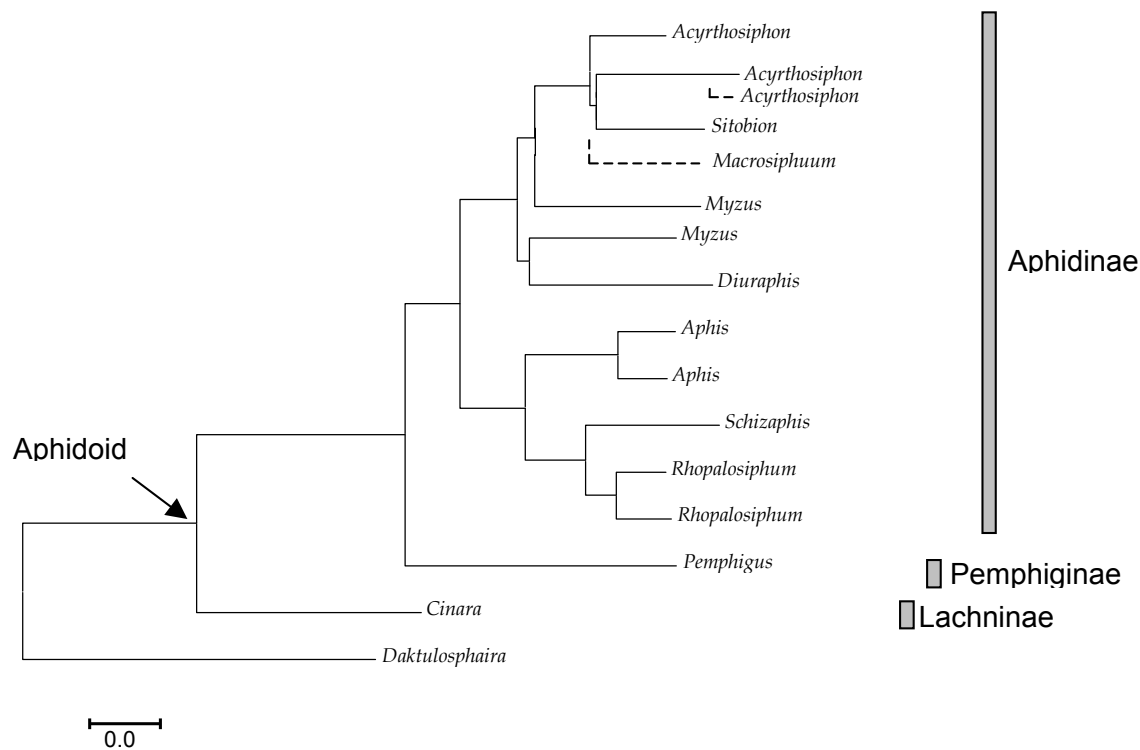
More than 34,000 genes were predicted in the Acyr 1.0 gene set (1) of which only 1,926 have been manually curated. For assembly Acyr 2.0, with the support of the IAGC community and an increased effort of the RefSeq, we aim to double the number of manually curated genes. This goal will be achieved through annotation jamborees where biologists meet for several days together with computer scientists and/or virtual jamborees that will mobilize community members on a monthly basis to annotate additional genes in their home labs. Manual curation data will be posted to AphidBase, and uploaded to GenBank once a year.

Aim 2: The sequencing of new aphid genomes

Sequencing of additional aphid genomes is important for four main reasons. Firstly, genome information from across all aphid groups will facilitate comparative genomics and an understanding of aphid genome evolution. Comparisons will determine whether features observed in the pea aphid genome are characteristic of all aphids, and thus whether these features are indeed prerequisites for unique aspects of aphid biology. For example, community access to several genomes will facilitate investigation of ubiquity of gene/genome duplications as observed in the pea aphid genome. Additional aphid genomes are also important for understanding aspects of aphid biology that are not found in the pea aphid, such as insecticide resistance (*e.g. Myzus persicae*), gall induction (*e.g. Daktulosphaira vitifoliae*), host-alternation (*e.g. M. persicae*), transmission of circulative viruses (*e.g. M. persicae*), acquisition of viviparity (*Daktulosphaira vitifoliae*), and soldier caste formation (*e.g. Pemphigus spyrothecae*).

With the accessibility of next generation sequencing technologies, and now that a first aphid genome has been released, the aphid community is gearing up for generating genome information for additional aphid species and the pursuit of high-resolution comparative genomic and evolutionary analyses. To facilitate this process, the IAGC is generating a list of priority target species based on three criteria: (i) species that occupy key phylogenetic positions, (ii) species of economic importance, and (iii) species that have developed specific adaptations (Figure 2). Guided by the priority list, the IAGC will strongly support new initiatives to obtain funding for sequencing new aphid genomes by supporting the preparation of dedicated white papers and by providing bioinformatic support and infrastructure (see “bioinformatics” section below).

Several projects, focused mostly on species of economic importance but including species occupying key phylogenetic positions, are in development. Current target species include the green peach aphid (*Myzus persicae*), the greenbug (*Schizaphis graminum*), the Russian wheat aphid (*Diuraphis noxia*), the potato aphid (*Macrosiphum euphorbiae*), the cotton aphid (*Aphis gossypii*) and more basal species such as *Cinara cedri* (Lachinae family) and a phylloxera species (*Daktulosphaira vitifoliae*, family Phylloxeridae).



EF1alpha NJ tree (Kimura 2-p) on partial DNA sequences

Figure 2. Phylogenetic relationship of the main aphid species for which IAGC members plan to develop genomic resources.

2.1 Progress on the sequencing and annotation of new aphid genomes

2.1.1 Additional *Acyrtosiphon* races (IAGC contact: Jean-Christophe Simon, Jing-Jiang Zhou, Roger Butlin, Julia Ferrari, Carole Smadja)

The process of speciation involves the progressive evolution of reproductive isolation between divergent populations. When this process happens in the face of gene flow, differentiation is expected to be variable across the genome reflecting the direct operation of natural selection and the barrier created for regions surrounding selected loci (8). The chemical senses are frequently involved in aspects of prezygotic isolation, especially host/habitat and mate choice and this implies that chemosensory genes are good candidates for a role in the early stages of speciation. IAGC members Roger Butlin (University of Sheffield, UK), Julia Ferrari (University of York, UK), Jing-Jiang Zhou (Rothamsted Research, UK) and Carole Smadja (University of Montpellier, France) recently obtained NERC (National Environment Research Council) funding to measure sequence and expression divergence between host races of the pea aphid for the entire known repertoire of chemosensory genes and some control genes using a mix of 454 and Solexa sequencing, which are then candidates for host preference adaptation. This study system is unique in having multiple races at different levels of divergence, excellent background information and a sequenced genome. This allows us to apply the latest approaches (Nimblegen capture arrays, ‘454’ sequencing and Illumina Digital Gene Expression) to this major problem in evolutionary genetics.

Whole genome resequencing from host-specialized pea aphid lineages is facilitating identification of polymorphisms associated with aphid diversification and adaptation. Coordinated by the IAGC with Illumina sequencing performed at the Baylor College of Medicine and Genoscope, France current whole genome resequencing is focused on elucidation of the genomic basis of host plant specialization and will include resequencing the genomes of six new pea aphid genomes belonging to four distinct host-specialized lineages. These initial resequencing efforts will pave the way for future population-level environmental genomic studies.

2.1.2 Additional *Acyrtosiphon* species

Within the genus *Acyrtosiphon* there exist species, and lineages within species, that differ in several key aspects of their biology. Additional genome projects are proposed that will facilitate comparisons of evolutionary processes at a genome-wide scale. The include sequencing: (1) *A. pisum* (temperate species, alternating reproductive modes driven by seasons, showing a continuum of host plant adaptation) (2) *A. kondoi* (temperate/warm temperate, introduced populations are predominantly asexual, polyphagous on Leguminosae) and (3) *A. svalbardicum* (arctic species, alternating reproductive modes driven by genetic adaptation, host-specialized).

Acyrtosiphon kondoi (IAGC contact: Owain Edwards)

This species is a pest in Australian pastures, and has been the target of plant resistance breeding in Australian research programs. Resistance to this species and to *A. pisum* has been extensively studied in a model plant, *Medicago truncatula* (e.g. 9). Upcoming transcriptomic studies on whole bodies and salivary glands of these pest aphids will provide novel insect into insect-plant interactions in this agriculturally important system. Funding for a full genome sequence is being sought.

Acyrtosiphon svalbardicum (IAGC contact: Jean-Christophe Simon)

This is an additional *Acyrtosiphon* species that has been proposed for a genome sequencing project based on its unusual adaptations to its environment. It is an arctic species, endemic to the Svalbard archipelago, Northern Europe, and has developed very specific adaptive traits: asexual reproduction is limited to one single generation and the switch from asexual to sexual reproduction occurs at 24h daylight in summer. Unlike other aphid species, phenotypic plasticity for sexual reproduction is no longer governed by environmental cues, but is instead driven by genetic adaptation. This species feeds exclusively on mountain avens, *Dryas octopetala*.

2.1.3 Green peach aphid, *Myzus persicae* (IAGC contact: Alex Wilson)

The highly polyphagous green peach aphid, *M. persicae*, is an important global agricultural pest that feeds on plants from over 40 different plant families. Additionally, *M. persicae* is the world's most important aphid vector of both persistent and non-persistent plant viruses (10), and is the aphid model for studying the acquisition and molecular basis of insecticide resistance (11). IAGC members Georg Jander and Alex Wilson recently obtained USDA funding to generate 60x Illumina sequence coverage for sequencing *M. persicae* lineage G006. Additional support for this project has been obtained from the BBSRC Genome Analysis Center in the UK by Saskia Hogenhout, in collaboration with Lin Field and Brian Fenton, to provide scaffolding information for genome assembly in the form of paired end 8-kb and 20-kb Roche GS-FLX (454) data and deep, full-length transcript sequencing to aid gene model prediction.

2.1.4 Greenbug, *Schizaphis graminum* (IAGC contact: Gerald Reeck)

This aphid species is an important pest in North America and is biologically interesting because it has genetically-differentiated host races and, unlike pea aphids, can cause damage to plants at low population densities. Additionally, the greenbug is currently the best insect model for studying circulative virus transmission (12). A group of investigators at Kansas State University (KSU), drawn from the departments of Plant Pathology, Entomology, and Biochemistry, is undertaking a genomics project on the greenbug. This effort will start with a large-scale EST project, targeting whole-body and organ cDNA libraries. Shotgun sequencing will then be done on the genome and an initial attempt at assembly undertaken. The goal is to obtain external funding to complete genome sequencing and assembly, and to conduct comparative studies on several greenbug biotypes.

2.1.5 Russian wheat aphid, *Diuraphis noxia* (IAGC contact: Owain Edwards)

Like the greenbug, the Russian wheat aphid (RWA) is a serious pest of cereals and also causes damage at low population densities. Plant damage at low aphid density is hypothesized to result from the action of bioactive molecules in RWA saliva. Additionally, the Russian wheat aphid is an important model for the development of plant resistance-breaking biotypes; an aphid trait that is also hypothesized to be regulated by salivary factors. For this reason RWA genomics efforts are focused on saliva. Salivary gland and alimentary canal transcriptomics on several biotypes using RNA-Seq is ongoing in 2010-2011.

2.1.6 Potato aphid, *Macrosiphum euphorbiae* (IAGC contact: Isgouhi Kaloshian)

With a wide host range including important cultivated plants like potato, tomato, eggplant, and an ability to transmit a number of viruses including potato virus Y and potato leaf roll virus, the potato aphid is an important pest throughout North America. The potato aphid and its tomato host are model system for studying host resistance and plant-aphid interactions (13). A cloned tomato resistance gene, *Mi-1*, exists and a complete tomato genome sequence has just been released. In addition, resistance breaking aphid biotypes also exists. Aphid and gland-specific transcriptome has been sequenced using Illumina technology and assembled both *de novo* and using the pea aphid genome as a reference. Additional aphid RNA-Seq sequencing is planned for aphids exposed to different biotic and abiotic conditions. Furthermore, the proteome of the aphid salivary secretions will be sequenced using liquid chromatography, nano-electrospray ionization and tandem mass spectrometry (nanoLC ESI MS/MS).

2.1.7 Cotton aphid, *Aphis gossypii* (IAGC contact: Jing-Jiang Zhou)

The cotton aphid, *Aphis gossypii* Glover, has become an important cotton pest after the introduction of Bt-transgenic cotton for control of Lepidoptera pests. The Shanghai Institute of Plant Physiology and Ecology, China in collaboration with the Beijing Genome Institute, Shenzhen, China and Rothamsted Research, UK are currently sequencing the genome of the cotton aphid.

2.1.8 *Cinara cedri* (IAGC contact: Andres Moya & Amparo Latorre)

The Aphididae family is composed of eleven subfamilies distributed in three main lineages, one of which is formed by the Lachninae subfamily. Molecular and morphological evidence demonstrate that the subfamily Lachninae occupies a key phylogenetic position between the Phylloxeridae (see 2.1.7) and the Aphididae (14) (Figure 2). As a representative species of the Lachninae subfamily IAGC member labs propose to sequence the genome of *Cinara cedri* for several reasons that include that it feeds on gymnosperms, does not host-alternate and that its *Buchnera aphidicola* possesses the most reduced of all the *Buchnera* genomes and thus it is now obligately associated with a second bacterial symbiont, *Serratia symbiotica* (15).

2.1.9 Phylloxera, Daktulosphaira vitifoliae (IAGC contacts: Astrid Forneck & Denis Tagu)

While not an aphid *sensu strictu*, phylloxera, because of its basal phylogenetic relationship to aphids and because it shares only a subset of the biological adaptations associated with aphids, provides an interesting model for comparative genomics. Of particular note in this regard is the lack of *Buchnera aphidicola* and associated aphid secondary bacterial symbionts in phylloxera. On the other hand, phylloxera is a model for studying the biology of obligate gall-forming species in hemipterans and thus will provide insight into aspects of aphid biology not featured in pea aphids. A major focus of phylloxera research concerns its interaction with its *Vitis* spp. hosts and thus the function of the salivary glands and gut. It is hypothesized that salivary factors are essential for gall induction and maintenance, and possibly also for host adaptation. Secretory and excretory functions are likely to be very specialized in this taxon, and therefore the secretome and metabolome subject to specific adaptive processes. Transcriptomic and metabolomic approaches are planned.

Aim 3: Development of an AphidAtlas

The need for an AphidAtlas, an inclusive database of descriptive information at all biological levels, was a key decision from the IAGC workshop in Barcelona in June 2009. The first phase in development of such an atlas is generation of a consensus vocabulary for describing of the morphology and anatomy of aphids, which will then form the foundation for the second phase: integration of molecular data from transcriptomic, proteomic, and metabolomic studies. The third and final phase in AphidAtlas development will involve inclusion of functional and regulatory data.

3.1 AphidAtlas Development (Phase I)

3.1.1 AphidAtlas Anatomy Group (IAGC contact: Yvan Rahbé and Guy Smagghe)

The first goal of the AphidAtlas Anatomy Group is to develop a consensus description and vocabulary (ontology) for aphid anatomy and morphology. The group will also lead the development of a bioinformatic resource (AphidAtlas Data Management Framework, in collaboration with Fabrice Legeai) to store and manage aphid morphological data. Information relating to gene expression arising from transcriptomic, proteomic, and metabolomic research (see below) will then be linked to the AphidAtlas Data Management Framework. The integration of expression information will be facilitated through AphidBase.

3.2 Transcriptome and Proteome Data for AphidAtlas (Phase II)

3.2.1 AphidAtlas Transcriptome Group (IAGC contact: Stefano Colella)

In order to obtain a comprehensive description of aphid gene function, it is necessary to collect mRNA expression data for the full complement of morphs, organs, tissues, cell types, developmental stages and biotic and abiotic conditions (Appendix 3). The goal of the AphidAtlas Transcriptome Group is to catalog expression-based gene evidence generated using high throughput sequencing of full-length cDNA libraries.

Microarrays. A Nimblegen microarray, developed by INRA Lyon, for the pea aphid is available to all IAGC members. This high-density four-plex array, contains 72,000 60-mer oligonucleotides probes, representing 24,011 pea aphid transcripts, corresponding to 23,855 genes.

Full length cDNAs. About 50,000 full-length cDNAs are already available for the pea aphid, representing about 9,000 predicted genes. Full-length cDNAs are a very strong annotation tool since they provide biological evidence for the complete length of

transcripts, and thus facilitate determination of the intron/exon boundaries and 5' and 3' UTRs of their corresponding genes. Further, full-length cDNAs facilitate unambiguous identification of splice variants. Additional full-length cDNAs have been recently obtained from male and sexual female whole body libraries (INRA Rennes and Genoscope, France). However, more full-length cDNAs are required in order to fully describe all predicted genes.

High throughput sequencing. Preliminary experiments performed in the pea aphid by IAGC members at INRA Rennes and CRG Barcelona indicate that 6 lanes of Illumina sequencing, generating approx 100 million reads, from dissected embryos covered ~80% of the predicted genes. With the decreasing cost of short-reads runs, we can expect large collections of transcript tags to be generated by IAGC members in the near future. Bioinformatic support is essential for analysis of raw data, mapping on a reference genome, and storage and availability of data. AphidBase (see "Bioinformatics" section) will provide this support.

Many IAGC members are currently employing short-read RNA-Seq to serve two main purposes:

1. To determine all possible pea aphid transcripts to achieve a nearly exhaustive view of the pea aphid transcriptome.
2. To facilitate development of genome sequencing projects in new aphid species.

3.2.2 AphidAtlas Proteome Group (IAGC contact: Jim Carolan)

The pea aphid genome sequence has facilitated the application of proteomics to aphid biology and a number of publications have already been published that directly utilize genome sequence or consensus protein databanks (*e.g.* 16, 17). The AphidAtlas Proteome Group was established to utilize published proteomic data in addition to generating large peptide and protein catalogues from different morphs, tissues, organs, and experimental conditions defined during Phase I of the AphidAtlas project (Appendix 3). These reference subproteomes will be linked to each morph, tissue etc. thus providing the community with access to spatial and temporal protein expression information. Protein and peptide catalogs will be generated using large-scale multidimensional protein identification technologies (MudPIT) by a number of IAGC members. Funding is being sought to conduct whole organism and specific tissue glycoproteomic and phosphoproteomic analyses for the pea aphid. These studies will represent the first attempts to characterize the post translational modifications (PTMs) for this organism on a high throughput level.

The protein and peptide catalogues generated in Phase II of AphidAtlas will provide proteomic data of sufficient volume to permit proteogenomic analyses for the pea aphid. Proteogenomics (improving genome annotation using proteomics; see (18, 19) involves utilizing identified peptides in an analogous manner to RNA derived expressed sequence tags (ESTs) to support gene prediction and annotation. Peptides identified by mass spectrometry (expressed protein tags, EPTs; (31) will be used to confirm gene coding status, improve automated gene prediction, validate transcription start sites, support intron/exon boundaries, confirm hypothetical proteins and provide evidence for genes lacking transcriptomic support. The pipelines for processing and utilizing proteomic data for proteogenomic purposes are currently under construction by AphidBase bioinformaticians.

3.3 Cataloguing epigenetic and non-protein coding genomic features AphidAtlas (Phase III)

3.3.1 AphidAtlas Epigenetic Modification Group (IAGC Contact: Owain Edwards & Denis Tagu)

The pea aphid genome contains the complete methylation machinery, suggesting that DNA methylation occurs in aphids. The role of methylation in regulating gene expression in insects is still debated and a description of the methylation status of the genome and genes of the pea aphid genome is necessary. In parallel to methylation, modifications of chromatin by histone post-translational modification has been examined in the pea aphid in order to identify genes located in chromatin regions under epigenetic regulation. Since aphids display high levels of phenotypic plasticity, one hypothesis is that genes involved in this adaptation could be located in chromatin regions that change accessibility status for transcription.

The Methylome – A collaborative agreement is being negotiated by Owain Edwards (CSIRO, Perth, Australia) with BGI in Shenzhen, China to complete the methylome of an asexual pea aphid. This initial project will be followed by a methylome comparison among morphs, using additional full-methylome sequencing or an enrichment method developed by CSIRO.

Chromatin - ChIP/Seq technology using antibodies against modified histones is envisaged to i) describe the chromatin regions on the pea aphid genome and ii) compare chromatin regions between different morphs.

3.3.2 AphidAtlas Small Non-Coding RNA Group (IAGC contact: Stéphanie Jaubert-Possamai)
Several small non-coding RNAs (sncRNA) such as miRNAs and piRNAs are important regulators of gene expression and genome integrity. Since different scnRNAs target specific mRNAs or groups of mRNA accumulation of parallel mRNA and scnRNA sequence data will provide insight into the co-transcription of mRNAs and their miRNA regulators. These parallel data will facilitate the deciphering of aphid regulatory networks.

A catalog of miRNAs for the pea aphid has already been developed using next-generation sequencing from aphid whole bodies. In addition, a catalogue of piRNAs (that regulate transposons mainly in germlines) is underway at INRA, Rennes from pea aphid whole body tissues. However, a deeper description of sncRNA expression is required in order to identify and define the expression profiles of a larger number of miRNAs. Several IAGC member groups involved in gene expression analyses will collaborate to generate parallel mRNA and miRNA data from different morphs, tissues, organs, and experimental conditions defined during Phase I of the AphidAtlas project (Appendix 3).

3.3.3 AphidAtlas Mobilome Group (IAGC contacts: Andres Moya and Carlos Llorens)
In annotating Acyr 1.0 the IAGC paid particular attention to manually curate and annotate the diversity of LTR retroelements, their related transposases and host genes. A wide variety of new retroviruses and retrotransposons belonging to the most important taxonomic groups (*Bel/Pao*, *Ty1/Copia* and *Ty3/Gypsy*) together with many widely distributed DDE integrase/transposase clades were identified. Since all mobile genetic elements are prone to horizontal transfer, IAGC member labs propose to reexamine transposable elements in the improved assembly of Acyr 2.0 with the aim of elucidating the aphid *mobilome*. Importantly, a comprehensive description of the aphid mobilome may shed light on the mechanistic basis of genomic expansion in the form of extensive gene duplication in the aphid genome. All aphid mobile genetic element data will be integrated into GyDB Mobilomics of the Gypsy Database (GyDB) (20).

Aim 4: Bioinformatics and development of tools and services for the IAGC community

The IAGC community requires access to molecular and genetics tools and services to fully realize development of the aphid model. Aphids will not be accepted as a genetic model in the absence of forward and reverse genetic tools, such as RNAi, transgenesis and mutagenesis. The IAGC aims to organise and coordinate research through several geographic hubs where genetic resources can be maintained and distributed. The following tools and services must be further improved to raise aphids to the status of model organism. Bioinformatics is central to the storage, management and accessibility of all aphid genome resources.

4.1 Bioinformatics (IAGC contact: Fabrice Legeai)

Bioinformatics, even if presented here as one of the last features of this white paper, is key for the success of aphid genomics and post-genomics. Bioinformatics is necessary for i) analyses of the raw data produced by biologists (genomes, transcript sequences, protein/peptide sequencing, expression profiles), ii) annotation of genes, gene products and genomic features and iii) storage and mining of data for the whole community. Thus, the IAGC strongly supports bioinformatics efforts at all steps of the post-genomics research and will always encourage IAGC members to include expenses necessary to cover bioinformatics personnel and equipment within their research proposals.

The IAGC has developed a centralised tool called AphidBase as a support for gene annotation and storage of data (<http://www.aphidbase.com>). Furthermore, two satellite databases are also available to the IAGC community, a phylogenetic analysis of the pea aphid genome stored in PhylomeDB (<http://phylomedb.org/?phylomeid=16>) and AcypiCyc (<http://pbil.univ-lyon1.fr/software/cycads/acypicyc/home>), a BioCyc based reconstruction of the pea aphid metabolism. People involved in AphidBase (INRA Rennes, France) are developing analysis tools, supporting different research initiatives, and will be involved in the different work packages:

AphidAtlas: this central bioinformatics framework will allow the IAGC to link to AphidBase all available data generated by members of the consortium using *omics*-type approaches (see above, the AphidAtlas section for more details).

AphidBase: this is the centralized database (www.aphidbase.com) that contains all aphid genomics resources available in a browsable format for users. One of the next challenges is to create a gene report for each of the predicted genes containing information on sequence, structure, corresponding coding sequence, splices variants and expression profiles in relation with AphidAtlas. New search tools such as Biomart (<http://www.biomart.org/>) and Galaxy (<http://main.g2.bx.psu.edu/>) will be added in order to facilitate browsing all the features displayed in the database.

AcypiCyc: This is a BioCyc database including the reconstruction of the metabolism of the pea aphid (<http://acypicyc.cycadsys.org/home.html>). Metabolic pathways and reactions can be explored in this online resource and direct links to AphidBase and PhylomeDB are provided. Newly-sequenced aphid species will be added to the AcypiCyc database once the data will become available.

Genome assembly and automated annotation: Genomic data from newly-sequenced aphid species will be assembled and made available through a centralized location. The assembly of new genomes requires the application and development of *de novo*

assembly tools, including those necessary for short-read data, as well as development of automated *in silico* annotation pipelines.

Genome annotation: AphidBase has already adapted the Apollo annotation tool for manual annotation of predicted genes. Further manual annotation will be required for the second version of the pea aphid genome, and for new aphid genomes as they become available. Jamboree sessions will be organized and all data will be used to update gene models presented in AphidBase.

Genome mapping: Following re-sequencing of different genotypes, will we search for polymorphisms (SNPs) which will be mapped to a reference genome.

PhylomeDB: The complete collection of evolutionary histories of all aphid genes and their homologs in sequenced arthropods (i.e. the phylome), was generated for the Acyr 1.0 assembly (21). This facilitated the automated functional annotation of predicted genes base on phylogeny-base orthology assignments to characterized insect genes. The pea aphid genome was the first to be annotated with such a sophisticated pipeline. All phylogenetic trees and alignments are stored at Phylome DB (www.phylomedb.org), which is fully linked to the corresponding entries in AphidBase. This evolutionary information is a key resource for manual curators. Similar phylogenomic analyses will be performed for the Acyr 2.0 assembly and for upcoming genome sequences from additional aphid species.

4.1 Transgenesis and Mutagenesis (IAGC contact: Denis Tagu and Navdeep Mutti).

Functional analyses of genes rely on generation of mutations to identify gene function either by forward or reverse genetic approaches, requiring the production of mutants or transgenics. Techniques for this are not yet available in aphids. Thus, the IAGC is faced with a true bottleneck for gene expression analyses and elucidation of aphid gene function. Developing mutagenesis and transgenesis techniques adapted to the physiology and genetics of aphids is high risk and thus, the IAGC will support any prospective initiative aimed at developing collaborative efforts between IAGC members so that the risk will be distributed among different groups. Below are several options that could be pursued:

Cell culture - Simplified systems such as a cell culture can provide an indispensable tool for deciphering regulatory pathways. Tissue culture facilitates the application and delivery of agonistic and/or antagonistic drugs and the cells are often more suitable for RNAi experimentation and gene knock-down.

RNAi - This is the only tool currently available for manipulating gene expression in aphids. Several reports describe the knock-down of pea aphid genes by transient RNAi after either injection into hemolymph, or ingestion in artificial diets (e.g. 22-24). However, a comparative study between different protocols is lacking and most importantly, recent data, obtained collaboratively by INRA-Lyon and INRA-Rennes, indicate that not all aphid tissues are equally sensitive to RNAi. Thus, development of novel methods of dsRNA delivery such as delivery via the phloem of transgenic host plants and investment in the improvement of RNAi silencing protocols are necessary.

Transgenics - The ability to produce stable transgenic insects remains the most powerful

approach to decipher gene function. To date, production of stable transgenic aphids has proven problematic. First, generation of stable transgenics requires injection of recombinant DNA into germ cells which is difficult because aphids produce low numbers of non-synchronously laid sexual eggs per female rendering it difficult to synchronously inject the large numbers of eggs necessary for the successful production of a handful of transgenic lines. Second, embryonic development in sexually produced aphid eggs occurs through a 2-month-long diapause. Despite these limitations newer, more efficient transformation methods such as the ph1C31 (ϕ C31) based integration system (25) and zinc-finger nucleases (ZFNs) mediated gene disruption for generating gene knockout lines (26) offer hope for the development of aphid stable transgenic lines. Additionally, while not yet explored, aphid parthenogenesis offers the promise that production of stably transformed aphid lines may be achieved via injection of recombinant DNA into parthenogenetic ovarioles (unlike their sexual clone-mates, parthenogenetic females produce large numbers of embryos, with no arrested development). Finally another possible and untested alternative is the use of aphid viruses to vector transgenes; a method that necessitates identification of an aphid virus that can be manipulated *in vitro*, is able to bear integration of a large piece of foreign DNA and is characterized by high replication rates.

Mutagenesis - Forward genetics has been used for many years to identify through mutagenesis key genes responsible for an organism's development. The emergence of modern genetic and genomic techniques has facilitated more rapid association of mutated genes with phenotypes of interest. Mutants can be generated by chemical mutagens such as ethyl-methane sulfonate (EMS), which induces base substitution mutations in DNA. However, the majority of these mutations, when affecting a gene sequence, are recessive and balanced by the non-mutated wild type allele. Thus, after mutagenesis, isolation of homozygous individuals by several self-crosses is necessary to observe a new phenotype induced by the mutation(s). Since the chemicals used for mutagenesis do not differentiate between non-coding and coding DNA, and eukaryotic gene density is low, the frequency of mutants with an observable phenotype is very low. This implies that a large number of individuals need to be mutagenized, and that a large number of them need to be manipulated to achieve homozygosity and screened for a given phenotype. All these reasons have restricted the use of forwards genetics mainly to model species with a small body size that are easy to rear and that have a short generation (*e.g. Arabidopsis thaliana, Caenorhabditis elegans, or Drosophila melanogaster*). While aphids meet these requirements in some respects, the inability to backcross rapidly (due to the long length of the sexual phase of the reproductive cycle) precludes the use of traditional methods.

First, we plan to set up a protocol to determine the conditions for an effective application of EMS treatment on the pea aphid. Second, we open the discussion on the feasibility of forward genetics on the pea aphid in the framework of the IAGC in order to coordinate an approach to generate and characterize pea aphid mutants on phenotypic plasticity.

4.2 Aphid Virtual Stock Center Database (IAGC contact: Jean-Christophe Simon)

Sharing of aphid clonal lines amongst IAGC members will improve the power of functional analyses by facilitating data collection on a focused subset of genetic backgrounds. Individual IAGC labs can make their lines available to community members through the newly established

Aphid Virtual Stock Center. Hosted at AphidBase, the Aphid Virtual Stock Center provides a database of reference lines, together with archived reference material and a DNA fingerprint for each clonal lineage. DNA fingerprints are developed by the laboratory providing the material and are each associated with a detailed protocol describing DNA fingerprinting methods. Guidelines for the preparation and submission of clones to the Aphid Virtual Stock Center can be found in Appendix 4.

4.3 Recombinant Protein Centers (IAGC contacts: Owain Edwards and Jim Carolan)

Expression platforms now exist for the high throughput expression of hundreds of recombinant proteins that can then be incorporated into enzyme and physical interaction/binding or inhibitor assays in an attempt to determine function. Within the IAGC, two groups will develop and contribute to a panel of assays that will allow the rapid determination of biological function of newly expressed proteins. An established suite of functional assays (based on pea aphid proteins) will contribute to the determination of orthologous and functionally homologous genes when recombinant proteins arising subsequent genome sequenced aphid species become available. Funds are being sought to establish an aphid recombinant protein capability for the IAGC as part of a recently established center at CSIRO in Melbourne, Australia.

Concluding Remarks

The IAGC was established on a foundation of open global scientific collaboration. From its initial meeting in Paris, France in June 2003 through to the present our policy has been to utilize web-based tools that facilitate community-based building of the aphid as a scientific model. Our open collaborative approach is evidenced by our public aphidgenomics listserver, the web-based sign-up for pea aphid genome annotation groups, the AphidWiki and our regular teleconferences that all IAGC members are invited to attend. The IAGC is currently committed to meet annually. Overall, our approach has facilitated the growth and strengthening of the IAGC and it is in this spirit that we will continue.

Indeed, collaboration is necessary as the IAGC community faces the exciting challenges brought about by our rapid production of genomic data and the growing need for tools to manipulate, integrate and study these data. With this in mind, the IAGC is focused on four short-term goals: (1) Improvement of the current assembly and annotation of the pea aphid genome, (2) The sequencing of new aphid genomes, (3) Development of Aphid Atlas and (4) Bioinformatics and development of tools and services for the IAGC community.

References

1. IAGC (2010) Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biology* 8(2):e1000313.
2. Tagu D, Klingler JP, Moya A, & Simon JC (2008) Early progress in aphid genomics and consequences for plant-aphid interactions studies. *Molecular Plant-Microbe Interactions* 21(6):701-708.
3. Carolan JC, Fitzroy CIJ, Ashton PD, Douglas AE, & Wilkinson TL (2009) The secreted salivary proteome of the pea aphid *Acyrtosiphon pisum* characterised by mass spectrometry. *Proteomics* 9(9):2457-2467.
4. Legeai F, Rizk G, Walsh T, Edwards O, Gordon K, Lavenier D, Leterme N, Mereau A, Nicolas J, Tagu D, & Jaubert-Possamai S (2010) Bioinformatic prediction, deep sequencing of microRNAs and expression analysis during phenotypic plasticity in the pea aphid, *Acyrtosiphon pisum*. *BMC Genomics* 11(1):281.
5. Dear PH & Cook PR (1989) Happy Mapping - a proposal for linkage mapping the human genome *Nucleic Acids Research* 17(17):6795-6807.
6. Hawthorne DJ & Via S (2001) Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature* 412:904-907.
7. Braendle C, Friebe I, Caillaud MC, & Stern DL (2005) Genetic variation for an aphid wing polyphenism is genetically linked to an naturally occurring wing polymorphism. *Proc Roy Soc Lond B Biol Sci* 272:657-664.
8. Peccoud J, Ollivier A, Plantegenest M, & Simon JC (2009) A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences of the United States of America* 106(18):7495-7500.
9. Gao LL, Kamphuis LG, Kakar K, Edwards OR, Udvardi MK, & Singh KB (2010) Identification of potential early regulators of aphid resistance in *Medicago truncatula* via transcription factor expression profiling. *New Phytologist* 186(4):980-994.
10. Blackman RL & Eastop VF (2000) *Aphids on the world's crops: An identification and information guide* (John Wiley & Sons Ltd., Chichester), 2 Ed p 466.
11. Field LM & Foster SP (2002) Amplified esterase genes and their relationship with other insecticide resistance mechanisms in English field populations of the aphid, *Myzus persicae* (Sulzer). *Pest Manag. Sci.* 58(9):889-894.
12. Burrows ME, Caillaud MC, Smith DM, Benson EC, Gildow FE, & Gray SM (2006) Genetic regulation of polerovirus and luteovirus transmission in the aphid *Schizaphis graminum*. *Phytopathology* 96(8):828-837.
13. Kaloshian I & Walling LL (2005) Hemipterans as plant pathogens. *Annual Review of Phytopathology* 43:491-521.
14. Ortiz-Rivas B & Martinez-Torres D (2010) Combination of molecular data support the existence of three main lineages in the phylogeny of aphids (Hemiptera: Aphididae) and the basal position of the subfamily Lachninae. *Molecular Phylogenetics and Evolution* 55(1):305-317.
15. Pérez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A, & Latorre A (2006) A small microbial genome: The end of a long symbiotic relationship? *Science* 314(13 October 2006):312-313.
16. Le Trionnaire G, Francis F, Jaubert-Possamai S, Bonhomme J, De Pauw E, Gauthier J-P, Haubruge E, Legeai F, Prunier-Leterme N, Simon J-C, Tanguy S, & Tagu D (2009) Transcriptomic and proteomic analyses of seasonal photoperiodism in the pea aphid. *BMC Genomics* 10(1):456.
17. Wang YL, Carolan JC, Hao FH, Nicholson JK, Wilkinson TL, & Douglas AE (2010)

- Integrated Metabonomic-Proteomic Analysis of an Insect-Bacterial Symbiotic System. *Journal of Proteome Research* 9(3):1257-1267.
18. Tanner S, Shen ZX, Ng J, Florea L, Guigo R, Briggs SP, & Bafna V (2007) Improving gene annotation using peptide mass spectrometry. *Genome Research* 17(2):231-239.
 19. Castellana NE, Payne SH, Shen ZX, Stanke M, Bafna V, & Briggs SP (2008) Discovery and revision of Arabidopsis genes by proteogenomics. *Proceedings of the National Academy of Sciences of the United States of America* 105(52):21034-21038.
 20. Llorens C, Futami R, Bezemer D, & Moya A (2008) The Gypsy Database (GyDB) of mobile genetic elements. *Nucleic Acids Research* 36:D38-D46.
 21. Huerta-Cepas J, Marcet-Houben M, Pignatelli M, Moya A, & Gabaldon T (2010) The pea aphid phylome: a complete catalogue of evolutionary histories and arthropod orthology and paralogy relationships for *Acyrtosiphon pisum* genes. *Insect Molecular Biology* 19:13-21.
 22. Mutti NS, Park Y, Reese JC, & Reeck GR (2006) RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrtosiphon pisum*. *Journal of Insect Science* 6:38.
 23. Jaubert-Possamai S, Le Trionnaire G, Bonhomme J, Christophides GK, Rispé C, & Tagu D (2007) Gene knockdown by RNAi in the pea aphid *Acyrtosiphon pisum*. *BMC Biotechnol.* 7.
 24. Shakesby AJ, Wallace IS, Isaacs HV, Pritchard J, Roberts DM, & Douglas AE (2009) A water-specific aquaporin involved in aphid osmoregulation. *Insect Biochemistry and Molecular Biology* 39(1):1-10.
 25. Bischof J, Maeda RK, Hediger M, Karch F, & Basler K (2007) An optimized transgenesis system for *Drosophila* using germ-line-specific phi C31 integrases. *Proceedings of the National Academy of Sciences of the United States of America* 104(9):3312-3317.
 26. Santiago Y, Chan E, Liu PQ, Orlando S, Zhang L, Urnov FD, Holmes MC, Guschin D, Waite A, Miller JC, Rebar EJ, Gregory PD, Klug A, & Collingwood TN (2008) Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proceedings of the National Academy of Sciences of the United States of America* 105(15):5809-5814.

Appendix 1: International Aphid Genomics Consortium Member List (May 22nd 2010)

1. Mohatmed Abdel-latif	mohatmed@zedat.fu-berlin.de
2. Francisca Almeida	falmeida@ub.edu
3. Boran Altincicek	Boran.Altincicek@agrار.uni-giessen.de
4. Joanna Amey	joanna.amey@bbsrc.ac.uk
5. Melanie Andrews	melanie.andrews@syngenta.com
6. Peter Ashton	pda2@york.ac.uk
7. Hagop Atamian	hatam001@ucr.edu
8. Stéphanie Bankhead-Dronnet	stephanie.bankhead@univ-orleans.fr
9. Seth Barribeau	seth.barribeau@gmail.com
10. Chris Bass	chris.bass@bbsrc.ac.uk
11. Stéphane Blanc	blanc@supagro.inra.fr
12. Véronique Brault	brault@colmar.inra.fr
13. Jennifer Brisson	jennifer.brisson@gmail.com
14. Roger Butlin	r.k.butlin@shef.ac.uk
15. Marina Caillaud	mcaillaud@ithaca.edu
16. Federica Calevro	Federica.Calevro@insa-lyon.fr
17. Salvador Capella Gutierrez	scapella@crg.es
18. Jim Carolan	james.carolan@ucd.ie
19. Chun-che Chang	chunche@ntu.edu.tw
20. TePin Chang	tepin@nhri.org.tw
21. Hubert Charles	hubert.charles@insa-lyon.fr
22. Michelle Cilia	mlc68@cornell.edu
23. John Colbourne	jcolbour@cgb.indiana.edu
24. Stefano Colella	stefano.colella@jouy.inra.fr
25. Teresa Cortes	teresa.cortes@uv.es
26. Richard Dale	richard.dale@syngenta.com
27. T.G. Emyr Davies	tdavies@bbsrc.ac.uk
28. Gregory Davis	gdavis@brynmawr.edu
29. Martin De Vos	mdv@keygene.com
30. Catherine Dogimont	catherine.dogimont@avignon.inra.fr
31. Angela Douglas	aes326@cornell.edu
32. Aviv Drombrovsky	aviv@volcani.agri.gov.il
33. Owain Edwards	Owain.Edwards@csiro.au
34. Jay Evans	jay.evans@ars.usda.gov
35. Gerard Febvay	gerard.febvay@jouy.inra.fr
36. Brian Fenton	brain.fenton@scri.ac.uk
37. Julia Ferrari	jf557@york.ac.uk
38. Linda Field	lin.field@bbsrc.ac.uk
39. Christian Figueroa	christianfigueroa@uach.cl
40. Carol Fitzroy	carol.fitzroy@ucdconnect.ie
41. Toni Gabaldón	tgabaldon@crg.es
42. Filipe Garrett Vieira	fgarrett@ub.edu
43. Jena-Pierre Gauthier	jean-pierre.gauthier@rennes.inra.fr
44. Nicole Gerardo	nicole.gerardo@emory.edu
45. Charles Godfray	charles.godfray@zoo.ox.ac.uk
46. Karl Gordon	karl.gordon@csiro.au

47. Gazi Gorur	gazigorur@yahoo.com, ggorur@nigde.edu.tr
48. Cornelis (Cok) Grimmelikhuijzen	cgrimmelikhuijzen@bio.ku.dk
49. Karina Hansen	kakhansen@bio.ku.dk
50. Frank Hauser	fhauser@bio.ku.dk
51. Abdelaziz Heddi	Abdelaziz.Heddi@insa-lyon.fr
52. Dale Hedges	dhedges@med.miami.edu
53. Etienne Herrbach	etienne.herrbach@colmar.inra.fr
54. Saskia Hogenhout	saskia.hogenhout@bbsrc.ac.uk
55. Jamie Huerta Cepas	jhuerta@crg.es
56. Laura Hunnicutt	LHunnicutt2010@gmail.com
57. Wayne Hunter	Wayne.hunter@ars.usda.gov
58. Asano Ishikawa	asano@ees.hokudai.ac.jp
59. Georg Jander	gj32@cornell.edu
60. Stéphanie Jaubert-Possamai	stephanie.jaubert@rennes.inra.fr
61. Ho-Jong Ju	hj97@cornell.edu
62. Isgouhi Kaloshian	isgouhi.kaloshian@ucr.edu
63. Lars Kamphuis	lars.kamphuis@csiro.au
64. Philip Kushlan	pkushlan@bio.miami.edu
65. Nora Lapitan	Nora.Lapitan@ColoState.EDU
66. Amparo Latorre	amparo.latorre@uv.es
67. Terence-Neil Ledger	Terence.Ledger@sophia.inra.fr
68. Fabrice Legeai	fabrice.legeai@rennes.inra.fr
69. Zheng-Xi Li	zxli@cau.edu.cn
70. Gee-way Lin	d95632002@ntu.edu.tw
71. Sijun Liu	sliu@iastate.edu
72. Carlos Llorens	carlos.llorens@uv.es
73. Hsiao-Ling Lu	d96632001@ntu.edu.tw
74. Sandy Macdonald	ajm513@york.ac.uk
75. Gian Carlo Manicardi	giancarlo.manicardi@unimore.it
76. Marina Marcet Houben	mmarcet@crg.es
77. David Martinez Torres	david.martinez@uv.es
78. Toru Miura	miu@ees.hokudai.ac.jp
79. Baptiste Monsion	baptiste.monsion@colmar.inra.fr
80. Nancy Moran	nmoran@u.arizona.edu
81. Mizue Morioka	mizue@mbr.nifty.com
82. Andres Moya	andres.moya@uv.es
83. Atsushi Nakabachi	bachi@riken.jp
84. Justin Pachebat	jap64@cam.ac.uk
85. Benjamin Parker	bparke4@emory.edu
86. Emmanuelle Permal	emmanuelle.permal@versailles.inra.fr
87. Miguel Pignatelli	miguel.pignatelli@uv.es
88. Dan Price	dan.price@bio.miami.edu
89. Hadi Quesneville	Hadi.Quesneville@versailles.inra.fr
90. Yvan Rahbé	Yvan.Rahbe@jouy.inra.fr
91. Claudio Ramirez	clramirez@utalca.cl
92. John Ramsey	jsr47@cornell.edu
93. Karen Reardon	karenreardon@ucdconnect.ie

94. Stephen Richards	stephenr@bcm.tmc.edu
95. Dean Rider	drider@cvm.tamu.edu
96. Alain Robichon	Alain.Robichon@sophia.inra.fr
97. Julio Rozas	jrozas@ub.edu
98. Calum Russell	cwr7@cornell.edu
99. Panagiotis Sapountzis	panagiotis.sapountzis@jouy.inra.fr
100. Shuji Shigenobu	sshigeno@princeton.edu
101. Jean-Christophe Simon	jean-christophe.simon@rennes.inra.fr
102. Carole Smadja	c.smadja@sheffield.ac.uk
103. Guy Smagghe	guy.smagghe@ugent.be
104. Dayalan Srinivasan	dsriniva@princeton.edu
105. Elisabeth Stafflinger	EStafflinger@bio.ku.dk
106. David Stern	dstern@princeton.edu
107. Denis Tagu	denis.tagu@rennes.inra.fr
108. Cécilia Tamborindeguy	ctamborindeguy@ag.tamu.edu
109. Sophie Tares	Sophie.Tares@sophia.inra.fr
110. Gavin Thomas	ght2@york.ac.uk
111. Eric Van Fleet	evanfleet1@gmail.com
112. Jan Veenstra	j.veenstra@cnic.u-bordeaux1.fr
113. Rodrigo Velarde	velardra@wfu.edu
114. Augusto Vellozo	avellozo@gmail.com
115. Andreas Vilcinskis	Andreas.Vilcinskis@agrar.uni-giessen.de
116. Thomas Walsh	tom.walsh@csiro.au
117. Tom Wilkinson	tom.wilkinson@ucd.ie
118. Michael Williamson	mrwilliamson@bio.ku.dk
119. Alex Wilson	acwilson@bio.miami.edu
120. Jing-Jiang Zhou	jing-jiang.zhou@bbsrc.ac.uk

Appendix 2. A full list of publications associated with publication of the pea aphid genome. (32 Papers)

- Brisson JA, Ishiawa A, Miura T. (2010) Wing development genes of the pea aphid and differential gene expression between winged and unwinged morphs. *Insect Mol Biol.* 19(s2):63-73.
- Carolan JC, Fitzroy CI, Ashton PD, Douglas AE, Wilkinson TL. (2009) The secreted salivary proteome of the pea aphid *Acyrtosiphon pisum* characterised by mass spectrometry. *Proteomics.* (9):2457-67.
- Christiaens O, Iga M, Velarde RA, Rougé P, Smaggh G. (2010) Halloween genes and nuclear receptors in ecdysteroid biosynthesis and signaling in the pea aphid. *Insect Mol Biol.* 19(s2):187-200.
- Cortés T, Ortiz-Rivas B, Martínez-Torres D. (2010) Identification and characterization of circadian clock genes in the pea aphid *Acyrtosiphon pisum*. *Insect Mol Biol.* 19(s2):123-139.
- Dale RP, Jones AK, Tamborindeguy C, Davies TGE, Amey JS, Williamson S, Wolstenholme A, Field LM, Williamson MS, Walsh TK, Sattelle DB. (2010) Identification of ion channel genes in the *Acyrtosiphon pisum* genome. *Insect Mol Biol.* 19(s2):141-153.
- Degnan PH, Leonardo TE, Cass BN, Hurwitz B, Stern D, Gibbs RA, Richards S, Moran NA. (2010) Dynamics of genome evolution in facultative symbionts of aphids. *Environ Microbiol.* 2009 Oct 16.
- Gerardo NM, Altincicek B, Anselme C, Atamian H, Barribeau SM, de Vos M, Duncan EJ, Evans JD, Gabaldón T, Ghanim M, Heddi A, Kaloshian I, Latorre A, Moya A, Nakabachi A, Parker BJ, Pérez-Brocá V, Pignatelli M, Rahbé Y, Ramsey JS, Spragg CJ, Tamames J, Tamarit D, Tamborindeguy C, Vincent-Monegat C, Vilcinskis A. Immunity and other defenses in pea aphids, *Acyrtosiphon pisum*. *Genome Biol.* 11(2):R21. (2010)
- Huang TY, Cook CE, Davis GK, Shigenobu S, Chen RP-Y, Chang CC. (2010) Anterior development in the parthenogenetic and viviparous form of the pea aphid, *Acyrtosiphon pisum*: hunchback and orthodenticle expression. *Insect Mol Biol.* 19(s2):75-85.
- Huerta-Cepas J, Marcet-Houben M, Pignatelli Moya MA, Gabaldón T. (2010) The pea aphid phylome: a complete catalogue of evolutionary histories and arthropod orthology and paralogy relationships for *Acyrtosiphon pisum* genes. *Insect Mol Biol.* 19(s2):13-21.
- Huybrechts J, Bonhomme J, Minoli S, Prunier-Leterme N, Dombrovsky A, Abdel-Latif M, Robichon A, Veenstra JA, Tagu D. (2010) Neuropeptide and neurohormone precursors in the pea aphid, *Acyrtosiphon pisum*. *Insect Mol Biol.* 19(s2):87-95.
- International Aphid Genomics Consortium. (2010) Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.* 8(2):e1000313.
- Jaubert-Possamai S, Rispe C, Tanguy S, Gordon K, Walsh T, Edwards O, Tagu D. (2010) Expansion of the miRNA pathway in the hemipteran insect *Acyrtosiphon pisum*. *Mol Biol Evol.* (5):979-87.
- Legeai F, Rizk G, Walsh T, Edwards O, Gordon K, Lavenier D, Leterme N, Mereau A, Nicolas J, Tagu D, Jaubert-Possamai S. (2010) Bioinformatic prediction, deep sequencing of microRNAs and expression analysis during phenotypic plasticity in the pea aphid, *Acyrtosiphon pisum*. *BMC Genomics.* 11(1):281.
- Legeai L, Shigenobu S, Gauthier J-P, Colbourne J, Rispe C, Collin O, Richards S, Wilson ACC, T. Murphy, D. Tagu. (2010) AphidBase: a centralized bioinformatic resource for annotation of the pea aphid genome. *Insect Mol Biol.* 19(s2):5-12.
- Nakabachi A, Miyagishima S. (2010) Expansion of genes encoding a novel type of dynamin in the genome of the pea aphid, *Acyrtosiphon pisum*. *Insect Mol Biol.* 19(s2): 165-173.
- Nakabachi A, Shigenobu S, Miyagishima S. (2010) Chitinase-like proteins encoded in the genome of the pea aphid, *Acyrtosiphon pisum*. *Insect Mol Biol.* 19(s2): 175-185.

- Le Trionnaire G, Francis F, Jaubert-Possamai S, Bonhomme J, De Pauw E, Gauthier JP, Haubruge E, Legeai F, Prunier-Leterme N, Simon JC, Tanguy S, Tagu D. (2009) Transcriptomic and proteomic analyses of seasonal photoperiodism in the pea aphid. *BMC Genomics*. 10:456.
- Moran NA, Jarvik T. (2010) Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science*. 328(5978):624-7.
- Nikoh N, McCutcheon JP, Kudo T, Miyagishima SY, Moran NA, Nakabachi A. (2010) Bacterial genes in the aphid genome: absence of functional gene transfer from *Buchnera* to its host. *PLoS Genet*. 6(2):e1000827.
- Ollivier M, Legeai F, Rispe C. (2010) Comparative analysis of the *Acyrtosiphon pisum* genome and expressed sequence tag-based gene sets from other aphid species. *Insect Mol Biol*. 19(s2): 33-45.
- Price DRG, Tibbles K, Shigenobu S, Smertenko A, Russell ACW, Douglas AE, Fitches E, Gatehouse AMR, Gatehouse JA. (2010) Sugar transporters of the major facilitator superfamily in aphids; from gene prediction to functional characterization. *Insect Mol Biol*. 19(s2): 97-112.
- Ramsey JS, MacDonald SJ, Jander G, Nakabachi A, Thomas GH, Douglas AE. (2010) Genomic evidence for complementary purine metabolism in the pea aphid, *Acyrtosiphon pisum*, and its symbiotic bacterium *Buchnera aphidicola*. *Insect Mol Biol*. 19(s2): 241-248.
- Ramsey JS, Rider DS, Walsh TK, De Vos M, Gordon KHJ, Ponnala L, Macmil SL, Roe BA, Jander G. (2010) Comparative analysis of detoxification enzymes in *Acyrtosiphon pisum* and *Myzus persicae*. *Insect Mol Biol*. 19(s2): 155-164.
- Rider SD, Srinivasan DG, Hilgarth RS. (2010) Chromatin-remodelling proteins of the pea aphid, *Acyrtosiphon pisum* (Harris). *Insect Mol Biol*. 19(s2): 201-214.
- Shigenobu S, Bickel RD, Brisson JA, Butts T, Chang C-c, Christiaens O, Davis GK, Duncan EJ, Ferrier DEK, Iga M, Janssen R, Lin G-W, Lu H-L, McGregor AP, Miura T, Smaghe G, Smith JM, van der Zee M, Velarde RA, Wilson MJ, Dearden PK, Stern DL. (2010) Comprehensive survey of developmental genes in the pea aphid, *Acyrtosiphon pisum*: frequent lineage-specific duplications and losses of developmental genes. *Insect Mol Biol*. 19(s2): 47-62.
- Shigenobu S, Richards S, Cree AG, Morioka M, Fukatsu T, Kudo T, Miyagishima S, Gibbs RA, Stern DL, Nakabachi A. (2010) A full-length cDNA resource for the pea aphid, *Acyrtosiphon pisum*. *Insect Mol Biol*. 19(s2): 23-31.
- Smadja C, Shi P, Butlin RK, Robertson HM. (2009) Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrtosiphon pisum*. *Mol Biol Evol*. (9):2073-86.
- Srinivasan DG, Fenton B, Jaubert-Possamai S, Jaouannet M. (2010) Analysis of meiosis and cell cycle genes of the facultatively asexual pea aphid, *Acyrtosiphon pisum* (Hemiptera: Aphididae). *Insect Mol Biol*. 19(s2): 229-239.
- Tamborindeguy C, Monsion B, Brault V, Hunnicutt L, Ju HJ, Nakabachi A, Van Fleet E. A genomic analysis of transcytosis in the pea aphid, *Acyrtosiphon pisum*, a mechanism involved in virus transmission. *Insect Mol Biol*. 19(s2): 259-272. (2010)
- Walsh TK, Brisson JA, Robertson HM, Gordon K, Jaubert-Possamai S, Tagu D, Edwards OR. (2010) A functional DNA methylation system in the pea aphid, *Acyrtosiphon pisum*. *Insect Mol Biol*. 19(s2): 215-228.
- Wilson ACC, Ashton PD, Calevro F, Charles H, Colella S, Febvay G, Jander G, Kushlan PF, Macdonald SJ, Schwartz JF, Thomas GH, Douglas AE. (2010) Genomic insight into the amino acid relations of the pea aphid, *Acyrtosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. *Insect Mol Biol*. 19(s2): 249-258.

Zhou J-J, Vieira FG, He X-L, Smadja C, Liu R, Rozas J, Field LM. (2010) Genome annotation and comparative analyses of the odorant-binding proteins and chemosensory proteins in the pea aphid *Acyrtosiphon pisum*. *Insect Mol Biol.* 19(s2): 113-122.

Appendix 3. List of morph, developmental and anatomical features requiring gene expression evidence in aphids.

Foreword: This is conceived as the start of a controlled vocabulary for **aphid genome projects**. The **vocabulary** (*i.e.* glossary) will be followed by a **dictionary** (*i.e.* definitions), and included in an **ontology** (*i.e.* structured space, knowledge model of the systems-approach representation of aphid physiology and anatomy). The primary objective of this ontology is to allow the **integration of data and models** between the aphid genome database (**AphidBase**) and the numerous **omics** experiments being produced by the community. In between AphidAtlas and the gene-centered database AphidBase, there will be numerous **data repositories** (raw or first-phase data from *omics* experiments; *e.g.* dbEST from Genbank) and full-size **databases** (AcypiCyc, PhylomeDB *etc*). All this knowledge will be coordinated by AphidBase.

Phase 0 included below is a tentative incomplete though accurate description of the vocabulary needed to describe the present state of data (all tissue libraries available as of May 2010). The vocabulary should not change but of the ontology, and even definitions, will. All boldfaced terms should be fixed and shared by the community as soon as possible. The **objects analysed** by this **ontology** are either **insect individuals** (epigenetic, genetic, development-time scale) or **cells/group of cells —tissues or organs—** (anatomical space *i.e.* individual sub-geometric space).

Appendix 3 cont. List of morph, developmental and anatomical features requiring gene expression evidence in aphids.

Term 0	Term 1	Term 2	Term 3	Term 4
Dimensions of the ontology ^a (defined as space of objects analysed)	Object Classes (level 1)	level 2	level 3	level 4
Morphs (epigenetic space) ^b —except sex, which is not an epigenetic character in <i>A.pisum</i> —	Parthenogenetic virginoparae female	Unwinged (Apterous ?) Winged (Alate ?) \$	could be color morph level if needed, etc ... \$\$ green, pink (or rose)	aposymbiotic ? \$\$ (or simply used as additional keyword)
	Parthenogenetic sexuparae female Parthenogenetic gynoparae females Parthenogenetic fundatrix Parthenogenetic fundatrigen Sexual female Male	Unwinged ,Winged Unwinged ,Winged Unwinged ,Winged Unwinged ,Winged Unwinged ,Winged Unwinged ,Winged	(or simply used as additional keyword)	
Developmental processes (time-development space) ^c	Larval stage (question: should we use larval or nymphal) \$\$	L1 (first instar) ?\$ L2 (second instar) L3 (third instar) L4 (fourth instar) adult	stages after Miura <i>et al.</i> 2003	
	Parthenogenetic Embryo	early development mid development late development		
	Sexual egg/embryo	egg early development mid development late development		
Anatomical features (anatomical space <i>s</i>) ^d	Whole body (may be level 1 see how drosophila holds this) \$\$ Head	central nervous system	brain protocerebrum corpora allata corpora cardiaca	
	Thorax Abdomen	eyes antennae		
		bacteriocytes gut	foregut midgut hindgut honeydew	
		ovarioles	germarium ... ovariole sheath trophic cord embryo	terms after Bermingham <i>et al.</i> 2009 trophocytes primary oocytes
	Transmetameric organs and cells (?) \$\$	hemolymph	plasma hemocytes	...
		salivary gland	saliva , primary salivary gland , secondary salivary gland	
	Appendages (note: antenna is an appendage...) \$\$			

a: additional dimensions could be: **Genetic** space (*A.pisum* clone LSR1 is default); **Symbiotic** space (or status); **Language** space (English is default); **Taxonomic space** (Aphididae is default), insect could be the “target” group ?

b: vocabulary as defined by Blackman (1994) The simplification of aphid terminology, *European Journal of Entomology* 91:139-141.

c: reference(s) to be defined; Miura *et al.* 2009, Chang *et al.* Xx).

d: reference(s) to be defined (Ponsen, 1972, Bermingham *et al.* 2009, Chang *et al.* 2009, and “compatible” with Drosophila ontology as in Flybase).

e: **boldfaced red-colored** items are used/needed for already-performed/published transcriptomic experiments (EST, microarrays, RNAseq), **boldfaced blue color** items are used/needed for already-performed/published proteomic analyses; **violet** for both.

Appendix 4. Aphid Virtual Stock Center Submission Guidelines

STEP 1: Generation of Clone Fingerprint

Using published markers, preferably microsatellites, generate a clone fingerprint for all aphid lines to be submitted to the Aphid Virtual Stock Center.

STEP 2: Archiving Reference Material

On submission of a clone to the Aphid Virtual Stock Center, 10 ug of high-quality reference DNA should be sent to AphidBase (Jean-Christophe Simon, INRA Rennes, UMR BiO3P, 35327, 35653 Le Rheu cedex France, jean-christophe.simon@rennes.inra.fr), and a further 10 ug of high-quality reference DNA should be archived in the host lab. To ensure DNA quality, submission of reference DNA needs to be accompanied by an agarose gel image of the reference DNA relative to a suitable DNA size marker.

STEP 3: Submission to AphidBase: Aphid Virtual Stock Center

The following information is required for clone submission to the Aphid Virtual Stock Center:

Species:

Clone Name:

Color:

Collection Date:

Collection Locality: town, city, country.

Latitude:

Longitude:

Collector's Name:

Host Plant:

Secondary symbionts if known:

Lifecycle: Holocyclic/Intermediate/Anholocyclic/Unknown

Publications:

Additional Information:

Date of Submission to Aphid Virtual Stock Center:

DNA fingerprint: Link to pdf file containing information about DNA fingerprint and the fingerprinting protocol

Host Lab:

Host Lab Contact Information:

STEP 4: Distribution of Aphid Clonal Lines

It is the responsibility of the lab requesting an aphid line to obtain the necessary importation permits. It is the responsibility of the host lab to ensure the clone integrity. When distributing aphid lines, host labs should send aphids together with 100 ng of archived reference DNA (sufficient DNA for 10-20 PCR reactions). Following establishment of the aphid lines in the new lab, the aphid line should be genotyped against the reference DNA.

QUALITY CONTROL

Host laboratories are responsible to ensuring the integrity of their clonal lines prior to distribution to new labs. It is recommended that aphid lines be genotyped on a bi-yearly basis to ensure their integrity.