



Molecular biology and genetics/Biologie et génétique moléculaires

The anatomy of an aphid genome: From sequence to biology

Anatomie du génome d'un puceron : de la séquence à la biologie

Denis Tagu^{a,*}, Sébastien Dugravot^b, Yannick Outreman^c, Claude Rispe^a,
Jean-Christophe Simon^a, Stefano Colella^d

^a INRA, UMR BiO3P, BP 35327, 35653 Le Rheu cedex, France

^b UMR BiO3P, campus Beaulieu, université Rennes 1, 35042 Rennes cedex, France

^c Agrocampus Ouest, UMR BiO3P, centre de Rennes, 65, rue de Saint-Brieuc CS 84215, 35042 Rennes cedex, France

^d UMR203 biologie fonctionnelle insectes et interactions (BF2I), INRA, INSA-Lyon, université de Lyon, 69621 Villeurbanne, France

ARTICLE INFO

Article history:

Available online 15 May 2010

Keywords:

Acyrtosiphon pisum
Aphid
Genome
Hemiptera
Insect

Mots clés :

Acyrtosiphon pisum
Génome
Insecte
Hémiptère
Puceron

ABSTRACT

In 2009, the genome of the pea aphid (*Acyrtosiphon pisum*) was sequenced and partially annotated. This means that the scientific community has now access to the genetic material of aphids that are serious pests of plants. The description of an aphid genome is a first step to go deeper in the understanding of the biology of these insects. In this article, we first describe how the pea aphid genome has been sequenced and its major characteristics. However, a genome is only a part of what an individual is and describing a genome without taking into account the role of the environment makes no sense. Thus, in the second part of the review, we envisage how this genomic resource will fuel many other disciplines such as ecology, evolutionary biology, population genetics and symbiosis.

© 2010 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

R É S U M É

Le génome du puceron du pois (*Acyrtosiphon pisum*) a été séquencé et partiellement annoté en 2009. La communauté scientifique a donc accès à la description du matériel génétique de pucerons qui sont des ravageurs des plantes. Cette description est une première étape indispensable pour améliorer nos connaissances sur la biologie de ces insectes. Dans une première partie, nous décrivons comment ce génome a été obtenu ainsi que ses principales caractéristiques. Cependant, un génome n'est qu'une composante partielle de ce qui forme un individu et décrire un génome sans tenir compte de l'environnement local n'aurait pas de sens. Aussi, dans une seconde partie, nous proposons des pistes sur l'utilisation de ces données génomiques dans différentes disciplines comme l'écologie, la biologie évolutive, la génétique des populations et les interactions symbiotiques.

© 2010 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

* Corresponding author.

E-mail address: denis.tagu@rennes.inra.fr (D. Tagu).

Abbreviations

bp	base pair
EST	Expressed Sequence Tags
IAGC	International Aphid Genomics Consortium
IMD	immunodeficiency
Mb	megabases
NCBI	National Center for Biotechnology Information
QTL	Quantitative Trait Loci
SNPs	Single Nucleotide Polymorphisms
WGS	Whole Genome Shotgun

1. Sequencing “a” genome?

The pea aphid (*Acyrtosiphon pisum* [*A. pisum*]) genome has been recently sequenced [1] becoming one of the 32 available insect genomes (Table 1) and the first assembled and annotated genome for a Hemiptera. Hopefully, the pea aphid will quickly loose this leadership position with the sequencing of other Hemiptera genomes (several sequencing projects are in progress including the bug *Rhodnius prolixus*, white flies...). Why is this a good thing? Because generating a better understanding of

Table 1
Thirty-two insect genome sequencing projects.

Order	Organism	Genome size (Mb)	# Chr	Depth	Release date	Sequencing center	Country
Coleoptera	<i>Tribolium castaneum</i>	200	10	7.3X	08/17/2005	Baylor College of Medicine	USA
Diptera	<i>Aedes aegypti</i> Liverpool	800	3	8X	02/11/2005	TIGR	USA
	<i>Anopheles gambiae</i> M	300	5	5.6X	02/22/2008	GSC at WashU	USA
	<i>Anopheles gambiae</i> S	300	5	5.7X	02/22/2008	J. Craig Venter Institute	USA
	<i>Anopheles gambiae</i> str. PEST	300	5	10X	03/22/2002	Celera Genomics/Genoscope	USA and France
	<i>Culex pipiens quinquefasciatus</i> JHB	580	N/A	8X	04/19/2007	Broad Institute (more)	USA
	<i>Drosophila ananassae</i>	231	4	9X	04/11/2006	Agencourt bioscience corporation	USA (private)
	<i>Drosophila erecta</i>	153	4	10X	04/11/2006	Agencourt bioscience corporation	USA (private)
	<i>Drosophila grimshawi</i>	N/A	N/A	8X	04/11/2006	Agencourt bioscience corporation	USA (private)
	<i>Drosophila melanogaster</i>	180	4	N/A	22/06/1995	Celera genomics	USA
	<i>Drosophila mojavensis</i>	194	6	8X	04/11/2006	Agencourt bioscience corporation	USA (private)
	<i>Drosophila persimilis</i>	188	5	4X	09/28/2005	Broad institute	USA
	<i>Drosophila pseudoobscura</i>	131	5	9.1X	12/24/2003	Baylor college of medicine	USA
	<i>Drosophila sechellia</i>	167	4	5X	09/28/2005	Broad institute	USA
	<i>Drosophila simulans</i> C167.4	150	4	1X	03/23/2005	GSC at WashU	USA
	<i>Drosophila simulans</i> MD106TS	150	4	1X	03/23/2005	GSC at WashU	USA
	<i>Drosophila simulans</i> MD199S	150	4	1X	03/23/2005	GSC at WashU	USA
	<i>Drosophila simulans</i> New Caledonia 48S	150	4	1X	03/23/2005	GSC at WashU	USA
	<i>Drosophila simulans</i> SIM4	150	4	1X	03/23/2005	GSC at WashU	USA
	<i>Drosophila simulans</i> SIM6	150	4	1X	03/23/2005	GSC at WashU	USA
	<i>Drosophila simulans</i> mosaic	150	4	4X	08/16/2006	GSC at WashU	USA
<i>Drosophila simulans</i> white501	150	4	4X	03/23/2005	GSC at WashU	USA	
<i>Drosophila virilis</i>	206	6	8X	04/11/2006	Agencourt bioscience corporation	USA (private)	
<i>Drosophila willistoni</i>	236	3	8.4X	04/21/2006	J. Craig Venter institute	USA	
<i>Drosophila yakuba</i>	166	4	9X	07/07/2004	GSC at WashU	USA	
Hemiptera	<i>Acyrtosiphon pisum</i> LSR1	525	4	6X	04/01/2008	Baylor college of medicine	USA
	<i>Rhodnius prolixus</i>	670	11	8X	06/17/2009	GSC at WashU	USA
Hymenoptera	<i>Apis mellifera</i> DH4	200	16	7–8X	12/19/2003	Baylor college of medicine	USA
	<i>Nasonia vitripennis</i>	350	5	6.2X	05/04/2007	Baylor college of medicine	USA
Lepidoptera	<i>Bombyx mori</i> Dazao	530	28	5.9X	10/12/2004	Southwest agricultural university	China
	<i>Bombyx mori</i> p50	530	28	3X	04/27/2004	International Lepidopteran Genome Project	Japan
	<i>Bombyx mori</i> p50T (= Dazao)	530	28	10X	04/23/2008	The international silkworm genome sequencing consortium	Japan and China
Phthiraptera	<i>Pediculus humanus</i>	120	N/A	8X	04/18/2007	J. Craig Venter institute	USA

Source July 2009: <http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi>.

Mb: megabase; N/A: not available; # Chr: number of haploid chromosomes; X: sequencing coverage.

species evolution and biological functions requires multiple genome complete sequences. Recently, thanks to the development of novel high-throughput sequencing technologies (for a review, see [2]), a democratization of the access to full genome sequencing has taken place and this will allow novel genomics research approaches for several organisms. Because of the plant pest status of aphids, and because of their numerous original life history traits (e.g. host plant adaptation, symbiosis. . .) [3], the sequencing of this genome represents a first essential step for the development of genomics based aphid biology research. The aim of this article is to explain how this is true, but also to discuss the issues that remain to be solved. And one can ask: what is the meaning of a sequenced genome? What are the advantages to get such information for a given organism?

Sequencing a eukaryotic genome is getting all (or most of) the ordered nucleotides of the DNA present in the cell, thus the sequence of the different haploid chromosomes. Even if this is true, can we say today that we have access to the human genome? (Box 1) In reality, the DNA to be sequenced to generate the sequence of the “human genome” has been extracted from a few individuals (at least for the initial human genome project) and the final sequence is thus a consensus or an “average genome” that is supposed to represent the genome of the species [4,5]. Nowadays, mainly for economical reasons under the impulse of pharmaceutical companies, new technologies offer the access to “personal genomics”, and thus the

genome sequence for a single individual can be generated [6,7]; also, in this case, we have a consensus genome since the two haploid sets of chromosomes are not identical due to allelic polymorphism. In the case of the pea aphid, things get “easier”. Because of their ability to reproduce clonally [8], several individuals belonging to the same genetic lineage (and containing a supposedly identical genome) can be reared and used for DNA extraction before sequencing. Thus, for aphids, the genome of a colony represents in fact the genome of an individual. However, the two sets of haploid chromosomes are still not identical. In order to partially overcome this limitation, the degree of heterozygosity has been reduced by one round of self-cross: the result was the LSR1 pea aphid clone used for the DNA sequencing. Heterozygosity should ideally be reduced even further, through several rounds of self-cross, but this strategy has not been chosen for *A. pisum* for fear of reduced fertility associated with strong inbreeding in aphids.

Sequencing of a new eukaryotic genome is often seen as an important event, as well as a somewhat disappointing “first step”. This is a major event in that access to the genome offers new tools that can be immediately useful: list of genes to approach species adaptations and specificity, list of molecular markers to assist genetics analyses, tools for comparative genomics. . . But this is also only a first step, in that access to the genome is not sufficient to answer critical social, health or environmental questions: the human genome did not immediately solve health problems of human populations. And of course sequencing of the pea aphid genome will not immediately allow protecting crops from damage caused by these insect pests. In this article, after a description of the state-of-the-art of the pea aphid genome data and resources, we will discuss how access to the genome will help different areas of aphid research, from gene function, to population genetics and evolutionary biology.

Box 1. Analogies

A genome sequence could be compared to the description of the human anatomy made several centuries ago: an essential step for knowing the structure of the human body. But anatomy does not solve human health alone: physiology is required to understand how organs function and how they interact. The anatomy (sequence) of a genome is essential but not sufficient to understand cell functions. Large-scale functional analyses for every gene in the genome are required. They represent the post-genomics steps that follow the genome sequence and that are characterized by the development and use of “omics” technologies: transcriptomics (concerning the RNAs), proteomics (concerning the proteins), metabolomics (concerning the metabolites) and furthermore, when looking at interactions between all these different cellular components (interactomics). Another example of analogy concerns ecology: sequencing a genome leads to identifying genes coding proteins (and not only), and assigning a name to them. This is a taxonomy of genes. Taxonomy is essential to study interactions between species (ecology) and naming genes is essential to study interactions among molecular components of cellular pathways. Genome accessibility is also essential for evolutionary biology, by comparing the genome structure between species and inferring evolution pathways. Such as anatomy of several species allowed comparative embryology to decipher key developmental pathways common or specific to different species.

2. The pea aphid genome: the movie

2.1. The fellowship of the genome

Obtaining a genome sequence for a species for which no related genomes are available is a long task that requires the involvement of a motivated and collaborative community. In the case of the pea aphid, the IAGC (founded in 2003) wrote a white paper describing the advantage of getting the genome sequence of an aphid and rapidly agreed to focus on one model species: *A. pisum*. The pea aphid is considered as a model aphid species for several reasons: a low number of chromosomes ($n=4$) and a medium-size genome (525 Mb, six times less than the *Homo sapiens* genome, but three times that of *Drosophila melanogaster* [*D. melanogaster*]) (Table 1). Furthermore, the pea aphid has been often used for physiological studies – mostly for nutrition physiology and endosymbiosis interaction. *A. pisum* is a quite large aphid (several millimeters long for an adult) amenable to dissections. This is a species easy to rear; complete life cycle and genetic crosses can be performed in controlled room on a single plant species (*Vicia fabae*). Finally, *A. pisum* represents an extraordinary example of host-races adapted to different host plants (among the *Fabaceae*) [9].

The financial support was obtained from the National Human Genome Research Institute (USA): it is noteworthy that a large majority of sequencing projects (not only for insects) are funded by the USA, a kind of new quest such as the moon project in the 1960s. The sequencing of the pea aphid genome was performed at the Baylor College of Medicine (USA). The strategy used was the Whole Genome Shotgun (WGS) which consists in:

- extracting genomic DNA;
- cutting it by different restriction enzymes and generating small DNA fragments;
- cloning the cut fragments into different vectors (plasmids for small fragments, Bacterial Artificial Chromosomes for large fragments);
- mass sequencing.

In the case of the pea aphid, three libraries were constructed corresponding to approximately 2, 10 and 100 kb fragments, respectively (Baylor College of Medicine, Houston, USA). Sequencing was performed by the regular Sanger technique that provides long reads (> 600 bp) with a high level of accuracy. In the WGS strategy, the following step is to reconstruct the genome. This is the assembly step, one of the most critical for a genome project. Genome assembly is performed by specialized algorithms that work by alignments of overlapping fragments from the different libraries. Because of the relatively small size of the sequences (600 pb) and the presence of numerous and long stretches of repeated sequences within a eukaryotic genome, it is difficult, even impossible in some cases, to assemble correctly all the raw sequences. To partially overcome this limitation of WGS, one approach consists in sequencing a very large number of DNA fragments, in order to increase the probability that each DNA fragment is sequenced at least once, and to get more overlapping fragments. That is why for a genome project, the final number of sequences is largely higher than the estimated size of a genome. This is the so-called “coverage”. The pea aphid genome has, for example, a 6X coverage, meaning that the final number of sequenced nucleotides corresponds to six times the number of nucleotides in the pea aphid genome. Thus, for the 525 Mb pea aphid genome, we obtained $525 \times 6 = 3150$ million sequenced bases.

The assembly (performed by the Baylor College of Medicine using the Atlas assembly pipeline) of the pea aphid genome ended with approximately 22,000 scaffolds; this means 22,000 different stretches of genomic DNA still with gaps among the different scaffolds. We are thus quite far from the four expected haploid chromosomes. More sequencings, as well as genetic maps, are required to improve the assembly. Among all the sequences, 25% were not assembled, probably corresponding to repeated regions of the genome. Nevertheless, despite these imperfections, the newly assembled pea aphid genome already provides a gold mine of information for gene and other features analyses (see below).

2.2. Drowning by numbers

After sequencing and assembly, the third essential step for a genome project is the annotation. The aim is to

describe the different features of a genome: genes encoding proteins, non-coding RNAs, repeats, transposons, telomeric and centromeric sequences... The strategy is based first on a scan of the genome with specialized algorithms able to detect these different features, and a second phase where experts check manually what the algorithms found. For the pea aphid genome, several algorithms were used to find genes encoding proteins (find a start and stop codon, intron/exon junctions...) and using homology search with other species. These programs also used other genomic resources that were available for the pea aphids, such as ESTs that correspond to sequences of mRNAs, thus putative protein encoding genes [10]. For the pea aphid, a first set of approximately 10,000 putative genes was defined, with strongly supported predictions due to their homology to well described genes (from *Drosophila melanogaster* for instance) or the detection of pea aphid ESTs. This constitutes a set of high-confidence genes, similar to the so-called “RefSeq” catalog provided by the NCBI [11]. A second set of approximately 24,000 other genes was defined by combining several of the prediction algorithms results using the GLEAN method, used for the first time to define the gene set of the *Apis mellifera* [12], but these predicted genes have less biological evidence supporting their existence and function. Thus, a total of approximately 34,000 predicted genes were found on the pea aphid genome and this is the Official Gene Set associated with the first version of the assembly.

This number seems high when compared to other insect species, nearly the double of what was described for Diptera, Lepidoptera, Coleoptera and Hymenoptera. At first, such a high number of predicted genes was thought to correspond to sequencing artifacts that could have affected the overall quality of the genome assembly. Furthermore, allelic polymorphisms might also cause the generation of different assembled scaffolds that in reality correspond to the same locus (a factor increased by the choice of a single generation of self-cross prior sequencing). These possibilities were carefully evaluated for the pea aphid genome, but it was shown that such potential errors could not account for the high number of genes. A detailed analysis of the genes in the pea aphid genome led to the identification of many gene duplications. Analysis of the evolutionary distances among duplicated genes clearly shows that duplication occurred at a unusually high level over a long evolutionary history (dating back to the diversification of the aphid group) rather than at a precise time. Even though the genome sequence is scattered with 22,000 scaffolds, which limits the physical description of gene expansions (i.e. some blocks of duplication involving many genes – as, for example, duplication of chromosomes or parts of chromosomes – could have occurred and not be detected), the very low frequency or synteny among duplicated genes (paralogons) strongly suggest that duplications did mostly arise as repeated events involving single genes. Why the frequency of duplication (or the propensity to conserve duplicated copies once they arise) is so much higher in aphids than in other insects still remains to be explained.

At the same time as the pea aphid genome sequencing, the genome of *Daphnia pulex* (<http://wfleabase.org/>) was released: this arthropod has the peculiarity to share with

aphids both the capacity to express phenotypic plasticity towards local environmental changes, and the cyclical parthenogenesis ability [8]. It was puzzling to notice that the *Daphnia* genome also possessed a large number of predicted genes (approximately 40,000) with many duplications. Whether this similitude between the two organisms is a purely random event or it reflects the adaptation to phenotypic plasticity is an intriguing question [13]. This interesting comparison clearly shows one of the advantages of global genome comparisons between species with evolutionary clues and how this can open new research perspectives.

2.3. Private functions

What are the functions of these 34,000 predicted genes? Nowadays, putative gene functions rely mainly on homology search; sequence-A of the pea aphid resembles sequence-B of *Drosophila* that is known to have function X. In the best-case scenario, function X in *Drosophila* has been experimentally demonstrated. But in most cases, the relation of homology is indirect: sequence-A of the pea aphid resembles sequence-B of *Bombyx* that resembles sequence-C of *Drosophila* that resembles sequence-D of mice that has been showed to have the same function as sequence-E from human. In such cases, one needs to keep in mind that gene function annotation based on homology, albeit essential, only provides putative functions that have to be demonstrated experimentally for a given species of interest (here, the pea aphid). Thus, most of the pea aphid annotated genes have “putative functions” since very little experimental data demonstrating a specific function are available for aphid proteins. That is the reason why a manual annotation step is critical: this is a long procedure by which researchers (i.e. human beings and not machines) carefully check the predictions made by different algorithms. Approximately 2000 of the 34,000 genes were manually examined for the pea aphid genome [14]. This manual annotation required a strong and large community of specialists. A list of nearly 30 annotation groups were defined within the aphid community, each specialized on a given gene family functional role (e.g. meiosis, neuropeptides, immune response...) and each annotation group defined the gene of interests to be analysed. Such careful analysis allowed the identification of several features in different gene families. Manual annotation never ends since new descriptions and demonstrations might become available over time: as an example, the fifth version of the annotated *D. melanogaster* genome has been released in 2009, 9 years after the first draft of the genome and further annotations efforts are still ongoing in the community.

The main characteristic of the gene functions of the pea aphid have been described and reviewed [1]. Briefly, large duplications of genes affect many protein families involved in different functions, including microRNA synthesis, chromatin modification [14] and sugar transport [15]. In all these cases, it is still too early to fully explain the adaptive processes behind these duplications, but further research will help understanding this intriguing phenomenon. Some duplicated copies are under positive accelerated selection, suggesting they are acquiring new functions

that remain to be characterized. Besides gene duplications, gene losses were also identified in some other pathways, as for example the missing selenoproteins (necessary for processing rare codons encoding selenocysteine). This is also the case for several genes involved in immune function: some genes from the IMD (immunodeficiency) pathway were not detected in the pea aphid genome whereas they are present in genomes of other sequenced insects. Complementary expression data indicated that the immune response of the pea aphid to different biotic challenges is reduced compared to other insects [16]. Several hypotheses can be raised such as a relatively sterile feeding source (the phloem sap) or the presence of a cortege of bacterial symbionts that might have affected during evolution the capacity of aphids to reject bacterial intruders. In fact, symbiosis is central to aphids biology, in particular the primary symbiont of the pea aphid *Buchnera aphidicola* lives in specialised aphids cells the bacteriocytes, and it is vertically transmitted to the offspring. This obligatory association had been known to provide essential amino acids lacking in the phloem sap. *Buchnera* genome sequence is available [17] and the comparison with the aphid genome data has revealed a clear integration and complementarity between the two organisms in several pathways, in particular as expected in the amino acids synthesis and degradation [18].

All genomic resources require solid, efficient and stable tools to store, analyze, organize, distribute and display all the available data. This is the aim of AphidBase, a centralized bioinformatics resource that was developed to collect all genome related data and to facilitate community annotation of the pea aphid genome by the IAGC (<http://www.aphidbase.com>) [19]. This essential genomic repository is complemented for the pea aphid by two other resources: the PhylomeDB (<http://phylo-medb.org>) containing the phylogenomic analysis of the pea aphid gene set [20,21] and the AcypiCyc database (<http://pbil.univ-lyon1.fr/software/cycads/acypicyc/>), a database reconstructing the metabolism of the pea aphid and its symbiont bacteria *Buchnera aphidicola*.

3. A genome for post-genomic studies

3.1. A genome for insect and aphid evolutionary biology

One of the most interesting aspects of obtaining a new genome is not so much its static description than its comparative study with the genomes from other organisms, as it may reveal major changes in gene repertoires and genome organization, which can be related to shifts in ecological traits and biological novelties. For example, the evidence discussed above of a massive number of duplicated genes in the pea aphid requires phylogenetic studies to characterize the relationships among the duplicated genes: are some of those ancient duplicates that were lost in other insects, or did most of these gene duplications arise in an ancestor of modern aphids, after they diverged from other insects? It is already clear that the second scenario prevails in general, as most gene expansions seen in *A. pisum* appear to be monophyletic (they are specific to aphids). One limitation of current

comparisons between genomes of aphids and other insects is, however, the considerable evolutionary distance between the different insect orders (i.e. at least 300 My separate *A. pisum* and all other completely sequenced insect genomes, whereas only a few My separate humans and other vertebrates). It will be rapidly necessary to develop genomic data, if possible complete genomes from other aphid species to allow finer scale studies: this has already been undertaken using EST data sets obtained in different species (in particular in the peach potato aphid *Myzus persicae*, another pest species found on many different crops). The advantage is that the sequences of two not too distant aphid species are more easily comparable in many ways: assessing their status as orthologous genes, aligning them, and estimating their evolutionary rates (especially the rates of synonymous and non-synonymous mutations indicate the selective pressure that characterizes the gene, and can only be accurately estimated between not too distant organisms). The comparison between *A. pisum* and *M. persicae* partial gene sets have already shown that some genes appear to evolve at very high rates, possibly due to diversifying (or “positive”) selection [22]. In the near future, extensive genomic data will be collected in several other aphid species, which will provide a dynamic framework of the evolution of the pea aphid genome. It will, in particular, allow one to date much more precisely the many duplications seen in *A. pisum* and to determine if and how gene duplications affect sequence evolutionary rates (an important question in genome biology, [23]) and if these can be related with different adaptations of the different aphid species: for example *M. persicae* has been shown to have a recent large expansion of esterases associated with insecticide resistance. Many more species-specific expansions or losses of genes will likely be detected once more extensive genomic data is obtained from different aphids.

3.2. A genome for functional genomics/systems biology study of symbiosis

The pea aphid and *Buchnera aphidicola* constitute the first example of a vertically transmitted endosymbiosis where both the genomes of the host and its primary symbiont are available. This unique resource opens the way to functional genomics research to better understand the regulatory networks underlying symbiosis. Much is already known about the pea aphid – *Buchnera* symbiosis from different studies and several experimental approaches have been used to study this intimate relationship. Genomics level work have been performed in the past few years thanks to the availability of the bacterial symbiont genome sequence, using both *in silico* evolutionary studies [24] and experimental approaches [25]. At present, the genomic status of both organisms opens the way to global integrated functional genomics approaches that will generate data allowing a systems level analysis of this endosymbiosis.

As previously discussed, the relationship between the pea aphid and its primary symbiont *Buchnera* is known to be centered on the exchange of essential amino acids. Even if

we know that this association is obligate and long lasting, little is known about the molecular cross talk underlying and regulating the symbiotic relationship. A first step towards a better understanding of the metabolic relationship has been achieved through the analysis of the reduced symbiont genome [17]; this is now complemented by the integration of knowledge about the host genome that it has allowed the full symbiosis metabolic network reconstruction in the AcypiCyc database. This BioCyc database was built using an *ad hoc* annotation management system (including a database and several software tools), as part of the pea aphid genome annotation effort. A comparison of the integrated annotation of AcypiCyc with manual annotation work showed very good consistency and has allowed a first characterization of the amino acid metabolism [18], as previously discussed. This example is a clear indication that thanks to the availability of both genomes global studies of the metabolic network of symbiosis will be possible in the future and these will complement the knowledge on the *Buchnera* metabolic network [26]. Modelling approaches will also lead to the design of better experiments to dissect the complex metabolic partnership. It is also clear that future integration of secondary symbiont sequenced genomes data in this kind of database are key to perform analyses to further our understanding of more complex and composite metabolic symbiotic relationships.

In recent years, several studies of the gene expression patterns for the pea aphid have been performed using both EST sequencing [10,27,28] and custom microarrays platforms [29–31]. Inference of regulatory genetic network methods, starting from high-throughput experimental “omics” data and transcription factors binding site mining, has developed in recent years [32]. The genome sequence of the pea aphid is a stepping-stone to transcriptome analysis leading to the reconstruction of the genetic networks underlying the fascinating biology of this organism. Beyond metabolism, for which both the metabolic and genetic networks could be now analysed at the same time, transcriptomic and proteomic experiments could be integrated in modelling studies to better understand other aspects of pea aphid biology, as for example the unique characteristics of its reduced immune system [16].

3.3. A genome for aphid population and quantitative genetics

The genomic resources generated for functional analysis on the pea aphid may also serve studies on population and quantitative genetics in several ways. High-throughput sequencing allows developing, without spending time and money in the laboratory, polymorphic markers that could be used to analyze population structure, to construct genetic maps and to conduct genome-wide association studies aimed at identifying genes involved in a given biological function. Indeed, sequencing of a genome generally requires sequencing more than one copy for assembly (see above) and usually more than one specimen of the target species. Thus, sequencing traces contain allelic variants of the same individual (intragenomic variation due to heterozygosity) and of different indivi-

duals (intergenomic variation due to population polymorphism). This diversity can be extracted through automated procedure at the scale of the whole genome by blast comparisons and sequence alignments. Variation can be used to detect polymorphic loci such as microsatellites (stretch of di-, tri- or tetranucleotide repeats), SNPs or indels (insertion/deletion sites). Therefore, an almost unlimited number of polymorphic markers can be derived from high-coverage genome-sequencing projects. EST databases, that usually accompany whole genome sequencing projects, can be also mined for polymorphic loci.

Analysing nucleotide diversity across genomes of several to many individuals provides a deep understanding of the evolutionary forces acting either at whole genome level or at specific genomic regions or sites. It can notably help detecting loci under selection and involved in adaptive divergence as well as reconstructing historical demographic events such as colonization, extension and bottleneck events. Also, numerous polymorphic markers are needed to built genetic maps and search for QTL underlying complex traits, e.g. reproduction, development, immunity and defense.

So far, population genetic studies of the pea aphid have been conducted with a limited amount of markers (< 20) isolated in the laboratory and used to measure population differentiation at neutral loci among host-adapted races or biotypes [9,33]. Recently, population genomic approaches have been launched for resolving the genetic architecture of complex traits of the pea aphid such as reproductive mode variation [8], ecological specialization and dispersal [34]. QTL maps have been elaborated to locate chromosomal position of performance and preference loci underlying host adaptation [35].

We expect more dense genetic maps to be constructed in a very near future, facilitating the assembly of the current 22,000 scaffolds into four linkage groups corresponding to the four haploid chromosomes of the pea aphid. We also anticipate the genome sequencing of other pea aphid genotypes with distinct geographic or host plant origins. Combined with functional genomics and other post-genomic methodologies, population genomics and genome-wide association studies will allow deciphering the mechanisms and the evolutionary histories of some of the peculiar and complex biological adaptations of the pea aphid.

3.4. A genome for aphid ecology

3.4.1. Genomics and behavioural ecology

Animals exhibit different activities during the course of their lives: they forage in their habitat for finding food and mates, defend themselves against their natural enemies, and eventually, take care of their relatives. These behaviours, like many other phenotypic traits, can be understood at different levels of causation, from absolutely proximal to the ultimate, or evolutionary [36]. The study of the proximal causation of animal behaviour led to the historical and controversial “nature versus nurture” debate: it concerns the relative importance of an individual’s genes (nature) versus environment (nurture) in determining or causing individual differences in

behavioural traits. Even if this debate is still present, many biologists have accepted that behaviours are orchestrated by interplay between inherited and environmental influences acting on the same substrate; the genome [37].

Rapid advances in molecular genetic studies of insect species genome result in the identification of genes associated with complex behavioural traits. Various techniques (mapping of genomic areas, identification of candidate genes and characterisation of causative mutations) have been indeed used to successfully identify genes associated with a number of adaptive behaviours like courtship and mating in *Drosophila melanogaster* [38], foraging in *D. melanogaster* [39,40] and honeybees [41] and social interaction [42], all of which have been evolutionarily conserved.

A genomic approach to aphid behaviour is still an open question. In aphids, essential behaviours such as behavioural defences against their natural enemies or dispersal for colonization of new habitat have different causes. Faced with natural enemies, aphids defend themselves either by exhibiting behavioural defences (e.g., kicking, jerking, escaping) or emitting a sticky substance containing an alarm pheromone. This emission elicits escape or defences in other individuals of the colony, reducing enemy efficiency. Kunert et al. [43] show that response towards alarm pheromone strongly varies between aphid clones within *A. pisum* species: while some genotypes exhibit escaping responses others stay on the host plant. The defensive behaviours can also be affected by environment conditions: when the temperature increases, the pea aphids reduce their tendency to drop from the plant in presence of natural enemies [44]. Concerning their dispersal, many aphid species are able to produce two alternative dispersal phenotypes: winged or wingless [45]. The winged dispersal morph is mainly responsible for the colonization of new plants and is generally produced under adverse environmental conditions such as crowding, poor plant quality [46,47] or risk of predation. This is one of the first cases of a natural enemy-induced morphological shift in a terrestrial antagonist system [48].

The identification of the genes associated to these behavioural variations and/or phenotypic plasticity and changes in their level of expression can offer a great expansion of our understanding of aphid behaviour from an evolutionary perspective. Significant progress in molecular biology and genomics, and the output of the pea aphid genome-sequencing project makes this an opportune time for such new behavioural programmes of research.

3.4.2. A genome for aphid chemical ecology

Chemical compounds produced by organisms are fundamentally involved in numerous intra- or interspecific interactions between individuals, from bacteria to mammals. Insects such as aphids are no exception. In aphids, essential components of their life cycle such as the selection of a suitable trophic resource, the sexual partner encounter or defence behaviours towards their natural enemies are closely linked to the production or the

presence of chemical signals in the environment [49]. Up to now, studies performed in the field of chemical ecology attempt to describe the complex interactions between individuals mediated by chemical signals, using behavioural observations associated with (bio-) chemical analysis. However, all the mechanisms linked to chemical information transfer between organisms, whether the production or perception, have a molecular basis and are the result of a particular gene expression. Therefore, the recent advances in genomic techniques and the possibility to work on species in which the genomes have been entirely sequenced provide new opportunities for chemical ecologists to better understand the interactions between organisms mediated by chemical cues [50]. Moreover, one of the main objectives of the annotation of a sequenced genome and more globally of the postgenomic biology is to determine the function of the identified genes. In order to achieve that goal, the function of numerous genes, such as environmental response genes, can only be understood within the context of chemical ecology [51]. Functional genomics has already been used in the field of chemical ecology in plant-insect interactions but essentially from a plant perspective [52]. However, very few chemical ecologists have used genomic technologies in insects. The determination of the genetic and molecular basis of the component elements of chemical communication systems could shed light on important ecological phenomena. For instance, the availability of the *Drosophila melanogaster* genome led to the identification of the first insect olfactory receptor (Or) genes [53]. These genes encode for transmembrane proteins, which detect chemical stimuli in the environment leading to the activation of secondary messaging systems and nerve impulses. After their identification in the *D. melanogaster* genome, special efforts have been made to annotate this gene family in the genome of several insect species including the pea aphid *A. pisum*. In addition to traditional chemical ecology studies focusing on the role of semiochemicals in aphid behaviour, the complete annotation of the aphid genome will undoubtedly bring crucial information to the understanding of the complex interactions between aphids and their host plants. Post-genomic tools could be used to better understand how specialist species developing at the expense of only a few host plant species are able to recognize a suitable trophic substrate. It would, therefore, be interesting to compare different aphid species to determine how the olfactory system detects the specific chemical cues with which they are confronted. For instance, what are the olfactory receptor genes involved and how do they govern the response spectra of their olfactory receptor neurons? This is also true at the intraspecific level. It has recently been discovered in the pea aphid, *A. pisum* that several host races can be identified, differing in their adaptation to different host plants [9]. In this case, it would be particularly interesting to investigate whether the differentiation between host races could be attributed to differences in the expression or structure of genes encoding for chemoreceptor proteins. This could be one of the mechanisms leading to sympatric speciation.

4. Conclusions

As stated in the first part of this review, having obtained an aphid genome will not easily free crops from aphid infestation. More research will be needed to reach agro-genomic based innovations such as the development of new insecticides that would target aphids, which will not be active on other non-targeted species and that could thus be used at low concentration in a time-regulated application to prevent environment pollution. The better understanding of population dynamics could also be used, for example, to develop novel input parameters in decision-making tools in order to improve modelling and sustainable agriculture.

The availability of the pea aphid genome is a starting point to functional genomic research and it opens the way to the development of more aphid genomic resources for research. Firstly, as already discussed, re-sequencing pea aphid genomes corresponding to different populations can fuel association genetics approaches to identify important loci that are under selective pressure in aphids and corresponding to major life history traits (reproduction, host adaptation...). Aphids can now enter the “personal genomic” era, a necessary step to develop successful strategies towards different aphid pests. Secondly, getting sequences from other aphid species will help to confirm or reject hypotheses made starting from the pea aphid genome studies. It is becoming more and more clear how important it is to get access to genome sequences of other important aphid pests, such as *Myzus persicae* (a generalist aphid that transmit many plant viruses) or different cereal aphids that also transmit key plant viral diseases.

Starting to describe the “anatomy” of a pea aphid genome has been a fantastic challenge. However, the next steps in research that will be fuelled by this achievement are even more important. The future research years will bring much new information about this genome: by giving a function to the different gene and protein complements of the pea aphid, by identifying interaction networks among these genes and proteins and by integrating this knowledge at the level of intra- and interspecies comparison. These new challenges are now being considered by the IAGC¹ whose members will continue to work together for the development of shared tools that will allow them to reach these challenging research goals.

Acknowledgements

The authors would like to thank all the members of the IAGC for the access to unpublished data from the annotation of the pea aphid genome. This work has been supported by INRA SPE for the “Réseau de la biologie adaptative des pucerons” and the ANR “Aphicibles” project and the ANR-BBSRC System Biology Program METNET4-SyBio. Thanks are given to Hubert Charles for a critical reading of the manuscript.

¹ To have access to the different activities of the IAGC, please register at the mailing list at <http://www.aphidbase.com/>.

References

- [1] International Aphid Genomics Consortium, Genome sequence of the pea aphid *Acyrtosiphon pisum*, *PLoS Biology* (2010) 8(2):e1000313.
- [2] E.R. Mardis, The impact of next-generation sequencing technology on genetics, *Trends in Genetics* 24 (2008) 133–141.
- [3] D. Tagu, J.P. Klingler, A. Moya, J.C. Simon, Early progress in aphid genomics and consequences for plant-aphid interactions studies, *Molecular Plant Microbe Interactions* 6 (2008) 701–708.
- [4] J. Venter, M. Adams, E. Myers, P. Li, R. Mural, G. Sutton, H. Smith, M. Yandell, C. Evans, R. Holt, J. Gocayne, P. Amanatides, R. Ballew, D. Huson, J. Wortman, Q. Zhang, C. Kodira, X. Zheng, L. Chen, M. Skupski, G. Subramanian, P. Thomas, J. Zhang, G. Gabor Miklos, C. Nelson, S. Broder, A. Clark, J. Nadeau, V. Mckusick, N. Zinder, A. Levine, R. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. Heiman, M. Higgins, R.-R. Ji, Z. Ke, K. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. Merkulov, N. Milshina, H. Moore, A. Naik, V. Narayan, B. Neelam, D. Nusskern, D. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M.-H. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferreira, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. Mccawley, T. Mcintosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y.-H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. Abril, R. Guigo, M. Campbell, K. Sjolander, B. Karlak, A. Kejarival, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yoosheph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y.-H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Eparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, X. Zhu, The sequence of the human genome, *Science* 291 (2001) 1304–1351.
- [5] International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome, *Nature* 409 (2001) 860–921.
- [6] D. Pushkarev, N.F. Neff, S.R. Quake, Single-molecule sequencing of an individual human genome, *Nature Biotechnology* 27 (2009) 847–850.
- [7] B. Yngvadottir, D.G. MacArthur, H. Jin, C. Tyler-Smith, The promise and reality of personal genomics, *Genome Biology* 10 (2009) 237.
- [8] J.C. Simon, S. Stoeckel, A.C.C. Wilson, D. Tagu, Evolutionary and functional insights into reproductive strategies of aphids, *C. R. Biologies* 333 (2010) 488–496.
- [9] J. Peccoud, A. Ollivier, M. Plantegenest, J.-C. Simon, A continuum of genetic divergence from sympatric host races to species in the pea aphid complex, in: *Proceedings of the National Academy of Sciences of the United States of America* 106, 2009, pp. 7495–7500.
- [10] B. Sabater-Muñoz, F. Legeai, C. Rispe, J. Bonhomme, P. Dearden, C. Dossat, A. Duclert, J.P. Gauthier, D.G. Ducray, W. Hunter, P. Dang, S. Kambhampati, D. Martinez-Torres, T. Cortes, A. Moya, A. Nakabachi, C. Philippe, N. Prunier-Leterme, Y. Rahbé, J.C. Simon, D.L. Stern, P. Wincker, D. Tagu, Large-scale gene discovery in the pea aphid *Acyrtosiphon pisum* (Hemiptera), *Genome Biology* 7 (2006) R21.
- [11] K.D. Pruitt, T. Tatusova, W. Klimke, D.R. Maglott, NCBI Reference Sequences: current status, policy and new initiatives, *Nucleic Acids Research* 37 (2009) D32–36.
- [12] C.G. Elsik, A.J. Mackey, J.T. Reese, N.V. Milshina, D.S. Roos, G.M. Weinstock, Creating a honey bee consensus gene set, *Genome Biology* 8 (2007) R13.
- [13] D.G. Gilbert, Aphid and waterflea have a high rate of gene duplications compared to other arthropods, *PLoS ONE*, submitted.
- [14] S.D. Rider Jr., G. D.G. Srinivasan, R.S. Hilgarth, Chromatin remodeling proteins of the pea aphid, *Acyrtosiphon pisum*, *Insect Molecular Biology* 19 suppl 2 (2010) 201–214.
- [15] D.R.G. Price, K. Tibbles, S. Shigenobu, A. Smertenko, C.W. Russell, A.E. Douglas, E. Fitches, A.M.R. Gatehouse, J.A. Gatehouse, Sugar transporters of the major facilitator superfamily in aphids: from gene prediction to functional characterisation, *Insect Molecular Biology* 19 suppl 2 (2010) 97–112.
- [16] N.M. Gerardo, B. Altincicek, C. Anselme, et al., Immunity and defense in the pea aphid, *Acyrtosiphon pisum*, *Genome Biology* 2010, 11(2):R21.
- [17] S. Shigenobu, H. Watanabe, M. Hattori, Y. Sakaki, H. Ishikawa, Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera sp.* APS, *Nature* 407 (2000) 81–86.
- [18] A.C.C. Wilson, P.D. Ashton, F. Calevro, H. Charles, S. Colella, G. Febvay, G. Jander, P. Kushlan, S.J. Macdonald, J. Schwartz, G.H. Thomas, A.E. Douglas, Genomic insight into the amino acid relations of the pea aphid *Acyrtosiphon pisum* with its symbiotic bacterium *Buchnera aphidicola*, *Insect Molecular Biology* 19 suppl 2 (2010) 249–258.
- [19] F. Legeai, S. Shigenobu, J.P. Gauthier, J. Colbourne, C. Rispe, O. Collin, S. Richards, A.C.C. Wilson, D. Tagu, AphidBase: A centralized bioinformatic resource for annotation of the pea aphid genome, *Insect Molecular Biology* 19 suppl 2 (2010) 5–12.
- [20] J. Huerta-Cepas, M. Marcet-Houben, P.M.A. Moya, T. Gabaldón, The pea aphid phylome: a complete catalogue of evolutionary histories and arthropod orthology and paralogy relationships for *Acyrtosiphon pisum* genes, *Insect Molecular Biology* 19 suppl 2 (2010) 13–21.
- [21] J. Huerta-Cepas, A. Bueno, J. Dopazo, T. Gabaldón, PhylomeDB: a database for genome-wide collections of gene phylogenies, *Nucleic Acids Research* 36 (2008) D491–496.
- [22] M. Ollivier, F. Legeai, C. Rispe, Comparative analysis of the *Acyrtosiphon pisum* genome and EST-based gene sets from other aphid species, *Insect Molecular Biology* 19 suppl 2 (2010) 33–45.
- [23] S. Ohno, Evolution by gene duplication, Springer, New York, 1970.
- [24] N.A. Moran, P.H. Degnan, Functional genomics of *Buchnera* and the ecology of aphid hosts, *Molecular Ecology* 15 (2006) 1251–1261.
- [25] Symbiotic bacteria enable insect to use a nutritionally inadequate diet. Akman Gündüz E, Douglas AE. *Proc Biol Sci.* 2009 Mar 7;276(1658):987–91.
- [26] G.H. Thomas, J. Zucker, S.J. Macdonald, A. Sorokin, I. Goryanin, A.E. Douglas, A fragile metabolic network adapted for cooperation in the symbiotic bacterium *Buchnera aphidicola*, *BMC Systems Biology* 3 (2009) 24.
- [27] A. Nakabachi, S. Shigenobu, N. Sakazume, T. Shiraki, Y. Hayashizaki, P. Carninci, H. Ishikawa, T. Kudo, T. Fukatsu, Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium *Buchnera*, in: *Proceedings National Academy Science USA* 102, 2005, pp. 5477–5482.
- [28] D. Tagu, N. Prunier-Leterme, F. Legeai, J.P. Gauthier, A. Duclert, B. Sabater-Muñoz, J. Bonhomme, J.C. Simon, Annotated expressed sequence tags for studies of the regulation of reproductive modes in aphids, *Insect Biochemistry Molecular Biology* 34 (2004) 809–822.
- [29] A.C.C. Wilson, H.E. Dunbar, G.K. Davis, W.B. Hunter, D.L. Stern, N.A. Moran, A dual-genome microarray for the pea aphid, *Acyrtosiphon pisum*, and its obligate bacterial symbiont *Buchnera aphidicola*, *BMC Genomics* 7 (2006) 50.
- [30] G. Le Trionnaire, S. Jaubert, B. Sabater-Muñoz, A. Benedetto, J. Bonhomme, N. Prunier-Leterme, D. Martinez-Torres, J.C. Simon, D. Tagu, Seasonal photoperiodism regulates the expression of cuticular and signalling protein genes in the pea aphid, *Insect Biochemistry Molecular Biology* 37 (2007) 1094–1102.
- [31] G. Le Trionnaire, F. Francis, S. Jaubert-Possamai, J. Bonhomme, E. De Pauw, J. Gauthier, E. Haubruge, F. Legeai, N. Prunier-Leterme, J. Simon, S. Tanguy, D. Tagu, Transcriptomic and proteomic analyses of seasonal photoperiodism in the pea aphid, *BMC Genomics* 10 (2009) 456.
- [32] S. Hannenhalli, Eukaryotic transcription factor binding sites – modeling and integrative search methods, *Bioinformatics* 24 (2008) 1325–1331.
- [33] A. Frantz, M. Plantegenest, L. Mieuze, J.C. Simon, Ecological specialization correlates with genotypic differentiation in sympatric host-populations of the pea aphid, *Journal Evolution Biology* 19 (2006) 392–401.
- [34] C. Braendle, I. Friebe, M.C. Caillaud, D.L. Stern, Genetic variation for an aphid wing polyphenism is genetically linked to a naturally occurring wing polymorphism, in: *Proceedings of the Royal Society B-Biological Sciences* 272, 2005, pp. 657–664.
- [35] D.J. Hawthorne, S. Via, Genetic linkage of ecological specialization and reproductive isolation in pea aphids, *Nature* 412 (2001) 904–907.
- [36] A. Joshi, Behaviour genetics in the post-genomic era: From genes to behaviour and vice versa, *Current Science* 89 (2005) 1128–1135.
- [37] G.E. Robinson, Genomics: Beyond nature and nurture, *Science* 304 (2004) 397–399.

- [38] H. Ito, K. Fujitani, K. Usui, K. Shimizu-Nishikawa, S. Tanaka, D. Yamamoto, Sexual orientation in *Drosophila* is altered by the satori mutation in the sex-determination gene fruitless that encodes a zinc finger protein with a BTB domain, in: Proceedings of the National Academy of Sciences 93, 1996, pp. 9687–9692.
- [39] K.A. Osborne, A. Robichon, E. Burgess, S. Butland, R.A. Shaw, A. Coulthard, H.S. Pereira, R.J. Greenspan, M.B. Sokolowski, Natural behaviour polymorphism due to a cGMP-dependent protein kinase of *Drosophila*, *Science* 277 (1997) 834–836.
- [40] M.B. Sokolowski, *Drosophila*: Genetics meets behaviour, *Nature Reviews Genetics* 2 (2001) 879–890.
- [41] Y. Ben-Shahar, A. Robichon, M.B. Sokolowski, G.E. Robinson, Influence of gene action across different time scales on behaviour, *Science* 296 (2002) 741–744.
- [42] G.E. Robinson, R.D. Fernald, D.F. Clayton, Genes and social behaviour, *Science* 322 (2008) 896–900.
- [43] G. Kunert, E. Belz, J.C. Simon, W.W. Weisser, Y. Outreman, Differences in defensive behaviour between host adapted races of the pea aphid, *Ecology Entomology* 35 (2010) 147–154.
- [44] D.A. Stacey, M.D.E. Fellowes, Influence of temperature on pea aphid (*Acyrtosiphon pisum*) resistance to natural enemy attack, *Bulletin of Entomological Research* 92 (2002) 351–357.
- [45] A.F.G. Dixon, *Aphid ecology*, Chapman & Hall, London, 1998.
- [46] O.R.W. Sutherland, The role of crowding in the production of winged forms by two strains of the pea aphid, *Acyrtosiphon pisum*, *Journal of Insect Physiology* 15 (1969) 1385–1410.
- [47] O.R.W. Sutherland, The role of the host plant in the production of winged forms by two strains of the pea aphid, *Acyrtosiphon pisum*, *Journal of Insect Physiology* 15 (1969) 2179–2201.
- [48] G. Kunert, S. Otto, U.S.R. Rose, J. Gershenson, W.W. Weisser, Alarm pheromone mediates production of winged dispersal morphs in aphids, *Ecology Letter* 8 (2005) 596–603.
- [49] J.A. Pickett, L.J. Wadhams, C.M. Woodcock, The chemical ecology of aphids, *Annual Review of Entomology* 37 (1992) 67–90.
- [50] T. Eisner, M.R. Berenbaum, *Chemical Ecology: missed opportunities*, *Science* 295 (2002) 1973.
- [51] M.R. Berenbaum, Postgenomic chemical ecology: from genetic code to ecological interactions, *Journal of Chemical Ecology* 28 (2002) 873–896.
- [52] C. Tittiger, Functional genomics and insect chemical ecology, *Journal of Chemical Ecology* 30 (2004) 2335–2358.
- [53] Q. Gao, A. Chess, Identification of candidate *Drosophila* olfactory receptors from genomic DNA sequence, *Genomics* 60 (1999) 31–39.