Comparison of Gene Repertoires and Patterns of Evolutionary Rates in Eight Aphid Species That Differ by Reproductive Mode

M. Ollivier1,4, T. Gabaldón2, J. Poulain3, F. Gavory3, N. Leterme1, J.-P. Gauthier1, F. Legeai1, D. Tagu1, J. C. Simon1, and C. Rispe1,*

1INRA Rennes UMR BIO3P, Domaine de la Motte, Le Rheu, France
2Bioinformatics and Genomics Program, Centre for Genomic Regulation, Universitat Pompeu Fabra, Barcelona, Spain
3Genoscope and CNRS UMR 8030, Centre National de Séquençage, Evry, France
4Present address: Institut de Génomique Fonctionnelle de Lyon, Université Lyon 1, CNRS, INRA, Ecole Normale Supérieure de Lyon, Lyon, France.

*Corresponding author: E-mail: claude.rispe@rennes.inra.fr.

Accepted: 23 December 2011

Abstract

In theory, the loss of sexual reproduction is expected to result in the accumulation of deleterious mutations. In aphids, two main types of life cycle, cyclic and obligate parthenogenesis, represent respectively “sexual” and “asexual” reproductive modes. We used the complete pea aphid genome and previously published expressed sequence tags (ESTs) from two other aphid species. In addition, we obtained 100,000 new ESTs from five more species. The final set comprised four sexual and four asexual aphid species and served to test the influence of the reproductive mode on the evolutionary rates of genes. We reconstructed coding sequences from ESTs and annotated these genes, discovering a novel peptide gene family that appears to be among the most highly expressed transcripts from several aphid species. From 203 genes found to be 1:1 orthologs among the eight species considered, we established a species tree that partly conflicted with taxonomy (for Myzus ascalonicus). We then used this topology to evaluate the dynamics of evolutionary rates and mutation accumulation in the four sexual and four asexual taxa. No significant increase of the nonsynonymous to synonymous ratio or of nonsynonymous mutation numbers was found in any of the four branches for asexual taxa. We however found a significant increase of the synonymous rate in the branch leading to the asexual species Rhopalosiphum maidis, which could be due to a change in the mutation rate or to an increased number of generations implied by its change of life cycle.

Key words: reproductive mode, aphids, evolutionary rates, phylome, orthologs, EST.

Introduction

The reproductive mode of organisms, and more precisely their degree of commitment to recombination and sexual reproduction, is known to determine the level of genetic variation, how polymorphisms are distributed among individuals, populations, and in the genome, and in the long term, how the species may evolve (Charlesworth and Wright 2001; Glémin 2007). At the two extremities of variations in that trait are respectively sexual organisms (the majority of animal species) and asexual organisms, which have essentially lost recombination. Despite being the dominant reproductive mode, sexual reproduction is costly in the short term (Maynard Smith 1978; Otto and Lenormand 2002) but possesses two advantages regarding mutations. First, meiotic segregation and recombination facilitate the ability of natural selection to combine advantageous mutations (Muller 1932). Second, sex reduces the accumulation of deleterious mutations on sequences (Muller 1964). On the contrary, asexual species are unable to purge deleterious mutations that irreversibly tend to accumulate (Muller 1964). The predicted signature of accumulation of deleterious mutations is an increased rate of evolution at the amino acid level for genes subject to purifying selection (assuming that most amino acid replacements represent slightly deleterious mutations; Nachman 1998). Deleterious mutation accumulation
accumulation has indeed been evaluated in aphid endosymbiotic bacteria (Moran 1996; Funk et al. 2001), free-living bacteria (Andersson and Hughes 1996), protozoan (Bell 1988), fungi (Nygren et al. 2011), Daphnia species (Paland and Lynch 2006), and snails (Johnson and Howard 2007; Neiman et al. 2010). Comparisons of mutation accumulation in gene sequences have also been made for aphids (Normark 1999; Normark and Moran 2000), using one nuclear and one mitochondrial gene, and showing a slight excess of mutations on the former in some asexual taxa. Many of the recent studies concern mitochondrial coding sequences (CDS): Although mitochondrial genomes are essentially recombination free whether the host species is sexual or asexual, authors indeed argued that the tight linkage of mitochondrial and nuclear genomes in asexual taxa should result in similar effects for both genomes, that is, increased mutation accumulation (Birky and Walsh 1988; Paland and Lynch 2006; Neiman 2010). In contrast, we here chose to focus on nuclear CDS, expecting that a larger number of different genes might give us more power to detect changes in mutational patterns specific to asexual species and that nuclear genes should more directly reflect consequences from the suppression of recombination.

Aphids provide a choice model for studying the mutational effects of losing sexuality (Wilson et al. 2003). The typical life cycle of these insects is cyclic parthenogenesis: Over 1 year, aphids normally alternate between viviparous parthenogenesis (several generations without sex) and oviparous sexual reproduction (one generation). It has been shown that the single sexual generation results in genetic parameters (in terms of diversity and heterozygosity) equivalent to that of an organism that would reproduce only sexually (Delmote et al. 2002; Simon et al. 2002), so aphid species that reproduce by cyclical parthenogenesis can represent “sexual” organisms. A few aphid species are known to have lost entirely the sexual phase (they reproduce by obligate parthenogenesis; Moran 1992). In theory, these asexual taxa should then show traces in their genomes of shifts in molecular evolution rates, with, in particular, an accumulation of nonsynonymous mutations (Muller 1964).

The genome of the pea aphid, Acyrthosiphon pismum (Aphididae, Aphidinae, Macrosiphini), comprising more than 34,000 predicted genes, has been recently described (International Aphid Genomics Consortium 2010). The result of the pea aphid

### Materials and Methods

**Acyrthosiphon pismum Complete Genome and EST-Based Sequences from Other Species**

The complete genome sequencing and preliminary annotation of the pea aphid *A. pismum* have been recently achieved (International Aphid Genomics Consortium 2010). The resulting reference set of protein-coding sequences comprises 34,603 predicted genes. We designed our sequencing of ESTs to compare evolutionary rates among several related sexual and asexual taxa (if possible in the same genus), using previously published EST sequences for *Myzus persicae* (Figueroa et al. 2007; Ramsey et al. 2007) and *Aphis gossypii*, and acquiring sequences from other species for relevant comparisons. The lineages used for EST sequencing in *M. persicae* (Ramsey et al. 2007), *A. pismum*, and *Rhopalosiphum padi* (this study) were all cyclical parthenogens, as shown experimentally by their response to inducing conditions. For *Pemphigus syriacae*, the species is known to be always cyclically parthenogenetic (Pike et al. 2007). Therefore, all these lineages represent sexual taxa. In contrast, *A. gossypii* is known as essentially asexual worldwide (Carletto et al. 2009), although a few sexual forms have rarely been observed. In fact, a strict conversion to complete asexuality is always difficult to demonstrate, but we considered that this species did represent an asexual taxon. We here also obtained sequences for *Acyrtho-
siphon kondoi* and *Myzus ascalonicus*, which are described as asexual taxa worldwide (MacKay and Lamb 1988; Blackman et al. 2007).
and Eastop 2000). Another species chosen for this study, *Rhopalosiphum maidis*, is known to be asexual worldwide, with the exception of a very localized sexual population in the Himalayas (Remaudière and Naumann-Etienne 1991). Because the lineage sampled was from a European permanently parthenogenetic population, we therefore considered this lineage as "asexual." To complete the sexual/asexual comparisons within each genus, we also obtained sequences from the dominantly cyclical parthenogen M. persicae. As an outgroup, we used *P. syrrothecae*, which belongs to a different family (Pemphigidae) and which forms galls on poplar trees. For *A. kondoi*, a laboratory clone (initially collected from a field in Australia) was used. The *M. ascalonicus*, *R. maidis*, and *R. padi* strains were in each case a laboratory clone reared in Rennes (initially collected from a field in France)—the *R. padi* line was a clonal lineage known for its ability to produce sexual forms. For the gall-forming aphid *P. syrrothecae*, a single gall—with a shape characteristic of the species—was sampled from a poplar tree near Rennes, France (in 2009), and insects at various nymphal stages were collected—they were expected to descend from a single mother and then to represent a genetic clone.

**RNA Extraction and Sequencing of cDNA**

In the present study, five new cDNA libraries were then built for five aphid species: *A. kondoi*, *M. ascalonicus*, *R. maidis*, *R. padi*, and *P. syrrothecae*. Depending on the size of the individuals in each species, about 20–40 individuals (adults and various larval stages for each species) were collected and kept at −80 °C until use. Total RNA was extracted using the RNaseasy Plant Mini Kit (Qiagen, Hilden, Germany) in the RNA extraction buffer, following the manufacturer’s instructions. Plasmid cDNA libraries were constructed with the Creator Smart cDNA library construction kit (BD Biosciences Clontech, Palo Alto, CA). The bacterial glycerol stocks are archived at the INRA Rennes laboratory (France). Sequences were obtained at the sequencing center Genoscope (Eawy, France). The libraries were plated, arrayed robotically, and bacterial clones had their plasmid DNA amplified using phi29 polymerase. The plasmids were end sequenced at the sequencing center Genoscope using BigDye Terminator kits on Applied Biosystems 3730xl DNA Analyzers. The sequences were published in databanks and have the following accession numbers for each species: *A. kondoi* (FO000003 to FO017614 and FQ994572 to FQ999999), *R. maidis* (FQ976923 to FQ994571), *R. padi* (FO059144 to FO076577), *M. ascalonicus* (FO017615 to FO040556), and *P. syrrothecae* (FO040557 to FO059143).

**Reconstruction of Protein-Coding Genes from ESTs**

The same following protocol was used for all species studied here, that is, for the newly obtained sequence collections and for the ones that have been published previously (with methods used in Ollivier et al. 2010). First, ESTs corresponding to potential contaminants (mitochondrial proteins or ribosomal RNA) were filtered out; however, ESTs matching mitochondrial genes were assembled separately to provide further phylogenetic information. We could, for example, reconstruct the near entire cytochrome oxidase 1 (CO1) sequence for each species and check if it were identical to the sequence of the same species in a reference DNA barcoding database (Ratnasingham and Hebert 2007). The retained ESTs were assembled using Tgicl++ (Pertea et al. 2003). Unique consensus transcripts were then compared with predicted *A. pisum* proteins trough BlastX (Altschul et al. 1990). This information helped to identify potential homology and was used for CDS reconstruction with FramedP, a pipeline developed at the INRA of Toulouse (Schiex et al. 2003).

**Functional Annotation**

To annotate the five new EST gene sets (and also all the *A. pisum* genes), we used Blast2GO (Conesa et al. 2005) (http://www.blast2go.org/). Each sequence was blasted against the National Center for Biotechnology Information sequence. Gene Ontology (GO) terms were then mapped on the blast results using annotation files provided by the GO Consortium. For each GO categories, molecular function, cellular component, and biological process, we compiled GO terms of the same ontological level in each species. We compared the distribution of GO terms for each category between these six species.

**Phyloyme Reconstruction**

We reconstructed the complete collection of phylogenetic trees for all available genes of the eight aphid species using a pipeline similar to that used for the reconstruction of the human phylome (Huerta-Cepas et al. 2007) and the pea aphid phylome (Huerta-Cepas, Marcet-Houben, et al. 2010). We performed a blast against *A. pisum* predicted proteins (BlastP, e-value cutoff $<1 \times 10^{-10}$) for each of the seven EST-based partial genomes. Sequences that aligned with a continuous region longer than 50% of the query sequence were retained and aligned using Muscle 3.6 (Edgar 2004). The protein alignments were used to guide corresponding nucleic alignments and all columns with gaps were removed using trimAL (Capella-Gutiérrez et al. 2009) (http://trimal.cgenomics.org/). Phylogenetic trees were inferred using Neighbor-Joining (NJ) with scoredist distances as implemented in BioNJ (Gascuel 1997) and by maximum likelihood (ML) as implemented in PhyML v2.4.4 (Guindon and Gascuel 2003). A general time reversible (GTR) evolutionary model was used to construct all trees assuming a discrete gamma distribution model with four rate categories and invariant sites. The model was selected using jModeltest (Posada 2008).
Orthology Determination

Orthology relationships among aphid genes were inferred using the species overlap algorithm implemented in ETE (Huerta-Cepas, Dopazo, et al. 2010) using a species overlap score (SOS) of 0.0. In brief, this algorithm (Huerta-Cepas et al. 2007) uses the level of species overlap between the two branches of a given node to define a duplication (SOS ≥ 0.0) or a speciation (SOS = 0.0). After mapping all speciation events in a tree, all orthology relations can be predicted, according to the original definition of orthology (Fitch 1970). Often, for genes present in all species, we found several sequences representing one species. We checked manually the alignments to determine if these copies were true paralogs or could represent artifacts of EST assembly. The latter case was very likely when these potential copies had identical nucleotidic sequences, and these extra copies were suppressed from the data set (this allowed to increase the number of one-to-one orthologs).

Species Tree Reconstruction

We identified a total of 203 nuclear single-copy orthologs between the eight species, which were used to build the species tree. All alignments were concatenated in a super alignment of 82,950 base pairs. An ML tree was reconstructed as implemented in PhyML v2.4.4 (Guindon and Gascuel 2003). Parameters of the substitution model were determined by running jModeltest (Posada 2008) to test 56 different models of substitution. The best-fit model selected was the GTR + gamma + proportion invariant model, and a bootstrap analysis of 100 replicates was performed (Felsenstein 1985). NJ and Bayesian trees were also generated (not shown). MEGA4 (Kumar et al. 2008) was used to build the NJ tree, and a bootstrap analysis of 5,000 replicates was performed. We used MrBayes (Huelsenbeck and Ronquist 2001) to generate the Bayesian tree: 20,000 trees were sampled every 100 generation states totaling 2,000,000 generations. An additional set of 52 genes, found to be 1:1 orthologs in the data set but only absent from the outgroup species (P. syrothecae), was identified and used for comparisons of evolutionary rates and mutational patterns within the in-group.

Substitution Rates and Computational Estimation of Nonsynonymous Substitutions

Substitution rates were estimated for genes identified in the phylome analysis as 1:1 orthologs (203 nuclear genes present in all species), 52 additional nuclear genes present in all species except the outgroup, and finally 10 partial mitochondrial sequences. The methods for alignments were the same as described for the species tree reconstruction. We then estimated synonymous (dS) and nonsynonymous (dN) evolutionary rates using a codon-based model (CodeML from PAML 3.15; Yang 1997). The ratio dN/dS is commonly used as an indicator of variable evolutionary pressures among protein-coding genes: Low ratios are typical of highly constrained sequences, whereas values close to unity reflect relaxed selection and ratio above unity, positive selection. We used a free-ratio model to evaluate potential differences among branches (particularly those corresponding to related sexual and asexual taxa). We also obtained estimates of the numbers of nonsynonymous substitutions on each branch by parsing the “rst” result file, which evaluates the probability of ancestral states and of mutations occurring along the branches (PAML 3.15; Yang 1997).

Results and Discussion

Unique Transcripts Catalogs Based on ESTs

ESTs were sequenced for five aphid species, A. kondoi, M. ascalonicus, R. maidis, and P. syrothecae. We obtained the following numbers of ESTs after filtration by quality and removal of potential contaminants (in parentheses we indicate numbers of tentative unique transcripts): 19,425 (7,011) for A. kondoi; 20,256 (5,855) for M. ascalonicus; 16,459 (4,558) for R. padi; 15,807 (6,670) for R. maidis; and 16,073 (8,488) for P. syrothecae (fig. 1). Between 56% and 65% of tentative unique transcripts were composed of only one EST, a figure comparable with what has been found in previous studies (Sabater-Mun˜oz et al. 2006; Ramsey et al. 2007; Ollivier et al. 2010). A large fraction of tentative unique transcripts did not have a hit in the UniProt database (61–75%), nor in the A. pisum reference gene set (43–66%). Either these sequences corresponded to genes unique to aphids in general or unique to some aphid species, or they represented noncoding regions (untranslated transcribed regions) (Whitfield et al. 2002; Sabater-Mun˜oz et al. 2006)—the latter explanation was most likely as their was a strong correlation between having no hit in A. pisum and no CDS reconstructed, as seen below. We sorted the transcripts by the number of sequences in the contig supporting that transcript, and reported the first 21 of these in table 1. For A. kondoi, the first contig (2,521 ESTs) corresponded to a Buchnera structural RNA (tmRNA). Also, one of the top unique transcripts corresponded to mitochondrial rRNA (large subunit): We had included a filter of ESTs matching mitochondrial rRNAs from aphids but the highly stringent criteria used prevented many ESTs matching this sequence to be filtered out. Other top unique transcripts matched ribosomal proteins and muscle actin (often highly expressed genes), or had similarity to hypothetical proteins from A. pisum. For M. ascalonicus, the top contig did not match any gene in A. pisum, nor did it have any similarity to CDS or non-CDS in GenBank, whereas in R. padi, the top contig (3,532 ESTs) corresponded to a viral polyprotein and was then most likely a contaminating RNA from a virus contained in the aphid body. For M. ascalonicus, R. padi, R. maidis, and P. syrothecae, we found patterns globally similar to A. kondoi: 1) presence of
contaminant material (16S rRNA, tmRNA from Buchnera); 2) presence of several ribosomal proteins, muscle actin, the elongation factor 1-alpha, a putative ADP/ATP translocase, heat shock proteins, and so on; and 3) presence of genes matching hypothetical proteins from *A. pisum*. Interestingly, several of the top unique transcripts in three species (*A. kondoi*, *M. ascalonicus*, and *P. spyrothecae*) matched peptides from *A. pisum* with no known similarity outside aphids (e.g., ACYPI30077 and others). These genes belong to a family of short peptides (38 or 39 residues) also well represented in ESTs from the pea aphid. We manually checked and corrected the annotations for this gene family in the *A. pisum* genome, resulting in the identification of 29 copies, which are organized in several clusters of a few copies on different scaffolds (due to the short CDS and to this organization in clusters most of the initial automatic annotations were incorrect, with chimeras being formed between adjacent copies, or with no annotation at all). Because the alignment length was short, phylogenetic analyses yielded trees in which most nodes were not robust; we however evaluated pairwise evolutionary rates and found nonsynonymous to synonymous rates falling in the 0.20–0.30 range, well below unity and not suggestive of positive selection (results not shown).

We also counted the gene copy numbers from all other species (retaining only transcripts supported by at least two ESTs) and found that this number ranged from 6 in *A. gossypii* to 47 in *M. ascalonicus* (table 2); the three species belonging to Aphidini seemed to be characterized by lower copy numbers, compared with either the Macrosiphini or the outgroup species, *P. spyrothecae*. Although our data were based only on partial transcriptomes, this could suggest a reduction of gene copy number in an ancestor of Aphidini. This new gene family has then been identified in all the species studied here and represents some of the most abundant transcripts in several of them. Given the absence of similarity outside aphids, it is difficult to determine the potential function of this highly expressed gene family, which will then deserve further specific studies. For example, microarrays or transcriptomic approaches (as RNA-seq) could point to quantitative variations of these genes in different tissues and conditions, and also could show with which known proteins they are coregulated. This should help determine the pathways in which members of this novel family are involved.

**Identification of CDS and Non-CDS**

For these five species, we predicted a CDS for 45–65% of the unique transcripts. A very small fraction (0.5–2%) of unique transcripts with a hit in the *A. pisum* gene set had no CDS predicted. These consisted of very short sequences containing a short coding region or with several frameshifts. Overall, we could reconstruct 4,113, 3,502, 3,761, 4,275, and 2,353 different CDS for *A. kondoi*, *M. ascalonicus*, *P. spyrothecae*, *R. maidis*, and *R. padi*, respectively (fig. 1): About 10% of unique transcripts are predicted to contain a complete CDS (fig. 2). In addition, with the same methods, we obtained 6,652 and 15,810 CDS for *M. persicae* and *A. gossypii*, respectively. Across the different collections of CDS for the different aphid species, we found between 18% and 35% of sequences with a hit in *A. pisum* genes but no hit in UniProt. This suggests that a relatively high fraction of these transcriptomes correspond to genes apparently found only in aphids.

**Functional Annotation**

The numbers of sequences annotated with Blast2GO were 1,641, 1,520, 1,169, 1,576, and 785 for *A. kondoi*, *M. ascalonicus*, *P. spyrothecae*, *R. maidis*, and *R. padi*, respectively.

---

**Fig. 1.**—Number of ESTs (dark), unique transcripts (gray), and reconstructed CDS (open) for five aphid species: *Acyrthosiphon kondoi*, *Myzus ascalonicus*, *Rhopalosiphum maidis*, *Rhopalosiphum padi*, and *Pemphigus spyrothecae*. 

Downloaded from https://academic.oup.com/gbe/article-abstract/4/2/155/557881 on 29 July 2018
| Contigs               | Length | ESTs Number | Corresponding Acyrthosiphon pisum Gene | Similarity                      |
|----------------------|--------|-------------|----------------------------------------|---------------------------------|
| **Acyrthosiphon kondoi** |        |             |                                        |                                 |
| CL2Contig2           | 323    | 2,521       | ACYPI30077                              | tmRNA (Buchnera)                |
| TCL6_1_CONTig5       | 423    | 287         | ACYPI24917                              | Hypothetical protein ("sp" family) |
| CL4Contig1           | 416    | 241         | ACYPI30077                              | Hypothetical protein ("sp" family) |
| TCL6_203             | 710    | 235         | ACYPI30077                              | Hypothetical protein ("sp" family) |
| TCT6_1_CONTig2       | 419    | 180         | ACYPI30077                              | Hypothetical protein ("sp" family) |
| TCT6_86              | 349    | 163         | ACYPI20392                              | Hypothetical protein ("sp" family) |
| CL1Contig7           | 353    | 152         | ACYPI30077                              | 16S rRNA (mitochondrial)        |
| CL1Contig5           | 359    | 140         | ACYPI20392                              | Hypothetical protein ("sp" family) |
| CL12Contig1          | 704    | 137         | ACYPI004796                             | Hypothetical protein            |
| TCT6_4               | 4,559  | 133         | ACYPI58320                              | Hypothetical protein ("sp" family) |
| CL5Contig3           | 357    | 131         | ACYPI58108                              | Hypothetical protein ("sp" family) |
| TCT6_307             | 275    | 97          | ACYPI063423                             | Ribosomal protein L41           |
| CL38Contig1          | 1,456  | 93          | ACYPI58228                              | Muscle actin                    |
| NRCL4_14             | 466    | 92          | ACYPI30077                              | Hypothetical protein ("sp" family) |
| CL1Contig6           | 471    | 88          | ACYPI30077                              | Hypothetical protein ("sp" family) |
| NRCL4_18             | 389    | 83          | ACYPI30077                              | Hypothetical protein ("sp" family) |
| NRCL3_81             | 310    | 76          | ACYPI58128                              | Hypothetical protein ("sp" family) |
| CL159Contig1         | 1,295  | 74          | ACYPI57336                              | Ribosomal protein L18           |
| TCT6_517             | 897    | 72          | ACYPI000294                             | Hypothetical protein            |
| TCT6_68              | 371    | 72          | ACYPI30077                              | Hypothetical protein ("sp" family) |
| **Myzus ascalonicus** |        |             |                                        |                                 |
| CL5Contig2           | 585    | 509         | No hit                                 | No similarity                   |
| CL4Contig1           | 700    | 386         | No hit                                 | 16S rRNA (mitochondrial)        |
| TCT5_77              | 536    | 358         | ACYPI30077                              | Hypothetical protein ("sp" family) |
| TCT5_39              | 536    | 273         | ACYPI30077                              | Hypothetical protein ("sp" family) |
| NRCL4_14             | 466    | 92          | ACYPI30077                              | Hypothetical protein ("sp" family) |
| CL1Contig6           | 471    | 88          | ACYPI30077                              | Hypothetical protein ("sp" family) |
| NRCL4_18             | 389    | 83          | ACYPI30077                              | Hypothetical protein ("sp" family) |
| NRCL3_81             | 310    | 76          | ACYPI58128                              | Hypothetical protein ("sp" family) |
| CL159Contig1         | 1,295  | 74          | ACYPI57336                              | Ribosomal protein L18           |
| TCT6_517             | 897    | 72          | ACYPI000294                             | Hypothetical protein            |
| TCT6_68              | 371    | 72          | ACYPI30077                              | Hypothetical protein ("sp" family) |
| **Rhopalosiphum maidis** |      |             |                                        |                                 |
| TCT6_216             | 2,061  | 133         | ACYPI58208                              | Elongation factor 1-alpha       |
| TCT6_236             | 750    | 102         | ACYPI063423                             | Ribosomal protein L41           |
| CL79Contig1          | 1,703  | 80          | ACYPI58227                              | Muscle actin                    |
| CL6Contig1           | 1,521  | 75          | ACYPI58228                              | Muscle actin                    |
| CL10Contig1          | 472    | 138         | ACYPI24917                              | Hypothetical protein ("sp" family) |
| TCT6_15              | 391    | 137         | ACYPI30077                              | Hypothetical protein ("sp" family) |
| CL1Contig10          | 629    | 130         | ACYPI30077                              | Hypothetical protein ("sp" family) |
| CL283Contig1         | 2,202  | 126         | ACYPI58208                              | Elongation factor 1-alpha       |
| TCT6_26              | 318    | 116         | ACYPI20392                              | Hypothetical protein ("sp" family) |
| TCT6_17              | 528    | 104         | ACYPI20392                              | Hypothetical protein ("sp" family) |
| TCT6_343             | 1,009  | 100         | ACYPI0006075                            | Ribosomal protein L19           |
| TCT5_69              | 356    | 100         | ACYPI24917                              | Hypothetical protein ("sp" family) |
| **Ollivier et al.**   |        |             |                                        |                                 |
| **GBE**              |        |             |                                        |                                 |

Table 1
Description of Top 21 Contigs (Contigs with the Highest EST Support) in Each Data Set
respectively. For the *A. pisum* gene reference set, 7,886 sequences could be annotated in the same way. We compiled for these species the 10 GO terms describing the most represented molecular functions and then combined all these terms across all the species studied (fig. 3). Because many genes have multiple GO terms assigned for them, many of which are parents or daughters of other terms, we decided to represent the GO terms belonging to the same ontological level (level 3). The two sources of sequences, genomic (*A. pisum*) and ESTs (for other species), may

| Contigs         | Length | ESTs Number | Corresponding *Acrystosiphon pisum* Gene | Similarity                        |
|-----------------|--------|-------------|----------------------------------------|----------------------------------|
| TCL6_40         | 1,047  | 55          | ACYPI005092                             | Ribosomal protein S6             |
| TCL5_155        | 740    | 55          | ACYPI06773                              | Ribosomal protein L26            |
| CL2Contig1      | 1,304  | 53          | ACYPI001035                             | Protein take out                 |
| TCL2_67         | 766    | 53          | ACYPI006075                             | Ribosomal protein L19            |
| TCL6_254        | 1,044  | 52          | ACYPI003993                             | Ribosomal protein L5             |
| TCL6_348        | 779    | 51          | ACYPI08369                              | Cuticular protein 49Aa           |
| **Rhopalosiphum padi** |   |              |                                        |                                  |
| CL1Contig1      | 9,396  | 3,532       | No hit                                  | Virus clone RhPV6                |
| CL129Contig1    | 1,270  | 1,464       | No hit                                  | 16S rRNA (mitochondrial)         |
| TCL6_434        | 417    | 133         | ACYPI063423                             | Ribosomal protein L41            |
| TCL6_305        | 1,336  | 130         | ACYPI08228                              | Muscle actin                     |
| CL15Contig1     | 1,370  | 85          | ACYPI08208                              | Elongation factor 1-alpha        |
| TCL6_52         | 561    | 68          | ACYPI00896                              | Ribosomal protein L44            |
| TCL6_5          | 750    | 66          | ACYPI06857                              | Ribosomal protein L18            |
| CL19Contig1     | 1,509  | 61          | ACYPI08050                              | Putative ADP/ATP translocase      |
| CL235Contig1    | 537    | 59          | ACYPI06754                              | Ribosomal protein L36            |
| CL121Contig1    | 691    | 57          | ACYPI005604                             | Non structural polyprotein       |
| CL264Contig1    | 813    | 52          | ACYPI06766                              | Ribosomal protein L10            |
| TCL5_67         | 592    | 52          | ACYPI000455                             | Ribosomal protein L23e           |
| TCL6_384        | 1,056  | 48          | ACYPI00030                              | ATP synthase c-subunit           |
| TCL6_218        | 785    | 47          | ACYPI006075                             | Ribosomal protein L23e           |
| TCL6_380        | 758    | 47          | ACYPI001578                             | Ribosomal protein S8e            |
| TCL6_356        | 749    | 47          | ACYPI000783                             | Ribosomal protein S9             |
| CL205Contig1    | 831    | 44          | ACYPI06769                              | Ribosomal protein L15            |
| TCL6_36         | 1,003  | 43          | ACYPI07262                              | Cuticular protein (RRI1)         |
| CL118Contig1    | 641    | 43          | ACYPI06765                              | Ribosomal protein L9             |
| TCL6_220        | 479    | 43          | ACYPI00048                              | Ribosomal protein S28            |
| **Pemphigus sypyrotecae** |   |              |                                        |                                  |
| TCL6_1_Contig3  | 1,379  | 458         | ACYPI00030                              | ATP synthase c-subunit like       |
| CL220Contig1    | 1,425  | 423         | No hit                                  | 16S rRNA (mitochondrial)         |
| CL13Contig1     | 2,103  | 182         | No hit                                  | Hslu and lbp (*Buchnera*)        |
| CL14Contig1     | 341    | 149         | ACYPI03077                              | Hypothetical protein (“sp” family) |
| TCL6_353        | 351    | 145         | ACYPI03077                              | Hypothetical protein (“sp” family) |
| CL4Contig1      | 1,127  | 121         | No hit                                  | Likely bacterial contaminant      |
| TCL6_8_Contig2  | 385    | 120         | ACYPI03077                              | Hypothetical protein (“sp” family) |
| NRCL4_1         | 291    | 114         | No hit                                  | tmRNA (*Buchnera*)               |
| TCL2_19         | 208    | 95          | No hit                                  | No similarity                    |
| TCL6_259        | 397    | 89          | ACYPI03077                              | Hypothetical protein (“sp” family) |
| CL6Contig1      | 917    | 88          | No hit                                  | No similarity                    |
| CL2Contig1      | 594    | 75          | ACYPI009263                             | Similar to Hsp60                 |
| TCL6_121        | 1,390  | 74          | ACYPI08228                              | Muscle actin                     |
| TCL6_446        | 1,465  | 69          | No hit                                  | Hypothetical protein             |
| TCL6_412        | 644    | 57          | No hit                                  | No similarity                    |
| TCL6_296        | 661    | 54          | ACYPI009454                             | Ribosomal protein S24            |
| CL5Contig1      | 542    | 52          | No hit                                  | No similarity                    |
| TCL2_9          | 310    | 45          | ACYPI03077                              | Hypothetical protein (“sp” family) |
| TCL2_71         | 234    | 39          | No hit                                  | No similarity                    |
| TCL6_70         | 998    | 35          | No hit                                  | No similarity                    |

**Note.—** Several of these contigs in *A. kondoi*, *M. ascalonicus*, and *P. sypyrotecae* matched a short-peptide gene family (“sp” family), containing 29 genes (including ACYPI03077) in the pea aphid genome, with no known similarity outside aphids.
result in biased representation of genes in the latter: This can be seen in the overrepresentation of the “structural constituent of ribosome” category (mostly, ribosomal proteins) for EST-based gene sets. However, overall, the distributions of GO terms were similar between A. pism genes and genes from other species, with the exception of P. sphyrothecae: Indeed, several GO categories were either over- or underrepresented in that species. This may be explained by the fact that P. sphyrothecae is relatively distant phylogenetically from all other species, or because it has a specific gall-feeding life style, which is likely to affect expression globally. Additionally, differences in gene content (with different patterns of gene family expansions or gene loss) could account for that difference.

Reconstruction of the Aphid Phylome and Orthology Relationships across Aphid Genomes

The reconstructed phylome for the present multispecific gene collection contained 14,345 phylogenetic trees. To determine orthology, we parsed this aphid phylome using the ETE tool (Huerta-Cepas, Dopazo, et al. 2010). A total of 203 one-to-one orthologs in the eight aphid species supported the same topology (fig. 4). As expected, the two Rho- palosiphum species formed a monophyletic group, and the Aphi- din tribe (Rhopalosiphum species and A. gossypii) was also strongly supported. The remaining species (all belonging to the Macrosiphini tribe) also formed a strongly supported group, within which the two Acrithosiphon species were grouped together. But a major surprise was that M. ascalonicus did not group with M. persicae and instead appeared as basal to the three other species in the same tribe.

| Species                  | Tribe/Family | Copy Number of “sp” Family |
|-------------------------|-------------|----------------------------|
| Acrithosiphon pism       | Macrosiphini| 29                         |
| Acrithosiphon kondoi     | Macrosiphini| 22                         |
| Myzus persicae           | Macrosiphini| 23                         |
| Myzus ascalonicus        | Macrosiphini| 47                         |
| Rhopalosiphum padi       | Aphi- din   | 9                          |
| Rhopalosiphum maidis     | Aphi- din   | 9                          |
| Aphis gossypii           | Aphi- din   | 6                          |
| Pemphigus sphyrothecae   | Pemphiginae | 21                         |

Note.—For A. pism, number of copies that we annotated in the genome (version 1). For other species, number of different EST-based sequences (different tentative CDS).

Phylogenetic Species Tree

ML, Bayesian, and NJ trees for the concatenated sequences of 203 one-to-one orthologs in the eight aphid species supported the same topology (fig. 4). As expected, the two Rhopalosiphum species formed a monophyletic group, and the Aphidini tribe (Rhopalosiphum species and A. gossypii) was also strongly supported. The remaining species (all belonging to the Macrosiphini tribe) also formed a strongly supported group, within which the two Acrithosiphon species were grouped together. But a major surprise was that M. ascalonicus did not group with M. persicae and instead appeared as basal to the three other species in the same tribe.

Table 2

| Species                  | Tribe/Family | Copy Number of “sp” Family |
|-------------------------|-------------|----------------------------|
| Acrithosiphon pism       | Macrosiphini| 29                         |
| Acrithosiphon kondoi     | Macrosiphini| 22                         |
| Myzus persicae           | Macrosiphini| 23                         |
| Myzus ascalonicus        | Macrosiphini| 47                         |
| Rhopalosiphum padi       | Aphi- din   | 9                          |
| Rhopalosiphum maidis     | Aphi- din   | 9                          |
| Aphis gossypii           | Aphi- din   | 6                          |
| Pemphigus sphyrothecae   | Pemphiginae | 21                         |

Table 3

| Mitochondrial Genes     | Full Gene Length in Base Pairs (from the Complete Mitochondrial Genome Sequence from Schizaphis graminum, an Aphid Species) |
|-------------------------|------------------------------------------------------------------------------------------------------------------|
| CO1                     | 1,531 1,500                                                                                                         |
| CO2                     | 669 651                                                                                                              |
| ATPF08                  | 165 —                                                                                                               |
| ATPF06                  | 651 651                                                                                                              |
| CO3                     | 783 753                                                                                                              |
| ND2                     | 351 228                                                                                                              |
| ND5                     | 1,113 471                                                                                                            |
| ND4                     | 1,309 669                                                                                                            |
| ND4L                    | 1,308                                                                                                                |
| ND6                     | 288 —                                                                                                                |
| CYT8                    | 1,113 1,080                                                                                                           |
| ND1                     | 927 864                                                                                                              |
| ND2                     | 975 690                                                                                                              |

Note.—A fraction of the sequenced ESTs matched mitochondrial sequences in all species, allowing us to reconstruct partial CDS in all species for 10 of these genes, with high EST support. The resulting alignment length for each gene—with no gaps—is shown in base pairs (overall 67.5% of the mtDNA CDS was therefore obtained for these eight species).

Adding the 52 in-group orthologs, we analyzed the 255 one-to-one orthologs, focusing on genes for which all Macrosiphini species corresponded to mitochondrial DNA, in particular, the CO1 sequence, which has been widely used as a molecular barcode to distinguish species (Hebert et al. 2003; table 3). Our reconstructed near entire CO1 sequences were identical (or near identical) to CO1 sequences from DNA barcode databases for all species, including M. ascalonicus. This recently described species, which was unknown until the 1940s (Doncaster 1946), has been placed into the Myzus genus. Obviously, our phylogenetic data conflict with taxonomic classification and suggest that the few morphological characters relating it to the Myzus genus could result from convergent evolution, or simply that the taxonomical placement of this species needs revision. Because of the strong molecular phylogenetic support for M. ascalonicus as basal to other Macrosiphini species, we retained this hypothesis in the analyses of mutational patterns: We then compared evolutionary rates in the M. ascalonicus branch not just with M. persicae but also with the two sexually reproducing Macrosiphini (M. persicae and A. pism) from this data set.

Compared Rates of Substitution between Sexual and Asexual Taxa

To test theoretical predictions on the accumulation of non-synonymous mutations in asexual lineages, we compared...
rates of substitution (dN, dS, and the ratio dN/dS) between sexual and asexual taxa. We did so for the 255 genes that were identified as 1:1 orthologs through our phylogenomic analyses, using Wilcoxon signed rank test, and for 10 partial mitochondrial genes obtained also from the ESTs (table 4). For instance, we compared the two *Acyrthosiphon* and the two *Rhopalosiphum* species, assumed to represent (relatively) closely related sexual/asexual taxa. For *A. pisum* (sexual) versus *A. kondoi* (asexual), we detected no significant difference in any of these types of substitution rates both for nuclear and for mitochondrial genes. Also, the total of all nonsynonymous mutations for the 255 nuclear genes was 239 for *A. pisum* and 221 for *A. kondoi*—numbers of mutations were not significant with the Wilcoxon test.

This suggests that there has been no effect of the loss of sexuality on rates in *A. kondoi*, in these sets of genes. For *R. padi* (sexual) versus *R. maidis* (asexual), no significant differences in the nonsynonymous rate was found, but a significant difference was found for synonymous rates (*P* = 8.10^−9, Wilcoxon test) of nuclear genes; *R. maidis* tended to show higher dS rates, the medians being 0.021 and 0.035 for *R. padi* and *R. maidis*, respectively. We did also find a significant difference in dN/dS rates (*P* = 0.005, Wilcoxon test), which tended to be higher in *R. padi* (however, the medians were identical in both species and equal to zero). The difference appears to result from relatively few genes, where one or very few nonsynonymous mutations were seen in each species, whereas only the *R. maidis* branch also had

**Fig. 2.**—Reconstruction of CDS from unique transcripts for five aphid species. Dashed lines, percentage of unique transcripts with no predicted CDS; dark, partial CDS in both 5’ and 3’; light gray, partial CDS in 5’; dark gray, partial CDS in 3’; open, predicted complete CDS.

**Fig. 3.**—GO annotations (ontological level 3) for the genes of six aphid species. Frequencies of GO category among annotated predicted genes (*Acyrthosiphon pisum*) or unique transcripts (all other species) on the y axis. For each GO category, the species are, from left to right, *A. pisum*, *Acyrthosiphon kondoi*, *Myzus ascalonicus*, *Rhopalosiphum maidis*, *Rhopalosiphum padi*, and *Pemphigus spyrothecae*.
synonymous substitutions (R. padi having none): This resulted in very high estimates of the dN/dS ratio for the genes in R. padi lineage. Yet, overall, the total numbers of mutations were 78 (R. padi) and 89 (R. maidis)—this difference being nonsignificant. Therefore, there is clearly no sign of accumulation of nonsynonymous mutations in either of these lineages but rather an unexpected difference in synonymous rates. Because synonymous changes are normally close to neutral, synonymous rates are assumed to be relatively constant within biological groups (Kimura 1968; Graur and Li 2000) and to depend mostly on the mutation rate. A possible explanation for this observation would be that the

Table 4
Median of Evolutionary Rate Parameters and Number of Nonsynonymous Mutations for n = 255 Nuclear Coding and n = 10 Mitochondrial Coding Genes, Based on Complete Genomic Information (Acyrthosiphon pisum) or Transcriptomes (Other AphidSpecies)—Values for Asexual Taxa in Bold

| Asexual/Sexual Taxa | Median dN/dS | Median dN | Median dS | No. of NS Mutations |
|---------------------|--------------|-----------|-----------|--------------------|
| Nuclear CDS (1:1 orthologs in the in-group, n = 255) | | | | |
| Rhopalosiphum maidis/Rhopalosiphum padi | 0.000/0.000, P = 0.01 | 0.000/0.000, ns | 0.035/0.021, P < 10⁻⁸ | 89/78, ns |
| Acyrthosiphon kondoi/Acyrthosiphon pisum | 0.000/0.000, ns | 0.000/0.000, ns | 0.041/0.046, ns | 221/239, ns |
| Aphis gossypii/R. padi | 0.003/0.000, ns | — | — | — |
| Myzus ascalonicus/Myzus persicae + A. pisum | 0.000/0.000, ns | — | — | — |
| Mitochondrial genes (n = 10) | | | | |
| R. maidis/R. padi | 0.006/0.004, ns | 0.004/0.005, ns | 0.722/0.782, ns | 30/24, ns |
| A. kondoi/A. pisum | 0.018/0.010, ns | 0.005/0.009, ns | 0.718/0.314, ns | 34/42, ns |
| A. gossypii/R. padi | 0.009/0.004, ns | — | — | — |
| M. ascalonicus/M. persicae + A. pisum | 0.018/0.010, ns | — | — | — |

NOTE.—NS, not significant. Rates were estimated by a free-ratio model on each branch, and are here compared for sexual–asexual species pair for each gene (next to medians, result of a signed rank Wilcoxon test). The branch estimates for M. ascalonicus were compared with the mean of terminal branches of sexual taxa in the same subfamily (Macrosiphini) and A. gossypii was compared with a sexual species from a different genus but from the same subfamily (Aphidini); in both cases, only dN/dS ratios were then compared. For the two other comparisons, dN and dS, their ratio dN/dS, and numbers of NS mutations were compared. For numbers of NS mutations, totals combined for all the genes are shown.
mutation rate has specifically increased in *R. maidis*. Alternatively, we propose that the loss of sexual reproduction in this species potentially increased the number of generations per unit of time. Indeed, the single sexual generation in *R. padi* in temperate regions covers nearly half of the year (Dixon 1976), whereas *R. maidis* can achieve several asexual generations during the wintertime, likely resulting in a larger number of generations per year. Therefore, this could result in a faster evolutionary rate, which could be more apparent for the less constrained synonymous changes as compared with the non-synonymous changes. For mitochondrial genes, no significant differences between *R. maidis* and *R. padi* were found in any of the ratios, although a slight excess of total nonsynonymous mutations (30 vs. 24) was found in the asexual/sexual comparison, which went in the same direction as nuclear genes. This however does not explain why no such acceleration in synonymous rates was observed for the other asexual taxa studied here, and for which the same argument could be advanced. Possibly, the rate increase could depend on the age of the loss of sex, which is likely to be the most recent for *R. maidis*, as it still has a sexual population in its region of origin, in central Asia (Remaudier and Naumann-Etienne 1991). We finally compared the dN/dS ratio in the *M. ascalonicus* branch with the average of the two sexual taxa from the Macrosiphini, and no significant difference was found. No comparisons were made for dN or dS rates because the branches covered different times of divergence, which might bias the result. We also compared *A. gossypii* with the sexual taxon in the same tribe (*R. padi*), and again, no significant difference in dN/dS was found. Therefore, for these other two asexual taxa, *M. ascalonicus* and *A. gossypii*, no acceleration of the dN/dS ratio was detected in this set of genes. For these comparisons, mitochondrial genes also did not show significant differences in dN/dS ratios among asexual and sexual taxa. We noted, however, that medians of this parameter tended to be higher in the asexual than in the sexual taxa, but this effect seemed to result from lower values of dS. The estimates of dS for the different branches and for the 10 mitochondrial genes were relatively high and widely variable, so these estimates are likely to have large standard errors. We therefore conclude that there is no sign of increased rates of accumulation of mutations in the mitochondrial sequences for *A. gossypii* and *M. ascalonicus*.

**Conclusions**

**Gene Reconstruction from ESTs, Specificities of the Transcriptomes, and Novel Gene Family**

We have reconstructed CDS in seven aphid species: These sequences were often partial; yet, they constitute a significant sample of the coding genome comprising several thousands of CDS in each species. The fact that we obtained these sequences from similar stages and conditions allowed us to compare the contents of these transcriptomes. Two of the most interesting conclusions we could draw from these comparisons and that will deserve further explorations are thus: First, *P. sprogthecae*, the outgroup species, which has a different life style as it lives in galls, showed a distinct profile in expression (GO terms comparisons). It will be interesting to evaluate the impact of living in a gall on expression profiles, using larger data sets at different stages of the life cycle. Second, the compared genome or transcriptomes from eight aphid species helped identify a novel gene family of short peptides. In several of the species, some of these genes were among the most highly expressed genes. Through reannotation of the pea aphid genome for this family, we found 29 gene copies, although ESTs from other species suggest rather variable copy numbers. It will therefore be interesting to further explore the functional role of this gene family.

**Quality of the Sequences and Orthology Assessment Issues**

We determined the evolutionary relationships among thousands of CDS in the different aphid species. To this end, we adapted a phylome reconstruction pipeline (normally used for complete genomes) to this data set. This helped us determine some limitations and possible caveats of EST-based sequences. First, some of the sequences (usually those that had the weakest EST support, sometimes a single EST) could contain errors, which were not entirely corrected through the CDS reconstruction process. These could result in frame-shifts, which were seen by manual inspection of alignments. Second, EST assembly sometimes resulted in multiple contigs that contained an identical CDS. Rather than paralogs (as they were categorized in the automatic phylome pipeline), we therefore reclassified these identical sequences as being the same gene and obtained a much higher number of 1:1 orthologs. Third, data sets comprising only a subset of the species could have been studied separately to provide further gene sequence comparisons. We however identified many difficulties associated with this objective. Often, the main problem was orthology assessment. When studying groups of sequences limited to a few species, we often detected anomalies in trees or evolutionary distances that were consistent with problems of orthology assignment (some EST-based sequences corresponding to out-paralogs). This is a significant risk with partial gene collections, especially given that aphid genomes are rich in duplicated genes (Huerta-Cepas, Marcet-Houben, et al. 2010). For that reason, we decided to analyze evolutionary rates only on a “golden” set of 1:1 orthologs, for which all alignments were checked manually.

**Influence of the Reproductive Mode on the Evolution of Genes**

In this study, our aim was to test theories predicting increased numbers of nonsynonymous mutations for asexual...
species. Our results did not support this prediction for any of the four species we considered as asexual. Although we studied a relatively large set of CDS (255), it may be that this set of genes represents a biased sample of the genome. Indeed, these genes tend to correspond to ubiquitous essential genes, which should be under stronger-than-average selective constraints, so that deleterious mutations might be strongly selected against in all species. However, the range of values in evolutionary rates for these genes remained large enough that at least some of the genes should reflect a global decrease in the efficacy of selection (supplementary fig. S1, Supplementary Material online). Another reason why no effect was found overall could be the recent loss of sex in the cases studied (Delmotte et al. 2003). The fact that a localized sexual population persists in R. maidis (Remaudière and Naumann-Etienne 1991) and that a few sexuals have been observed in A. gossypii (Ebert and Cartwright 1997) indeed argue for recent losses; in addition, rare events of sexual reproduction (or “covert sex”—Simon et al. 2002) could have a strong effect on restoring genetic parameters (in terms of genetic diversity and mutation accumulation) close to regularly sexually reproducing species. A last observation is that several asexual taxa have been shown to result from hybridization among two species (Johnson and Bragg 1999; Delmotte et al. 2003; Morgan-Richards and Trewick 2005; Lundmark and Saura 2006; Lunt 2008). Even if hybridization events were recent, this also would not explain the increased rates. More detailed studies for these species, including information on genetic variation among populations, would be interesting to determine if genetic diversity bears the signature of relatively ancient asexuality or of hybridization. Also, larger sequence data sets would allow the exploration of a larger sample of functions. It may be expected, for example, that genes involved in sexual reproduction would be a primary target of mutation accumulation in newly asexual taxa (this could be checked by targeted sequencing). Finally, asexual aphid species could be able to resist mutation accumulation through compensatory changes; for example, an increase in population size, often associated with clonal organisms, could slow down mutation accumulation. Given that the in-group species (sexual or asexual) considered here are all pests and characterized by large population sizes, it could be even advanced that large population sizes came first in these taxa and were followed by the loss of sex (Normark and Johnson 2011); a larger population size is indeed expected to result in enhanced natural selection and would make populations more resistant to mutation accumulation, so it would allow asexual mutants to persist. Also, duplications could represent a possible buffering effect, saving genes from mutation accumulation (e.g., bdelloid rotifers: Lundmark and Saura 2006; Mark Welch 2008), because duplicated genes could represent backups and would attenuate the effects of mutations on one copy. Asexuality in organisms is often accompanied by increased ploidy levels, but this is not the case in aphids (Simon et al. 2002); because only a fraction of the aphid genome is duplicated, this buffer effect would be limited to gene families, so whether this phenomenon could provide enough resistance to mutation accumulation is not obvious. Whole-genome comparisons among related sexual and asexual taxa will be needed to more fully evaluate the dynamics and the consequences of the loss of the sexual reproduction.

**Supplementary Material**

Supplementary figure 1 is available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

**Acknowledgments**

This work was supported by the French Ministry of Research (Genoscope), INRA (Department: Santé des plantes), Région Bretagne, Rennes Métropole, and the Spanish Ministry of Science and Innovation (BFU2009-09168).

**Literature Cited**

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Evol. 215:403–410.

Andersson DI, Hughes D. 1996. Muller’s ratchet decreases fitness of a DNA-based microbe. Proc Natl Acad Sci U S A. 93:906–907.

Artacho P, Figueroa CC, Cortes PA, Simon JC, Nespolo RF. 2011. Short-term consequences of reproductive mode variation on the genetic architecture of energy metabolism and life-history traits in the pea aphid. J Insect Physiol. 57:986–994.

Bell G. 1988. Recombination and the immortality of the germ line. J Evol Biol. 1:67–82.

Birky CW Jr, Walsh JB. 1988. Effects of linkage on rates of molecular evolution. Proc Natl Acad Sci U S A. 85:6414–6418.

Blackman RL, Eastop VF. 2000. Aphids on the world’s crops. An identification and information guide. Chichester (United Kingdom): Wiley.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.

Carletto J, et al. 2009. Ecological specialization of the aphid Aphis gossypii Glover on cultivated host plants. Mol Ecol. 18:2198–2212.

Charlesworth D, Wright SI. 2001. Breeding systems and genome evolution. Curr Opin Genet Dev. 11:685–690.

Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676.

Delmotte F, Leterme N, Gauthier JP, Rispe C, Simon JC. 2002. Genetic architecture of sexual and asexual populations of the aphid Rhopalosiphum padi based on allozyme and microsatellite markers. Mol Ecol. 11:711–723.

Delmotte F, et al. 2003. Phylogenetic evidence for hybrid origins of asexual lineages in an aphid species. Evolution 57:1291–1303.

Dixon AFG. 1976. Reproductive strategies of alate morphs of bird cherry-oat aphid Rhopalosiphum padi L. J Anim Ecol. 45:817–830.

Doncaster JP. 1946. The shallot aphis. Myzus ascalonicus sp. n. (Hemiptera, Aphididae). Proc R Entomol Soc Lond. 15:27–31.

Ebert TA, Cartwright B. 1997. Biology and ecology of Aphis gossypii Glover (Homoptera: Aphididae). Southwestern Entomol. 22:116–153.
Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:1–19.

Felsenstein J. 1985. Confidence-limits on phylogenies—an approach using the bootstrap. Evolution 39:783–791.

Figueroa CC, et al. 2007. Annotated expressed sequence tags and xenobiotic detoxification in the aphid Myzus persicae (Sulzer). Insect Sci. 14:29–45.

Fitch WM. 1970. Distinguish homologous from analogous proteins. Syst Zool. 19:99–106.

Funk DJ, Wernegreen JJ, Moran NA. 2001. Intraspecific variation in symbiotic genomes: bottlenecks and the aphid-Buchnera association. Genetics 157:477–489.

Gascuel O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol Biol Evol. 14:685–695.

Glémin S. 2007. Mating systems and the efficacy of selection at the molecular level. Genetics 177:905–916.

Graur D, Li WH. 2000. Fundamentals of molecular evolution. 2nd ed. Sunderland (MA): Sinauer.

Guindon S, Gascuel O. 2003. PhyML—a simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. Science 275:777–781.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR. 2003. Biological identifications through DNA barcodes. Proc Biol Sci. 270:313–321.

Huelsenbeck J, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

Huerta-Cepas J, Dopazo H, Dopazo J, Gabaldon T. 2007. The human genome. Phyloinformatics 18.

Huerta-Cepas J, Dopazo J, Gabaldon T. 2010. ETE: a python environment for tree exploration. BMC Bioinformatics 11.

Huerta-Cepas J, Marcelet-Houben M, Pignattelli M, Moya A, Gabaldon T. 2010. The pea aphid phylome. Genome Biol. 10.

Johnson SG, Howard RS. 2007. Contrasting patterns of synonymous and nonsynonymous sequence evolution in asexual and sexual freshwater snail lineages. Evolution 61:2728–2735.

Kimura M. 1968. Evolutionary rate at the molecular level. Nature 217:624–626.

Kumar S, Nei M, Dudley J, Tamura K. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform. 9:299–306.

Lundmark M, Saura A. 2006. Asexuality alone does not explain the success of clonal forms in insects with geographical parthenogenesis. Hereditas 143:23–32.

Lunt DH. 2008. Genetic tests of ancient asexuality in root knot nematodes reveal recent hybrid origins. BMC Evol Biol. 8.

MacKay PA, Lamb RJ. 1988. Genetic variations in asexual populations of sea aphids (Aphidoidea). Biol J Linn Soc. 79:115–135.

Maynard Smith J, editor. 1978. The evolution of sex. Cambridge University Press.

Normark BB, Johnson NA. 2011. Niche explosion. Genetic 139:551–564.

Normark BB, Moran NA. 2000. Testing for the accumulation of deleterious mutations in asexual eukaryote genomes using molecular sequences. J Nat Hist 34:1719–1729.

Nguyen N. 2010. A comprehensive phylogeny of Neurospora reveals a link between reproductive mode and molecular evolution in fungi. Mol Phylogenetic Evol. 59:649–663.

Ollivier M, Legueai F, Rispe C. 2010. Comparative analysis of the Acrystosiphon pisum genome and expressed sequence tag-based gene sets from other aphid species. Insect Mol Biol. 19:33–45.

Otte SP, Lenormand T. 2002. Resolving the paradox of sex and recombination. Nat Rev Genet. 3:252–261.

Paland R, Lynch M. 2006. Transitions to asexuality result in excess amino acid substitutions. Science 311:990–992.

Pereza G, et al. 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19:651–652.

Pike N, Whitfield IA, Foster WA. 2007. Ecological correlates of sociality in Pemphigus aphids, with a partial phylogeny of the genus. BMC Evol Biol. 7:185.

Posada D. 2008. jModelTest: phylogenetic model averaging. Mol Biol Evol. 25:1253–1256.

Ramsey JS, et al. 2007. Genomic resources for Myzus persicae: eST sequencing, SNP identification, and microarray design. BMC Genomics 8:423.

Ratnasingham S, Hebert PDN. 2007. BOLD: the Barcode of Life Data System. Mol Ecol Notes. 7:355–364.

Remaudiere G, Naumann-Etienne K. 1991. Discovery in Pakistan of the sea aphid Acyrthosiphon pisum f. sp. tropicalis (Fitch)(Hom. Aphididae). C R Acad Agric Fr. 77:61–62.

Sabater-Muñoz B, et al. 2006. Large-scale gene discovery in the pea aphid Acrystosiphon pisum (Hemiptera). Genome Biol. 7.

Schiex T, Gouzy J, Moisan A, de Oliveira Y. 2003. FrameD: a flexible program for quality check and gene prediction in prokaryotic genomes and noisy matured eukaryotic sequences. Nucleic Acids Res. 31:3738–3741.

Simon JC, Rispe C, Sunnucks P. 2002. Ecology and evolution of sex in aphids. Trends Ecol Evol. 17:34–39.

Whitfield CW, et al. 2002. Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. Genome Res. 12:555–566.

Wilson ACC, Sunnucks P, Hales DF. 2003. Heritable genetic variation and potential for adaptive evolution in asexual aphids (Aphidoidea). Biol J Linn Soc. 79:115–135.

Yang ZH. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 13:555–556.

Associate editor Nancy Moran