Highly conserved genes expressed in aphid saliva are candidates for host plant adaptation in Aphis gossypii

Nathalie Boissot (nathalie.boissot@inrae.fr)
INRA https://orcid.org/0000-0002-8266-9386
Leslie Dutartre-Fricaux
INRA
Coralie Beucher
INRA
Flavie Vanlerberghe
INRAE

Research article

Keywords: Salivary genes, Phloem feeder, Cucurbits, Aphids

Posted Date: March 15th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-17211/v1

Title

Highly conserved genes expressed in aphid saliva are candidates for host plant adaptation in *Aphis gossypii*

Short Title

Host plant adaptation in aphids

Keywords

Salivary genes, Phloem feeder, Cucurbits, Aphids

Authors

Nathalie Boissot[1], Leslie Dutartre-Fricaux[1], Coralie Beucher[1], Flavie Vanlerberghe[2]

[1] INRAE, GAFL, F-84143 Montfavet, France
[2]INRAE, UMR CBGP, IRD, CIRAD, Montpellier SupAgro, F-34988 Montferrier-sur-Lez, France

Corresponding author

Nathalie Boissot : INRAE, GAFL, F-84143 Montfavet, France
Phone: +33 432 72 27 10
E-mail: nathalie.boissot@avignon.inrae.fr

Word counts

Paper text 4838
Abstract 243

Five figures
Three tables
Nine files for supporting information

Declarations

Ethics approval and consent to participate: Not Applicable.
Consent to publish: Not Applicable
Availability of data and materials: All data are available in the supplementary files
Competing interests: No
Funding: This work was partly supported by funding from ANR (VirAphid project ANR-2010-STRA-001-01)

Authors' Contributions

Nathalie Boissot designed the work, analyzed substantial part of data, interpreted the data, substantively revised the manuscript and approved the submitted version.
Leslie Dutartre-Fricaux acquired and analyzed substantial part of data, drafted the manuscript and approved the submitted version.
Coralie Beucher acquired and analyzed substantial part of the data and approved the submitted version.

Flavie Vanlerberghe interpreted the data, substantively revised the manuscript and approved the submitted version.

Acknowledgements
We thank Dr Claude Rispe from INRA Rennes for sending the CDS sequences reported in [1] and the BIPAA team and the GenOuest Platform at Rennes for assistance with bioinformatics. We thank Cécile Desbiez and Astrid Cruaud for advices concerning phylogenetic analyses of the megagene. We thank Virginie Chareyron, Pascale Mistral and Vincent Rittener for highly efficient technical assistance.

Abstract

Background

Aphids are major crop pests, most species attacking crops specialize on a narrow range of plant species from a single family. By contrast, *Aphis gossypii* is a highly polyphagous species, for which host races specializing on particular crops have been clearly described. Salivary components, which aphids inject into the phloem via their stylets, play a key role in establishing compatible interactions between plants and aphids, and are probably involved in specialization.

Results

We used the extensive resources available for *Myzus persicae* and *Acyrthosiphon pisum* to identify putative salivary proteins expressed in *Aphis gossypii,* despite the lack of genomic resources for this species. *In silico*, we identified 51 putative salivary proteins; we focused on 17 genes with orthologs in at least one aphid species, assuming that some of the conserved genes expressed in salivary glands are involved in host specialization. We amplified and sequenced 10 coding sequences in full, from 17 clones of *Aphis gossypii* specialising on plants from Malvaceae, Cucurbitaceae or Solanaceae. We reconstructed the phylogenetic tree for these genes, on which we identified a clade corresponding to all clones specializing on cucurbits. Three of these genes were under positive selection.

Conclusions

Full adaptation to a particular host plant may require a combination of alleles at quantitative trait loci in aphids. The three genes we identified could potentially be part of a cocktail of effectors manipulating the immune system of cucurbits and therefore responsible for *A. gossypii* specialization on that plant family.

Background

The genetic basis of adaptation is a key topic in evolutionary biology. Specialization on different hosts is a very common adaptation of plant pests and pathogens. Such specialization has been studied in detail for a number of pathogens, including fungi, bacteria, viruses and protists [2]. Several genes have been shown to be evolving rapidly in *Microbotryum* fungal pathogens specializing on different hosts [3]. By contrast, the genetic basis of insect pest adaptation to host plants has received much less attention, possibly due to the difficulties involved in developing genomic resources for insects [4].

Hemipterans and especially aphids are plant pests that can be considered to resemble pathogens in several ways. Indeed, they penetrate the plant tissues discreetly, until they reach the phloem the phloem, and the plant response to this penetration overlaps substantially with the responses mounted against microbial pathogens [5]. Four thousand aphid species have been identified, mostly in temperate regions, where they colonize 25% of existing plant species [6]. Many aphid species are restricted to one or a few host plants. Others are polyphagous, but differentiated host races within apparently polyphagous species are frequently identified, suggesting an ongoing process of speciation by adaptation to different host plants in several groups [7]. In the pea aphid *Acyrthosiphon pisum,* which specializes on several plant species from the Fabaceae, a genome scan survey highlighted 11 loci displaying genetic differentiation not consistent with a hypothesis of neutral evolution that might be responsible for host plant specialization in this aphid species [8].

Aphids have highly modified mouthparts, enclosing two flexible needle-like canals (stylets), which they use to feed. The alimentary canal is used to ingest plant fluids, whereas the salivary canal is used to deliver saliva along the path of the stylet, up to the feeding site. The aphid secretes a small amount of gelling saliva on the plant surface before stylet insertion. The stylets penetrate the plant epidermis at the border between two cells and they follow a pathway passing between the fibers of the secondary cell wall of one of these cells [9]. The stylets continue to produce gelling saliva whilst tracing an

intercellular pathway through the mesophyll to the phloem. Along the way, the stylets very briefly puncture numerous cells, into which they inject a watery saliva [9]. The role of aphid saliva has been studied in detail, particularly in terms of its capacity to modulate plant defense responses [10-12]. Aphid saliva contains a number of proteins, including oxidases, proteases and pectinases (see, for example, [13-20]). Some of the proteins present in aphid saliva prevent the clogging of vessels with plant phloem proteins, a naturally occurring calcium-triggered response to injury [9, 21]. The C002 protein, which has been shown to play an important role in the survival of the pea aphid on its host plant, bears no similarity to any protein outside the family Aphididae [22]. By contrast, a protein known to be conserved among parasites, a macrophage migration inhibitory factor secreted in aphid saliva, inhibits the plant immune response [23].

Aphis gossypii, the cotton-melon aphid, has a worldwide distribution and has been observed on more than a hundred different types of host plant [24]. *Aphis gossypii* populations display a host plant-based genetic structure, whatever their geographic origin [25]. Thomas et al., in their 2012 study, highlighted host plant selection *in natura* by demonstrating a large decrease in the genetic diversity in aphid populations during melon infestation. The spring-migrant populations visiting melon crops consist of alate aphids from several genetic clusters, whereas the populations infesting melon, which consist of their apterous offspring, belong exclusively to the cucurbit genetic clusters. This pattern strongly suggests that these highly specialized aphids can counteract cucurbit defenses.

We investigated the genetic basis of this host adaptation, with the following strategy: (i) we established a list of putative salivary candidate proteins in *A. gossypii,* making use of the salivary cDNA sequences available for two aphid species, *Acyrthosiphon pisum* and *Myzus persicae ;* (ii) we focused on orthologous candidates, due to the high degree of similarity in salivary function between members of the aphid family [13]; (iii) we studied the variability of these candidates in a collection of 17 *A. gossypii* clones for which we characterized host-plant adaptation.

Methods

Bioinformatic identification of salivary candidate genes in Aphis gossypii

Identification of candidate salivary proteins in *M. persicae*

We used sequences from salivary gland libraries or that had been identified as corresponding to salivary proteins [26, 27]. Expressed sequence tag (EST) data were retrieved from the NCBI EST database, with the accession numbers reported in previous studies [20, 26, 27]. We excluded redundant accessions, and then removed identical ESTs from further analyses by local reciprocal blastn searches (Score≥100, e-value≤1 x 10^{-5}, coverage≥50% of the length of at least one sequence) [28]. The remaining ESTs were locally blasted against *M. persicae* coding sequences (CDS) available from Aphidbase 2.1 (blastn, Score≥100, e-value≤1 x 10^{-5}, https://bipaa.genouest.org/is/aphidbase/). We then removed redundant CDS accessions from the list of CDSs identified.

Search for salivary candidate proteins in *A. pisum*

We used sequences from salivary gland libraries or that were reported to encode salivary proteins [16, 18, 27]. Sequences associated with the NCBI EST accession numbers reported in a previous study [27] were retrieved and locally blasted against *A. pisum* CDSs version 2, downloaded from AphidBase 2.1. The ACYPI accessions reported in various studies correspond to the first released version of the *A. pisum* genome annotation. As some ACYPI accessions were absent from the second version of this annotation, the corresponding sequences from the first annotation were locally blasted against the second (tblastx, Score≥100, e-value≤1 x 10^{-5}[28]). Additional sequences were then retrieved with new ACYPI accession numbers. Sequences corresponding to identical accessions were removed.

Putative candidate sequences common to *M. persicae* and *A. pisum*

The *M. persicae* and *A. pisum* sequences identified as described above were locally blasted against each other (tblastx, Score≥100, e-value≤1 x 10^{-5}, coverage≥50% of the length of at least one sequence) [28] to retrieve candidates common to both species.

Searches for similarity among *A. gossypii* CDSs

The putative candidate sequences common to *M. persicae* and *A. pisum* were used in local reciprocal tblastx approaches (Score≥100, e-value≤1 x 10^{-5}, coverage≥50% of the length of at least one sequence) with 15810 CDSs from *A. gossypii* [1, 28]. The *A. gossypii* candidates identified by screening with sequences from the other two aphid species were compared and candidate common to the different species were retrieved for further analysis after the removal of redundant CDSs.

Identification of orthologs

The shared *A. gossypii* candidates were used as queries in tblastx approaches (Score≥100, e-value≤1 x 10^{-5}, coverage≥50% of the length of at least one sequence) [28] against non-redundant aphid sequences from 10 species with the blast tool of NCBI, or locally against data downloaded from AphidBase 2.1 or from a previous study [1]. Phylogenies were constructed for each candidate and sequences similar to it, by applying two methods to the protein data. Sequences were aligned with ClustalW [29], with gaps managed by "pairwise-deletion", and phylogenies were reconstructed with *p*-distance for the substitution model and by neighbor-joining for the reconstruction method, with MEGA5 [30]. The second reconstruction was performed on the phylogeny.fr website [31]. Sequences were aligned with MUSCLE [32]. No curation of the alignment was permitted. Phylogenies were then reconstructed with PhyML as the reconstruction method, with the other options set as "default parameters" [33]. Reconstruction methods were optimized if the tree topologies were discordant. Maximum-likelihood analyses were performed in MEGA5 to identify the best-fitting substitution model [30]. Two new phylogenies were reconstructed as described above, but with the best-fit model for protein substitution. Orthologous relationships were identified on the basis of comparisons between the candidate tree topologies and the putative "aphid tree" (Figure 1). Putative candidates with at least one ortholog previously identified in a salivary gland cDNA library and presenting an excretion signal, or identified as encoding a salivary protein were chosen for further analyses.

Candidate sequence corrections, annotations and signal peptide predictions

A. gossypii sequences were corrected and extended based on available *A. gossypii* ESTs identified and downloaded from AphidBase 2.1 by blastn approaches [28]. When extension at the 3' or 5' end of the sequence was not possible, alignments with orthologs were generated with ClustalW2 [29]. Candidate sequences were extended on the basis of an orthologous sequence if the percentage identity between the two sequences exceeded 99%.

The corrected and extended sequences were annotated with Blast2GO software version 2.6.5 [34]. Briefly, after blastx on the nr database (e-value≤1.10^{-3}), mapping was performed to link BLAST hits to functional information (Gene Ontology -GO- terms). Annotation was performed and extended with the Annotation Expander tool. GO InterProScan terms were then merged with the annotations.

We investigated the presence and location of signal peptide cleavage sites in the candidate sequences with SignalP 4.1 Server [35]. The subcellular distribution of the candidate proteins and the presence of transmembrane domains were predicted with TargetP1.1 and TMHMM1.0, respectively [36, 37].

Rearing and characterization of aphid clones

Seventeen *A. gossypii* clones, collected from various species from the Cucurbitaceae, Malvaceae or Solanaceae, in many areas of the world, between 1978 and 2012, were maintained on melon (*Cucumis melo*) for clones collected from cucurbits, okra (*Abelmoschus esculentus*) for clones collected on Malvaceae, eggplant (*Solanum melongena*) for clones collected on *Solanum* sp., and *Capsicum annuum* for clones collected from this species (Table 1). The clones were reared in insect-proof chambers, with an eight- hour night at 18°C and a 16-hour day at 24°C.

Adaptation of aphid clones to plant families

For each clone, we collected seven five- to seven-day-old apterous aphids from this rearing, which we transferred onto a young plantlet of an alternative host. The presence of nymphs and new adults was scored 12 to 15 days later (no, few, yes). At least three replicates were performed for each transfer.

Characterization of aphid clones on the basis of mitochondrial and nuclear SSR sequences

The DNA sequences for the mitochondrial gene encoding cytochrome *b* and the barcode region of cytochrome oxidase I (COI) distinguished unambiguously between *A. gossypii* and its sister species *A. frangulae* [38]. We extracted DNA from 17 aphid clones with a 5% (w/v) Chelex resin solution, as previously described [39]. The sequences of cytochrome *b* and COI (749 and 661 bases, respectively) were obtained for the 17 clones, as described by Carletto et al. 2009, and were aligned with an *Aphis frangulae* sequence [38]. The two sequences were concatenated for each clone, to build a phylogenetic tree. The best model for invertebrate mitochondrial DNA was identified for construction of the phylogeny tree with the Mega X platform [40].

We amplified eight microsatellite loci specific to the *A. gossypii* genome [41] in two PCRs, as previously described [38]. We determined the size of the allele at each locus by comparison with a molecular size standard, with GeneMapper v3.7 software (Applied Biosystems, Foster City, California, USA), and a multilocus genotype (MLG) was assigned to each aphid. The 17 MLGs were assigned to genetic clusters by comparison with 44 MLGs known to represent host race structuring in *A. gossypii* [38], nine of which were common. Mitochondrial DNA analysis suggested that one clone belonged to the sister species *Aphis frangulae*. We therefore added 77 MLGs identified in 289 *A. frangulae* individuals collected from their primary host *Frangula alnus* in South-East France in the spring of 2006. The 129 different MLGs were subjected to Bayesian clustering [42], using an admixture model with a burn-in of 500,000 iterations and a subsequent Markov Chain of 250,000 iterations. For each putative number of clusters (*K*, ranging from 1 to 12), we compared 10 replicate runs, to assess the consistency of the estimated

values. We used the Evanno method to determine the most likely number of genetic clusters [43]. For this number of clusters, we then performed one run of the admixture model with a burn-in of 500,000 iterations and a Markov Chain of 1,000,000 iterations.

Diversity of candidate genes in A. gossypii

Sequencing of candidate genes

RNA was extracted from about 100 aphids of each clonal population with the RNeasy Plant Minikit (Qiagen) and RLC lysis buffer, according to the manufacturer's instructions. DNase digestion was performed with the RNase-Free DNase kit (Qiagen). At the end of the purification process, RNA was eluted in 40 µL of RNase-free water. Nanodrop spectroscopy measurements and electrophoresis in a 1% agarose gel were performed to quantify the RNA and check its quality. RNA was stored at -80°C until use. Reverse transcription was performed on 2 µg of purified RNA with 50 µM oligoT21, 10 mM of each dNTP and AMV Reverse Transcriptase, according to the manufacturer's instructions (Promega). We checked the quality of the cDNA produced by subjecting 1 µL of the product of the reverse transcription reaction to electrophoresis in a 1% agarose gel stained with ethidium bromide.

Primer3 [44], NetPrimer -http://www.premierbiosoft.com/netprimer/) and local blast approaches [28] were used to design specific primers for amplifying full-length candidate coding sequences. PCR was performed according to a standard protocol with the primers at a concentration of 10 µM, 35 cycles of denaturation for 4 min at 94°C, annealing for 30 s at 94°C, hybridization for 30 s at 50°C<Tm<60°C, depending on the primer characteristics (Additional file 1.xls), and elongation for 1 min/kb at 72°C, with a final elongation for 10 min at 72°C. PCR was performed with the Hot Start Taq DNA polymerase (TaKara kit) according to the manufacturer's instructions. PCR products were subjected to electrophoresis in a 1% agarose gel stained with ethidium bromide to check the size of the amplicons, and were stored at -20°C until Sanger sequencing (GenoScreen).

Heterozygosity and detection of polymorphisms

Chromatograms were analyzed with Mixed Sequence Reader software (MSR) [45], which created nucleotide sequences with heterozygous sites coded according to the IUPAC nomenclature to avoid bias when building the phylogenetic tree [46]. The log ratio of intensity value was set to 1. The protein sequences corresponding to each gene were aligned with MUSCLE [32].

Phylogeny reconstruction and positive selection fingerprint

Sequences beginning with an ATG and containing equal numbers of base triplets were concatenated to construct a megaCDS for each *A. gossypii* clone. Phylogenetic trees were reconstructed from this megaCDS with the GTR + Γ model and parameters specific for each gene calculated with RAXML [47].

The action of positive selection on the megaCDS was investigated with the Datamonkey suite of web tools [48], by fast unconstrained Bayesian approximation (FUBAR) [49], to ensure robustness against model misspecification by averaging over a large number of predefined site classes.

Results

Identification and annotation of conserved salivary genes in *A. gossypii*

The complete strategy for establishing a list of candidate genes putatively expressed in the salivary glands of *A. gossypii* is presented in Additional file 2. We retrieved 575 unique salivary ESTs from *M. persicae* and 343 from *A. pisum*, resulting in 181 CDSs for *M. persicae* and 315 CDSs for *A. pisum* after the elimination of redundancy (Additional files 3 and 4). A reciprocal tblastx between the 181 salivary CDSs from *M. persicae* and the CDSs from *A. gossypii* retrieved from databases identified 63 putative candidates in *A. gossypii*. With the 315 CDSs from *A. pisum,* the same approach identified 78 putative *A. gossypii* candidates. In total, 59 *A. gossypii* sequences were common to the two approaches, and 51 unique candidates were identified after the removal of redundant CDSs (Additional file 5). Tree

topologies including *A. gossypii* candidates and similar sequences from other aphid species were reconstructed by two different methods. When topologies were discordant, parameter optimization was performed, with the substitution model in particular. Orthologs for all candidates were identified in *A. pisum, M. persicae, Rhopalosiphum padi, Rhopalosiphum maidis, Myzus ascalonicus, Acyrthosiphon kondoi, Toxoptera citricida, Diuraphis noxia, Aphis fabae* and *Pemphigus spyrothecae* (Additional file 5). Only 17 *A. gossypii* candidates were retained for further analyses on the basis of their having at least one ortholog previously described as “salivary”. The tree obtained for one of the candidates, TCL2_2311, is shown in Figure 2.

The CDS was complete for only seven of these 17 candidates. For five of the remaining 10 sequences, we were able to complete the CDS sequence using EST sequences from *A. gossypii* and orthologous sequences (Additional file 6).

The 12 full-length sequences from *A. gossypii* were annotated with Blast2GO software, using the accession number of their *A. pisum* ortholog (Table 2, Additional file 7). Six sequences (CL1251, TCL2_2311, TCL5_3354, TCL6_1068, TCL6_147 and TCL6_4857) were unknown. CL373 was annotated as "γ-interferon-inducible lysosomal thiol reductase" and TCL5_430 was annotated as a "CD63 antigen". Two other sequences, TCL2_1813 and TCL5_4105, were identified as corresponding to cuticle proteins, both involved in molecular functions, particularly in structural molecule activity. TCL5_3248 and TCL2_1247 were identified as RAS-related proteins, and both sequences were implicated in biological processes (level 2), including cellular processes, response to stimulus, localization, signaling, and metabolic (TCL5_3248) and developmental processes (TCL2_1247). Both were also involved in molecular function (level 2), including binding and catalytic activity.

Signal peptide cleavage sites were detected in eight of the 12 genes (CL_1251, CL373, TCL2_1813, TCL2_2311, TCL_4857, TCL5_3354, TCL6_1068, TCL6_147). The subcellular distribution was predicted for CL373 and TCL_430, and the presence of transmembrane domains was predicted for CL373.

Biological and genetic characterization of a set of 17 *A. gossypii* clones

The host plant and country from which each clone was collected are listed in Table 1 for the 17 *A. gossypii* clones studied here. Fifteen clones survived the entire study: seven collected on cucurbits and reared on melon, five collected on cotton or *Hibiscus syriacus* and reared on okra, two collected on potato and eggplant and reared on eggplant and one collected and reared on sweet pepper. All the clones were able to colonize okra, whatever the original host plant and the host plant used for rearing in the laboratory. By contrast, the clones collected/reared on cucurbits were unable to colonize eggplant or sweet pepper). The clones collected on plants from the Malvaceae were unable to colonize melon, but some were able to colonize eggplant, in an erratic manner. The clones collected/reared on *Solanum* sp. and the clone collected on *Capsicum* were unable to colonize melon. Thus, okra could serve as the host plant for any clone, whereas melon could be used only for clones collected on cucurbits.

According to mitochondrial DNA sequences encoding cytochrome *b* and COI, all clones except GEL6 belonged to the *A. gossypii* species (Figure 3). GEL6 was assigned to the *A. frangulae* clade.

The eight microsatellite loci specific to the nuclear DNA of the *A. gossypii* genome were amplified from all 17 clones, but with different multilocus genotypes (MLGs) according to the combination of alleles for these eight markers. We determined the position of the 17 clones relative to the diversity previously reported for *A. gossypii*, by performing Bayesian clustering of the 17 MLGs with a complementary set of 35 other references MLGs from *A. gossypii*, identified in individuals collected on a wide range of host plants. We checked the specific position of GEL6, by adding 77 MLGs identified in

A. frangulae collected on *Frangula alnus*. We identified three to 32 alleles per locus for these 129 MLS (Additional file 8).

The 129 MLGs were assigned to eight clusters. All individuals collected on *Frangula alnus* clustered in three groups shown in green in the upper part of Figure 4. The sum of the probabilities of belonging to these three clusters was greater than 0.70. These MLGs were considered representative of *A. frangulae* diversity. The other five clusters are shown in the lower part of Figure 4. We considered these MLGs to be representative of *A. gossypii* diversity. The MLG GEL6, which had an *A. frangulae* mitochondrial genome, clustered with these MLGs; it was therefore considered to belong to the species *A. gossypii* species and was retained for further analyses. Eight of the 52 *A. gossypii* MLGs were not clearly assigned (i.e. with a probability greater than 0.70) to any of the five clusters; all were located on the left side of the graph. These MLGs were considered to represent the admixture in the species *A. gossypii*. They were collected from plants of the Solanaceae (Tom1 and PsP4), Malvaceae (Hib1, Hib9, Hib10, Burk6) or Cucurbitaceae (GEL6 and GWD). Conversely, two clusters grouped together the MLGs identified in individuals collected from a single plant family. They were located on the right side of the graph, and were collected from plants of the Malvaceae for the blue cluster, or *Solanum* spp. for the purple cluster. The three gray clusters contain MLGs identified in individuals collected from diverse host plants: Solanaceae (PsPx), Malvaceae (Hibx and Ivo), or Cucurbitaceae (NM1) for the light gray cluster, *Hibiscus* (Hibx), *Citrus* (Cit1) or strawberry (Fresca) for the mid-gray cluster, and Cucurbitaceae (Cx, CUC1, GWD2) or *Hisbicus* (Hibx) for the dark gray cluster.

Phylogeny and positive selection on conserved salivary genes in 17 *A. gossypii* clones

The 12 candidate CDSs were Sanger sequenced for the 17 MLGs. Two CDSs, TCL5_4105 and TCL6_1068, could not be sequenced correctly in all clones, possibly because they belong to a multigene family; they were discarded from subsequent analyses.

The 10 remaining CDSs were 402 to 885 bp long. Each CDS had at least three polymorphic positions. GEL6 had the most SNPs and, for some CDSs, it had an allele not present in the other clones. The 10 CDSs represented 6054 bp when concatenated in a megaCDS (Additional file 9). As expected for orthologous genes, few polymorphic positions (only 65), were detected between the 17 megaCDSs.

The megaprotein contained 2018 amino acids, 31 of which were polymorphic. Two CDSs, TCL2_1247 and TCL2_2311, were monomorphic at the protein level. For TCL 1251 and TCL2_3248, only GEL6 displayed polymorphism, with one putative allele common to all the other clones and a specific allele. For two other CDSs, CL373 and TCL_4857, polymorphisms were observed in several clones, but a putatively specific allele was identified in GEL6.

The phylogenetic tree reconstructed from the megaCDS had a remarkable topology, with one clade grouping together specifically the clones collected on cucurbits, including GEL6, and supported by a high bootstrap value (> 0.85) (Figure 5).

We then investigated the positive (diversifying) selection and negative (conservative) selection acting on each site of the megaCDS (Table 3). Positive selection was detected for sites in three CDSs, TCL5_3354, TCL6_147 and TCL6_4857. Posterior probabilities exceeded 0.9 for one site in TCL5_3354 and two sites in TCL6_4857. Five more sites in TCL6_4857 and one site in TCL6_147 had probabilities of positive selection exceeding 0.8. All other had a probability of positive selection below 0.5. Negative selection was detected for sites in five CDSs, CL1251, CL363, TCL2_2311, TCL5_3354 and TCL6_4857. . Posterior probabilities exceeded 0.9 for one site in CL363, TCL2_2311 and TCL5_3354, three in CL1251, and four in TCL6_4857. All other sites had probabilities of negative selection below 0.5.

Discussion

Host-plant specialization is related to the ability of pathogens to manipulate host-plant responses to facilitate their attack. Pathogens encounter an active immune system in the host plant. Two main types of immunity have been described in plants, both locking out the microorganisms trying to invade the tissues [50]. The first type of immunity in plants is PAMPS-triggered immunity (PTI), which is induced by pathogen-associated molecular patterns (PAMPs) detected in the extracellular space. Organisms able to infect host plants deploying this type of immunity release a core effector into the intracellular space to block PTI. However, some of these effectors are recognized by plant resistance genes, triggering a second type of immunity known as effector-triggered immunity (ETI). In 2011, hemipterans were shown to be among the organisms triggering PTI, and aphid saliva has since been considered to contain the equivalent of PAMPs and effectors [10, 51]. Even aphid extracts can trigger PTI [52]. In 2013, hemipterans were proposed to trigger ETI [53], and it was suggested that the saliva introduced into plant cells during furtive puncturing contains effectors interacting with plant resistance genes. The aphid *Aphis gossypii* triggers ETI on melon plants harboring a specific resistance gene [54]. These findings suggest that the PTI/ETI framework is relevant for *A. gossypii*.

A. gossypii is a highly polyphagous species with clones specializing on particular plant species [25]. We made use of the various *A. gossypii* clones sampled on various host plants and belonging to different MLGs available to confirm, in the laboratory, that okra is an universal host and to demonstrate that the observed specialization on cucurbits is biologically clearcut. None of the clones collected from cucurbits could colonize any of the solanaceous hosts tested. None of the clones collected from plants from Malvaceae or Solanaceae could colonize melon. The GEL6 clone, in particular, highlights this point. The GEL6 clone had an *A. frangulae* mitochondrial genome and a 'chimeric' *A. frangulae*/*A. gossypii* nuclear genome, as shown by its alleles for the SSR markers and CDSs characterized here. GEL6 was collected from cucurbits and observed on melon crops at different sites in France [55]. It had the biological features of clones specializing on cucurbits.

Many winged clones of *Aphis gossypii* visit cucurbit fields during spring, but few are able to colonize cucurbits [56]. These aphids clustered mostly in a single genetic group (dark gray in Figure 5), but this group also contained clones collected on *Hibiscus syriacus*. Some clones specializing on cucurbits displayed admixture (e.g. GWD or GEL6), or belonged to another genetic group (e.g. NM1). *Aphis gossypii* clones able to infect cucurbits may deploy a core effector that attacks PTI during exploratory puncture. We therefore investigated the diversity of salivary proteins in a set of 17 *Aphis gossypii* clones. There is currently no reference genome or transcriptome available for *A. gossypii.* We therefore adopted an *in silico* approach. By using salivary data for *A. pisum* and *M. persicae*, we established a list of 51 CDSs putatively encoding proteins present in *A. gossypii* saliva. Most of these loci encode odorant and gustatory receptor genes, suggesting a role for such receptors in the host plant specialization and speciation of the pea aphid [57].

We focused on CDSs encoded by orthologous genes in aphid species that had been confirmed to encode a salivary protein in at least one aphid species. This reduced the list to 17 CDSs, only 10 of which could be fully sequenced in all 17 aphid clones. These CDSs were short, consistent with findings for most of the effectors fully characterized in fungi and oomycetes [58]. We reconstructed a phylogenetic tree from the ten concatenated CDSs. The resolved tree highlighted differences in evolutionary history between the megaCDS for clones collected from cucurbits and that of those for clones collected from other plants, such as *Capsicum*, *Solanum* sp. or members of the Malvaceae. Inferences regarding selection may provide important functional information [59] and findings for selection at megagene level are consistent with those for the positive selection of each of the genes considered separately [2]. We therefore investigated whether positive or negative selection acted on the 10 CDSs, by analyzing the megaCDS. Positive selection was detected at several sites of TCL5_3354, TCL6_4857 and TCL6_147. Signal peptide cleavage sites were detected in these three genes. We

reconstructed a phylogenetic tree from the three concatenated CDSs, which again contained a separate clade corresponding to all the clones collected on cucurbits.

This suggests that the genes expressed in *Aphis gossypii* gland salivary may encode proteins with an effector function for attacking PTI in cucurbits. A role for TCL6_4857 in host plant family selection by aphids is supported by the characterization of its ortholog in *A. pisum*, ACYPI000472, a candidate protein responsible for host-plant specialization in pea aphid. Indeed, the gene encoding this protein was located in the environment of an outlier locus displaying a higher level of genetic differentiation between races than would be expected under a neutral evolution model [8].

Our knowledge of the molecular functions of fungal and oomycete effectors is increasing, but we still know very little about these aspects in aphids and, unfortunately, no function has yet been described for the three genes highlighted here. In a recent large genomics study investigating effectors in *A. pisum* [60], only 10 single-copy genes of the 3603 genes expressed in saliva had already been characterized in terms of their role in the plant/insect interaction. Salivary proteins, or other effectors triggering plant defenses, such as GroEL, only slightly reduced fecundity or survival when expressed one-by-one *in planta* [27, 51]. This suggests that a cocktail of effectors is required for the full expression of immunity. Conversely, a macrophage migration inhibitory factor (MIF) highly conserved among parasites and secreted in aphid saliva has been shown to manipulate the plant immune system [23] and its downregulation decreases the survival and fecundity of aphids. Full adaptation to a particular host plant may require a combination of alleles at quantitative trait loci in aphids, as suggested by mapping analyses of host acceptance and performance in two North American *A. pisum* host races, one specializing on alfalfa and the other on clover [61]. The three genes identified here may be part of this cocktail in *A. gossypii*.

Abbreviations

EST expressed sequence tag
CDS CoDing Sequence
cDNA Complementary DNA
nr database no redundant database
SSR Simple sequence repeats
COI cytochrome oxidase I
MLG multilocus genotype
IUPAC International Union of Pure and Applied Chemistry
PCR Polymerase chain reaction
Bp base pair
SNP single-nucleotide polymorphism
PAMPs pathogen-associated molecular patterns
ETI effector-triggered immunity

Figure Legends

Figure 1 Schematic representation of the evolution of 15 aphid species.

The cladogram was reconstructed based on [1, 62-64].

Figure 2 Phylogenetic tree for TCL2_2311 from *Aphis gossypii* and its orthologs in other aphid species.

The tree was reconstructed after MUSCLE alignment and PhyML reconstruction. Only nodes with bootstrap values greater than 75 are shown (100 iterations). The sequences highlighted in red have already been identified as salivary proteins or were obtained from salivary gland cDNA libraries and present an excretion signal.

Figure 3 Phylogenetic tree for the cytochrome *b* and cytochrome oxidase I (COI) barcode regions of 17 clones of *Aphis gossypii*

The tree was reconstructed by the maximum likelihood method with the Hasegawa-Kishino-Yano model [65]. Branches corresponding to partitions reproduced in less than 50% of bootstrap replicates (500) were collapsed. Evolutionary analyses were conducted on the MEGA Platform [40].

Figure 4 Clustering of 129 multilocus genotypes (MLGs) detected in *Aphis gossypii* and *Aphis frangulae* populations

Percentage assignation of each MLG to eight clusters. Results of Bayesian clustering after analyzing the assignment of 129 MLGs with *K* values from 1 to 12, with 10 simulations for each value of *K* [42, 43]. The expected MLGs for *Aphis frangulae* clustered in three groups in green (A), the expected MLGs for *Aphis gossypii* clustered in five groups (B). MLGs that were not clearly assigned ($p < 0.7$) for any of the five clusters are on the left side of the graph (B). The two clusters on the right side of the graph grouped together MLGs identified in individuals collected on plants from a single family, cotton (Malvaceae) for the blue cluster, eggplant and potato (*Solananum* spp.) for the purple cluster. The gray clusters contained MLGs identified in individuals collected on various hosts (sweet pepper (PsPx), Hibiscus (Hibx), cotton (Ivo), melon (NM1, Cx, CUC1, GWD2), *Citrus* (Cit1) and strawberry (Fresca).

Figure 5 Phylogenetic tree for a megaCDS of 10 candidate salivary genes in 17 *A. gossypii* clones. The tree was reconstructed after the concatenation of 10 CDSs with RAXML [47], using the GTR + Γ model with parameters specific for each gene. One clade contained all clones collected on cucurbits (C4, C6, GWD2, GWD, NM1, CUC1, GEL6 and C9).

Literature cited

1. Ollivier M, Gabaldon T, Poulain J, Gavory F, Leterme N, Gauthier JP, Legeai F, Tagu D, Simon JC, Rispe C: **Comparison of Gene Repertoires and Patterns of Evolutionary Rates in Eight Aphid Species That Differ by Reproductive Mode**. *Genome Biology and Evolution* 2012, **4**(2):155-167.
2. Aguileta G, Refregier G, Yockteng R, Fournier E, Giraud T: **Rapidly evolving genes in pathogens: Methods for detecting positive selection and examples among fungi, bacteria, viruses and protists Discussion**. *Infection Genetics and Evolution* 2009, **9**(4):656-670.
3. Aguileta G, Lengelle J, Marthey S, Chiapello H, Rodolphe F, Gendrault A, Yockteng R, Vercken E, Devier B, Fontaine MC *et al*: **Finding candidate genes under positive selection in Non-model species: examples of genes involved in host specialization in pathogens**. *Molecular Ecology* 2010, **19**(2):292-306.
4. Richards S, Murali SC: **Best practices in insect genome sequencing: what works and what doesn't**. *Current Opinion in Insect Science* 2015, **7**:1-7.
5. Kaloshian I, Walling LL: **Hemipterans as plant pathogens**. *Annual Review Phytopathol* 2005, **43**:491-521.
6. Dixon AFG, Kindlmann P, Leps J, Holman J: **Why there are so few species of aphids especially in the tropics?** *American Naturalist* 1987, **129**(4):580–592.
7. Peccoud J, Simon JC, von Dohlen C, Coeur d'Acier A, Plantegenest M, Vanlerberghe-Masutti F, Jousselin E: **Evolutionary history of aphid-plant associations and their role in aphid diversification**. *C R Biologies* 2010, **333**(6-7):474-487.

8. Jaquiery J, Stoeckel S, Nouhaud P, Mieuzet L, Maheo F, Legeai F, Bernard N, Bonvoisin A, Vitalis R, Simon JC: **Genome scans reveal candidate regions involved in the adaptation to host plant in the pea aphid complex.** . *Molecular Ecology* 2012, **21**:5251-5264.
9. Tjallingii WF: **Salivary secretions by aphids interacting with proteins of phloem wound responses**. *Journal of Experimental Botany* 2006, **57**(4):739-745.
10. Hogenhout SA, Bos JIB: **Effector proteins that modulate plant-insect interactions**. *Current Opinion in Plant Biology* 2011, **14**(4):422-428.
11. Will T, Kornemann SR, Furch ACU, Tjallingii WF, van Bel AJE: **Aphid watery saliva counteracts sieve-tube occlusion: a universal phenomenon?** *Journal of Experimental Biology* 2009, **212**(20):3305-3312.
12. Rodriguez PA, Bos JIB: **Toward Understanding the Role of Aphid Effectors in Plant Infestation**. *Molecular Plant-Microbe Interactions* 2013, **26**(1):25-30.
13. Miles PW: **Aphid saliva**. *Biological Reviews of the Cambridge Philosophical Society* 1999, **74**(1):41-85.
14. Urbanska A, Tjallingii WF, Dixon AFG, Leszczynski B: **Phenol oxidising enzymes in the grain aphid's saliva**. *Entolmol Exp Appl* 1998, **86**(2):197-203.
15. Ni XZ, Quisenberry SS, Pornkulwat S, Figarola JL, Skoda SR, Foster JE: **Hydrolase and oxido-reductase activities in Diuraphis noxia and Rhopalosiphum padi (Hemiptera : Aphididae)**. *Annals of the Entomological Society of America* 2000, **93**(3):595-601.
16. Carolan JC, Fitzroy CIJ, Ashton PD, Douglas AE, Wilkinson TL: **The secreted salivary proteome of the pea aphid *Acyrthosiphon pisum* characterized by mass spectrometry**. *Proteomics* 2009, **9**(9):2457-2467.
17. Cooper WR, Dillwith JW, Puterka GJ: **Salivary Proteins of Russian Wheat Aphid (Hemiptera: Aphididae)**. *Environmental Entomology* 2010, **39**(1):223-231.
18. Carolan JC, Caragea D, Reardon KT, Mutti NS, Dittmer N, Pappan K, Cui F, Castaneto M, Poulain J, Dossat C *et al*: **Predicted effector molecules in the salivary secretome of the pea aphid (*Acyrthosiphon pisum*): A dual transcriptomic/proteomic approach**. *Journal of Proteome Research* 2011, **10**(4):1505-1518.
19. Rao SAK, Carolan JC, Wilkinson TL: **Proteomic profiling of cereal aphid saliva reveals both ubiquitous and adaptive secreted proteins**. *Plos One* 2013, **8**(2).
20. Harmel N, Letocart E, Cherqui A, Giordanengo P, Mazzucchelli G, Guillonneau F, De Pauw E, Haubruge E, Francis F: **Identification of aphid salivary proteins: a proteomic investigation of *Myzus persicae***. *Insect Molecular Biology* 2008, **17**(2):165-174.
21. Will T, Tjallingii WF, Thönnessen A, van Bel AJE: **Molecular sabotage of plant defense by aphid saliva**. *Proceedings of the National Academy of Science* 2007, **104**:10536-10541.
22. Mutti NS, Louis J, Pappan LK, Pappan K, Begum K, Chen MS, Park Y, Dittmer N, Marshall J, Reese JC *et al*: **A protein from the salivary glands of the pea aphid, *Acyrthosiphon pisum*, is essential in feeding on a host plant**. *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**(29):9965-9969.
23. Naessens E, Dubreuil G, Giordanengo P, Baron OL, Minet-Kebdani N, Keller H, Coustau C: **A secreted MIF cytokine enables aphid feeding and represses plant immune responses**. *Current Biology* 2015, **25**(14):1898-1903.
24. Blackman RL, Eastop VF: **Taxonomic Issues**. In: *Aphids as crop pests.* Edited by van Emden HF, Harrigton R. Oxfordshire, UK: CABI; 2007: 1-29.
25. Carletto J, Lombaert E, Chavigny P, Brevault T, Lapchin L, Vanlerberghe-Masutti F: **Ecological specialization of the aphid *Aphis gossypii* Glover on cultivated host plants**. *Molecular Ecology* 2009, **18**(10):2198-2212.
26. Ramsey JS, Wilson ACC, de Vos M, Sun Q, Tamborindeguy C, Winfield A, Malloch G, Smith DM, Fenton B, Gray SM *et al*: **Genomic resources for Myzus persicae: EST sequencing, SNP identification, and microarray design**. *BMC Genomics* 2007, **8**.

27. Bos JIB, Prince D, Pitino M, Maffei ME, Win J, Hogenhout SA: **A Functional genomics approach identifies candidate effectors from the aphid species *Myzus persicae* (Green Peach Aphid)**. *Plos Genetics* 2010, **6**(11).
28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *Journal of Molecular Biology* 1990, **215**(3):403-410.
29. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al*: **Clustal W and clustal X version 2.0**. *Bioinformatics* 2007, **23**(21):2947-2948.
30. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods**. *Molecular Biology and Evolution* 2011, **28**(10):2731-2739.
31. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J, Guindon S, Lefort V, Lescot M *et al*: **Phylogeny.fr: robust phylogenetic analysis for the non-specialist**. *Nucl Acid Res* 2008, **36**:W465 - W469.
32. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput**. *Nucleic Acids Research* 2004, **32**(5):1792-1797.
33. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O: **New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0**. *Systematic Biology* 2010, **59**(3):307-321.
34. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research**. *Bioinformatics* 2005, **21**(18):3674-3676.
35. Petersen TN, Brunak S, von Heijne G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions**. *Nature Methods* 2011, **8**(10):785-786.
36. Emanuelsson O, Brunak S, von Heijne G, Nielsen H: **Locating proteins in the cell using TargetP, SignalP and related tools**. *Nature Protocols* 2007, **2**(4):953-971.
37. Moller S, Croning MDR, Apweiler R: **Evaluation of methods for the prediction of membrane spanning regions**. *Bioinformatics* 2001, **17**(7):646-653.
38. Carletto J, Blin A, Vanlerberghe-Masutti F: **DNA-Based discrimination between the sibling species *Aphis gossypii* Glover and *Aphis frangulae* Kaltenbach**. *Systematic Entomology* 2009, **34**(2):307-314.
39. Fuller SJ, Chavigny P, Lapchin L, Vanlerberghe-Masutti F: **Variation in clonal diversity in glasshouse infestations of the aphid, *Aphis gossypii* Glover in southern France**. *Mol Ecol* 1999, **8**(11):1867-1877.
40. Kumar S, Stecher G, Li M, Knyaz C, Tamura K: **MEGA X: Molecular evolutionary genetics analysis across computing platforms**. *Molecular Biology and Evolution* 2018, **35**(6):1547-1549.
41. Vanlerberghe-Masutti F, Chavigny P, Fuller SJ: **Characterization of microsatellite loci in the aphid species *Aphis gossypii* Glover**. *Molecular Ecology* 1999, **8**:685-702.
42. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data**. *Genetics* 2000, **155**(2):945-959.
43. Evanno G, Regnaut S, Goudet J: **Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study**. *Molecular Ecology* 2005, **14**(8):2611-2620.
44. Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers**. In: *Methods in Molecular Biology™.* Edited by S. M, S.A. K, vol. 132. Totowa, NJ: Humana Press; 2000.
45. Chang CT, Tsai CN, Tang CY, Chen CH, Lian JH, Hu CY, Tsai CL, Chao A, Lai CH, Wang TH *et al*: **Mixed Sequence Reader: A program for analyzing DNA sequences with heterozygous base calling**. *Scientific World Journal* 2012.
46. Birky CW: **Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes**. *Genetics* 1996, **144**(1):427-437.

47. Stamatakis A: **RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies**. *Bioinformatics* 2014, **30**(9):1312-1313.
48. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Pond SLK: **Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes**. *Molecular Biology and Evolution* 2018, **35**(3):773-777.
49. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K: **FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection**. *Molecular Biology and Evolution* 2013, **30**(5):1196-1205.
50. Jones DG, Dangl JL: **The plant immune system**. *Nature* 2006, **444**:323-329.
51. Chaudhary R, Atamian HS, Shen ZX, Brigg SP, Kaloshian I: **GroEL from the endosymbiont *Buchnera aphidicola* betrays the aphid by triggering plant defense**. *Proceedings of the National Academy of Sciences of the United States of America* 2014, **111**(24):8919-8924.
52. Prince DC, Drurey C, Zipfel C, Hogenhout SA: **The Leucine-Rich Repeat Receptor-Like Kinase BRASSINOSTEROID INSENSITIVE1-ASSOCIATED KINASE1 and the Cytochrome P450 PHYTOALEXIN DEFICIENT3 contribute to innate immunity to Aphids in *Arabidopsis***. *Plant Physiol* 2014, **164**:2207-2219.
53. Dangl JL, Horvath DM, Staskawicz BJ: **Pivoting the plant immune system from dissection to deployment**. *Science* 2013, **341**:746-751.
54. Boissot N, Thomas S, Chovelon V, Lecoq H: **NBS-LRR-mediated resistance triggered by aphids: viruses do not adapt; aphids adapt via different mechanisms**. *BMC Plant Biology* 2016, **16**:25.
55. Thomas S, Vanlerberghe-Masutti F, Mistral P, Loiseau A, Boissot N: **Insight into the durability of aphid resistance from the demo-genetic study of *Aphis gossypii* populations in melon crops**. *Evolutionary Applications* 2016, **9**(6):756-768.
56. Thomas S, Boissot N, Vanlerberghe-Masutti F: **What do spring migrants reveal about sex and host selection in the melon aphid?** *BMC Evolutionary Biology* 2012, **12**(47).
57. Smadja CM, Canback B, Vitalis R, Gautier M, Ferrari J, Zhou JJ, Butlin RK: **Large-scale candidate gene scan reveals the role of chemioreceptor gene in host plant specialization in the pea aphid**. *Evolution* 2012, **66**(9):2723-2738.
58. Selin C, de Kievit TR, Belmonte MF, Fernando WGD: **Elucidating the role of effectors in plant-fungal interactions: Progress and challenges**. *Frontiers in Microbiology* 2016, **7**.
59. Nielsen R: **Molecular signatures of natural selection**. *Annual Review of Genetics* 2005, **39**:197-218.
60. Boulain H, Legeai F, Guy E, Morliere S, Douglas NE, Oh J, Murugan M, Smith M, Jaquiery J, Peccoud J *et al*: **Fast evolution and lineage-specific gene family expansions of aphid salivary effectors driven by interactions with host-plants**. *Genome Biology and Evolution* 2018, **10**(6):1554-1572.
61. Hawthorne DJ, Via S: **Genetic linkage of ecological specialization and reproductive isolation in pea aphids**. *Nature* 2001, **412**(6850):904-907.
62. von Dohlen CD, Rowe CA, Heie OE: **A test of morphological hypotheses for tribal and subtribal relationships of Aphidinae (Insecta : Hemiptera : Aphididae) using DNA sequences**. *Molecular phylogenetics and evolution* 2006, **38**(2):316-329.
63. Kim H, Lee S: **A molecular phylogeny of the tribe Aphidini (Insecta: Hemiptera: Aphididae) based on the mitochondrial tRNA/COII, 12S/16S and the nuclear EF1 alpha genes**. *Systematic Entomology* 2008, **33**(4):711-721.
64. Kim H, Lee S, Jang Y: **Macroevolutionary Patterns in the Aphidini Aphids (Hemiptera: Aphididae): Diversification, Host Association, and Biogeographic Origins**. *Plos One* 2011, **6**(9).
65. Hasegawa M, Kishino H, Yano TA: **Dating the human age splitting by amolecualr clock of mitochondrial-DNA**. *Journal of Molecular Evolution* 1985, **22**(2):160-174.

Supporting information

Additional file 1.xls

Primers used in this study

Additional file 2.pdf

Schematic representation of the bioinformatic pipeline used to identify putative salivary genes in *Aphis gossypii*.

EST sequence identifiers were from [20, 26, 27] for *M. persicae* and [16, 18, 27] for *A. pisum* and were used to retrieve sequences from EST-NCBI. CDSs of *M. persicae* and *A. pisum,* and *A. gossypii* sequences (CDSs and ESTs), were downloaded from AphidBase 2.1. A blast result was selected for further analysis if the alignment had an e-value≤1 x 10^{-5}, a score≥100 and if the coverage represented at least 50% of the length of one sequence (query or subject). Blast2GO annotation, phylogeny reconstruction and ortholog identification were performed as described in the materials and methods. *: the tblastx approach was used on the NCBI nr aphid database, data downloaded from AphidBase 2.1 and data from a previous study [1].

Additional file 3.fasta

Sequences of the 181 unique salivary CDSs of *Myzus persicae*.

Additional file 4.fasta

Sequences of the 315 unique salivary CDSs *Acyrthosiphon pisum.*

Additional file 5.xls

Annotations and orthologs of the 51 putative salivary candidates from *Aphis gossypii.*

Annotations were performed with Blast2GO v2.6.5. Orthology relationships were established by comparing phylogenies of the tblastx (nr) results with the aphid phylogeny.

Additional file 6.fasta

Full-length sequences of the 12 putative salivary candidates from *Aphis gossypii*.

Additional file 7.xls

Annotations and orthologs of the 12 full-length putative salivary candidates of *Aphis gossypii.*

Annotations were performed with Blast2GO v2.6.5. Orthology relationships were established by comparing phylogenies of the tblastx (nr) results with the aphid phylogeny.

Additional file 8.csv

Allelic composition at 8 SSR markers for 44 *A. gossypii* and 77 *A. frangulae* clones.

The 44 *A. gossypii* clones were collected from various plant hosts, see [25] and Table 1. All *A. frangulae* clones called Frxxx were collected from *Frangula alnus,* the primary host of the species *Aphis frangulae*.

Additional file 9.fasta

Sequences of 10 concatenated CDSs (6054 bp) for 17 *A. gossypii* clones.

The CDSs were concatenated in the following order: CL373 (885 bp), CL1251 (468 bp), TCL2_1247 (600 bp), TCL2_1813 (465 bp), TCL2_2311 (600 bp), TCL5_430 (711 bp), TCL5_3248 (633 bp), TCL5_3354 (426 bp), TCL6_147 (402 bp) and TCL

_4857 (864 bp).

Figures

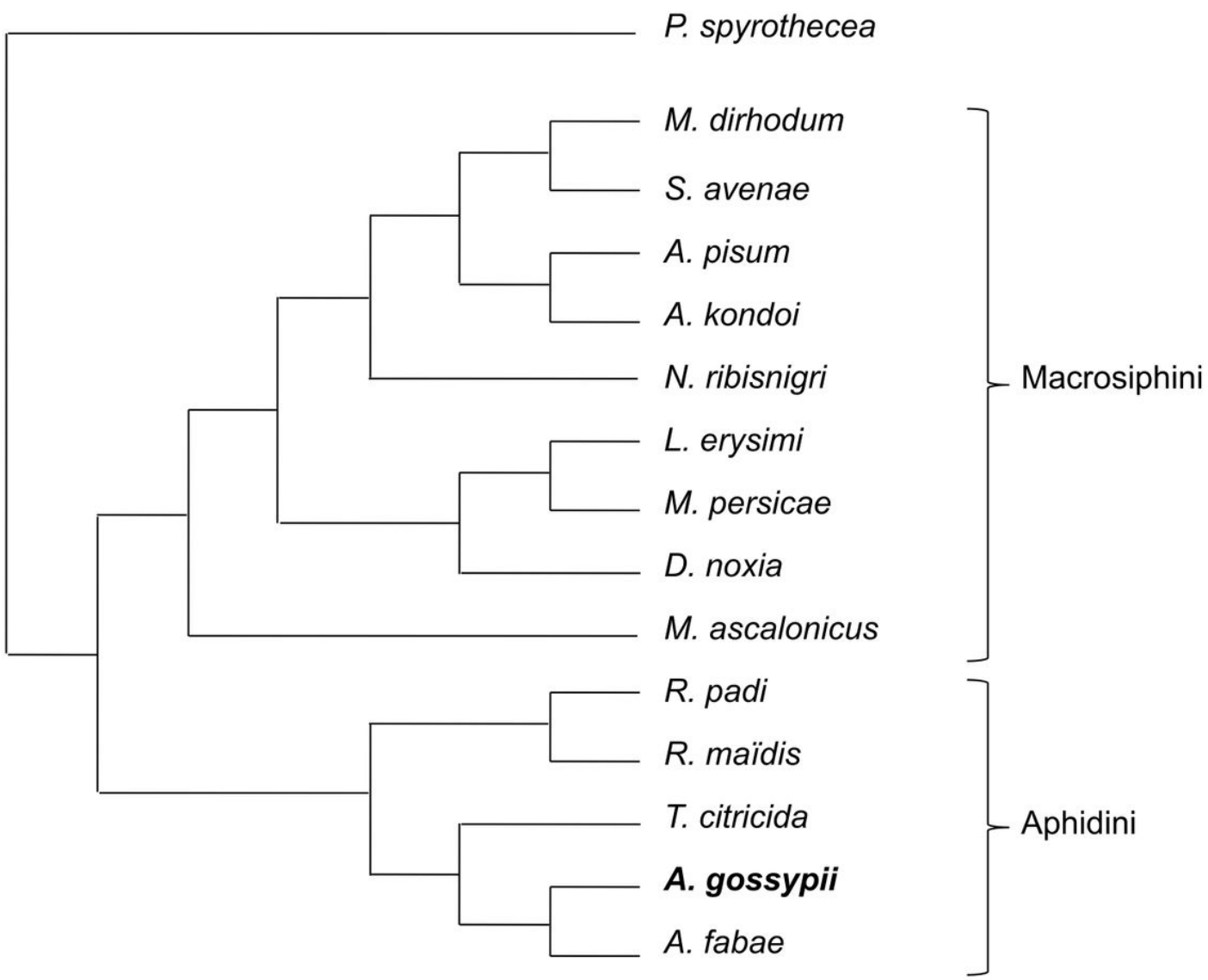

Figure 1

Schematic representation of the evolution of 15 aphid species. The cladogram was reconstructed based on [1, 62-64].

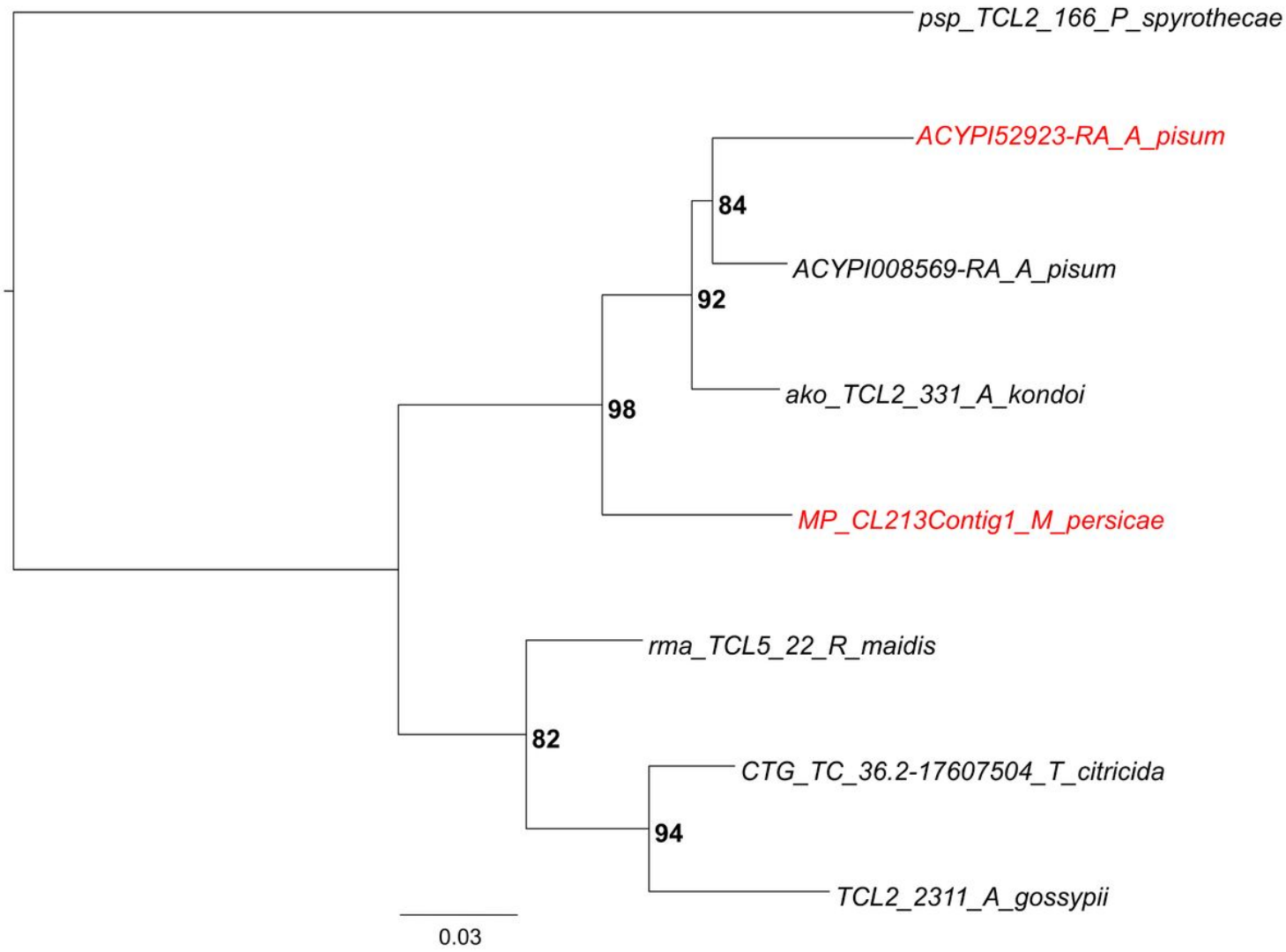

Figure 2

Phylogenetic tree for TCL2_2311 from Aphis gossypii and its orthologs in other aphid species. The tree was reconstructed after MUSCLE alignment and PhyML reconstruction. Only nodes with bootstrap values greater than 75 are shown (100 iterations). The sequences highlighted in red have already been identified as salivary proteins or were obtained from salivary gland cDNA libraries and present an excretion signal.

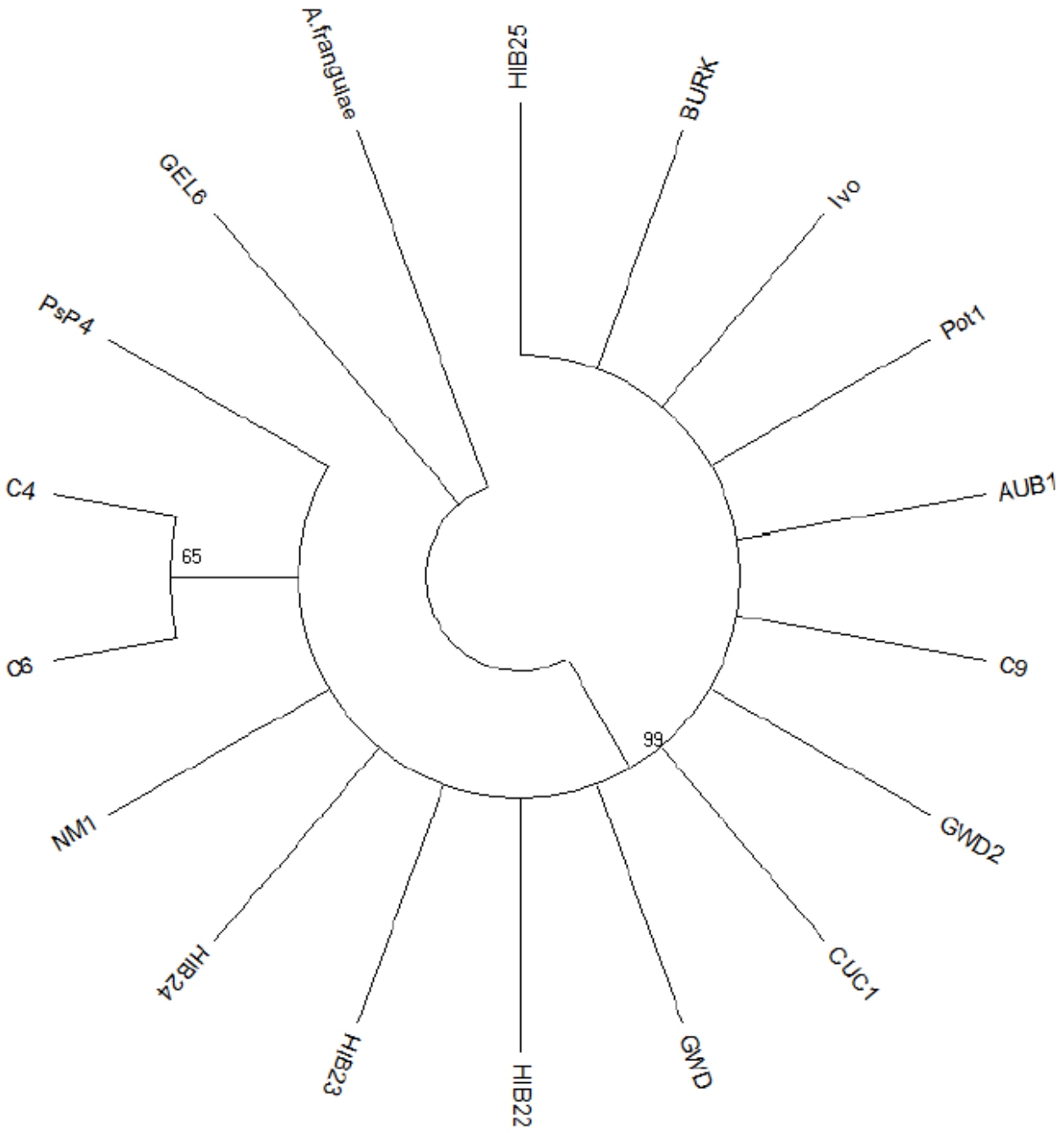

Figure 3

Phylogenetic tree for the cytochrome b and cytochrome oxidase I (COI) barcode regions of 17 clones of Aphis gossypii. The tree was reconstructed by the maximum likelihood method with the Hasegawa-Kishino-Yano model [65]. Branches corresponding to partitions reproduced in less than 50% of bootstrap replicates (500) were collapsed. Evolutionary analyses were conducted on the MEGA Platform [40].

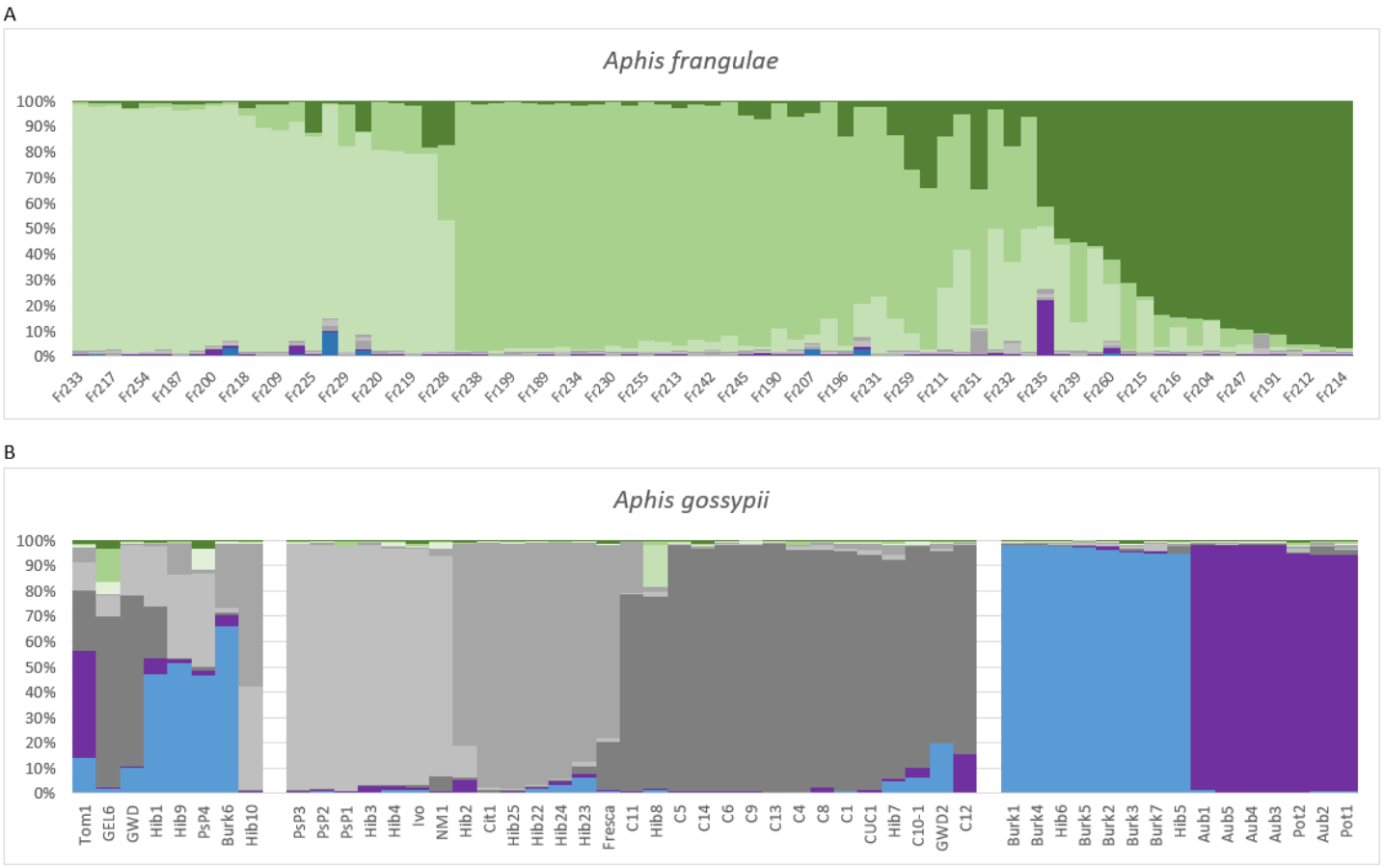

Figure 4

Clustering of 129 multilocus genotypes (MLGs) detected in Aphis gossypii and Aphis frangulae populations. Percentage assignation of each MLG to eight clusters. Results of Bayesian clustering after analyzing the assignment of 129 MLGs with K values from 1 to 12, with 10 simulations for each value of K [42, 43]. The expected MLGs for Aphis frangulae clustered in three groups in green (A), the expected MLGs for Aphis gossypii clustered in five groups (B). MLGs that were not clearly assigned ($p < 0.7$) for any of the five clusters are on the left side of the graph (B). The two clusters on the right side of the graph grouped together MLGs identified in individuals collected on plants from a single family, cotton (Malvaceae) for the blue cluster, eggplant and potato (Solananum spp.) for the purple cluster. The gray clusters contained MLGs identified in individuals collected on various hosts (sweet pepper (PsPx), Hibiscus (Hibx), cotton (Ivo), melon (NM1, Cx, CUC1, GWD2), Citrus (Cit1) and strawberry (Fresca).

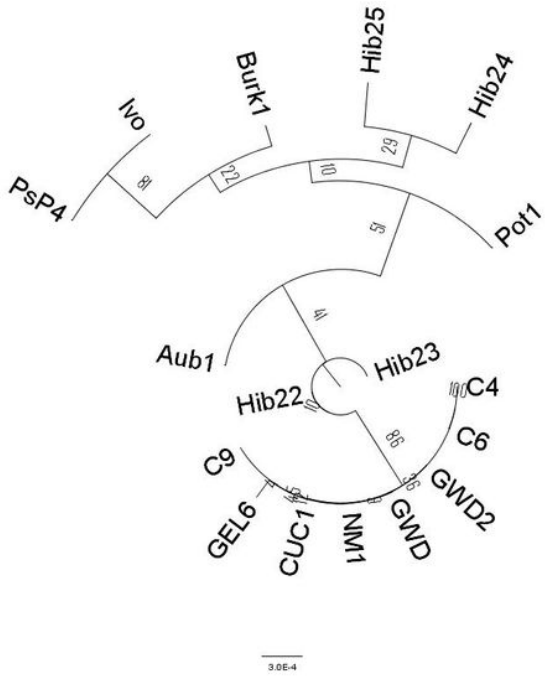

Figure 5

Phylogenetic tree for a megaCDS of 10 candidate salivary genes in 17 A. gossypii clones. The tree was reconstructed after the concatenation of 10 CDSs with RAXML [47], using the GTR + Γ model with parameters specific for each gene. One clade contained all clones collected on cucurbits (C4, C6, GWD2, GWD, NM1, CUC1, GEL6 and C9).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile2.pdf
- additionalfile5.xls
- Additionalfile8.csv
- Additionalfile1.xls
- Additionalfile9.fas
- Additionalfile7.xls
- Additionalfile6.fasta
- Additionalfile4.fasta
- Additionalfile3.fasta