Supporting Online Material for

Evolutionary Dynamics of Immune-Related Genes and Pathways in Disease-Vector Mosquitoes

Robert M. Waterhouse, Evgenia V. Kriventseva, Stephan Meister, Zhiyong Xi, Kanwal S. Alvarez, Lyric C. Bartholomay, Carolina Barillas-Mury, Guowu Bian, Stephanie Blandin, Bruce M. Christensen, Yuemei Dong, Haobo Jiang, Michael R. Kanost, Anastasios C. Koutsos, Elena A. Levashina, Jianyong Li, Petros Ligoxygakis, Robert M. MacCallum, George F. Mayhew, Antonio Mendes, Kristin Michel, Mike A. Osta, Susan Paskewitz, Sang Woon Shin, Dina Vlachou, Lihui Wang, Weiqi Wei, Liangbiao Zheng, Zhen Zou, David W. Severson, Alexander S. Raikhel, Fotis C. Kafatos,* George Dimopoulos, Evgeny M. Zdobnov,* George K. Christophides*

*To whom correspondence should be addressed.
E-mail: g.christophides@imperial.ac.uk (G.K.C.);
zdobnov@medecine.unige.ch (E.M.Z.); f.kafatos@imperial.ac.uk (F.C.K.)

This PDF file includes:

Materials and Methods
SOM Text
Figs. S1 to S22
Tables S1 and S2
References

Evolutionary dynamics of immune-related genes and pathways in disease vector mosquitoes

Supporting Online Material

Section 1: Materials and Methods

Gene Prediction: A comparative gene prediction pipeline (Zdobnov, unpublished) was applied to the selected insect genomes of *Aedes aegypti, Anopheles gambiae, Apis mellifera, Tribolium castaneum, Bombyx mori, Drosophila ananassae, Drosophila erecta, Drosophila grimshawi, Drosophila mojavensis, Drosophila pseudoobscura,* and *Drosophila virilis*. This approach relied on similarity to known Human and *D. melanogaster* proteins to identify genomic regions with protein-coding potential using tBLASTN (*S1*), followed by the homology-assisted gene prediction using Fgenesh+ (*S2*). The Ensembl *Anopheles* PEST3 annotation and the official *Aedes* Release1.0 gene build were also subsequently integrated into the analysis.

The gene prediction pipeline was applied to the following genome assemblies:

A. aegypti Aedes Genome Consortium aedes_aegyti_1

A. gambiae Anopheles Genome Consortium AgamP3

A. mellifera Baylor Amel_2.0

B. mori BGI 2003-10-01

D. ananassae ARACHNE assembly from Agencourt 20050801

D. erecta ARACHNE assembly from Agencourt 20050801

D. grimshawi ARACHNE assembly from Agencourt 20050801

D. mojavensis ARACHNE assembly from Agencourt 20050801

D. pseudoobscura Flybase release 1.04

D. virilis ARACHNE assembly from Agencourt 20050801

T. castaneum Baylor Tcas_1.0

Immunity Gene Families: We grouped genes into homologous clusters using several methods and at varying levels of stringency. Orthologous groups were identified using a COG-like strategy to cluster reciprocally best matching triangles from all-against-all Smith-Waterman (*S3*) comparisons and Inparanoid-like identification of co-orthologs (*S4*). Homologous groups at different similarity levels were identified using Blastclust (*S5*) on the basis of single-linkage (nearest neighbour) clustering of all-against-all BLAST (*S6*) sequence comparisons. Protein domain families were identified using InterProScan (*S7*) analysis on the basis of SMART (*S8*), Pfam (*S9*), and PROSITE-profiles (*S10*) signatures. Data on known immune-related genes and gene families were integrated with the data from the gene clustering methods to automatically define gene families across all species. For each known *D. melanogaster* 'seed' gene from a given immunity family, all members from the seed's orthologous group and blastclusters, as well as all genes having the same InterPro domain(s) as the seed protein, were putatively assigned to the family. Using these initial gene family results, sets of related genes were manually selected and ClustalW (*S11*) multiple sequence alignments were generated and refined where necessary. These were used to build and calibrate Hidden Markov Model (HMM) profiles with the HMMER package (*S12*) which were scanned against the gene sets again to identify additional gene family members. The results of the bioinformatics analysis are organized into a web-accessible resource (http://cegg.unige.ch/Insecta/immunodb/), which lists putative gene family members along with supporting evidence from orthologous groups, blastclusters, InterPro domains, and additional HMM profiles, as well as providing sequence and gene model data. Expert reviewing of these data for *D. melanogaster* and the two mosquitoes resulted in confirmation, refinement of gene models or rejection of putative gene family members based on knowledge of defining characteristics of a given immune-related gene family. The curated data for these three species, along with the results from phylogenetic analyses, are also available through the web-accessible resource; Immunodb.

Phylogenetics: The manually curated sets of genes were subjected to a phylogenetic analysis which aimed to reconstruct evolutionary relationships within each family as well as to identify genes or protein domains which are under similar evolutionary pressures and thereby likely to preserve functions. Multiple sequence alignments were computed using Muscle (*S13*), or HMM profiles (*S12*) where only particular domains were required to be aligned. These alignments were used to compute phylogenetic trees employing the Neighbour-Joining (NJ) algorithm implemented by ClustalW (*S11*), excluding all gap positions, correcting for

multiple substitutions, and assigning confidence with 1000 bootstrap samples. Confident orthologous relationships were assigned where the relevant bootstrap support was greater then 70%. To investigate functional relationships within clades shown in Fig. 2 and Fig. 3, and to compute *Dm-Ag* and *Dm-Aa* orthologous trio distances shown in Fig. 1B, confidently aligned conserved regions were extracted using Gblocks (*S14*), excluding all gap positions. The sequence relationships among these conserved protein cores were estimated in terms of amino acid substitutions using the Maximum Likelihood (ML) algorithm implemented by PhyML (*S15*) which allows for differential substitution rates between lineages and at different sites.

Orthologous Trios (Fig. 1A): Immunity single-copy orthologs were identified through our phylogenetic analyses, while all single-copy orthologs were identified automatically as outlined above using the Smith-Waterman algorithm. The gapless conserved core of the alignment of each trio was extracted and used to estimate amino acid substitution rates employing the ML algorithm. These are plotted as *Dm-Ag* distances versus *Dm-Aa* distances for all 1:1:1 orthologous trios. An increased level of sequence divergence is observed among the immunity (Im) 1:1:1 orthologs (red) compared to all 1:1:1 orthologs (blue): *Ag*Im mean =1.0120 and *Ag*Non-Im mean=0.8074, Wilcoxon rank sum test of *Ag*Im versus *Ag*Non-Im p=1.275e-04, *Aa*Im mean=0.9844 and *Aa*Non-Im mean=0.7812, Wilcoxon rank sum test of *Aa*Im versus *Aa*Non-Im p=8.298e-05. A greater accumulation of amino acid substitutions is observed in conserved protein cores of *Ag* compared to *Aa*: *Ag*All mean=0.8099, *Aa*All mean =0.7835, paired Wilcoxon rank sum test of *Ag*All versus *Aa*All p<2.2e-16. In order to examine different sets of functionally-related genes, we performed an analysis of genes grouped by Gene Ontology (GO, www.geneontology.org) classifications. GO annotations were retrieved for the *Dm* member (best annotated species) of each trio of the set of all 1:1:1 orthologs. Using the GO parent-child relationships, all terms with 10 or less associated proteins were successively merged with their parent terms until the set contained more than 10 members. Due to the nature of GO, any given protein may be associated with more than one term and thus overlap between the sets of proteins is expected. To reduce this redundancy, we then excluded all parent terms which had one or more child terms of more than 10 members, leaving 471 terms, with a mean of 21.7 members. Wilcoxon rank sum tests were performed as above for significance testing for each group and these data are presented in **Fig. S1A** and **Table S2**. The differences between our immune repertoire and all trios are also clearly shown through cumulative frequencies plots (**Fig. S1B**) of the *Dm-Ag/Aa* distances for the four datasets: *Ag*All, *Aa*All, *Ag*Im, and *Aa*Im.

Section 2: Text

Anti-Microbial Peptides: AMPs ([Fig S2](#))

AMPs fall into three major classes which can be described as peptides containing cysteine disulfide bonds, linear peptides which form α-helices, or proline- and/or glycine-rich peptides (*S16*). *Dm* possesses a wide range of AMPs including metchnikowin, drosocin, defensin, diptericins, attacins, cecropins, and drosomycins (*S17*). Four attacins have been identified in *Dm*, however only one of these glycine-rich AMPs has been found in each of the mosquitoes. An orthologous group of cysteine-rich defensins, with three peptides in *Aa*, is formed with *Dm*Def and *Ag*DEF1. Gambicins, which have so far only been identified in the two mosquitoes, are cysteine-rich peptides which form four disulfide bridges. Cecropins are much more widespread among insects; however these α-helical peptides are relatively divergent between the mosquitoes and *Dm*. *Ag* gambicin has been shown to be induced in the mosquito after challenge with *E. coli* but not *Plasmodium*, contrasting *Ag*CEC1 which shows the opposite effect (*S18*). In addition to the AMPs presented in Fig. S2 and the gambicin discussed above, a putative diptericin and a putative holotricin (Glycine-Rich Repeat Protein: GRRP) were identified in the *Aa* genome.

Caspases: CASPs ([Fig S3](#))

Caspases are proteolytic enzymes which employ a cysteine protease mechanism to cleave aspartic acid of target proteins. CASPs and IAPs (inhibitors of apoptosis) are essential for regulating cell death during development. In a nonapoptotic context, a *Dm* caspase (*Dm*Dredd) and an IAP (*Dm*IAP2) are required for Imd signaling (*S19-21*). *Dm*Dredd encodes an initiator caspase and is an effector of the apoptosis activators reaper, grim and hid (*S22*). The overexpression of *Dm*Dredd induces apoptosis in SL2 cells (*S22, S23*), and interacts via its death effector domain with the apoptotic adaptor *Dm*FADD (*S24*). Six additional CASPs have been identified in *Dm*, and the mosquitoes exhibit expansions leading to 14 in *Ag* and 10 in *Aa*. Single-copy orthologs of the long-form CASPs *Dm*Dredd and *Dm*Dronc are found in both mosquitoes. Independent expansions in *Ag* and *Aa* have given rise to three *Dm*Damm/*Dm*Dream-related CASPs in each mosquito. Mosquito CASPS7 and CASPS8 form orthologous pairs, with CASPS7 being most closely related to *Dm*Ice and *Dm*Dcp1. Expansions in the mosquitoes, particularly in *Ag*, have resulted in two *Aa* and seven *Ag* CASPs related to *Dm*Decay.

Catalases: CATs ([Fig S4](#))

Catalase is a tetrameric enzyme which efficiently converts H_2O_2 to water and oxygen. Catalase is a single copy gene in *Ag* (*Ag*CAT1), but a second gene is present in *Dm* (CG9314). CG9314 mRNAs are present in EST databases, but the predicted enzyme lacks three of the seven heme-binding residues and is probably not active. This is in agreement with the observation that flies in which *Dm*Cat is disrupted are extremely weak, lack catalase activity and die soon after enclosion (*S25*). In contrast, both catalase genes in *Aa* appear to code for active enzymes.

CLIP-domain Serine Proteases: CLIP-[A](#),-[B](#),-[C](#),-[D](#),-[E](#)s ([Fig S5](#))

Serine proteases containing one or more clip domains (*S26*) function in extracellular pathways that regulate some immune responses of insects. The CLIP proteases represent a protein architecture apparently unique to arthropods, and they form large gene families in the insect species studies so far (*S27, S28*). Infections can stimulate activation of CLIP protease zymogens present in hemolymph, resulting in proteolytic activation of prophenoloxidases (PPOs) (*S29, S30*) or activation of the Toll-ligand spätzle (*S31, S32*). The latter also occurs to regulate development of dorsal/ventral pattern in *Dm* embryos. Some members of the CLIP family contain a protease domain in which one or more of the catalytic triad residues has mutated such that they are no longer possess proteolytic activity. Such serine protease homologs (SPH) can function as cofactors required for PPO activation by an active CLIP protease (*S33*), and they can also negatively regulate the melanization response (*S34*). The CLIP family in *Aa* is quite large, with 68 genes (compared with 45 in *Dm* and 54 in *Ag*), including at least 12 SPHs and 5 genes predicted to encode dual protease and SPH domains. Phylogenetic analysis of the sequences reveals five main subfamilies, including previously described groups A-D (*S27*) and a new subfamily E containing previously unannotated *Ag*CLIPs. Nearly all of the CLIPA and CLIPE genes encode SPHs, whereas the CLIPB, CLIPC, and CLIPD groups are predominantly proteases with intact catalytic triads. Subfamily B is the largest, and its sequences are most similar to the clip proteases that can activate spätzle (Easter and SPE) and those known from studies in lepidopteran and coleopteran species to directly activate PPO. Members of subfamily C with known functions include *Dm*Snake (embryonic dorsal/ventral pattern) and *Dm*Persephone (involvement in innate immune Toll pathway). Functions of CLIP proteases in subfamily D have not yet been identified. Although CLIP-domain sequences are poorly conserved, the CLIP-domain sequences from active CLIP-proteases in *Dm* fall into three groups based on distance between

conserved Cys residues and on a few other conserved residues (*S28*). There is a very good correlation between branches of the CLIP tree based on alignment of only the protease domain and the type of associated clip domain. Subfamily C is associated with clip domain type 1a, subfamily D with clip domain type 1b, and subfamily B with clip domain type 2. Considering that a function is known for only a few CLIP proteases in any insect species, the complexity of this gene family in *Aa* presents a challenging and exciting prospect for future experimental studies of their roles in immune responses.

C-Type Lectins: CTLs ([Fig S6](#))

Glycans, complex polymers of sugar units that decorate proteins and lipids, have various biological roles in the development, growth, function or survival. Many of the roles ascribed to glycans involve specific recognition by lectins, carbohydrate-binding proteins of ubiquitous nature. Perhaps the largest and most diverse family of animal lectins is the C-type lectins (CTLs) which are Ca^+-dependent proteins that function largely outside the cells and are either secreted or membrane-bound (*S35*). Sugar-binding activity of CTLs is usually ascribed to a single module designated a CRD (carbohydrate-recognition domain), which forms a subset of a large family of protein modules that are denoted C-type lectin-like domains (CTLDs) many of which are Ca^+-independent and bind to non-sugar ligands (*S36*). Invertebrate CTLs mediate several immune responses including opsonisation and microbial clearance (*S37, S38*), hemocyte nodule formation (*S39*) and activation of prophenoloxidase leading to melanization (*S40-42*). Two CTLs from *Ag* were shown to be negative regulators of ookinete melanization, while they lacked regulatory impact on Sepadex bead melanization, indicating a potential specialized immune evasive function for *Plasmodium (S43, S44)*. Phylogenetic analysis of *Ag*, *Aa* and *Dm* immunity-realated genes identified 25, 39 and 34 CTLs, respectively. Interestingly 9 clear 1:1:1 orthologues exist between the three species. The *Aa* orthologue of *Ag*CTL7 and CG1576 was partially identified; Blasting the *Aa* contigs with *Ag*CTL7 identified exons 3 and 4 of *Aa*CTL7 however exons 1 and 2 of the gene could not be found possibly due to an error in gene assembly in that region. As a result *Aa*CTL7 was not included in the phylogenetic analysis. Those genes conserved in all three species may be implicated in conserved developmental processes. Only one of these genes, *furrowed,*(CG1500) has been functionally studied in *Dm* and showed indeed to be required for the proper development of the *Dm* eye and mechanosensory bristles (*S45*). There was only 1 orthologous pair (*Ag*CTL8 and *Aa*CTL8) for which no *Dm* orthologue could be identified. Several species-specific family expansions exist. The largest one is in *Aa* and includes 20 genes. A *Dm* expansion

includes 14 genes all of which, except one (CG7763), are clustered on chromosome 2L. The *Ag* expansion includes 5 potential mannose binding CTLs, all of which are clustered within 12 kb at 2L/25D. Interestingly, an *Aa* CTL (AaCTLMA14) showed strong sequence homology to AgCTL4, which was shown to inhibit *P. berghei* ookinete melanization in *Ag*. While these genes could not be considered 1:1 orthologues as the relevant bootstrap support was not strong enough, their striking sequence homology within and outside the CRD domains suggests that they might have similar functions. Interestingly, *Ag*CTL4 clusters also with *Ag*CTLMA2 both in the tree as well as on chromosome 2L/21F suggesting that these parasite agonists originated by gene duplication followed by diversification.

Fibrinogen-Related proteins: FREPs (Fig S7)

Members of the fibrinogen-related proteins (FREPs) or fibrinogen-domain immuno-lectin (FBN) family contain the evolutionary conserved fibrinogen domain that is also found in the mammalian ficolins (*S46-52*). The ficolins are implicated in phagocytosis and complement activation, while the horseshoe crab and snail FBN genes have been implicated in bacteria binding, enhancement of antimicrobial activity and interactions with parasite (*Schistosoma*) components, respectively. The FREPs contain a pathogen-binding fibrinogen-like domain at their C-terminus and the N-terminal sequence is implicated in interactions with the N-terminus of other FBN proteins resulting in the formation of multimeric protein bundles with potentially increased affinity and specificity to the pathogens. FREPs exhibit species-specific expansions with only two identifiable orthologous trios and three mosquito orthologous pairs. The most notable expansion is found in *Ag*, which harbors as many as 61 members, compared to *Aa* with 37 and *Dm* with only 14. However, due to the lack of confidence in the quality of certain *Aa* FREP pseudogene sequences, the gene family in this species might turn out to be somewhat larger that currently predicted. The reason for this expansion in the mosquitoes remains unknown, but it has been speculated to be partly linked to hematophagy. Several *Ag* FREPs have been shown to be up-regulated by bacterial challenge and malarial infection (*S27, S48, S53*) (Dimopoulos, unpublished). The ability to form multimers (dimers in *Ag* cell line supernatant) may enable an increase of the mosquito's pattern recognition receptor repertoire since different combinations are likely to possess different binding specificities (Dong & Dimopoulos, unpublished). The phylogenetic tree was built from the alignment of FREP sequences to the Pfam HMM profile of the Fibrinogen domain (PF00147). Several proteins were removed (3 from *Aa*, 9 from *Ag*) as they were lacking sequences mostly at the N or C-terminal which may be a result of gene predictions missing the first and/or last exons.

Galectins GALEs ([Fig S8](#))

Galectins are thiol-dependent, ß-galactoside-binding lectins. Evidence has shown that GALEs are implicated in innate immunity in both *Dm* and *Ag*. *Dm*GALEs composed of 2 carbohydrate recognition domains (CRDs) connected by a peptide link (tandem repeat type) (*S54*). The absence of putative transmembrane domain and a classical secretion signal peptide implies *Dm*GALEs are secreted by a non-classical secretion pathway (*S54*). *In vitro* experiments had demonstrated the ability of *Dm*GALEs to bind β -galactoside sugars (*S55*). It is possible that *Dm*GALEs participate in the innate immune system of the fly by facilitating microbial recognition and/or phagocytosis. *Ag*GALEs contained only one CRD (prototype), and was found to be up-regulated in the salivary glands and gut of Anopheles mosquitoes that were infected with malaria or bacteria (*S56-58*). It was proposed that *Ag*GALEs might function as PRRs by binding saccharide ligands on the microbial surface to trigger a host immune response, or agglutinate and opsonize bacteria in the midgut following blood-feeding (*S58*). In Fig S8, three 1:1:1 orthologous trios, and one mosquito-specific 1:1 orthologous pair can be identified, as well as possible recent duplications in *Aa* to form GALE6A/B and GALE8A/B. A more divergent group of GALEs consists of one *Dm* (CG14879), one *Ag* (GALE10), and three *Aa* (GALE12/13/14) GALEs. The conserved GALEs (shared by different insects) are most likely to play roles in processes such as cell fate determination and cell proliferation.

Gram-Negative Binding Proteins: GNBPs ([Fig S9](#))

The GNBP/β-1,3-Glucan recognition family of proteins comprises members from several insects (*S59*). These proteins contain an N-terminal glucan binding domain and a C-terminal domain similar to β-1,3- and β-1,4- bacterial glucanases (*S60*). The *Dm* genome contains three unique GNBPs, sharing a 36% sequence homology between each other. GNBP1 has been proved both genetically and biochemically to be the co-receptor for Gram-positive bacteria along with PGRP-SA(*S61, S62*). Both the genomes of *Ag* and *Aa* have 7 GNBPs. These GNBPs can be classified into two subgroups: GNBPA and GNBPB. GNBPA1 and GNBPA2 are gene duplications with a higher homology to *Dm* GNBP1. Annotated sequences of mosquito GNBPBs however, indicate that this is a distinct sub-family of GNBPs unique to mosquitoes. Interestingly, the mosquito GNBPs show considerable conservation both between each other and with bacterial glucanases over the glucanase-like domain. This may point to the significance of a hydrolytic function of GNBPs to Gram-positive bacteria cell wall

components in invertebrate innate immunity. In *Ag*, GNBPB1 has been linked with resistance to *P. berghei* and *E. coli* infection, while it lacked activity against *P. falciparum* and *S. aureus* (*S53*). Our studies indicate a similar function for GNBP1 in the fly(*S62*). The N-terminal part of all GNBPs is much less conserved and shows a significant variation even within species. The exact function of this domain still remains unknown. In *Dm*GNBP1, this region is speculated to have similar glucan binding activity as in its bacterial counterparts. Therefore, a variation of this region might constitute a diversity of the GNBPs in pathogen recognition. Indeed, from the microarray data, different *Ag* GNBPs responded to *S. aureus* infection with a significant up or down regulation compared to *E.coli* and *B. bassiana* challenge, indicating that *Ag* GNBPs might be also involved in the Gram-positive bacterial defense as in *Dm*. Aa GNBPs seemed to respond equally well to all three infections, making the roles of *Aa* GNBPs more illusive in its immune defense.

Inhibitors of Apoptosis: IAPs ([Fig S10](#))

IAPs are characterized by a 70-residue domain, the baculoviral IAP repeat (BIR). In *Dm*, four IAPs were found with distinct domain architecture. *Dm*IAP2 encodes a protein that has three N-terminal BIR (baculovirus IAP repeat) domains and a C-terminal RING-finger (Really Interesting New Gene) domain. Although the previously characterized IAPs were shown to be inhibitors of caspases (*S63*), it is unlikely that *Dm*IAP2 directly inhibits DREDD. The depletion of *Dm*IAP2 leads to disruption of the Imd pathway and did not result in an enhancement or constitutive expression of immune genes (*S20, S21*). In both *Ag* and *Aa* genomes, clear 1:1 orthologs of three *Dm* IAPs (*Dm*IAP1, *Dm*BRUCE and CG12265) were found. Whereas *Ag*IAP2, the closest homolog of *Dm*IAP2, was predicted to contain only one BIR domain, *Aa*IAP2 contains three BIR domains and a C-terminal RING-finger, showing clear 1:1 ortholog relationship to *Dm*IAP2. The phylogenetic tree is derived from the alignment of the BIR domains of IAPs from all three species, with domains labeled a, b, or c from the N to the C terminal of the respective proteins. The only BIR of *Ag*IAP2 is shown as AgIAP2_domain c because it clusters with the c domain of *Dm*IAP2 and *Aa*IAP2. Domain a of *Ag*IAP7 is not included in the tree as it is incomplete, although the partial sequence does cluster with *Ag*IAP3/4_domain a. *Ag*IAP8 and *Aa*IAP9 have one BIR domain each, they are excluded from the tree as they disrupt the alignment, although *Ag*IAP8 may cluster with domain c of IAP2.

Lysozymes: LYSs ([Fig S11](#))

Lysozymes are 14-16 kDa basic proteins that hydrolyze peptidoglycan of bacterial cell walls. In insects, both chicken-type (c-type) and invertebrate-type lysozymes occur. Many studies have shown increases in expression of c-type lysozyme genes and of increases in lysozyme activity following exposure to bacteria, especially in the Lepidoptera. Interestingly, *Dm* lysozymes show little response to bacterial infection (*S64*), although several genes were upregulated after infection by the microsporidian parasite, *Octosporea* (*S65*). Mosquito genes *Ag*LYSC1 and *Ag*LYSC2 and *Aa*Lys-A (LYSC11) are upregulated following bacterial challenge (*S20, S66*). Our analysis indicates that an expansion of the lysozyme gene family has occurred in dipteran flies. There are 13 genes in *Dm*, 8 in *Ag* and 7 in *Aa*. Expansion of the c-type lysozyme gene family also is found in foregut-fermenting vertebrates. For example, ten lysozyme genes occur in *Bos taurus*. In both invertebrates and vertebrates, expansion is partly due to the use of lysozymes for digestion of bacteria as a food resource. Dipteran insects are associated with moist habitats and often use bacteria as a food. Larval forms of the mosquitoes and larval and adult *Dm* fit this profile. Several of the *Dm* lysozymes are expressed in the gut and exhibit acidic pIs, which are presumed to be adaptations to a digestive function in the acidic gut environment(*S67*). *Ag*LYSC3 and *Ag*LYSC8 are expressed at much higher levels in larvae than in adults and are candidates for a role in digestion (*S68*), however, no orthologous enzymes were found in *Aa*. A second group of dipteran lysozymes is characterized by the loss of one of the two amino acids that are critical for muramidase activity. In five of these insect proteins, Asp52 (using vertebrate notation) is altered to Asn52, suggesting the intriguing possibility that these enzymes may become activated upon deglycosylation. Finally, each of the three species has one unusual, long lysozyme that contains 4 or 5 lysozyme domains, these are not shown on the tree. In *Ag*, LYSC6 is constitutively expressed throughout the life cycle and in many adult tissues (*S68*) but transcript levels were not altered following immune challenge.

MD2-like Proteins: MLs ([Fig S12](#))

The MD2-Like gene family (MLs) code putative secreted proteins containing a lipid recognition domain (*S69*). MLs are essential for LPS mediated activation of TLR-4 signal transduction in mammals, and probably interact with other antigens as well through their lipid recognition domain (*S70, S71*). All the essential sequence features that have been linked to LPS and TLR-4 interaction are conserved between the *Ag*ML1 and vertebrate MD2 genes (*S53*). Recent studies have indicated *Ag*ML1 as a specific anti-*P. falciparum* factor that is

induced in the midgut by *P. falciparum* but not *P. berghei* infection (*S53*). *Ag*ML1 can also influence *Ag* resistance to bacteria infection and is therefore a likely broad spectrum pattern recognition receptor of the *Ag* innate immune system. *Aa* and *Ag* have 17 and 11 MLs gene family members, respectively, while *Dm* has only 8 members. The expansion of the mosquito gene family may indicate a specialized function in the defense against a blood meal ingested pathogen. Between the two mosquito species, this gene family shows quite significant diversity with only 4 orthologous pairs.

Peptidoglycan Recognition Proteins: PGRPs ([Fig S13](#))

13 genes with 16 PGRP domains have been identified in *Dm*, *Ag* has seven genes with ten domains, and eight genes with nine domains were identified in *Aa*. Clear 3-way orthologous relationships could be determined for *Aa* genes with PGPPS1, PGRPLB, PGRPLC, PGRPLA, and PGRPLD. Surprisingly, there is an ortholog to PGRP-LE in *Aa*, but not in *Ag*. Due to lack of conservation outside the PGRP domain, most gene predictions of the PGRP family are restricted to the PGRP domain. *Aa*PGRPLA appears to be a single domain PGRP domain gene like PGRP-LA in *Dm* − and unlike PGRPLA in *Ag* that possesses two PGRP domains. Determination of the gene architecture of *Aa*PGRPLC is hampered by an unsequenced region of unknown size within the gene. As a result, *Aa*PGRPLC was initially only predicted to have one instead of three PGRP domains, however this unsequenced region is likely to contain further domains. At least one was retrieved from the unassembled sequence reads of the genome project. Likewise, a PGRP domain of a potential ninth *Aa* (short) PGRP gene could be found in the unassembled sequences. However, it could not be determined if it represents a pseudogene or haplotype due to lack of coverage and size of this region. The tree is built from the alignment of the PGRP domains so that genes with multiple PGRP domains have more than one branch.

Peroxidases: Heme[HPXs], Glutathione[GPXs], and Thioredoxin[TPXs] ([Fig S14](#))

Local generation of high levels of reactive oxygen species (ROS) is an important effector mechanism during an immune response. For example, myeloperoxidase (MPO) produces highly reactive ROS during the neutrophils respiratory burst. Increased systemic levels of hydrogen peroxide (H_2O_2) in *Ag* have been associated with melanotic encapsulation-mediated refractoriness to *Plasmodium* (*S72*). Heme-containing peroxidases (HPX) like MPO use H_2O_2 as an electron acceptor to catalyze a number of oxidative reactions. Five highly similar members of this family are found in vertebrates, which are most homologous to the *Dm*

haemocyte peroxidasin and its mosquito orthologues, *Ag*HPX4 and *Aa*HPX4, suggesting that this subgroup expanded in vertebrates after their divergence from insects. Dual oxidases (DUOX) represent another group of vertebrate HPXs that combine a peroxidase with a NADPH-oxidase domain. These enzymes are expressed in various epithelia and are though to participate in local defense responses (*S73*). DUOX orthologs are present in *Dm* and mosquitoes. *Dm*Duox silencing markedly increases the mortality rate of adult flies feeding on microbe-contaminated food (*S74*). The apoptotic response of *Ag* midgut cells to *Plasmodium* invasion involves peroxidase-mediated nitration, possibly associated with induced levels of AgDUOX (*S75*). The HPX family has greatly expanded in insects producing seven orthologous groups. A double peroxidase (DBLPX) is present in insects with two highly divergent N-terminal (DBLPX-N) and C-terminal (DBLPX-C) peroxidase domains. Each domain type has one-to-one orthologs in the different insect species. Most peroxidases have 1:1 orthologues, but species-specific expansions have taken place in mosquitoes. ROS generated during the immune response are also potentially toxic to the host; it is therefore important that they be kept localized and rapidly neutralized. The thioredoxin and glutathione systems are important for protection against oxidative stress by reducing peroxides such as H_2O_2 to harmless products. The *Dm* and *Ag* genomes lack a glutathione reductase gene (GR) (*S46*), and functional studies indicated that glutathione is reactivated by Thioredoxin (*S76*). GR is also absent from *Aa*, suggesting that insects may only use the thioredoxin system. Surprisingly, genes with sequence homology to classic glutathione peroxidases (GPXs) are present in insects; two in *Dm* and three in mosquitoes. However, recent functional studies in *Dm* demonstrated that at least one of the two fly homologues (Gtpx-1) uses thioredoxin instead of glutathione as a substrate and is responsible for increased resistance to paraquat-induced oxidative stress (*S77*). Orthologues of Gtpx-1 and a second *Dm* GPX homologue (GPXH) are also present in *Ag* and *Aa* and are likely to use thioredoxin as a substrate. A third mosquito-specific GPX is probably catalytically inactive.

Prophenoloxidases: PPOs (Fig S15)

We have assessed the transcription of 6 individual *Aa* PPO sequences. Most of them are transcribed during larval and pupal stages. Among them, *Aa*PPO1 seems to be transcribed in the entire stages (larvae, pupae and adults), *Aa*PPO2 and *Aa*PPO3 are transcribed during early larval stages, and *Aa*PPO4 and *Aa*PPO6 are transcribed during both larval and pupal stages, but *Aa*PPO5 seems expressed only in adults 12 hours after a blood meal. Individual PPOs are actually translated in the *Aa*. For example, *Aa*PPO1, *Aa*PPO2, *Aa*PPO3, *Aa*PPO4 and

*Aa*PPO6 are translated in mosquito larvae and *Aa*PPO5 is translated in adults following blood feeding. Transcriptional profiles of individual PPOs do not seem to be closely correlated to the level of their gene products. For example, transcripts of *Aa*PPO2 and *Aa*PPO3 were not detected in later stage larvae, but their proteins were isolated from 6-day old larvae. Some PPOs might be proteolytically processed at their carboxyl end, resulting in the cleavage of the last 120 some residues. Because an absence of these carboxyl side residues does not affect their catalytic function, the carboxyl end fragment in PPOs might be somewhat related to their tissue specific localization. The phylogenetic tree shows expansions with duplications in each of the mosquitoes with *Ag*PPO1 and *Aa*PPO6 being most closely related to the *Dm* PPOs**.**

Rel-like NFkappaB Proteins: RELs (Fig S16)

The REL transcription factors show 1:1:1 conservation for DmRelish/REL2, a duplication of REL1 in *Aa*, and the loss of Dif in the mosquitoes. Please refer to the main text where the roles of RELs in innate immunity are discussed in detail.

Scavenger Receptors Class-A, Class-B, and Class-C (Fig S17)

Scavenger receptors, originally defined by their ability to bind modified forms of low-density lipoprotein (*S78*), function also as pattern recognition receptors for infectious nonself (pathogens such as Gram-negative and Gram-positive bacteria through recognition of LPS and LTA by SCRA and Plasmodium-infected erythrocytes by SCRB) and modified self (apoptotic cells and modified LDL) (*S79-81*). The multidomain SCRs vary markedly in structure, including molecules with collagenous, cysteine-rich, C-type-lectin or other domains. Related molecules have been discovered in *Dm*, where they have been implicated in clearance of apoptotic cells and in innate immunity (*S82-87*). We considered three major classes named A, B and C. Macrophage class A scavenger receptors (SCRAs) contribute to host defence by binding polyanionic ligands such as lipopolysaccharide (LPS) and lipoteichoic acid (LTA). They have been implicated in the phagocytic recognition of unopsonized microorganisms (*S81, S88, S89*) and opsonized pathogens (*S90*). Some members of this subfamily contain a Scavenger Receptor Cysteine-Rich (SRCR) domains which, in a human protein (MARCO), binds both Gram+ and Gram- bacteria (*S81*). Like *Dm* and *Ag* (*S27*), our phylogenetic analysis identified five SRCR-containing proteins in *Aa*. Four orthologous trios exist between the fruitfly and the two mosquitoes. The numbers of SCRC domains of these molecules vary between different protein members. Additional domains like Lysyl oxidase (LOX), C-type lectin and others are also found among SCRAs. Interestingly,

many more SRCR-containing molecules are detected in mammals (e.g. 26 and 33 in human and mouse respectively) compared to insects. This is in contrast to SCRBs which are discussed below. *Aa* proteins SCRASP1, SCRASP2 and their *Dm* and *Ag* orthologs share a C-terminal domain related to coagulation and inflammatory serine proteases and multiple LDL domains. Additional multiple chitin binding domains (CBD) are only detected in SCRASP1. Interestingly, the third serine-protease containing SCRA, SCRASP3, appears to have been lost from *Dm*. SCRAC has a partial C-type lectin domain, and the fourth group of SCRAs, SCRAL, matches the Lysyl oxidase (lox) domain of human Lox proteins (copper-containing amine oxidases that convert primary amines to reactive aldehydes). In *Dm* there are two SCRAL genes in contrast to one in each mosquito species. In conclusion, our SRCR domain analysis reveal that the ancestral genes had similar domain architecture with genes currently present in the three species, and multiple domains have not been acquired separately in each species during evolution. The only exception is *Aa*SCRAL1, which appears to encode only one SRCR domain compared to its ortholog in the other two species; however, we believe that this is due to bad sequence quality. Class B scavenger receptors (SCRBs) are thought to be a novel class of scavenger receptors characterized by a CD36 domain. SCRBs have diverse ligand-binding properties including oxidized low density lipoproteins, long chain fatty acids, anionic phospholipids, collagen types I, IV and V, thrombospondin (TSP), *P. falciparum* infected erythrocytes and apoptotic cells. The interaction between the CD36-binding domain of PfEMP-1 (*P. falciparum* erythrocyte membrane protein 1) and CD36 receptor is important for parasite survival, and may be involved in modulation of the host's response (*S91-93*). Croquemort, a *Dm* haemocyte cell-surface receptor, is a member of the SCRB family and can bind apoptotic cells (*S85, S86*). In addition, *Dm*Peste (Pes) is another CD36 family member required for uptake of mycobacteria, but not *Escherichia coli* or *Staphylococcus aureus*, which suggests a conserved role for SCRBs in pattern recognition and innate immunity (*S87*). Interestingly, there is a large expansion of the insect SCRB family compared to mammals. A total of 13 SCRBs were identified in fruitfly and in mosquitoes, whereas only three CD36-containing proteins have been identified in human. As in mammalian systems, insect SCRBs are transmembrane proteins with two short transmembrane domains adjacent to their short N- and C-termini, respectively. One mosquito-specific phylogenetic clade (SCRBQ) includes six SCRBs, all of them being 1:1 orthologs. These genes are closely related to three *Dm* SCRBs including Croquemort. *Dm*Pes appears to have been lost in mosquitoes. The *Ag*SCRB17 is the previously identified *Ag*SCRBQ4 receptor (*S27*), which is now renamed according to the new phylogenetic analysis. Remarkably, *Ag*SCRB16 is a newly annotated gene resulted from the

combination of *Ag*SCRB2, *Ag*SCRB4, *Ag*SCRB12 (*S27*) and the ENSANGT00000031448 transcript. The third class of scavenger receptors (SCRC) is specific for insects and originally founded by the *Dm*SR-CI, *Dm*SR-CII, *Dm*SR-CIII and *Dm*SR-CIV proteins (*S83, S84*). Like croquemort, *Dm*SR-CI appears to be macrophage specific and recognizes a broad range of polyanionic ligands, much like the mammalian class A SR (*S83*). Nevertheless, there is no significant sequence homology between SCRAs and SCRCs. SCRC are transmembrane or secreted multidomain proteins that contain several sequence motifs, including two complement-control protein (CCP) domains followed by a MAM domain (Meprin A5 antigen and RPTP Mu), and usually a somatomedin-B-like (BO) domain. They are thought to function as PRRs in phagocytosis and innate immunity; CCP together with MAM can bind bacteria in vitro (*S83, S84*). *Ag*SCRC1, the only *Ag* member of this class resembles *Dm*SR-CI and *Dm*SR-CII but, surprisingly, has two transmembrane domains at its N- and C-terminus, respectively. Two SCRCs (*Aa*SCRC1 and *Aa*SCRC2) were identified in *Aa*; both contain two CCP domains and one transmembrane domain at their C-terminus (like *Dm*), neither of them has a MAM domain.

Serine Protease Inhibitors: SRPNs ([Fig S18](#))

Serpins are a very large family of serine protease inhibitors and are found in all higher eukaryotes as well as viruses and have a wide range of biological functions (*S94*). Serpins are structurally conserved suicide substrates. They can be found intra- as well as extracellularly, and are usually 350-400 amino acid residues long with a reactive center loop (RCL) that is located 30 to 40 residues from the C-terminal end. Their RCL binds to the active site of the specific target protease similar to the binding of a substrate. The protease cleaves the serpin at the scissile bond (indicated on the tree), and upon cleavage the serpin becomes covalently linked to the protease, which is irreversibly inhibited (*S95*). Serpins are involved in a large variety of biological processes such as blood-clotting and fibrinolytic cascades, inflammation and complement activation, tumor suppression, extracellular matrix maintenance or remodeling and apoptosis. In *Dm*, three serpins have been investigated in detail. Spn27A, an ortholog of *M. sexta* Serpin-3 controls the Toll pathway during early development (*S96*) and is an inhibitor of the melanization reaction, most likely by inhibiting PAPs in adult flies (*S97, S98*). Spn43Ac controls the Toll pathway during the immune response. (*S99*). The latest *Dm* serpin to be analysed is Spn4, which is an inhibitor of furin, a subtilisin-like convertase that is required for pro-protein maturation (*S100, S101*). A genome-wide analysis of immune response in *Dm* identified five additional serpins that are up-regulated and three that are down

regulated after septic injury. The *Dm*, *Ag* and *Aa* genomes contain coding sequences for 30, 17 and 23 SRPNs, respectively. Based on the amino-acid composition of their RCLs we predict 17, 12 and 14 serpins to be active protease inhibitors. Since the original *Ag* SRPN annotation in 2002, we eliminated *Ag*SRPN15 on the basis of misassembled sequence that had been corrected in later genome releases. *Ag*SRPN13 (agCP12957) originally annotated in 2002, is no longer present in the current assembly. However, RT-PCR results clearly indicate the gene to be present in the *Ag* genome (Kanost and Suwanchaichinda, unpublished). We also identified 4 additional serpins (*Ag*SRPN16-19), which brings the total number of *Ag* SRPNs to 18. Of these, five serpins occur as isolated genes in the genome, but most of serpins are organized into four clusters located at 2R8A, 2L26A, 2L28D, and 3R33C, each cluster containing three SRPNs. The most striking finding is that all but one *Ag* SRPN have a clear ortholog in the *Aa* genome. In contrast, only two 1:1:1 orthologs shared by all three species can be found, namely Spn100A/SRPN12, and Spn85F/SRPN19, both of which are non-inhibitory serpins. Unfortunately, the functions of these *Dm* genes remain to be identified. All *Ag* sequences within the SRPN2 and mosquito-specific expansion cluster have 1:1 orthologs in *Aa* (*S102, S103*) and do not contain any additional *Aa* serpins. SRPN4 is alternatively spliced in *Aa* and *Ag*, however the gene encodes two additional isoforms of which one has arisen through an additional duplication in the *Aa* lineage. The *Aa* genome contains six additional serpins, of which four are located within the same chromosomal region (contig 1.65) and physically cluster with *Aa*SRPN7 and *Aa*SRPN14. The corresponding chromosomal location in *Ag* is 2L28D, which contains the orthologs of SRPN7 and 14 as well as *Ag*SRPN18, the only gene without an ortholog in *Aa*. Serpins are well conserved within the mosquito lineage, which will help functional analysis of the family. Genes contained within the mosquito-specific expansion cluster as well as in the chromosomal clusters discussed above, will be of special interest, as their recent diversification might reflect adaptation to specific environmental challenges that mosquitoes encounter.

Späetzle-like Proteins: SPZs (Fig S19)

The cleavage of *Dm* Späetzle results in binding of the product to the Toll receptor and subsequent activation of the Toll pathway. The phylogenetic tree is built from the alignments of the conserved cystine knot domain (InterPro: IPR006208). *Ag*SPZ1 does not contain this domain and is therefore excluded from the tree. Among the SPZs there is a high level of conservation with 1:1:1 orthologous relationships for SPZ2, SPZ4, SPZ5, and SPZ6, and a duplication of SPZ3 in *Aa*. Interestingly, three *Dm*Spz-like proteins are found in *Aa*.

Superoxide Dismutatses: SODs ([Fig S20](#))

Superoxide anion ($\cdot O_2^-$) is generated by NADPH oxidase as part of the oxidative burst in activated macrophages. Superoxide dismutases (SODs) convert $\cdot O_2^-$ into hydrogen peroxide (H_2O_2), a less toxic product. Eukaryotic MnSODs are typically nuclear-encoded mitochondrial enzymes, while Cu/ZnSODs are cytosolic. These two SOD classes appear to have evolved independently, as they share no sequence homology. MnSODs are single copy genes in *Dm* and *Ag*, but *Aa* has a second gene that appears to be diverging faster than the ancestral one. There are three Cu/Zn SODs with one-to-one orthologues in mosquitoes and *Dm*. Interestingly, the *Ag*SOD3a and *Ag*SOD3b transcripts are products of a single gene that share a common N-terminal exon and differ in their second exon, which codes for the catalytic domain. This duplication of the second exon in SOD3 is unique to mosquitoes.

Thio-Ester Containing Proteins: TEPs ([Fig S21](#))

The family of thioester-containing proteins (TEPs) comprises vertebrate complement factors C3/C4/C5 and the pan protease inhibitors α2-macroglobulins, which are involved in pathogen recognition and activation of immune responses. Insect TEPs could be separated in two main groups: (i) the highly conserved orthologous trio (*Dm*Tep6/*Ag*TEP13/*Aa*TEP13) and (ii) much more divergent sequences mostly grouped in species-specific expansions. The group of the three orthologous TEPs is supported by the presence of additional sequence stretches absent from other TEPs. Interestingly these three TEPs do not have the thioester motif (marked with "TE" on the tree) that gives name to the family. The more divergent TEPs are split in two groups: one comprising both *Dm* and mosquito sequences, all of them bearing the thioester. The second group contains only mosquito TEPs which are divided in species-specific expansions. In this group, some TEPs have lost their thioester motif. The best characterized insect TEP, *Ag*TEP1, binds to pathogen surfaces and promotes phagocytosis of bacteria by mosquito blood cells and killing of *Plasmodium* parasites (*S104-106*). It belongs to the *Ag*-specific expansion, with no clear 1-1 ortholog in *Aa*.

Toll Receptors: TOLLs ([Fig S22](#))

Genes of the Toll family encode single-pass transmembrane proteins with leucine rich repeats (LRR) interspersed with cysteine knots in the extracellular domain and an intracellular Toll-interleukin 1 receptor (TIR) domain. Some members have carboxyl extensions with various sequence motifs. Toll was originally identified in *Dm* as a key player in dorsal-ventral patterning during embryonic development (*S107*). It was found later that *Dm*Toll was

required for antifungal immune responses (*S108*). Genomic analysis found eight other related genes in the *Dm* genome. In mammals, different Toll related genes have been implicated in responses to various bacterial pattern molecules. A total of nine Toll related genes have been found in *Dm*, with *Dm*Toll, *Dm*Toll-5 and *Dm*Toll-9 showing antifungal activities (*S109-111*). The same set of genes is apparently conserved in *D. simulans*, *D. virilis*, *D. mojavensis* and *D. erectus*. Duplications of T3/T4 have been observed in *D. pseudoobscura*, *D. persimilis*, and *D. ananasse*. Ten Toll related genes have been identified in *Ag* (*S112*). There are only four orthologs between these two dipterans. There are species specific expansions and apparent gene duplications, forming orthologous groups (*S27*). An important difference is the presence of four Toll1 related genes (TOLL1A/1B/5A/5B) in *Ag* and only two (Tol1 and Toll-5) in *Dm*. In *Aa*, there are a total of 12 Toll related genes. Between *Aa* and *Ag*, there is a complete 1:1 correspondence of TLR genes, except for an apparent gene duplication that results in two Toll-9 related genes and the presence of a homologue of *Dm*Toll-3/Toll-4. Interestingly, one of the Toll-9 genes, AaTOLL9B, has three introns in the TIR domain, compared to 2 in other dipteran Toll-9 genes. At least seven TLR genes are found in *Tribolium castaneum*. Two additional sequences may represent polymorphism in Toll-1 gene of *T. castaneum*. At least 11 TLR genes have been identified in *Bombyx mori*. And an addition TLR gene may be present in *B. mori*, although it may also be the results of bad assembly of TOLL-2 and TOLL-2B sequences. Five TLR genes are found in *Apis mellifera*. It is clear that all insect species sequenced thus far posses at least one Toll-1. Another widespread TLR orthologue group is made of the Toll-9 genes. Toll-9s, which are quite distinct from other insect TLRs, have been found in Lepidoptera (*Bombyx mori*) and Coleoptera (*Tribolium castaneum*). It was therefore likely present before the split between protostomia and deuterostomia and is probably present in all insect species.

Gene Nomenclature:

This study identified a substantial number of genes in *Aa*, and many previously un-named genes in *Ag*, and a small number of genes that required refinement and renaming. We named the genes in accordance with provisional nomenclature rules devised for the *Ag* immunity gene family analysis (*S27*) and additional rules devised to undertake the naming of *Aa* genes.

The rules from the *Ag* immunity gene family analysis are as follows:

1. The names are mnemonic symbols, designed for easy recall. They do not aim to summarize all current information, which in any case is incomplete and subject to errors (orthology, function, chromosomal location).

2. To avoid errors in electronic communication all names consist exclusively of capital letters of the Latin alphabet and Arabic numerals; no punctuation marks, dashes etc. are used.

3. To minimize the length the formal names do not include taxonomic initials. If similarly named genes of two organisms are being compared, taxonomic initials can be added for convenience, but do not constitute part of the name (e.g. AgTEP to be easily distinguished from DmTep).

4. Roman letters and numerals indicate protein, italics indicate gene or RNA.

5. The name is based on sequence similarities and carries no functional implications, this must be determined experimentally.

6. The name consists of two to three contiguous fields, as follows: - The first field includes three to five letters and is an abbreviation of the highest sequence grouping used, usually a protein family, e.g. CLIP (for Clip-domain serine protease). - The second field, if present, includes one or more letters identifying a subgroup such as subfamily (e.g. CLIPD), or class (e.g. SCRB). - The third field enumerates each gene by using consecutive numerals (e.g. SCRB1,… 12). - Sometimes the third field numeral can be preceded by letter(s) indicating gene types within a subgroup (e.g. SCRBQ1, for a gene belonging to the SCRB Class, and to the croquemort type). - For historical reasons, in certain families, the third field can also enumerate by letters rather than numerals (e.g. PGRPLA, for gene A of the Long subfamily in the PGRP family).

7. It is recommended that names previously used in the literature or in database submissions be gradually replaced by systematic names, following consultation with the original author. Historical names or names that may be developed eventually to indicate experimentally verified function or orthology can be used as synonyms.

Additional nomenclature rules devised to undertake the naming of *Aa* genes:

1. Where orthology is clear (robust bootstrap support) in the gene tree, *Aa* genes will be named according to their *Ag* orthologs. Where orthology is tentative in the gene tree, it can be further investigated by examining the specific clade to decide whether there is enough confidence to assign orthology.

2. Where *Aa* expansions relative to an *Ag* gene are clear (robust bootstrap support) in the gene tree, *Aa* genes will be named using the number from the corresponding *Ag* gene, suffixed with uppercase letters A, B, C etc. E.g. AgSPZ3 and AaSPZ3A, AaSPZ3B, etc.

3. The remaining *Aa* genes which are neither clear orthologs (point A above), nor clear expansions (point B above) relative to *Ag*, will be named according to the rules defined in points 1-7 above, starting with the number following the highest number assigned to an *Ag* gene.

Section 3: Figures and legends

Figure S1A: The Gene Ontology analysis indicated that different functional groups exhibit a wide spectrum of divergence levels similar to that observed for the individual trio orthologs presented in Fig. 1A (although not as extreme, due to the effects of averaging over each functional group). Testing for significance as we did for our immune repertoire we find many highly and significantly conserved functional groups and few that are highly and significantly divergent (Fig S1A, $p<0.05$). The data for the 30 most divergent and 30 most conserved groups are presented in **Table S2**. There are obvious differences between our intensive manual curation which has defined the evolutionary dynamics of the immune repertoire and the GO categorization of biological functions; nevertheless, this analysis confirms that immunity is remarkably divergent.

Figure S1B: The cumulative frequencies of the *Dm-Ag* and *Dm-Aa* distances are plotted for the four datasets of 1:1:1 trios: *Ag*All, *Aa*All, *Ag*Immunity, and *Aa*Immunity. This figure supports the elevated diversity reported for the immunity versus the non-immunity trios, and the lower substitution rate for *Aa* compared *Ag*.

Figures S2-22: The phylogenetic trees presented in figures S2-22 are constructed using the Neighbor-Joining algorithm (ClustalW) from full sequence or domain-only sequence alignments computed using Muscle or HMMER as required. Robustness of the resulting trees was estimated with bootstrap analysis of 1000 samples. *Drosophila melanogaster* proteins (blue) are presented with names or CG-numbers, while assigned names are presented for all *Anopheles gambiae* (red) and *Aedes aegypti* (yellow) proteins. Branches with low bootstrap support indicate uncertainty with respect to the placement of the branch and therefore only bootstrap values of 500 or above are shown. Bootstrap supported 1:1:1 orthologous trios are indicated with a filled black circle, and 1:1 mosquito orthologous pairs are indicated with an unfilled black circle. The supporting texts describe relevant additional information on each immune-related family presented in Figures S2-S22. Additional labels for particular figures are explained in the relevant texts.
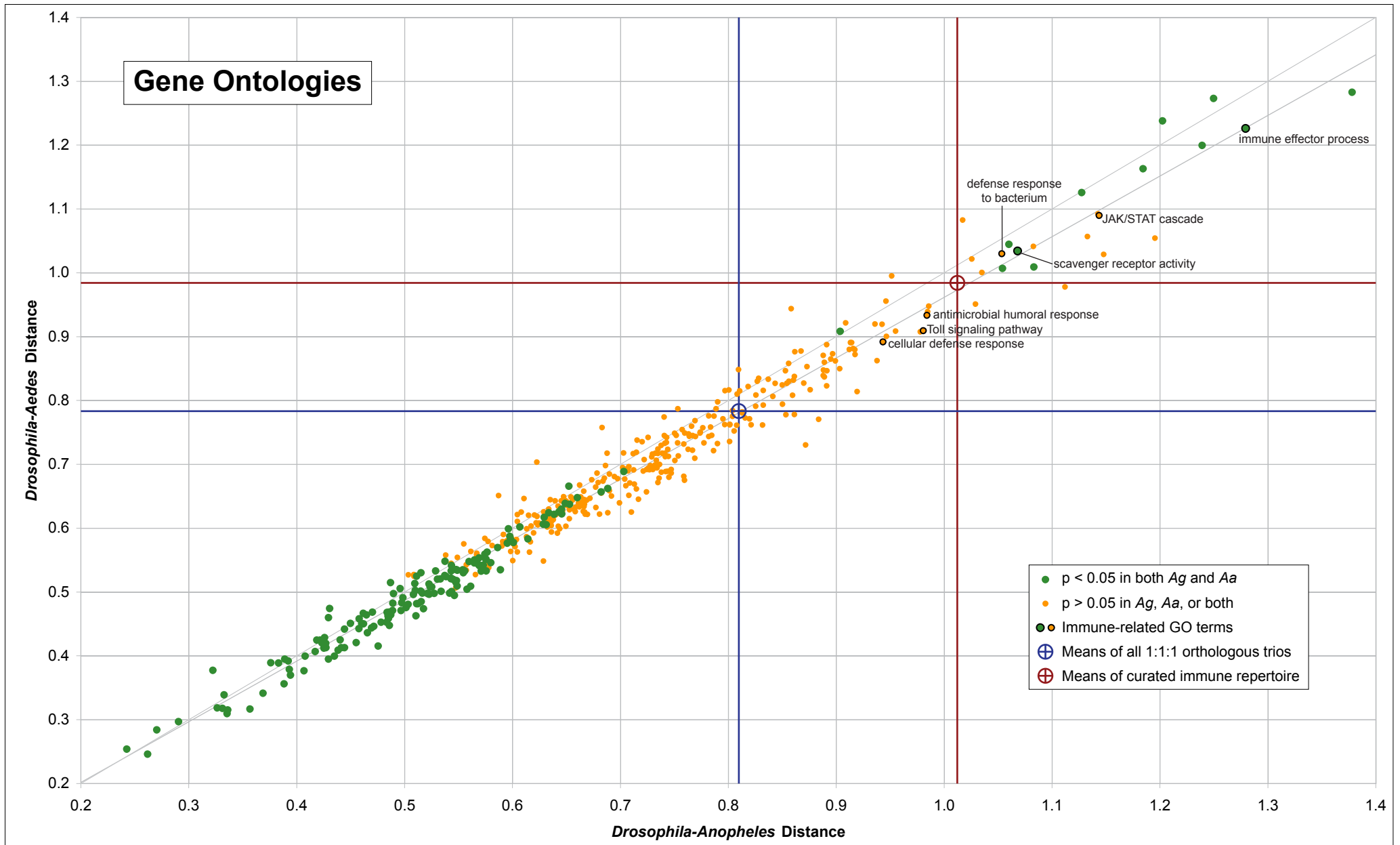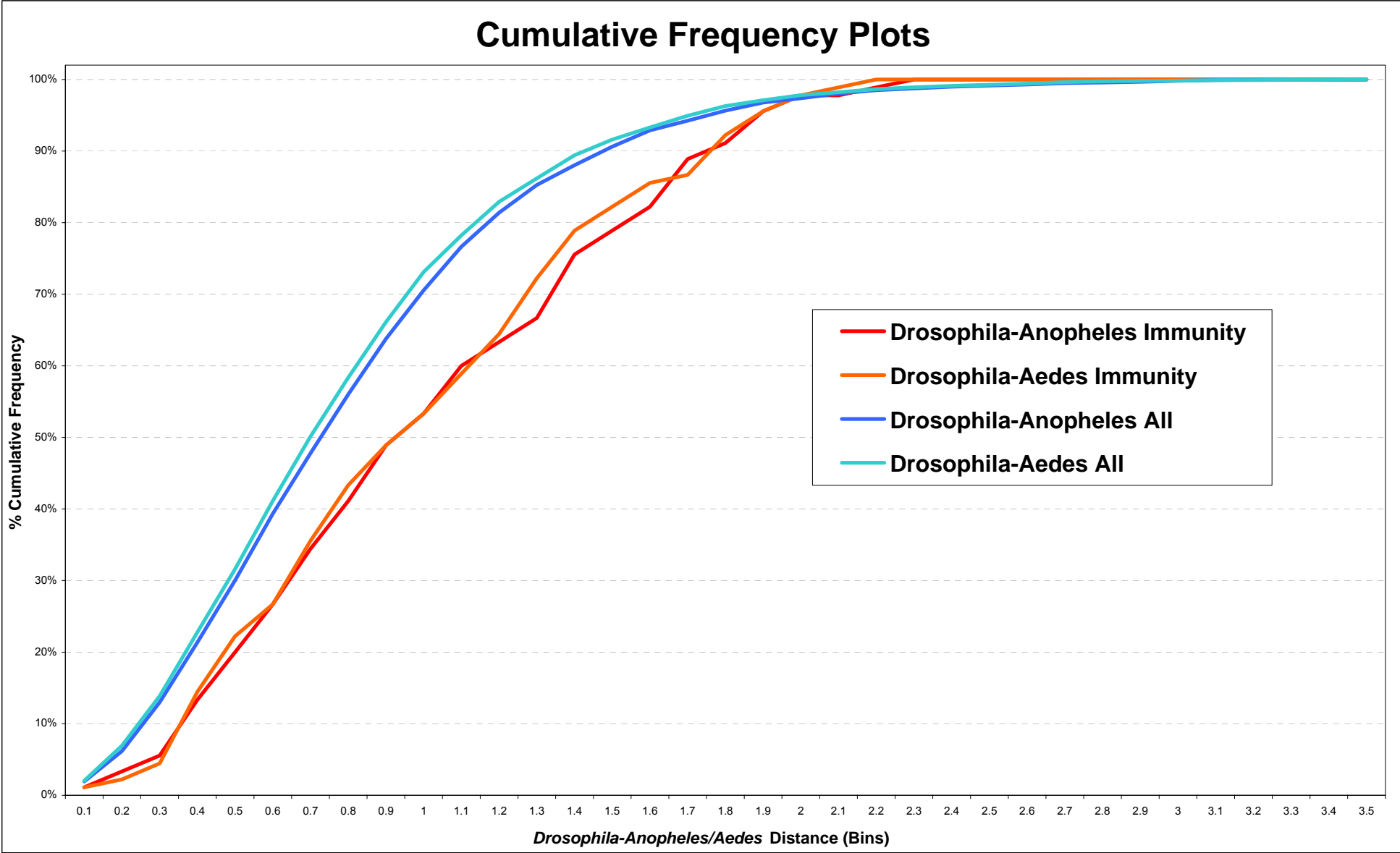
Fig. S1A

Fig. S1B

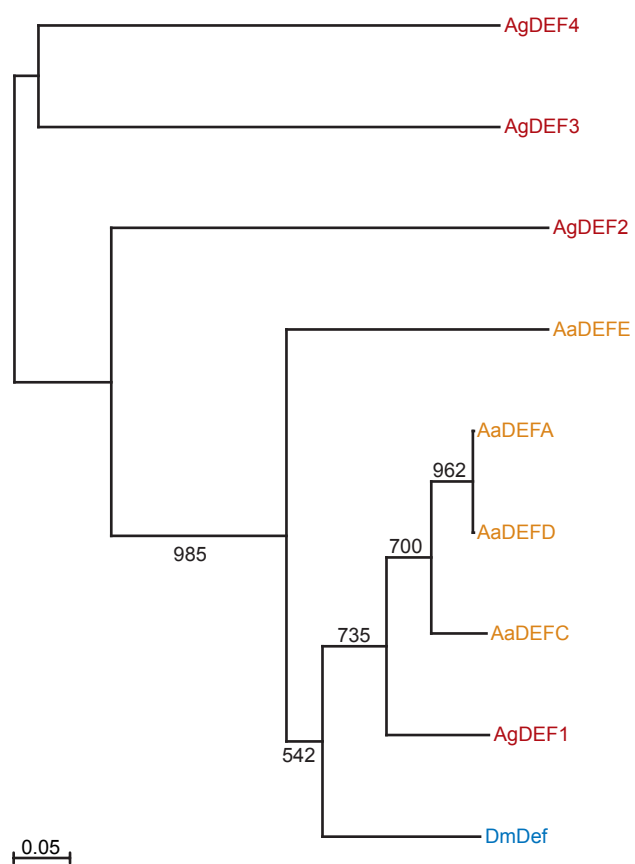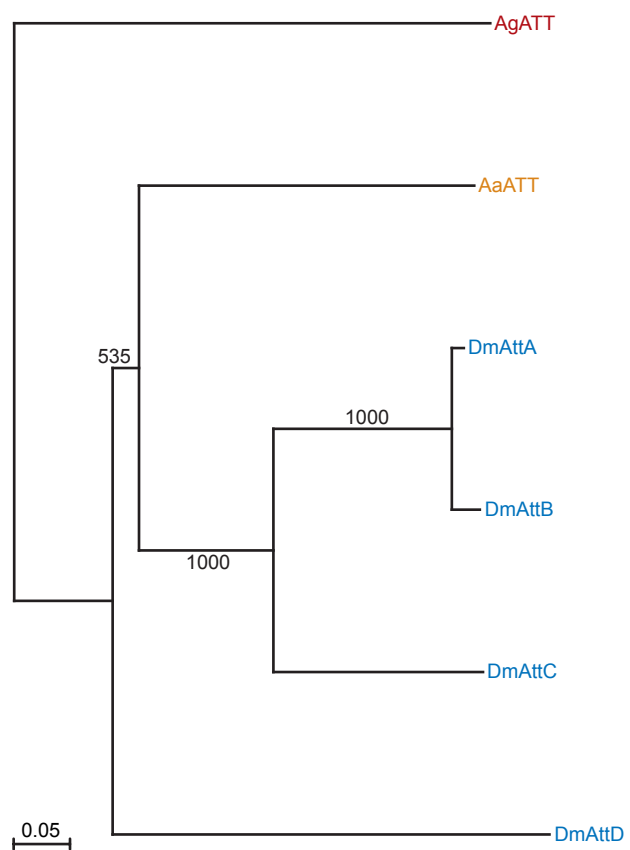Cumulative Frequency Plots

% Cumulative Frequency

Drosophila-Anopheles/Aedes Distance (Bins)

Legend:
- Drosophila-Anopheles Immunity
- Drosophila-Aedes Immunity
- Drosophila-Anopheles All
- Drosophila-Aedes All
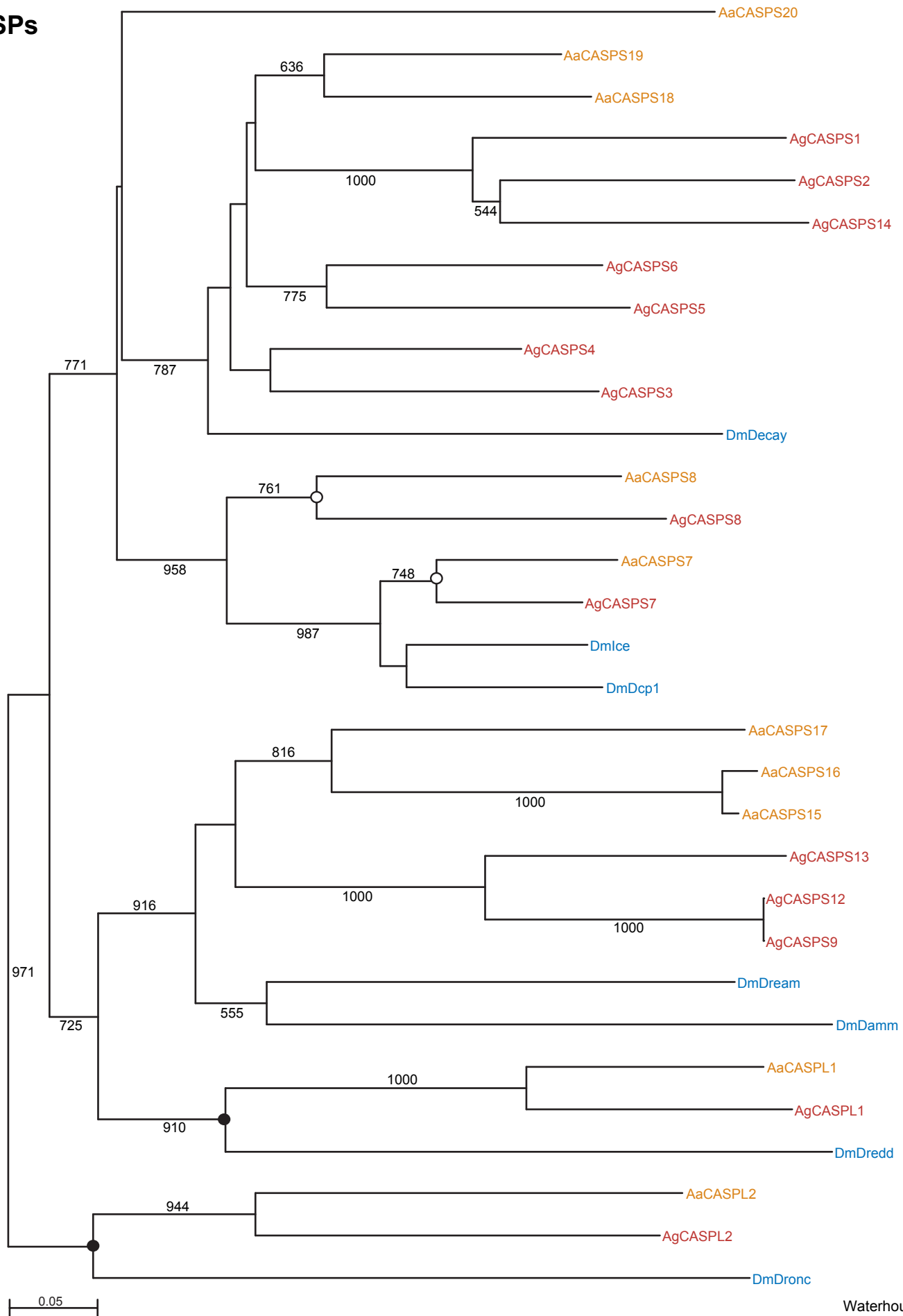
Waterhouse et al

Fig. S2
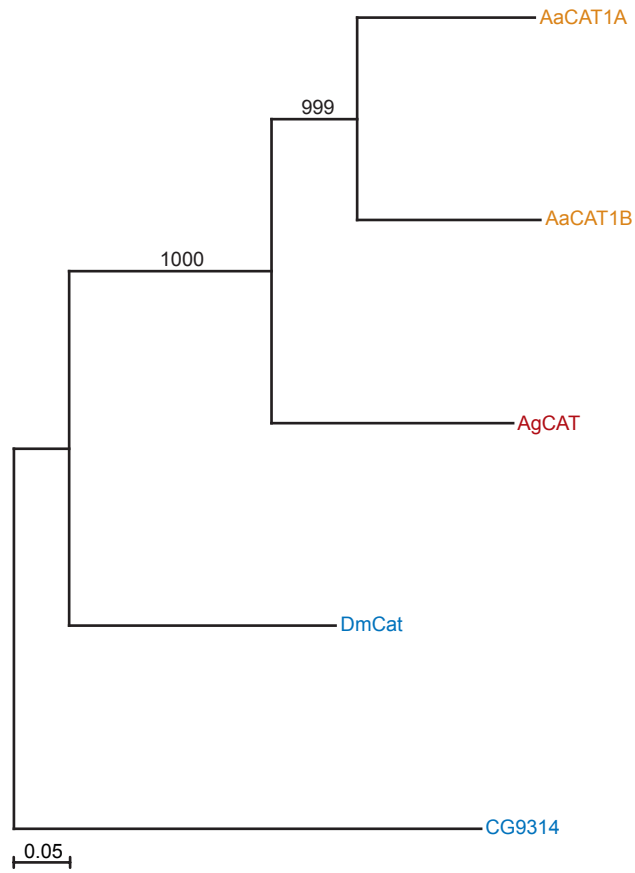**AMPs**
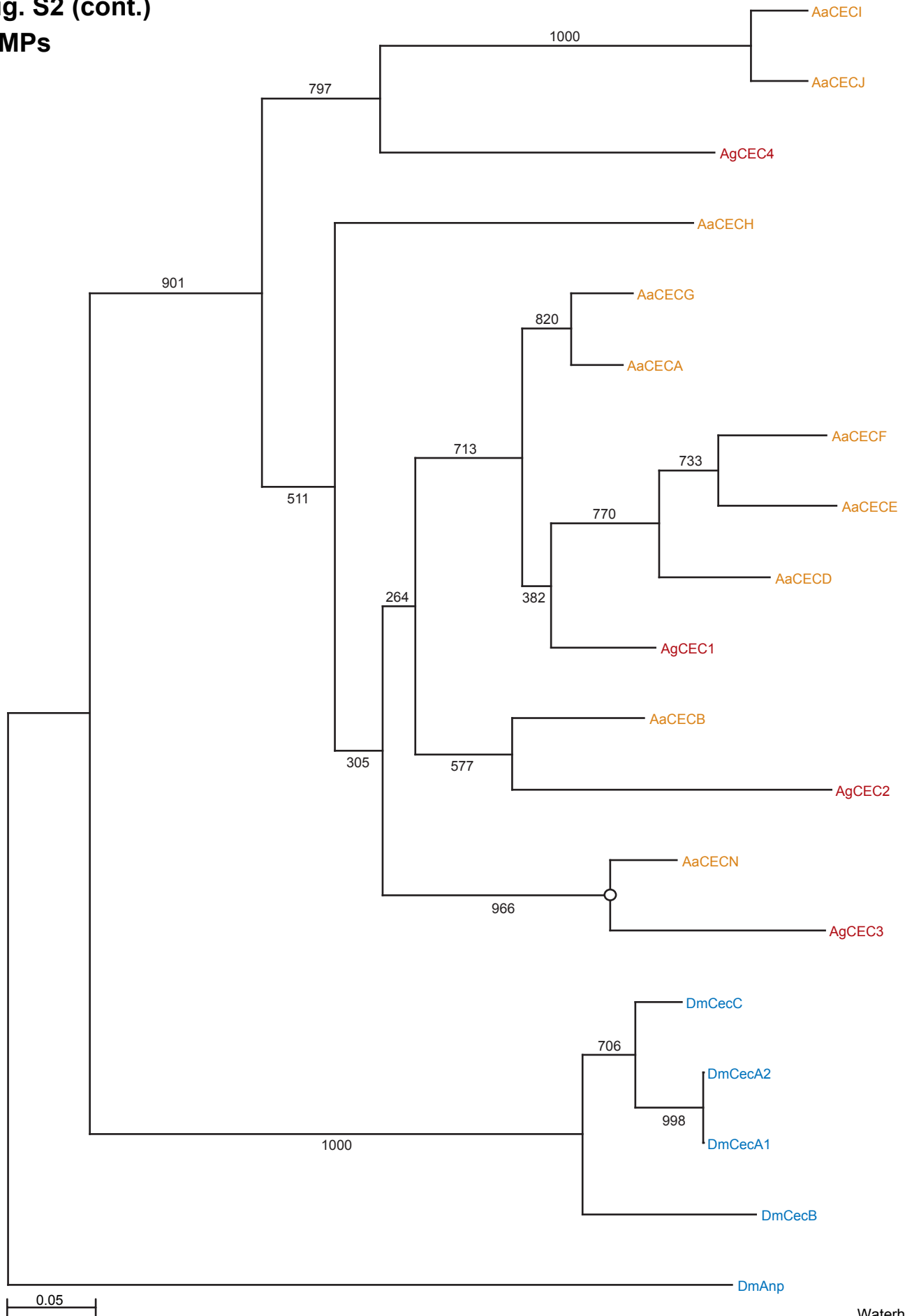
**Fig. S3
CASPs**

Fig. S4
CATs

Fig. S5 CLIPAs
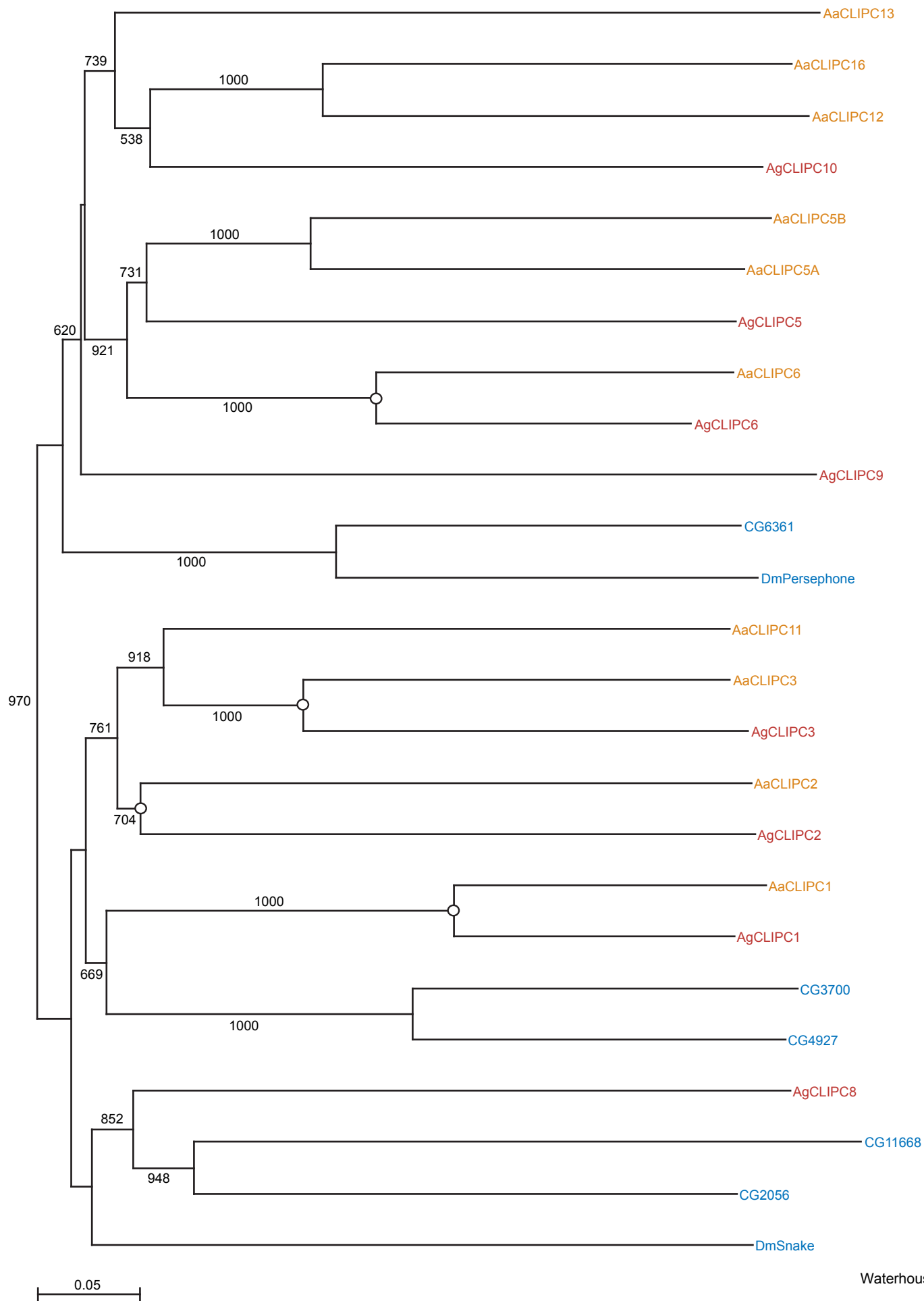
Fig. S5 (cont.) CLIPCs

Fig. S5 (cont.) CLIPEs

Fig. S6
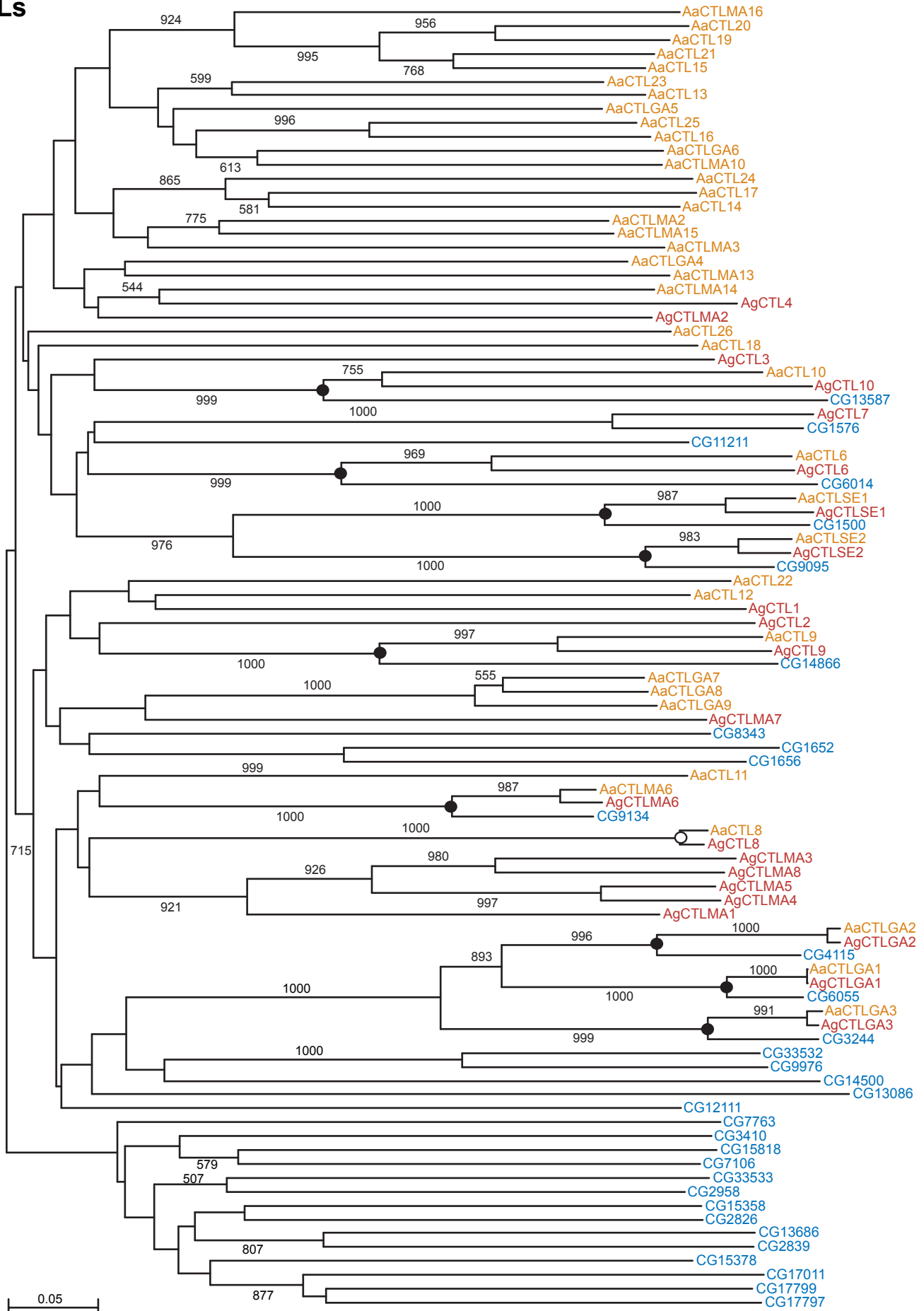## CTLs

Fig. S7 FREPs

AaFREP11
AgFREP9
AgFREP48
AgFREP64
AgFREP43
AaFREP19
AaFREP2
AgFREP3
AaFREP33
AaFREP34
AaFREP21
AaFREP22
AaFREP20
AgFREP8
AgFREP39
AgFREP46
AgFREP2
AaFREP36
AgFREP40
AgFREP44
AaFREP28
AaFREP8
AaFREP13
AaFREP7
AaFREP14
AaFREP9
AaFREP37
AaFREP27
AaFREP25
AgFREP65
AgFREP63
AgFREP11
AgFREP12
AgFREP10
AgFREP42
AgFREP34
AaFREP4
AaFREP26
AgFREP49
AgFREP14
AaFREP5
AgFREP33
AaFREP32
AaFREP29
AaFREP35
AaFREP31
AaFREP23
AaFREP10
AgFREP13
AaFREP24
AaFREP12
AaFREP16
AgFREP23
AgFREP38
AgFREP37
AgFREP30
AgFREP41
AgFREP28
AgFREP24
AgFREP25
AgFREP19
AgFREP31
AgFREP27
AgFREP5
AgFREP58
AgFREP18
AgFREP17
AgFREP6
AgFREP54
AgFREP55
AgFREP20
AgFREP7
AgFREP29
AgFREP4
AgFREP21
AgFREP32
AgFREP26
AgFREP22
AgFREP15
AaFREP18
CG7668
AaFREP17
AgFREP36
CG10359
CG1889
CG1791
CG32496
CG6788
CG31832
CG8642
CG5550
CG9500
CG30281
CG30280
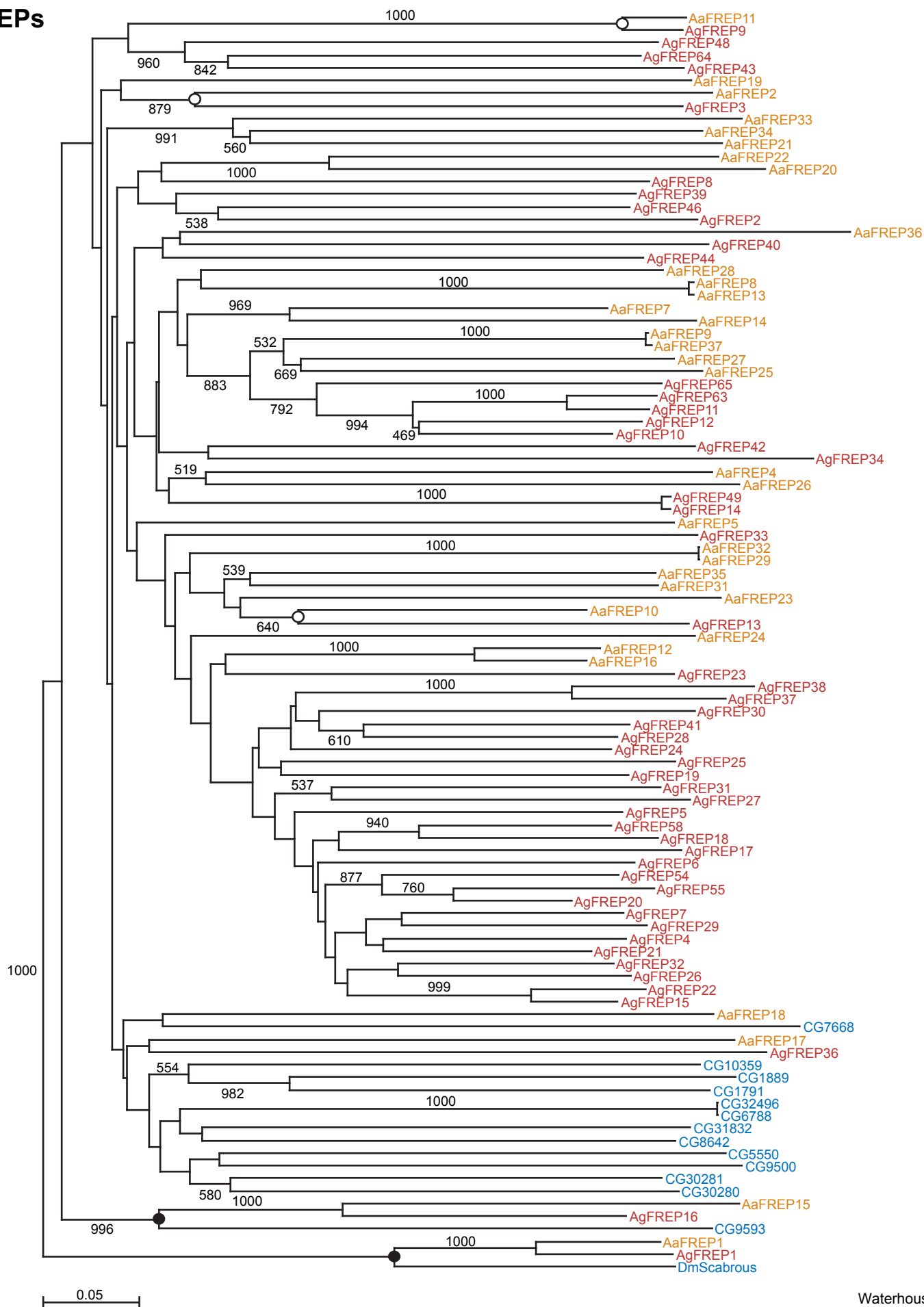AaFREP15
AgFREP16
CG9593
AaFREP1
AgFREP1
DmScabrous

0.05

34
Waterhouse et al

Fig. S8
GALEs

Fig. S9 GNBPs

Fig. S10
IAPs

AgIAP7_domain b
AgIAP4_domain b
AgIAP3_domain b
AaIAP1_domain b
AgIAP1_domain b
DmIAP1_domain b
AaIAP2_domain c
AgIAP2_domain c
DmIAP2_domain c
AaIAP5_domain a
AgIAP5_domain a
DETERIN_domain a
AaIAP6_domain a
AgIAP6_domain a
BRUCE_domain a
AgIAP4_domain a
AgIAP3_domain a
AgIAP1_domain a
AaIAP1_domain a
DmIAP1_domain a
AaIAP2_domain a
DmIAP2_domain a
AaIAP2_domain b
DmIAP2_domain b

996
998
808
998
788
991
733
1000
698
1000
561
1000
734
999
567
712
849
658

0.05

Waterhouse et al
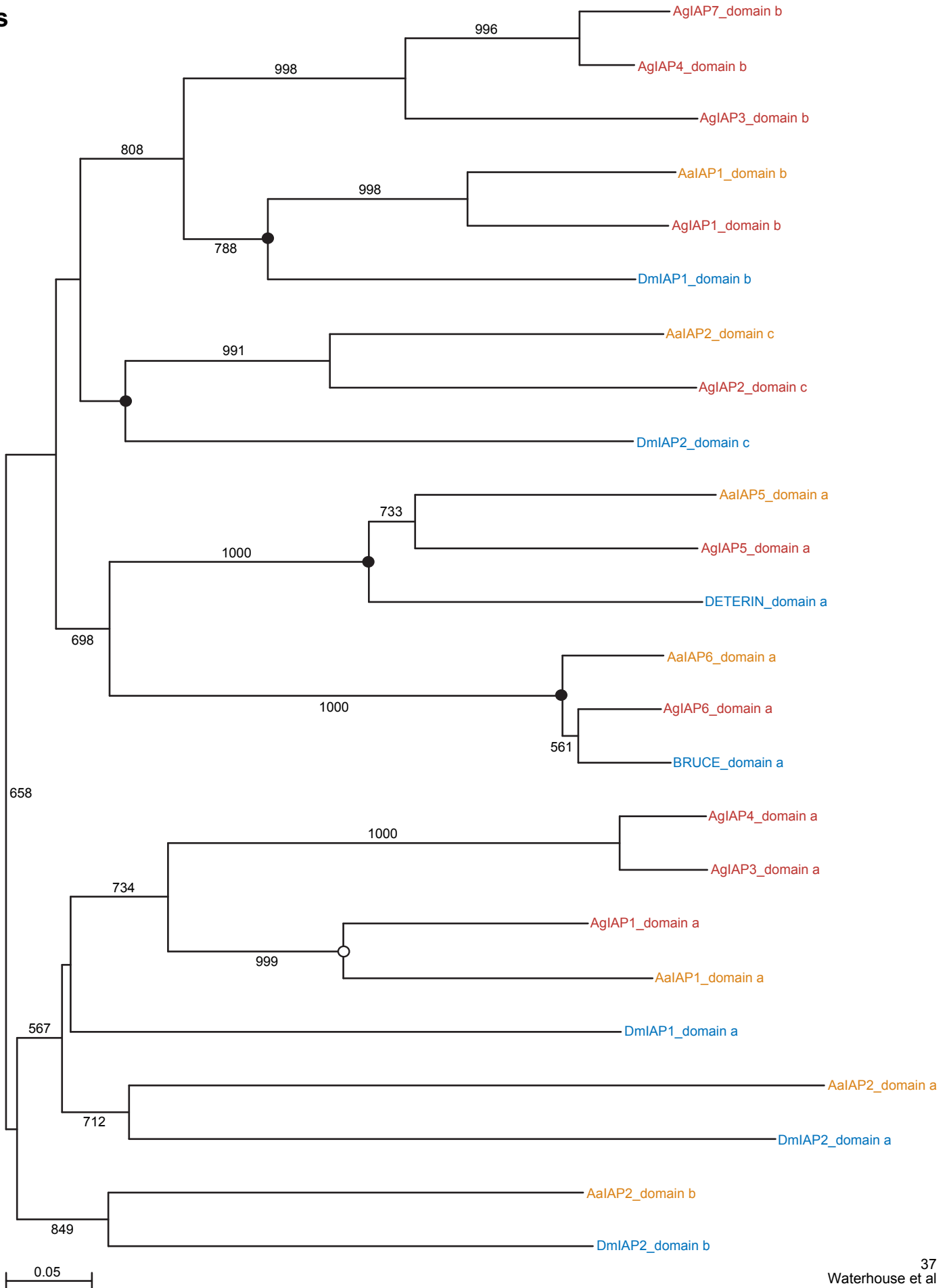
Fig. S11
LYSs

Fig. S12
MLs

Fig. S13
PGRPs

AaPGRPLE
DmPGRP-LE
AaPGRPS4
AaPGRPS5
AgPGRPS3
AgPGRPS2
DmPGRP-SD
DmPGRP-SC1a
DmPGRP-SC1b
DmPGRP-SC2
AaPGRPLB
AgPGRPLB
DmPGRP-LB
DmPGRP-SB1
DmPGRP-SB2
AaPGRPS1
AgPGRPS1
DmPGRP-SA
AaPGRPLC1
AgPGRPLC1
AgPGRPLC2
AaPGRPLC3
AgPGRPLC3
DmPGRP-LF1
DmPGRP-LCx
DmPGRP-LCy
AgPGRPLA1
AaPGRPLA2
AgPGRPLA2
DmPGRP-LA
DmPGRP-LCa
DmPGRP-LF2
AaPGRPLD
AgPGRPLD
DmPGRP-LD

0.05

40
Waterhouse et al

Fig. S14
HPXs

**Fig. S15
PPOs**

Fig. S16
RELs

**Fig. S17
SCRAs**

Fig. S18
SRPNs

Fig. S19
SPZs

AaSPZ3A
AaSPZ3B
AgSPZ3
DmSpz3
AaSPZ4
AgSPZ4
DmSpz4
AaSPZ5
AgSPZ5
DmSpz5
AaSPZ2
AgSPZ2
DmSpz2
AaSPZ1A
AaSPZ1B
AaSPZ1C
DmSpz
AaSpz6
AgSpz6
DmSpz6

980
807
1000
997
1000
769
1000
1000
1000
767
670
990
996
1000
776

0.05

50
Waterhouse et al

**Fig. S20
SODs**

AaMnSOD2

AaMnSOD1

1000

AgMnSOD1

DmSod2

0.05

AaCuSOD4

1000

AgCuSOD3_transcriptB

841

AaCuSOD3

840

AgCuSOD3_transcriptA

993

DmSod

AaCuSOD2

1000

AgCuSOD2

891

CG9027

AaCuSOD1

1000

AgCuSOD1

CG5948

0.05

Fig. S21
TEPs

AaTEP24 TE
AaTEP23
AaTEP22
AaTEP21 TE
AaTEP20 TE
AgTEP10
AgTEP9
AgTEP5
AgTEP1 TE
AgTEP6
AgTEP3
AgTEP4 TE
AaTEP15 TE
AgTEP15 TE
AgTEP2 TE
DmTep3 TE
DmTep4 TE
DmTep2 TE
DmTep1 TE
AaTEP13
AgTEP13
DmTep6

0.05

Fig. S22 TOLLs

Section 4: Tables and legends

Table S1: The table displays the number of genes in each species, the number of 1:1:1 orthologous trios and mosquito 1:1 orthologous pairs, and the total number of genes across the three species (*Drosophila melanogaster, Dm; Anopheles gambiae, Ag; Aedes aegypti, Aa*), for each immune-related gene (sub)family or signaling pathway.

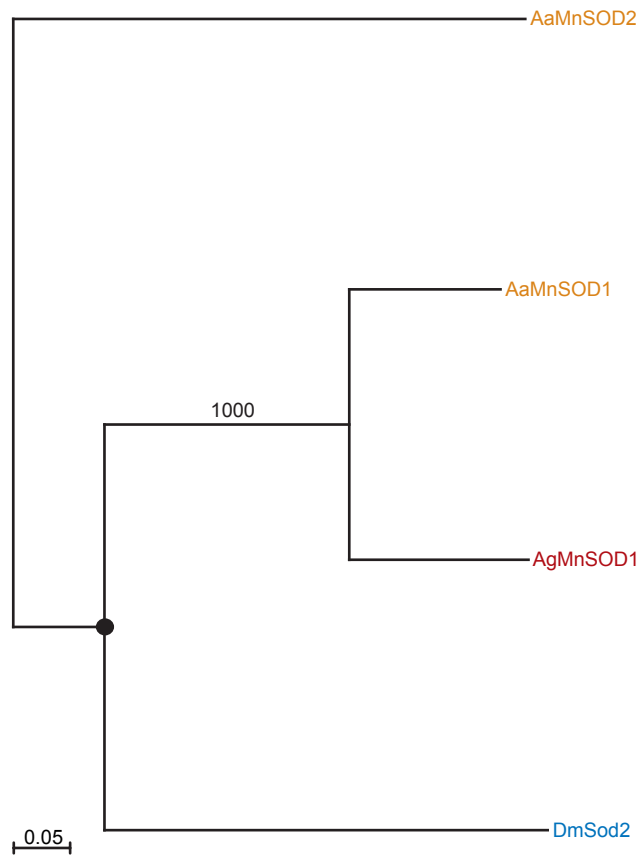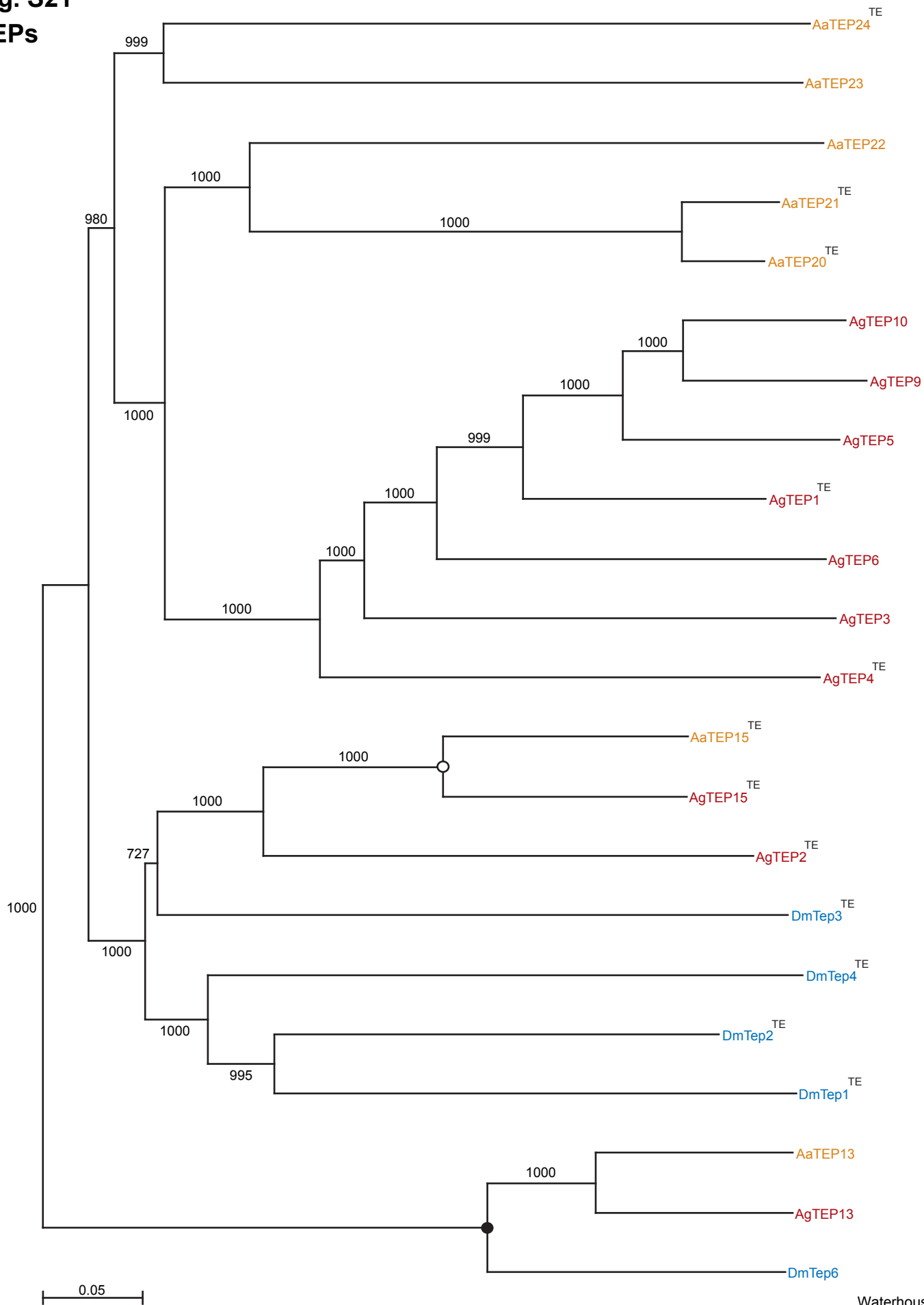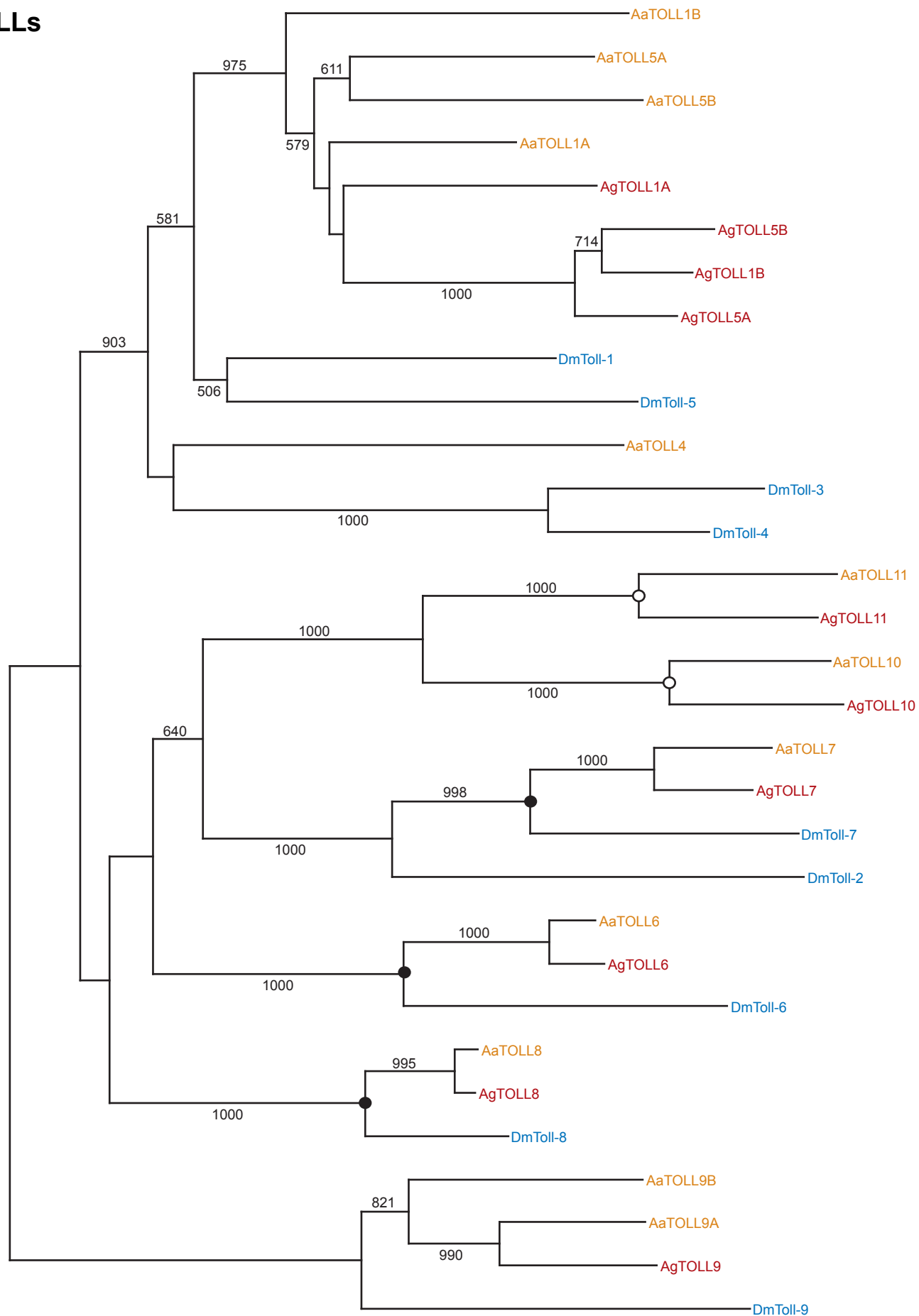| Gene Family or Pathway	Dm	Ag	Aa	1:1:1	1:1	Total
Attacins (ATTs)	4	1	1	0	0	6
Caspases (CASPs)	7	14	10	2	2	31
Catalases (CATs)	2	1	2	0	0	5
Cecropins (CECs)	5	4	10	0	1	19
CLIP-domain Serine Proteases A (CLIPAs)	11	15	6	3	1	32
CLIP-domain Serine Proteases B (CLIPBs)	14	20	36	3	5	70
CLIP-domain Serine Proteases C (CLIPCs)	7	8	12	0	4	27
CLIP-domain Serine Proteases D (CLIPDs)	10	7	8	5	0	25
CLIP-domain Serine Proteases E (CLIPEs)	3	6	5	0	0	14
C-Type Lectins (CTLs)	34	25	39	9	1	98
Defensins (DEFs)	1	4	4	0	0	9
Fibrinogen-Related proteins (FREPs)	14	61	37	2	3	112
Galectins (GALEs)	6	10	12	3	1	28
Glutathione Peroxidases (GPXs)	2	3	3	2	1	8
Gram-Negative Binding Proteins (GNBPs)	3	7	7	1	4	17
Heme Peroxidases (HPXs)	10	18	12	8	1	40
IMD Pathway Members	5	5	6	4	0	16
Inhibitors of Apoptosis (IAPs)	4	8	5	4	0	17
JAK/STAT Pathway Members	3	4	3	2	0	10
Lysozymes (LYSs)	13	8	7	1	1	28
MD2-like Proteins (MLs)	8	11	17	2	2	36
Other Anti-microbial Peptides (AMPs)*	11	1	3	0	1	15
Peptidoglycan Recognition Proteins (PGRPs)	13	7	8	5	0	28
Prophenoloxidases (PPOs)	3	9	10	0	3	22
Rel-like NFkappa-B Proteins (RELs)	3	2	3	1	0	8
Scavenger Receptors Class-A (SCRAs)	5	5	5	4	1	15
Scavenger Receptors Class-B (SCRBs)	13	13	13	8	5	39
Scavenger Receptors Class-C (SCRCs)	4	1	2	0	1	7
Serine Protease Inhibitors (SRPNs)	30	17	23	2	14	70
Späetzle-like Proteins (SPZs)	6	6	9	4	0	21
Superoxide Dismutatses (SODs)	4	5	6	4	1	15
Thio-Ester Containing Proteins (TEPs)	6	13	8	1	1	27
Thioredoxin Peroxidases (TPXs)	8	5	5	4	1	18
Toll Pathway Members	4	4	4	4	0	12
Toll Receptors (TOLLs)	9	10	12	3	2	30
Totals	**285**	**338**	**353**	**91**	**57**	**976**

* Diptericins, Drosomycins, Drosocin, Metchnikowin, Gambicin, Holotricin

Table S2: The data for the 30 most divergent and 30 most conserved groups from the Gene Ontology analysis presented in **Figure S1A**. GO accession numbers are provided with their descriptions, together with the number of proteins assigned to each term, the *Dm-Ag* and *Dm-Aa* averages, and the p-values from the Wilcoxon rank sum tests (see **methods**).

GO Accession	Number of Members	*Dm-Ag* Distance	*Dm-Aa* Distance	GO Term Description	p-value (*Ag*)	p-value (*Aa*)
GO:0051053	11	1.3782	1.2829	negative regulation of DNA metabolic process	7.076e-03	1.368e-02
GO:0002252	13	1.2794	1.2260	immune effector process	1.462e-02	1.955e-02
GO:0007131	11	1.2499	1.2732	meiotic recombination	4.977e-03	6.515e-03
GO:0006518	11	1.2392	1.1994	peptide metabolic process	8.306e-03	7.180e-03
GO:0000077	11	1.2024	1.2379	DNA damage checkpoint	2.835e-02	3.511e-02
GO:0007280	11	1.1956	1.0541	pole cell migration	8.456e-02	3.440e-01
GO:0005179	16	1.1845	1.1630	hormone activity	9.399e-03	4.870e-03
GO:0003725	15	1.1480	1.0288	double-stranded RNA binding	2.355e-02	8.486e-02
GO:0007259	11	1.1438	1.0895	JAK-STAT cascade	2.164e-01	3.037e-01
GO:0007291	12	1.1427	1.0922	sperm individualization	1.175e-01	1.773e-01
GO:0006275	13	1.1329	1.0565	regulation of DNA replication	3.137e-01	4.023e-01
GO:0046620	16	1.1276	1.1255	regulation of organ size	3.366e-02	2.969e-02
GO:0000059	12	1.1121	0.9778	protein import into nucleus, docking	5.294e-02	1.703e-01
GO:0000793	17	1.0832	1.0088	condensed chromosome	1.092e-02	4.460e-02
GO:0003684	12	1.0827	1.0411	damaged DNA binding	5.954e-02	6.862e-02
GO:0005044	12	1.0682	1.0341	scavenger receptor activity	4.327e-02	3.769e-02
GO:0007031	12	1.0600	1.0445	peroxisome organization and biogenesis	1.617e-02	2.252e-02
GO:0004520	15	1.0541	1.0070	Endodeoxyribonuclease activity	2.059e-02	2.508e-02
GO:0042742	16	1.0535	1.0299	defense response to bacterium	1.092e-01	1.079e-01
GO:0007530	22	1.0350	1.0005	sex determination	6.038e-02	6.151e-02
GO:0008083	14	1.0292	0.9510	growth factor activity	1.173e-01	2.151e-01
GO:0020037	21	1.0258	1.0215	heme binding	2.201e-01	2.092e-01
GO:0016705	15	1.0172	1.0824	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	9.485e-02	6.442e-02
GO:0016324	11	0.9860	0.9476	apical plasma membrane	5.356e-01	5.041e-01
GO:0005778	11	0.9848	0.9391	peroxisomal membrane	1.335e-01	1.586e-01
GO:0019730	25	0.9842	0.9334	antimicrobial humoral response	2.076e-01	1.879e-01
GO:0008063	15	0.9809	0.9092	Toll signaling pathway	3.043e-01	3.980e-01
GO:0000922	14	0.9780	0.9075	spindle pole	1.453e-01	2.725e-01
GO:0042078	16	0.9550	0.9085	germ-line stem cell division	9.798e-01	7.835e-01

GO:0016319	17	0.4296	0.4597	mushroom body development	3.631e-04	2.965e-03
GO:0017157	33	0.4272	0.4128	regulation of exocytosis	1.141e-07	5.927e-08
GO:0045211	15	0.4270	0.4194	postsynaptic membrane	3.174e-04	5.058e-04
GO:0006898	34	0.4260	0.4285	receptor-mediated endocytosis	1.656e-07	3.080e-07
GO:0043176	12	0.4254	0.4119	amine binding	2.625e-03	2.983e-03
GO:0009408	12	0.4233	0.4201	response to heat	1.790e-03	2.514e-03
GO:0007222	14	0.4219	0.4246	frizzled signaling pathway	1.330e-03	2.162e-03
GO:0005262	11	0.4188	0.4246	calcium channel activity	2.846e-03	5.688e-03
GO:0006936	53	0.4173	0.4067	muscle contraction	1.298e-12	1.364e-11
GO:0007270	24	0.4080	0.3994	nerve-nerve synaptic transmission	3.564e-06	6.360e-06
GO:0006073	11	0.4069	0.3766	glucan metabolic process	2.028e-03	1.431e-03
GO:0003924	79	0.3943	0.3698	GTPase activity	1.629e-02	1.629e-02
GO:0005525	107	0.3933	0.3787	GTP binding	3.705e-10	3.705e-10
GO:0006471	13	0.3922	0.3920	protein amino acid ADP-ribosylation	9.168e-04	1.417e-03
GO:0016917	11	0.3890	0.3950	GABA receptor activity	1.468e-03	2.539e-03
GO:0000271	14	0.3884	0.3561	polysaccharide biosynthetic process	1.958e-04	1.047e-04
GO:0008553	20	0.3833	0.3886	hydrogen-exporting ATPase activity, phosphorylative mechanism	8.107e-06	2.152e-05
GO:0016185	12	0.3760	0.3891	synaptic vesicle budding	1.059e-04	1.354e-04
GO:0010033	14	0.3689	0.3414	response to organic substance	1.932e-04	1.393e-04
GO:0009532	11	0.3568	0.3166	plastid stroma	3.279e-04	1.304e-04
GO:0008064	12	0.3361	0.3150	regulation of actin polymerization and/or depolymerization	1.582e-04	1.054e-04
GO:0006510	12	0.3355	0.3095	ATP-dependent proteolysis	1.165e-04	6.695e-05
GO:0004693	11	0.3329	0.3388	cyclin-dependent protein kinase activity	1.657e-04	3.101e-04
GO:0005838	17	0.3313	0.3175	proteasome regulatory particle (sensu Eukaryota)	4.404e-06	4.319e-06
GO:0050803	11	0.3264	0.3183	regulation of synapse structure and activity	1.404e-04	1.663e-04
GO:0044449	11	0.3225	0.3772	contractile fiber part	1.020e-04	1.083e-03
GO:0045172	13	0.2905	0.2970	germline ring canal	1.120e-05	2.190e-05
GO:0031202	15	0.2705	0.2838	RNA splicing factor activity, transesterification mechanism	7.358e-07	3.465e-06
GO:0005843	16	0.2620	0.2458	cytosolic small ribosomal subunit (sensu Eukaryota)	1.390e-07	1.074e-07
GO:0016471	12	0.2426	0.2538	hydrogen-translocating V-type ATPase complex	4.148e-06	9.518e-06

Section 5: References

S1. S. F. Altschul *et al.*, *Nucl. Acids Res.* **25**, 3389 (1997)
S2. A. A. Salamov, V. V. Solovyev, *Genome Res.* **10**, 516 (2000)
S3. T. F. Smith, M. S. Waterman, *Journal of Molecular Biology* **147**, 195 (1981)
S4. E. M. Zdobnov, P. Bork, *Trends Genet* (2006)
S5. I. Dondoshansky, Y. Wolf. (unpublished, http://biowulf.nih.gov/apps/blast/doc/blastclust.html).
S6. S. F. Altschul *et al.*, *Journal of Molecular Biology* **215**, 403 (1990)
S7. E. M. Zdobnov, R. Apweiler, *Bioinformatics* **17**, 847 (2001)
S8. I. Letunic *et al.*, *Nucl. Acids Res.* **34**, D257 (2006)
S9. R. D. Finn *et al.*, *Nucl. Acids Res.* **34**, D247 (2006)
S10. N. Hulo *et al.*, *Nucl. Acids Res.* **34**, D227 (2006)
S11. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res* **22**, 4673 (1994)
S12. S. R. Eddy, (2001)
S13. R. C. Edgar, *Nucl. Acids Res.* **32**, 1792 (2004)
S14. J. Castresana, *Mol Biol Evol* **17**, 540 (2000)
S15. S. Guindon, O. Gascuel, *Syst Biol* **52**, 696 (2003)
S16. P. Bulet, R. Stocklin, L. Menin, *Immunol Rev* **198**, 169 (2004)
S17. C. Hetru, L. Troxler, J. A. Hoffmann, *J Infect Dis* **187 Suppl 2**, S327 (2003)
S18. G. K. Christophides, D. Vlachou, F. C. Kafatos, *Immunol Rev* **198**, 127 (2004)
S19. F. Leulier *et al.*, *EMBO Rep* **1**, 353 (2000)
S20. V. Gesellchen *et al.*, *EMBO Rep* **6**, 979 (2005)
S21. A. Kleino *et al.*, *Embo J* **24**, 3423 (2005)
S22. P. Chen *et al.*, *Dev Biol* **201**, 202 (1998)
S23. A. Rodriguez *et al.*, *Nat Cell Biol* **1**, 272 (1999)
S24. S. Hu, X. Yang, *J Biol Chem* **275**, 30761 (2000)
S25. C. M. Griswold *et al.*, *Genetics* **134**, 781 (1993)
S26. H. Jiang, M. R. Kanost, *Insect Biochem Mol Biol* **30**, 95 (2000)
S27. G. K. Christophides *et al.*, *Science* **298**, 159 (2002)
S28. J. Ross *et al.*, *Gene* **304**, 117 (2003)
S29. M. R. Kanost, H. Jiang, X. Q. Yu, *Immunol Rev* **198**, 97 (2004)
S30. H. Tang *et al.*, *J Biol Chem* **281**, 28097 (2006)
S31. Z. Kambris *et al.*, *Curr Biol* **16**, 808 (2006)
S32. I. H. Jang *et al.*, *Dev Cell* **10**, 45 (2006)
S33. X. Q. Yu *et al.*, *Insect Biochem Mol Biol* **33**, 197 (2003)
S34. J. Volz *et al.*, *Cell Microbiol* **8**, 1392 (2006)
S35. K. Drickamer, M. E. Taylor, *Annu Rev Cell Biol* **9**, 237 (1993)
S36. K. Drickamer, A. J. Fadden, *Biochem Soc Symp*, 59 (2002)
S37. T. Jomori, S. Natori, *FEBS Lett* **296**, 283 (1992)
S38. X. Q. Yu, M. R. Kanost, *Dev Comp Immunol* **27**, 189 (2003)
S39. N. Koizumi *et al.*, *FEBS Lett* **443**, 139 (1999)
S40. C. Chen *et al.*, *Biochem J* **310 (Pt 1)**, 23 (1995)
S41. X. Q. Yu, H. Gan, M. R. Kanost, *Insect Biochem Mol Biol* **29**, 585 (1999)
S42. X. Q. Yu, M. R. Kanost, *J Biol Chem* **275**, 37373 (2000)
S43. M. A. Osta, G. K. Christophides, F. C. Kafatos, *Science* **303**, 2030 (2004)
S44. E. Warr *et al.*, *Insect Biochem Mol Biol* **36**, 769 (2006)
S45. L. A. Leshko-Lindsay, V. G. Corces, *Development* **124**, 169 (1997)
S46. E. M. Zdobnov *et al.*, *Science* **298**, 149 (2002)
S47. X. Wang *et al.*, *Mol Biochem Parasitol* **139**, 65 (2005)
S48. G. Dimopoulos *et al.*, *Proc Natl Acad Sci U S A* **99**, 8814 (2002)
S49. G. Dimopoulos *et al.*, *Proc Natl Acad Sci U S A* **97**, 6619 (2000)
S50. C. M. Adema *et al.*, *Proc Natl Acad Sci U S A* **94**, 8691 (1997)
S51. S. Gokudan *et al.*, *Proc Natl Acad Sci U S A* **96**, 10086 (1999)
S52. N. Kairies *et al.*, *Proc Natl Acad Sci U S A* **98**, 13519 (2001)
S53. Y. Dong *et al.*, *PLoS Pathog* **2**, e52 (2006)
S54. K. E. Pace, L. G. Baum, *Glycoconj J* **19**, 607 (2004)
S55. K. E. Pace *et al.*, *J Biol Chem* **277**, 13091 (2002)
S56. G. Dimopoulos *et al.*, *Proc Natl Acad Sci U S A* **93**, 13066 (1996)
S57. G. Dimopoulos *et al.*, *Embo J* **17**, 6115 (1998)
S58. G. Dimopoulos *et al.*, *Curr Opin Immunol* **13**, 79 (2001)

S59. R. Zhang *et al.*, *J Biol Chem* **278**, 42072 (2003)
S60. N. Yahata *et al.*, *Gene* **86**, 113 (1990)
S61. V. Gobert *et al.*, *Science* **302**, 2126 (2003)
S62. L. Wang *et al.*, *Embo J* **25**, 5005 (2006)
S63. Q. L. Deveraux, J. C. Reed, *Genes Dev* **13**, 239 (1999)
S64. E. De Gregorio *et al.*, *Proc Natl Acad Sci U S A* **98**, 12590 (2001)
S65. K. Roxstrom-Lindquist, O. Terenius, I. Faye, *EMBO Rep* **5**, 207 (2004)
S66. R. J. Ursic Bedoya *et al.*, *Insect Mol Biol* **14**, 89 (2005)
S67. S. Daffre *et al.*, *Mol Gen Genet* **242**, 152 (1994)
S68. B. Li *et al.*, *Gene* **360**, 131 (2005)
S69. N. Inohara, G. Nunez, *Trends Biochem Sci* **27**, 219 (2002)
S70. K. A. Fitzgerald, D. C. Rowe, D. T. Golenbock, *Microbes Infect* **6**, 1361 (2004)
S71. M. Fujihara *et al.*, *Pharmacol Ther* **100**, 171 (2003)
S72. S. Kumar *et al.*, *Proc Natl Acad Sci U S A* **100**, 14139 (2003)
S73. A. Donko *et al.*, *Philos Trans R Soc Lond B Biol Sci* **360**, 2301 (2005)
S74. E. M. Ha *et al.*, *Science* **310**, 847 (2005)
S75. S. Kumar *et al.*, *J Biol Chem* **279**, 53475 (2004)
S76. S. M. Kanzok *et al.*, *Science* **291**, 643 (2001)
S77. F. Missirlis *et al.*, *Biol Chem* **384**, 463 (2003)
S78. M. Krieger, J. Herz, *Annu Rev Biochem* **63**, 601 (1994)
S79. P. J. Gough, S. Gordon, *Microbes Infect* **2**, 305 (2000)
S80. L. Peiser, S. Gordon, *Microbes Infect* **3**, 149 (2001)
S81. L. Peiser, S. Mukhopadhyay, S. Gordon, *Curr Opin Immunol* **14**, 123 (2002)
S82. N. C. Franc, K. White, *Microbes Infect* **2**, 243 (2000)
S83. A. Pearson, A. Lux, M. Krieger, *Proc Natl Acad Sci U S A* **92**, 4056 (1995)
S84. M. Ramet *et al.*, *Immunity* **15**, 1027 (2001)
S85. N. C. Franc *et al.*, *Immunity* **4**, 431 (1996)
S86. N. C. Franc *et al.*, *Science* **284**, 1991 (1999)
S87. J. A. Philips, E. J. Rubin, N. Perrimon, *Science* **309**, 1251 (2005)
S88. L. Peiser *et al.*, *Infect Immun* **68**, 1953 (2000)
S89. S. Kim *et al.*, *Microb Pathog* **37**, 11 (2004)
S90. L. M. Pierini, *Cell Microbiol* **8**, 1361 (2006)
S91. L. Serghides *et al.*, *Trends Parasitol* **19**, 461 (2003)
S92. S. N. Patel *et al.*, *J Infect Dis* **189**, 204 (2004)
S93. F. M. Ndungu *et al.*, *Parasite Immunol* **27**, 373 (2005)
S94. P. G. Gettins, *Chem Rev* **102**, 4751 (2002)
S95. J. Potempa, E. Korzus, J. Travis, *J Biol Chem* **269**, 15957 (1994)
S96. P. Ligoxygakis, S. Roth, J. M. Reichhart, *Curr Biol* **13**, 2097 (2003)
S97. E. De Gregorio *et al.*, *Dev Cell* **3**, 581 (2002)
S98. P. Ligoxygakis *et al.*, *Science* **297**, 114 (2002)
S99. E. A. Levashina *et al.*, *Science* **285**, 1917 (1999)
S100. M. Oley, M. C. Letzel, H. Ragg, *FEBS Lett* **577**, 165 (2004)
S101. M. J. Richer *et al.*, *Proc Natl Acad Sci U S A* **101**, 10560 (2004)
S102. E. G. Abraham *et al.*, *Proc Natl Acad Sci U S A* **102**, 16327 (2005)
S103. K. Michel *et al.*, *EMBO Rep* **6**, 891 (2005)
S104. E. A. Levashina *et al.*, *Cell* **104**, 709 (2001)
S105. S. Blandin *et al.*, *Cell* **116**, 661 (2004)
S106. L. F. Moita *et al.*, *Immunity* **23**, 65 (2005)
S107. C. Hashimoto, K. L. Hudson, K. V. Anderson, *Cell* **52**, 269 (1988)
S108. B. Lemaitre *et al.*, *Cell* **86**, 973 (1996)
S109. C. Luo, L. Zheng, *Immunogenetics* **51**, 92 (2000)
S110. S. Tauszig *et al.*, *Proc Natl Acad Sci U S A* **97**, 10520 (2000)
S111. J. Y. Ooi *et al.*, *EMBO Rep* **3**, 82 (2002)
S112. C. Luna *et al.*, *Insect Biochem Mol Biol* **32**, 1171 (2002)