The Dscam Homologue of the Crustacean *Daphnia* Is Diversified by Alternative Splicing Like in Insects

Daniela Brites,[1] *Seanna McTaggart,*†[1] *Krystalynne Morris,*‡ *Jobriah Anderson,*‡ *Kelley Thomas,*‡ *Isabelle Colson,** *Thomas Fabbro,** *Tom J. Little,*† *Dieter Ebert,** *and Louis Du Pasquier**

*Zoologisches Institut, Evolutionsbiologie, Universität Basel, Vesalgasse 1, Switzerland; †Institute for Evolutionary Biology, School of Biological Sciences, Kings Buildings, University of Edinburgh, Edinburgh, United Kingdom; and ‡Hubbard Center for Genome Studies, University of New Hampshire

In insects, the homologue of the Down syndrome cell adhesion molecule (Dscam) is a unique case of a single-locus gene whose expression has extensive somatic diversification in both the nervous and immune systems. How this situation evolved is best understood through comparative studies. We describe structural, expression, and evolutionary aspects of a Dscam homolog in 2 species of the crustacean *Daphnia*. The Dscam of *Daphnia* generates up to 13,000 different transcripts by the alternative splicing of variable exons. This extends the taxonomic range of a highly diversified Dscam beyond the insects. Additionally, we have identified 4 alternative forms of the cytoplasmic tail that generate isoforms with or without inhibitory or activating immunoreceptor tyrosine–based motifs (ITIM and ITAM respectively), something not previously reported in insect's Dscam. In *Daphnia*, we detected exon usage variability in both the brain and hemocytes (the effector cells of immunity), suggesting that Dscam plays a role in the nervous and immune systems of crustaceans, as it does in insects. Phylogenetic analysis shows a high degree of amino acid conservation between *Daphnia* and insects except in the alternative exons, which diverge greatly between these taxa. Our analysis shows that the variable exons diverged before the split of the 2 *Daphnia* species and is in agreement with the nearest-neighbor model for the evolution of the alternative exons. The genealogy of the Dscam gene family from vertebrates and invertebrates confirmed that the highly diversified form of the gene evolved from a nondiversified form before the split of insects and crustaceans.

Introduction

The Down syndrome cell adhesion molecule (Dscam) belongs to a family of cell membrane molecules involved in the differentiation of the nervous system. As with some other members of the family (e.g., axonin, roundabout, neural cell adhesion molecule [NCAM], contactin, and L1 cell adhesion molecule [L1CAM]), the extracellular region of Dscam is made of immunoglobulin (Ig) and fibronectin (FN) domains. Throughout the metazoa, the bona fide Dscam domain composition and physical arrangement remains identical, namely, 9(Ig)-4(FN)-(Ig)-2(FN) (Shapiro et al. 2007).

For mammals and insects whose genome sequences are available, additional Dscam gene copies may be found. For example, humans have 2 gene copies, Dscam and the paralogue Dscam-like1 (Dscam-L1) (Yamakawa et al. 1998; Agarwala et al. 2001). Insects also have Dscam and several Dscam paralogs that have been named Dscam-L (Schmucker et al. 2000; Millard et al. 2007). In humans, the Dscam gene can generate 3 different transcripts through cryptic splicing sites in the gene (Yamakawa et al. 1998). In contrast, the *Drosophila* Dscam, but not Dscam-L, has the potential to generate over 38,000 different transcripts (Schmucker et al. 2000). This unprecedented repertoire of transcripts is due to 4 arrays of alternative exons that are spliced together in a mutually exclusive manner. The alternative exons encode the first half of the second and third Ig domains, the entire seventh Ig domain, and the transmembrane segment.

In insects, many different isoforms of Dscam play an essential role in growth and the directed extension of axon branches (Schmucker et al. 2000; Chen et al. 2006; Hattori et al. 2007). Biochemical studies support a model in which each isoform preferentially binds to the same isoform on opposing cell surfaces, providing neurons with a homolog interaction recognition system (Wojtowicz et al. 2004). In *Drosophila*, the diversity of Dscam isoforms is necessary for neural wiring specificity (Chen et al. 2006; Hattori et al. 2007) and is also thought to be important in insect immunity. For example, Dscam transcripts are found in hemocytes, in cells from the fat body, a central organ involved in immunity, and soluble Dscam molecules are present in the hemolymph serum (Watson et al. 2005). Additionally, the silencing of Dscam by interference RNA (RNAi) reduces the ability of *Drosophila* hemocytes to phagocytose by ~60% (Watson et al. 2005), whereas in mosquitoes it results in reduced survival after pathogen exposure (Dong et al. 2006). Watson et al. (2005) demonstrated that Dscam binds to bacteria and that this capacity varies among isoforms. Finally, different splice variant repertoires are expressed between pathogen-challenged and pathogen-unchallenged mosquitoes and cell lines (Dong et al. 2006).

A Dscam gene with alternative spliced exons generating 3 hypervariable Ig domains has evolved in several insect orders over ~250 Myr (Graveley et al. 2004; Watson et al. 2005). The origin of the alternative spliced exons remains elusive as, generally, no homology was found outside of insects (Crayton et al. 2006). Here we describe a homolog of a diversified Dscam in the branchiopod crustacean *Daphnia*. *Daphnia* reproduces mostly clonally, which permits us to study Dscam expression with strict control of the genetic background. The Dscam gene was studied in 2 different species, *Daphnia magna* and *Daphnia pulex*, which are thought to have diverged approximately 200 MYA (Colbourne and Hebert 1996). Recent studies suggest that hexapodes (arthropods having 6 legs, including insects) and branchiopod crustaceans are sister groups that shared a common ancestor around 420 MYA (Glenner et al.

[1] These authors contributed equally to this work.

Key words: Dscam evolution, immunoglobulin domains, somatic diversity, immunity, nervous system, arthropoda.

E-mail: d.brites@unibas.ch.

Mol. Biol. Evol. 25(7):1429–1439. 2008
doi:10.1093/molbev/msn087
Advance Access publication April 9, 2008

2006). Thus, the description and phylogenetic comparison of the Dscam gene across insects and crustaceans can provide insights into the evolution of the gene and the origin of its dual function in the nervous and immune systems. Furthermore, closer examination of the patterns of sequence evolution of the alternative exons within and between species provides insights into the evolution of the alternative exons.

Materials and Methods

Gene Recovery

We used insect Dscam protein sequences to probe the *D. pulex arenata* 10× scaffolding (http://daphnia.cgb.indiana.edu/) using TBlastN (Altschul et al. 1997). We extracted the region of scaffolding corresponding to significant matches, plus an additional 2,000 nt up- and downstream. This sequence was manually annotated in Artemis (http://www.sanger.ac.uk/Software/Artemis) using Blast high scoring segment pairs from the initial TBlastN search, in addition to those obtained from BlastP searches of the open reading frames of the target scaffold sequence in all 3 frames of the translated sequence, %GC content, and the identification of GT–AG boundaries that frame introns. We used the annotated gene as a new query amino acid sequence to search the *Daphnia* genome assembly for any additional copies.

We accepted genes as Dscam paralogs if, according to the SMART database, their extracellular Dscam domain structure was 9(Ig)-4(FN)-(Ig)-2(FN). The genome of *D. pulex* contains 2 regions with homology to nonvariable Dscam genes. One of these lacks 2 Ig domains, the transmembrane segment, the cytoplasmic tail, and the initiator methionine could not be identified. The second region lacks 1 Ig and 1 FN domain. The National Center for Biotechnology Information (NCBI) database was searched for additional putative Dscam homologs and paralogs (species GenBank accession numbers provided in the Supplementary Material online). In *Drosophila*, 4 Dscam members have been reported (Millard et al. 2007): the canonical variable Dscam (aaf71926.1) and the putative paralogues cg31190 (Dscam-L1), cg32387 (Dscam-L2), and cg33274. Only Dscam-L2 has a canonical Dscam domain structure and 2 alternatively spliced exons coding for the Ig7 domain of the molecule. The predicted structure of cg33274 lacks 1 Ig domain and thus was excluded from further analysis. The presence of the first FN domain of Dscam-L1 is ambiguous; however, the length of the gene is compatible with a full Dscam gene. Therefore, we included Dscam-L1 and Dscam-L2 in the Dscam paralog analysis.

We also sequenced Dscam from another *Daphnia* species, *D. magna*. Dscam genomic sequences were obtained from a fosmid library (for details, see Supplementary Material online). Additional genomic and cDNA data were generated from a single clonal line (clone Mu11, originally isolated from a pond near Munich, Germany). Further Dscam cDNA was obtained from hemocytes of individuals fom the genetic line HO2 (originally isolated from a pond in Hungary), which were infected with the pathogenic bacteria *Pasteuria ramosa* (Ebert et al. 1996).

RNA Extraction and cDNA Synthesis

Daphnia magna and *D. pulex* messenger RNA (mRNA) extractions were carried out with Dynalbeads technology (Dynalbeads mRNA Direct Micro kit) following the manufacturer's instructions. For whole-body mRNA preparation, mRNA was eluted in 6 μl of 10 mM Tris–HCl and used to synthesize cDNA directly or frozen at −80 °C. To obtain mRNA from hemocytes, single individuals were immobilized in microtest plates (Terasaki microtiter plates, Greiner Bio-One, Frickenhause, Germany) with a drop of 0.75% agar at 37 °C. Hemolymph was withdrawn by capillary action, with twice-pulled microcapillary glass tubes (Harvard apparatus GC100TF-10) inserted into the heart chamber, and brains were dissected. Both tissue types were immediately stored in RNA*later* (Ambion, Rotkreutz, Switzerland) solution.

To obtain the 5′ region of Dscam mRNA, we used SMART technology (SMART RACE cDNA Amplification Kit, Clontech, Sant-Germain-en-Laye, France) on mRNA samples extracted from whole *D. magna*. We used 3 μl of eluted mRNA with 2 reverse primers (primer sequences available upon request) specific to the Ig1 and Ig4 exons of *D. magna*. The remainders of the cDNA sequences were synthesized in a 20 μl reverse transcription (RT) reaction consisting of 2 μl of SuperScriptIII Reverse Transcriptase (Invitrogen) and 1 μl of oligo(dT) (50 μM), following the instructions of the manufacturer. In the RT reactions, either 3 μl of mRNA were used or, in the case of hemocyte and brain preparations, the whole mRNA samples were used directly to make solid-phase first–strand cDNA libraries.

Polymerase Chain Reaction, Cloning, and Sequencing

To obtain the first Dscam DNA sequences from *D. magna*, oligonucleotide primer pairs were designed using the *D. pulex* sequence in regions with high amino acid conservation among *D. pulex* and several insect species. Polymerase chain reaction (PCR) was carried out using the BD Advantage 2 PCR Kit on 1 μl of cDNA according to the manufacturer's directions. Several PCRs were required in order to complete the cDNA sequence (primer sequences and PCR conditions available upon request). To obtain the cDNA sequence of Ig2, Ig3, and Ig7 variable domains, we PCR amplified the first-strand cDNA libraries prepared with the mRNA isolated from hemocytes and brain. Fifteen microliters of the total 20 μl RT reaction were washed twice in 1× PCR buffer. The beads were combined with the PCR master mix, and the reactions were submitted to the following PCR conditions: 95 °C for 1 min, 2 cycles of: 57 °C for 30 s, 72 °C for 5 min, and 94 °C for 2 min. The beads were then removed from the reactions, and the PCR proceeded as above for 35 cycles, except that the 72 °C step was changed to 90 s. The PCR products were gel purified (QIAquick Gel Extraction kit; Qiagen, Basel, Switzerland) prior to cloning.

Most of the PCR products were cloned in the pCR 2.1-TOPO vector (Invitrogen). Due to the large size of the PCR product from the 3′ rapid amplification of cDNA ends, it was cloned into a pCR-XL-TOPO vector (Invitrogen, Basel,

Switzerland). All cloned products were sequenced under Big Dye terminator conditions, using the M13 reverse and/or M13 forward primers. For the PCR products that contained variable exons, several colonies were sequenced.

To test whether the exons from arrays 4, 6, and 11 are randomly expressed, we compared the observed frequency of the sequenced exons with the expected frequency using the Pearson chi-square statistic. The expected frequency was set to be equal for all exons present in the gene sequence. Simulations with the same number of replicates confirmed that the probability of a type I error was always very close to 5%.

Genealogy of Dscam

We constructed an amino acid multiple sequence alignment of the Ig and FN domains for selected organisms. We did not include the cytoplasmic tail sequence as it is too divergent to align with confidence. We then created a Bayesian inference phylogeny using MrBayes 3.1.2. We used the mixed model option to choose the amino acid substitution model from each data set, a gamma rate distribution estimated from our data set, and a burn-in equal to one-tenth the number of generations; after the burn-in phase, every 100th tree was saved. Two parallel Markov chains were run simultaneously in each of 2 runs. Tree length, amino acid model, log-likelihood score, and alpha value of the gamma distribution were examined in the program Tracer v1.3 prior to the termination of MrBayes to ensure that all parameters had reached stationarity. All variable exons from each exon array were extracted from the genome sequence and aligned using the default parameters of the ClustalW program in MacVector (v7.2.3), where they were corrected by eye. Bayesian genealogies of each of the 3 variable exon arrays were constructed as described above for *D. magna*, *D. pulex*, and *Apis melifera*.

To examine sequence divergence among exons within each array within and between the 2 *Daphnia* species, we computed the number of synonymous and nonsynonymous differences per synonymous (*ps*) and nonsynonymous site (*ps*), respectively. The calculations were performed using the Nei–Gojobori method (Zhang et al. 1998) estimating in all cases the transition/transversion ratio, using the pairwise deletion option and calculating standard errors by the bootstrap method (1,000 replicates). These analyses were performed using the software MEGA version 4 (Tamura et al. 2007).

Nomenclature

The major difference between Dscam family members is the presence or absence of arrays of alternatively spliced exons. For clarity, we shall refer to the gene with the alternative exon arrays as hypervariable Dscam and name it Dscam-hv.

Results and Discussion

Daphnia Dscam Gene Organization

The *Daphnia* Dscam-hv gene has a similar organization to its homolog in insects in that the exons coding for half of Ig domains 2 and 3 and the entire Ig7 of the Dscam-hv protein are present in arrays of multiple exons (fig. 1). The gene organization in both *Daphnia* species is very similar (GenBank accession numbers: *D. magna* EU307883 and *D. pulex* EU307884). There are 82 exons present in *D. pulex* and 81 in *D. magna*, of which 32 exons account for the mature mRNA in both species (fig. 1). They are organized as follows: the exon 4 array has 8 variants in both *Daphnia* species, the exon 6 array has 26 variants in *D. pulex* and 24 in *D. magna*, and the exon 11 array has 16 and 17 variants in *D. pulex* and *D. magna*, respectively (fig.1). There are 2 main differences in the Dscam-hv gene arrangement between insects and *Daphnia*. First, insects have 2 alternatively spliced exon variants coding for the transmembrane domains, whereas *Daphnia* has only 1 (fig. 1). Second, expression data revealed that 4 different cytoplasmic tails are expressed by both *Daphnia* species (fig. 2*A* and *B*), whereas, to date, insects express only 1 cytoplasmic tail isoform. The cytoplasmic tail of *Daphnia* can be coded by either exons 26 to 31 or exon 30 can be skipped, which results in exon 31 being translated in a different reading frame (fig. 2*A*). Furthermore, exon 27 may also be skipped accounting for 2 additional cytoplasmic tail possibilities. Altogether, the combined usage of the different alternatively spliced exons and cytoplasmic tail possibilities can potentially generate 13,312 different protein isomorphs in *D. pulex* and 13,056 in *D. magna*. This is the first finding of a Dscam-hv gene outside of the insects and the first identification of alternative cytoplasmic tails in Dscam-hv.

Ig, FN, and the Cytoplasmic Tail Domains of the Dscam Protein

Dscam-hv amino acid sequence conservation is high between insects and *Daphnia* for most of the Ig and FN domains, except for the regions coded by the alternative exons. Additionally, some highly conserved motifs are present in the cytoplasmic region of Dscam-hv in *Daphnia* and insects (fig. 3), which are absent from Dscam or Dscam-L in insects. Schmucker et al. (2000) identified some of these conserved motifs as SH2-/SH3-binding domains, which are involved in the binding of Pak to Dscam-hv via the adapter protein Dock, that could mediate changes in the cytoskeleton of cells to promote axon guidance. Although the strong similarity of these and other domains between *Daphnia* and insects (fig. 3) indicates that the molecules interacting with Dscam-hv are likely the same in the 2 groups, the different cytoplasmic tails expressed by *Daphnia* show that differences also exist. Although the functional role of the different cytoplasmic tails is as yet unknown, they are all expressed in both brain tissue and hemocytes.

The 47 amino acids that may or may not be present in the cytoplasmic tail of *Daphnia*, depending on whether exon 27 is skipped, contain several short regions that are highly conserved between *Daphnia* and insects, namely an endocytosis/phagocytosis motif (YXXL, fig. 3). In the 2 *Daphnia* species, this motif is part of a canonical ITAM, an immunoreceptor tyrosine–based activation motif (consensus: YXXL/V- 6 to 17 X-YXXL/V) (Barrow and

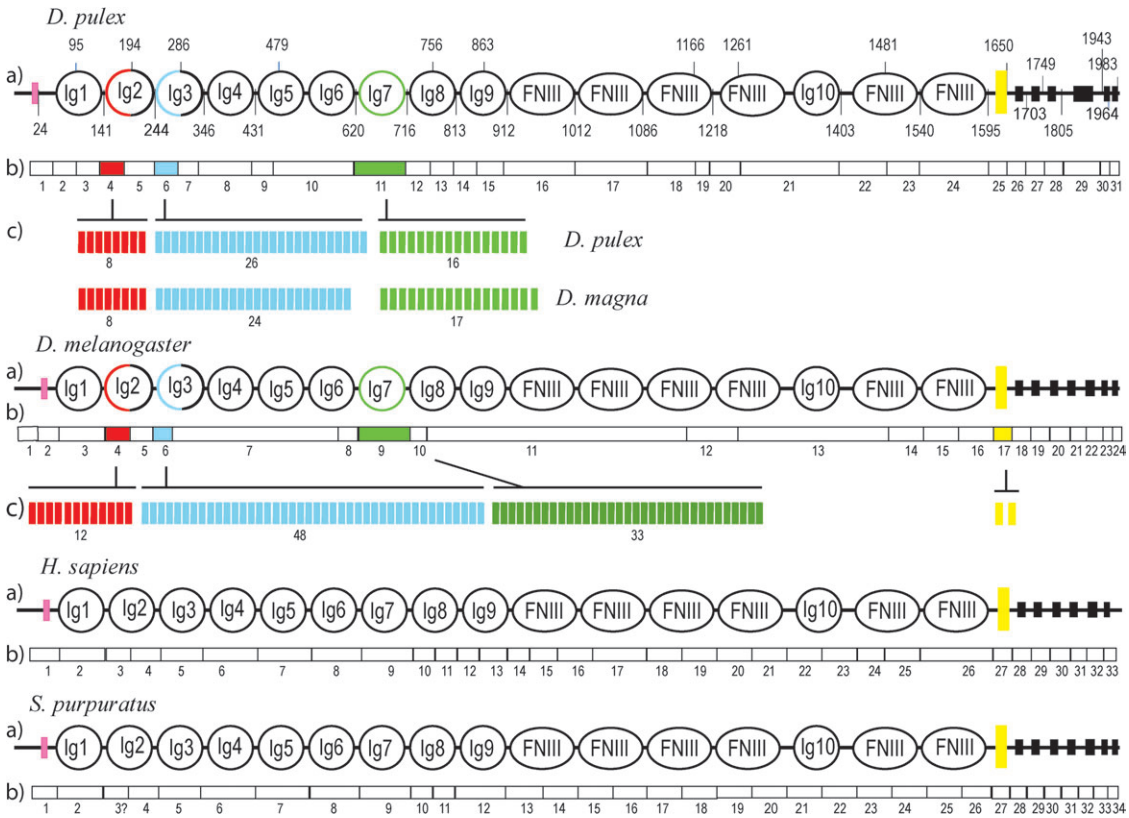

FIG. 1.—Dscam structure in *Daphnia*, *Drosophila melanogaster*, *Homo sapiens*, and the sea urchin *Strongylocentrotus purpuratus*. (*a*) Protein domains, in *Daphnia* exon boundaries in the mRNA are indicated by amino acid numbers, (*b*) mRNA structure, (*c*) arrays of exons coding for the N-terminal parts of Ig2 (red) and Ig3 (blue) and the complete Ig7 (green) domains in *Drosophila* and *Daphnia* represented by bars that correspond to the number of alternative exons present in each species. The transmembrane domain (yellow) in *D. melanogaster* is coded by 2 alternative exons. The cDNA structure of *S. purpuratus* between exon 2 and exon 4 is currently unclear.

Trowsdale 2006) (fig. 3). Isoforms with or without these motifs may have very important differences in their signaling capacity and in regulating the expression of surface membrane receptors (Indik et al. 1995). The cytoplasmic tail variants that result from the inclusion or exclusion of exon 30 and the subsequent reading of exon 31 in 2 different reading frames, differ in length and in the composition of the postsynaptic density, disc large, and zo-I protein (PDZ) domains motif (Fanning and Anderson 1999; Sheng and Sala 2001) that occurs at the very end of the carboxyl end of each form. The alternative PDZ domains (YDTV if exon 30 is included, and SLMV if exon 30 is excluded [fig. 2]) could involve Dscam-hv in different functions via different protein associations or localization in the cellular membrane (Fanning and Anderson 1999). The longest form of the cytoplasmic tail of *D. magna* and *D. pulex* harbors an immune tyrosine–based inhibition motif (ITIM) (consensus: I/S/V/LXYXXV/L) (figs. 2 and 3). After the interaction of the ligand with the extracellular part of the receptor, ITIM becomes phosphorylated on the tyrosine by Src kinases, which then allows it to recruit phosphotyrosine phosphatase that in turn decreases the activity of the cell (Barrow and Trowsdale 2006). The role of ITIM has not been investigated in any Dscam-hv, although the motif has been reported in mammalian Dscam (Staub et al. 2004). The fact that the alternative cytoplasmic tails in *Daphnia* may or may not encode an ITIM and ITAM (fig. 2) suggests that they have very different signaling capacities. *Daphnia* Dscam is therefore diverse in its recognition and effector capacities. The duality ITIM/ITAM in *Daphnia* Dscam reminds us of that observed in paired Ig receptors of vertebrates (Lanier 2001).

Expression of Dscam Transcript Diversity

To investigate how the potential exon diversity repertoire is expressed, we extracted mRNA from *D. magna* hemocytes, brain, and whole embryos using 10, 2, and 5 pooled *D. magna* individuals of the same clone, respectively. From each of these extractions, we amplified, cloned, and sequenced several reverse transcriptase–polymerase chain reaction products encompassing the 3 variable exon arrays. Variable expression of exons 4, 6, and 11 was detected in the hemocytes, brain, and embryos, respectively (fig. 4). All exons in the genomic sequence were expressed, except exons 6.3 and 6.10, demonstrating that *Daphnia* uses the full range of Dscam-hv diversity. The fact that various Dscam-hv isoforms are detected in both brain and hemocytes indicates that the Dscam-hv product diversity is exploited by both the nervous and immune systems of *Daphnia*, as it is in insects.

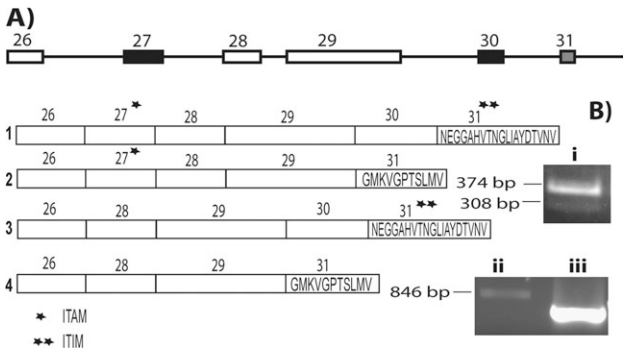

FIG. 2.—Schematic representation of *Daphnia* Dscam cytoplasmic tails (*A*) *Daphnia magna* tail structure and splicing possibilities result in 4 alternative forms. Exons 26–31 code for the cytoplasmic tail. Exons 27 and 30 can be included in the mRNA or skipped. C-terminal end of the cytoplasmic tail changes if exon 30 is included (1) or skipped (3). Two other forms, (2) and (4), are obtained through the inclusion or exclusion of exon 27. (*B*) *Daphnia magna* Dscam cytoplasmic tail expression in the whole-body mRNA. (i) The 2 bands correspond to the cDNA fragments that can be coded by exons 29–31. The bigger fragment includes exons 29, 30, and 31 and the smaller includes exons 29 and 31. (ii) Fragment correspondents to cDNA containing exons 27–31. Cloning and sequencing of this fragment revealed that exon 30 may or may not be transcribed. (iii) Control: whole-body mRNA actin expression.

Unlike *Drosophila*, which shows a more restricted expression of their exon 9 array (the equivalent to the exon 11 array in *Daphnia*), *Daphnia* has a restricted exon 6 array profile. Furthermore, more variants are expressed in brain tissue than in the hemocytes (fig. 4). The restricted exon expression observed in *Daphnia* hemocytes could stem from the fact that the individuals examined were infected with one parasite, however, this result is consistent with those obtained from uninfected *Drosophila* (Watson et al. 2005). If each hemocyte expresses on average 14 different Dscam-hv isoforms, as in *Drosophila* (Neves et al. 2004), the restricted expression in hemocytes results in individual isoforms being present at a higher concentration, which may increase their functional capacity. Additionally, Dscam expression in hemocytes can be rapidly modulated following exposure to diverse pathogens (Dong et al. 2006), which implies a rapid turnover of expressed molecules. The numerous destabilizing RNA motifs (Bevilacqua et al. 2003) encountered in the 3′ untranslated regions (UTR) of the *Daphnia* Dscam-hv could be related to this rapid turnover of the molecule (*D. magna*: 3 copies of ATTTA, 8 copies of TATT, and 10 copies of TAAA in 1,200 bp of 3′ UTR and *D. pulex*: 6 copies of ATTTA, 20 copies of TATT, and 15 copies of TAAA within 2,545 bp of the 3′ UTR).

The observed expression patterns of exon arrays 4 and 11 in the brain do not significantly deviate from random expectation ($P = 0.19$, $P = 0.74$), but the expression pattern for exon 6 array does ($P = 0.026$). In contrast, the expression patterns of exon arrays 4, 6, and 11 in hemocytes deviate strongly from random expectation ($P < 0.0001$, $P = 0.002$, $P < 0.0001$). In both brain and hemocytes, the observed combinations of the 3 variable exons from 1 mRNA molecule deviate strongly from a random expectation ($P < 0.0001$). Consistent with the hypothesis that the expression of Dscam-hv alternative exons is regulated, different exon combinations are preferred in the brain compared with hemocytes (fig. 4). Previously, changes in Dscam-hv expression patterns for each exon across time, tissue, and type of pathogen challenge have been demonstrated in both cell lines and in individuals of *Drosophila* and *Anopheles* (Celoto and Graveley 2001; Neves et al. 2004; Watson et al. 2005). Further

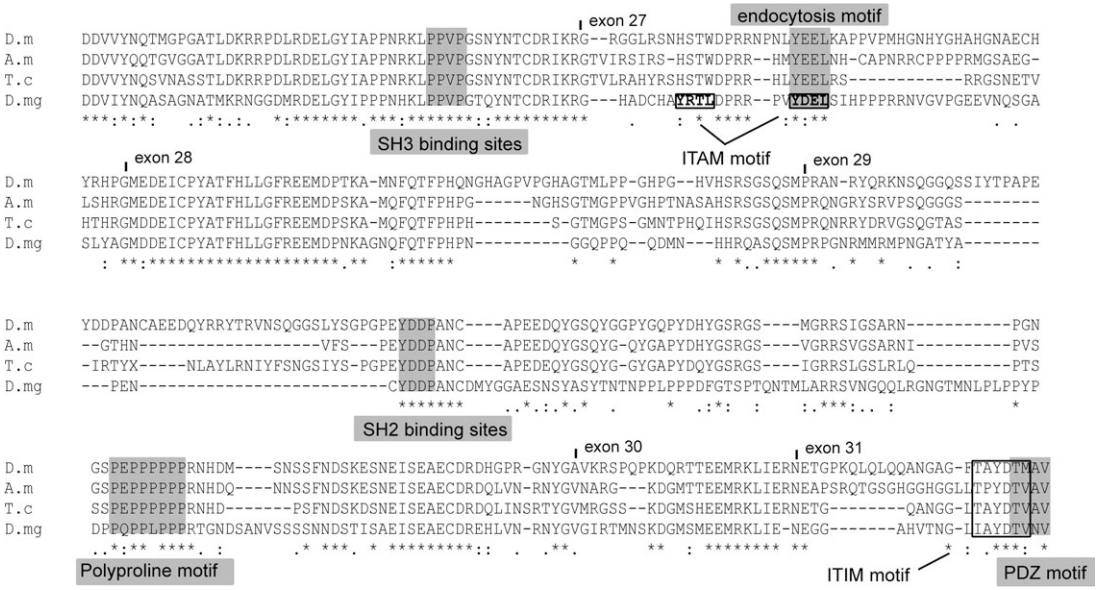

FIG. 3.—Amino acid sequence alignment of cytoplasmic tail segments of *Drosophila melanogaster* (D.m), *Apis mellifera* (A.p), *Tribolium castaneum* (T.c), and *Daphnia magna* (D.mg). Symbols represent levels of amino acid identity between species: (*) full identity, (:) strongly similar, (.) weakly similar, and () no similarity. Gray squares highlight homologous regions to motifs involved in signaling pathways: SH3- and SH2-binding motifs, a polyproline motif, an endocytosis, and a PDZ motif. Black squares highlight ITIM and ITAM. The PDZ motif presents in the alternative end of the cytoplasmic tail is shown in figure 2. Exon numbers correspond to *Daphnia*.

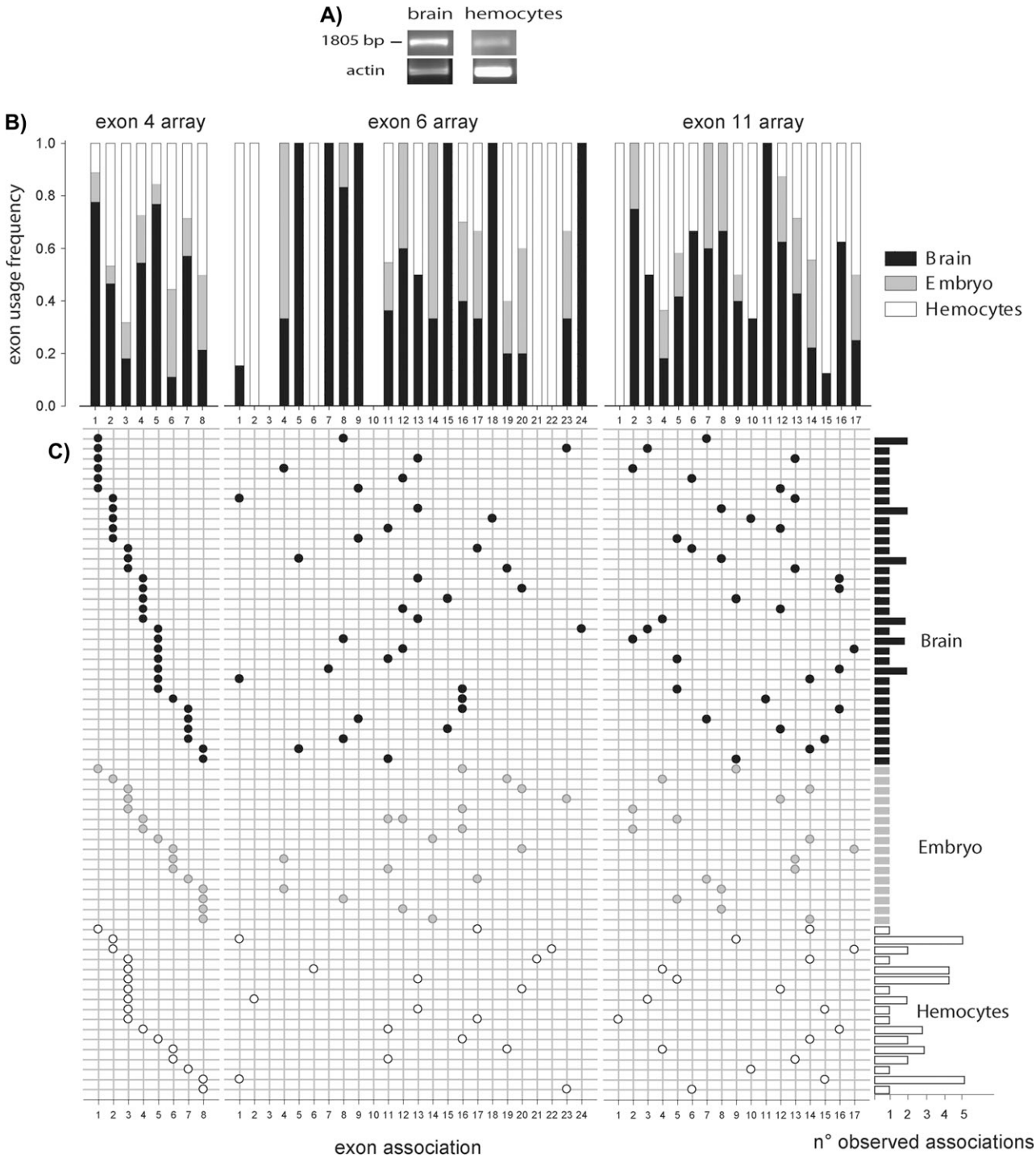

FIG. 4.—(*A*) *Daphnia magna* expression of a Dscam region encompassing Ig3 to Ig7 in the brain and hemocytes. Sequencing revealed that each band is composed of many different isoforms corresponding to the expression of exon variants from arrays 4, 6, and 11. (*B*) Exon usage frequency in different tissues in *D. magna*. Bars correspond to the expression of each exon in each tissue, relative to the total number of times the exon was observed in all tissues. (*C*) Association of exons from each array in single mRNA molecules from brain, embryos, and hemocytes. The bars on the right side of the graph represent the absolute number of times that each association was observed. Number of sequences: brain $n = 39$; embryo $n = 16$; and hemocytes $n = 37$. Exon 6.3 cannot be used because there is a mutation at the 3′ end of the exon that does not allow splicing with exon 7 (splicing law changed from type 2 to type 0).

immunological experiments will determine if this is also the case with *Daphnia*.

Although the mechanisms for mutually exclusive splicing of the variable exons are not fully understood, studies of *Drosophila* have identified 2 sequence motifs within the Dscam-hv gene that appear to be involved in regulating exons from arrays 4 and 6 (Graveley 2005; Kreahling and Graveley 2005). These sequence motifs

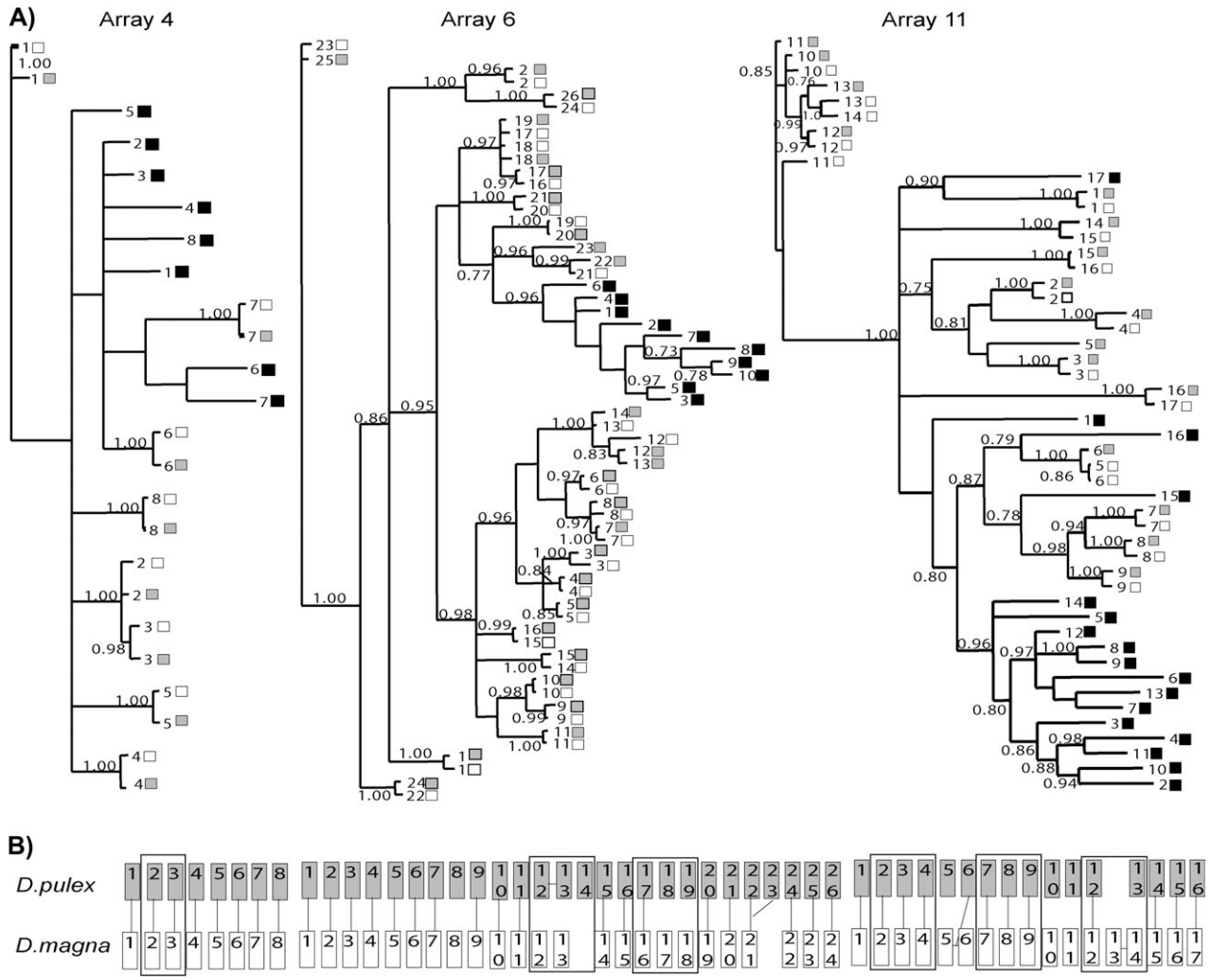

FIG. 5.—(*A*) Bayesian analysis of the exons from *Daphnia magna* (white), *Daphnia pulex* (gray), and *Apis mellifera* (black) contained in the 3 variable arrays of the *Daphnia* Dscam gene. In the exon 6 tree, only 10 representatives of *A. mellifera* were included. (*B*) Schematic representation of the exons depicting the orthologous pairing and synteny of the variable exons between the 2 *Daphnia* species. Boxes represent clustering among the nearest neighbors with a probability of 0.9 or more.

are also present in *Daphnia* (supplementary fig. S1, Supplementary Material online), suggesting that the regulatory machinery is evolutionarily conserved between these taxa.

Variable Regions within the Alternative Exons

A structural analysis of the first 4 Ig domains of 2 distinct Dscam-hv isoforms in *Drosophila* has demonstrated that the 5′ portions of the alternative exons 4 and 6 contribute to regions of the protein that are essential for Dscam-hv homophilic binding and reside on a region called epitope I (Meijers et al. 2007). Located on the opposite side of the 3D structure of the molecule is epitope II, defined by the 3′ region of exon 4 and the central region of exon 6. It does not participate in Dscam-hv homophilic binding (Meijers et al. 2007). A comparison of orthologous exons from arrays 4 and 6 from 12 *Drosophila* species revealed that the epitope II sequences are more variable than those of epitope I, suggesting that this region of the protein is under fewer selective constraints. Closer examination of the same sequences between *D. magna* and *D. pulex* is entirely consistent with the *Drosophila* observation, given that the regions of variability in crustaceans and insects are superimposable (supplementary fig. S2, Supplementary Material online).

Phylogenies of the Variable Exons

Clear orthologs exist between the 2 *Daphnia* species for the vast majority of exons in each of the arrays (fig. 5*A*), meaning that interspecific sequence similarity is higher than intraspecific. This suggests that the occurrence of concerted evolution is not affecting the evolution of the multiple exons of each array in a significant way (Nei and Rooney 2005). This relationship is strongest in exon 4 array, where 1:1 orthologous pairs were identified for every exon (fig. 5*B*). Similarly, almost all exon 6 array members have a clear pairing between the 2 *Daphnia* species (fig. 5*B*), despite having different numbers of exons. These results are

consistent with those obtained among 3 species of *Drosophila* from Graveley et al. (2004). Sites of recent gene duplication of exon 6 variants in *D. pulex*, or gene loss in *D. magna*, are exons 12, 13, or 14 and exon 23 according to the numbering of *D. pulex* (fig. 5*B*). Variation in exon 6 copy number also exists between *Drosophila melanogaster* and *Drosophila virilis* (48 and 52 copies, respectively), indicating that recombination leading to exon loss/gain in this portion of the gene may be more frequent than in the exon 4 region. Regarding the exon 11 array, there have been 2 exon duplication/loss events since the split between the *D. pulex* and *D. magna* (fig. 5*B*). In one case, *D. pulex* exon 11.5 does not have an orthologous match in *D. magna*. Because 1:1 orthologous pairings between the 2 Daphniids continue downstream, it is more likely that the *D. pulex* exon 11.5 is the result of an exon duplication event, as opposed to exon loss, in *D. magna*. In the other case, *D. magna* exons 11.13 and 11.14 are more closely related to each other than to any *D. pulex* exon and thus likely arose by exon duplication in *D. magna* after the split between these 2 species.

The fact that, generally, orthology of the alternative exons has been maintained between the 2 *Daphnia* species, coupled with their short branch lengths, suggests that at least part of the exon sequence variation may be functionally constrained. On the other hand, based on the lack of orthology between the alternative exons of *Daphnia* and insects (represented by *Apis mellifera*, the insect species with the highest Dscam sequence similarity to *Daphnia*) (fig. 5*A*), this constraint appears to be taxon specific. This contrasts with the high degree of sequence conservation in the constant domains of the molecule between these 2 groups of Arthropods. Furthermore, some characteristics of each of the 3 arrays are consistently shared among species. For example, the exon 4 array always has fewer variants than either of the other 2 arrays. Such shared characteristics among the arrays could reflect that they have experienced similar selective constraints in both insects and crustaceans.

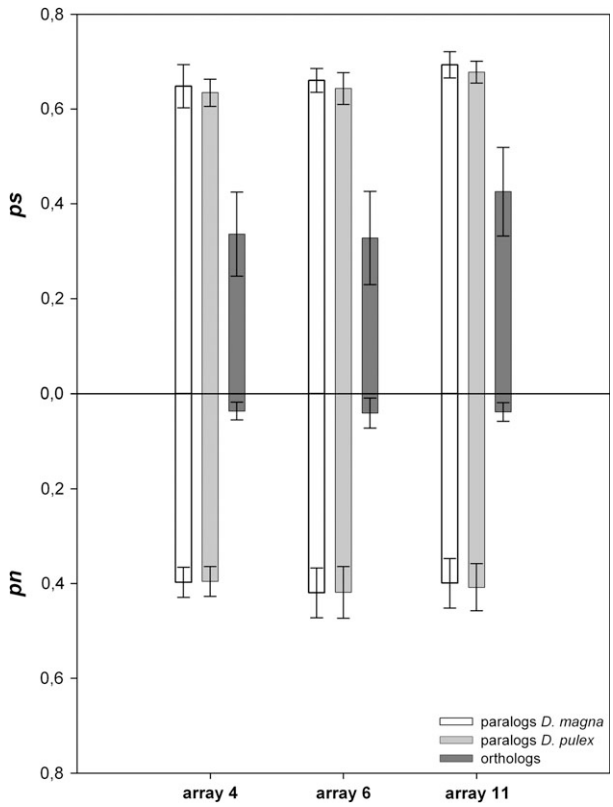

FIG. 6.—Average p.s. and p.n. of paralogs and orthologs from arrays 4, 6, and 11. The error bars correspond to the standard deviation of paralog and ortholog p.s. and p.n. values. The matrices of p.s. and p.n. values of all pairs of paralogs and orthologs and the estimated standard error are available by request.

The Evolution of the Duplicated Exons

It has been proposed that the alternative exons originated by duplication in a nearest-neighbor scenario, where exons closer to one another along the chromosome are more similar than exons that are further apart (Graveley et al. 2004). The phylogenies of the variable exon arrays 6 and 11 of the 2 *Daphnia* species are generally consistent with this model (fig. 5). For example, in the exon 6 array, some resolution beyond the orthologous pairings is obtained, where at least one large clade containing all the central exons in the array is strongly supported. Within this central exon clade, there are 2 additional clades that cluster exons 6.3–6.16 and 6.17–6.23 (numbering according to *D. pulex*) (fig. 5*A*). The resolved members within the exon 11 array also correspond with the nearest-neighbor hypothesis. However, in contrast, the exons present at the end and at the beginning of array 6 are more dissimilar to the central cluster. Furthermore, the relationship among paralogous exons is not well resolved for array 4, where only exon pairs 4.2 and 4.3 cluster together (fig. 5*A*), suggesting that the exons in this cluster evolved rapidly or that this array is older than the other 2.

The number of synonymous substitutions *ps* sites and nonsynonymous substitutions *pn* sites between alternative exons within each array is higher between than within the 2 *Daphnia* species (fig. 6 and supplementary fig. S3, Supplementary Material online). This suggests that paralogs largely evolved according to the birth-and-death model, which assumes that new genes are created by repeated duplication events and that some duplicates may stay in the genome for a long time, whereas others are deleted or become nonfunctional (Nei and Hughes 1992; Nei et al. 2000). The recent exon duplication and deletions described for arrays 6 and 11 give further support to the appropriateness of this model in explaining how the variable Dscam arrays are evolving. Only 1 nonfunctional exon was found (see caption of fig. 5). The *ps* values between paralogs in one array are generally near the saturation level with most values between 0.4 and 0.7, whereas *ps* of orthologs although high are lower (0.2–0.4) (see fig. 6 for average values and supplementary fig. S3, Supplementary Material online). The number of nonsynonymous differences between paralogous and orthologous exons indicates that there are many more nonsynonymous differences between paralogs (*pn*: 0.1 to 0.6) than orthologs (*pn*: 0 to 0.06) and this pattern is very consistent in the 3 arrays (fig. 6 for average values and supplementary fig. S3, Supplementary Material

online). This difference in the number of substitutions in orthologs and paralogs for the 3 arrays supports that the duplicated exons in each cluster had already diverged in the ancestor of the 2 *Daphnia* species. The *d*n and *d*s values were calculated for orthologous exons by correcting the *ps* and *pn* values with the Jukes–Kantor formula (Ota and Nei 1994). The *d*n/*d*s ratio of orthologous exons indicates that strong selection is acting to maintain the amino acid composition of each exon (average *d*n/*d*s: array 4 = 0.08; array 6 = 0.1; and array 11 = 0.06) (supplementary table S1, Supplementary Material online). Selection acting upon paralogs in each array seems to have been much weaker, allowing for more nonsynonymous substitutions (fig. 6) and subsequent diversification.

Dscam Family Evolution

Our searches for Dscam genes confirmed that, to date, only members of the insects (Crayton et al. 2006) and *Daphnia* have a Dscam-hv gene that contains at least 3 arrays of alternative exons (figs. 1 and 7). We found no *sensu stricto* Dscam-L paralogs in the current *D. pulex* genome assembly, even though 2 genes with homology were found with a different domain organization (see Materials and Methods). Our tree shows that the vertebrate Dscam and Dscam-L genes are clearly separate from those of insects, the sea urchin, and the flatworm *Dugesia*, despite the fact that the Dscam-L exon structure of insects lacks variable exon arrays and thus superficially more closely resembles the vertebrate homologs (fig. 7). Therefore, it seems that the ancestral Dscam gene duplicated in the 2 groups independently of one another or that concerted evolution within the 2 groups has destroyed the phylogenetic signal at this deep level. The intron/exon boundaries of both vertebrate and insect Dscam gene copies also support the hypothesis of independent duplication, with insect Dscam-L genes intron/exon boundaries being more similar to those of Dscam-hv than to human Dscam or Dscam-L. Furthermore, the motifs identified by Crayton et al. (2006) that discriminate the Dscam and Dscam-L of vertebrates were not found in any of the invertebrate Dscam genes. With respect to the timing of the duplication event within the invertebrates, both crustaceans and insects share the complex trait of alternative exon arrays and likely the same mechanisms of mutually exclusive splicing, suggesting that the duplication event in the invertebrate lineage must have occurred before the split of the Pancrustaceans (fig. 7). *Daphnia* appears to have strongly modified or lost its paralog of Dscam-hv. The 2 nematode genome sequences currently available (*Caenorhabditis elegans and Caenorhabditis briggsiae*) and the tunicate *Ciona* (a deuterostome) appears to lack Dscam altogether.

Differences between the Dscam-hv, Dscam, and Dscam-L can also been seen at the predicted properties of the respective proteins coded by these genes, like the number of gylocosylation sites. Glycosylation patterns suggest that there are fewer glycosylation sites in Dscam-hv compared with Dscam or Dscam-L (supplementary table S2, Supplementary Material online). This pattern holds true for the 3 insect species for which both forms of the gene

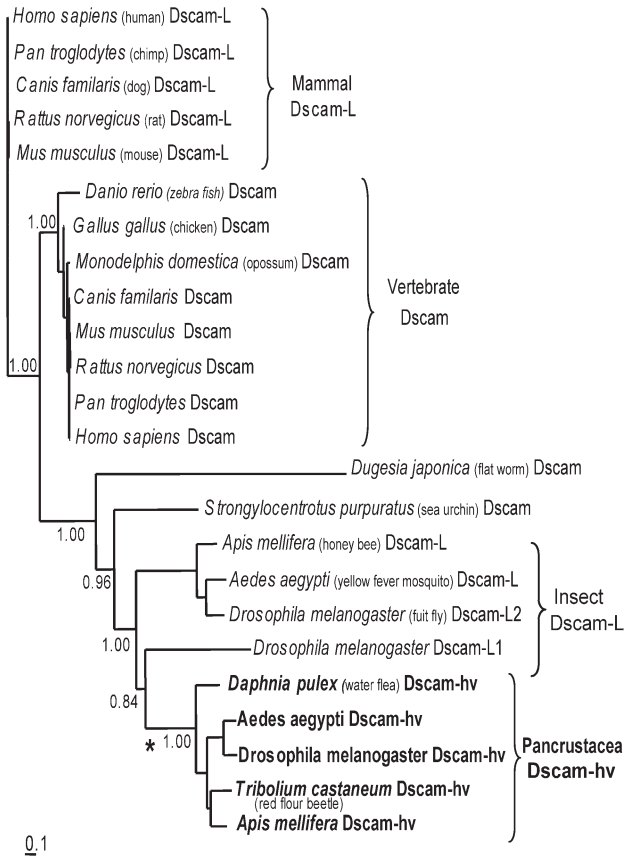

FIG. 7.—Bayesian topology of the extracellular regions of Dscam and Dscam-L genes from representative metazoan. Numbers at nodes are posterior probabilities. Only nodes relevant to the discussion are labeled. Asterisk represents the possible origin of mutually alternative splicing in Dscam.

occur and for which sequences are available. Carbohydrates mediate interactions between recognition molecules and a great variety of glycan chains and play a role in both the nervous and immune systems (Kleene and Schachner 2004). The higher number of glycosylation sites of the nonvariable and Dscam-L proteins might be a functional alternative or complement the Dscam-hv molecules diversified by mutually alternative splicing.

Concluding Remarks

Alternative exons coding for Dscam-hv Ig domains are present in insects and in the crustacean *Daphnia* but not in other invertebrates or vertebrates, suggesting that it evolved in the ancestor of the pancrustaceans. Dscam-hv amino acid conservation is high among divergent taxa, except in the regions that are coded by the alternative exons, which vary considerably in number and sequence between *Daphnia* and insects and even among insects. Another level of variability in the alternative exons is evident when comparing more closely related species in the regions of Dscam-hv suspected to play a role in heterologous recognition (Meijers et al. 2007). The structural position where this variability occurs seems to be conserved between *Daphnia* and several *Drosophila* species, despite the sequence

divergence of their alternative exons. Thus, the principles underlying Dscam-hv diversity are conserved between *Daphnia* and insects. Furthermore, as in insects, *Daphnia* expresses diverse repertoires of Dscam-hv isoforms in both brain tissue and hemocytes. It is not known whether Dscam-hv diversity originally evolved by selection on the nervous system, the immune system, or both (Du Pasquier 2005).

Two nonexclusive selective advantages may be conferred to both the nervous and immune systems as a result of Dscam-hv diversity. First, it is beneficial to have a large number of different isoforms present in either system, even if their sole property is that they undergo homologous binding. This benefit has been demonstrated in the nervous system (Chen et al. 2006; Hattori et al. 2007), where the structural basis for homologous interactions is understood (Meijers et al. 2007). Specifically, the homologous interactions and their variegated expression on the cell surface allow large numbers of cells to be distinguished from one another. Similarly, the immune system could benefit by creating individualized hemocytes that can patrol without aggregating. If this is the case, many exons with different sequences, but not the precise exon sequences, would confer a selective advantage.

A second hypothesis is that isoforms are selected for their ability to bind to heterologous ligands, for example, pathogens. In this scenario, specific exon sequences would be selected. Soluble forms of Dscam-hv circulate in the hemolymph of insects where they are unlikely to play any role in the nervous system but could act as opsonins. Supporting this idea, inhibition of their expression results in a lower phagocytosis capacity and Dscam-hv isoform expression changes after exposure to various antigens (Dong et al. 2006). Furthermore, a variable site on the molecule is oriented in a way that permits heterologous interaction (Meijers et al. 2007). All these suggest that the variability of Dscam-hv may be useful or even essential to the immune system. In fact, the pattern of rapid evolution of the alternative exons in different species is reminiscent of IgSF members involved in innate immunity in vertebrates (McQueen and Parham 2002), that is, a pattern modulated by the pathogen environment. If this is the case, selection acting on immune function would have been the driving force for maintaining an interesting form of alternative somatic diversification in the immune repertoire.

Supplementary Material

Supplementary information on the construction of the fosmid library, sequence accession numbers, supplementary figures S1–S3, and tables S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Brigitte Aeschbach for technical assistance and Dietmar Schmucker for support and helpful discussions. The *D. pulex* sequence data were produced by the US Department of Energy Joint Genome Institute http://www.jgi.doe.gov/in collaboration with the Daphnia Genomics Consortium http://daphnia.cgb.indiana.edu. D.B. is supported by the Portuguese Science Foundation. D.E. and I.C. were supported by the Swiss National Foundation. This work was done at the University of Basel and University of Edinburgh.

Literature Cited

Agarwala KL, Subramaniam G, Tsutsumi Y, Suzuki T, Kenji A, Yamakawa K. 2001. Cloning and functional characterization of DSCAML1, a novel DSCAM-like cell adhesion molecule that mediates homophilic intercellular adhesion. Biochem Biophys Res Commun. 285:760–772.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped Blast and PSI-Blast: a new generation of protein database search programs. Nucleic Acid Res. 25:3389–3402.

Barrow A, Trowsdale J. 2006. You say ITAM and I say ITIM, let's call the whole thing off: the ambiguity of immunoreceptor signalling. Eur J Immunol. 36:1646–1653.

Bevilacqua A, Ceriani MC, Capaccioli SNA. 2003. Post-transcriptional regulation of gene expression by degradation of messenger RNAs. J Cell Physiol. 195:356–372.

Celoto AM, Graveley B. 2001. Alternative splicing of the *Drosophila* Dscam pre-mRNA is both temporally and spatially regulated. Genetics. 159:599–608.

Chen BE, Kondo M, Garnier A, Watson FL, Püettmann-Holgado R, Lamar DR, Schmucker D. 2006. The molecular diversity of Dscam is functionally required for neuronal wiring specificity in *Drosophila*. Cell. 125:607–620.

Colbourne JK, Hebert PDN. 1996. The systematics of North American *Daphnia* (Crustacean: Anomopoda): a molecular phylogenetic approach. Philos Trans R Soc Lond B Biol Sci. 351:349–360.

Crayton M, Powell B, Vision T, Giddings M. 2006. Tracking the evolution of alternatively spliced exons within the Dscam family. BMC Evol Biol. 6:1–15.

Dong Y, Taylor HE, Dimopoulos G. 2006. AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. PLoS Biol. 4:e229.

Du Pasquier L. 2005. Diversify one molecule to serve two systems. Science. 309:1826–1827.

Ebert D, Rainey P, Embley TM, Scholz D. 1996. Development, life cycle, ultrastructure and phylogenetic position of *Pasteuria ramosa* Metchnikoff 1888: rediscovery of an obligate endoparasite of *Daphnia magna* strauss. Philos Trans R Soc Lond B Biol Sci. 351:1689–1701.

Fanning AS, Anderson JM. 1999. PDZ domains: fundamental building blocks in the organization of protein complexes at the plasma membrane. J Clin Invest. 103:767–772.

Glenner H, Thomsen PF, Hebsgaard MB, Sørensen MV, Willerslev E. 2006. The origin of insects. Science. 314:1883–1884.

Graveley B, Kaur A, Gunning D, Zipursky SL, Rowen L, Clemens JC. 2004. The organization and evolution of the dipteran and hymenopteran Down syndrome cell adhesion molecule (*Dscam*) genes. RNA. 10:1499–1506.

Graveley BR. 2005. Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. Cell. 123:65–73.

Hattori D, Demir E, Kim HW, Virahg E, Zipursky SL, Dickson BJ. 2007. Dscam diversity is essential for neuronal wiring and self-recognition. Nature. 449:223–228.

Indik ZK, Park JG, Hunter S, Schreiber AD. 1995. Structure/function relationships of Fc gamma receptors in phagocytosis. Semin Immunol. 7:45–54.

Kleene R, Schachner M. 2004. Glycans and neural cell interactions. Nat Rev Neurosci. 5:195–208.

Kreahling JM, Graveley B. 2005. The iStem, a long-range RNA secondary structure element required for efficient exon inclusion in the *Drosophila Dscam* pre-mRNA. Mol Cell Biol. 25:10251–10260.

Lanier LL. 2001. Face off—the interplay between activating and inhibitory immune receptors. Curr Opin Immunol. 13:326–331.

McQueen KL, Parham P. 2002. Variable receptors controlling activation and inhibition of NK cells. Curr Opin Immunol. 14:615–621.

Meijers R, Puettmann-Holgado R, Skiniotis G, Liu J-h, Walz T, Wang J-h, Schmucker D. 2007. Structural basis of Dscam isoform specificity. Nature. 449:487–491.

Millard SS, Flanagan JJ, Pappu KS, Wu W, Zipursky L. 2007. Dscam2 mediates axonal tiling in the *Drosophila* visual system. Nature. 447:720–724.

Nei M, Hughes AL. 1992. Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. In: Tsuji K, Aizawa M, Sasazuki T, editors. Proceedings of the 11th Histocompatibility Workshop and Conference. Oxford (UK): Oxford University Press. p. 27–38.

Nei M, Rogozin IB, Piontkivska H. 2000. Purifying selection and birth-and-death evolution in the ubiquitin gene family. Proc Natl Acad Sci USA. 97:10866–10871.

Nei M, Rooney AP. 2005. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet. 39:121–152.

Neves G, Zucker J, Daly M, Chess A. 2004. Stochastic yet biased expression of multiple Dscam splice variants by individual cells. Nat Genet. 36:240–246.

Ota T, Nei M. 1994. Variance and covariances of the numbers of synonymous and nonsynonymous substitutions per site. Mol Biol Evol. 11:613–619.

Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zypursky SL. 2000. *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell. 101:671–684.

Shapiro L, Love J, Colman DR. 2007. Adhesion molecules in the nervous system: structural insights into function and diversity. Annu Rev Neurosci. 30:451–474.

Sheng M, Sala C. 2001. PDZ domains and the organization of supramolecular complexes. Annu Rev Neurosci. 24:1–29.

Staub E, Rosenthal A, Hinzmann B. 2004. Systematic identification of immunoreceptor tyrosine-based inhibitory motifs in the human proteome. Cell Signal. 16:435–456.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol. 24:1596–1599.

Watson LF, Püttmann-Holgado FT, Thomas F, Lamar DL, Hughes M, Kondo M, Rebel VI, Schmucker D. 2005. Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science. 309:1874–1878.

Wojtowicz WM, Flanagan JJ, Millard SS, Clemens J, Zipursky SL. 2004. Alternative splicing of *Drosophila* Dscam generates axon guidance receptors that exhibit isoform-specific binding. Cell. 118:619–633.

Yamakawa K, Huo Y-K, Haendel MA, Hubert R, Chen X-N, Lyons GE, Korenberg JR. 1998. DSCAM: a novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system. Hum Mol Genet. 7:227–237.

Zhang J, Rosenbergdagger HF, Nei M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci USA. 95:3708–3713.

Billie Swalla, Associate Editor

Accepted April 6, 2008