doi:10.1111/evo.12179

MORE THAN ONE WAY TO PRODUCE PROTEIN DIVERSITY: DUPLICATION AND LIMITED ALTERNATIVE SPLICING OF AN ADHESION MOLECULE GENE IN BASAL ARTHROPODS

Daniela Brites,[1,2,3] Carlo Brena,[4] Dieter Ebert,[1] and Louis Du Pasquier[1]

[1]*Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel Switzerland*

[2]*Swiss Tropical and Public Health Institute, Socinstrasse 57, PO Box 4002, Basel*

[3]*E-mail: d.brites@unibas.ch*

[4]*Department of Zoology, University Museum of Zoology, Cambridge CB2 3EJ, United Kingdom*

Received January 13, 2013
Accepted May 27, 2013

Exon duplication and alternative splicing evolved multiple times in metazoa and are of overall importance in shaping genomes and allowing organisms to produce many fold more proteins than there are genes in the genome. No other example is as striking as the one of the Down syndrome cell adhesion molecule (Dscam) of insects and crustaceans (pancrustaceans) involved in the nervous system differentiation and in the immune system. To elucidate the evolutionary history of this extraordinary gene, we investigated *Dscam* homologs in two basal arthropods, the myriapod *Strigamia maritima* and the chelicerate *Ixodes scapularis*. In both, Dscam diversified extensively by whole gene duplications resulting in multigene expansions. Within some of the *S. maritima* genes, exons coding for one of the immunoglobulin domains (Ig7) duplicated and are mutually exclusively alternatively spliced. Our results suggest that Dscam diversification was selected independently in chelicerates, myriapods, and pancrustaceans and that the usage of Dscam diversity by immune cells evolved for the first time in basal arthropods. We propose an evolutionary scenario for the appearance of the highly variable Dscam gene of pancrustaceans, adding to the understanding of how alternative splicing, exon, and gene duplication contribute to create molecular diversity associated with potentially new cellular functions.

KEY WORDS: **Dscam, gene family, immunity, immunoglobulin domains, nervous system, protein diversification.**

Throughout the tree of life, molecules involved in host–pathogen interactions provide some of the best evidences of diversifying selection acting at the molecular level (Endo et al. 1996; Pal et al. 2006). This possibly results from relatively strong and continuous selection acting on both surface antigens of parasites and receptors of the host immune system (Woolhouse et al. 2002). The necessity of molecular recognition has also imposed similar selection in systems where specification of the identity of multiple single cells and specific adhesion is important, such as in the nervous system. In both cases, this resulted in the diversification of very large numbers of extra- or intracellular proteins. The genetic mechanisms exploited by different organisms, tissues, or cells to generate protein diversity are various: diversity at the population level has been achieved via balanced polymorphisms (e.g., hundreds of MHC alleles in mammals; Trowsdale and Parham 2004), whereas gene duplication resulted in diversity at the individual level (e.g., dozens of cadherin genes throughout metazoa [Hulpiau and van Roy 2009], hundreds of olfactory receptors genes in human [Malnic et al. 2004], hundreds of Toll-like receptors, and Nod-like receptors in the sea urchin [Rast et al. 2006]). But mechanisms for achieving even higher magnitude orders of protein diversity evolved, such as that of somatic diversification

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

through which each individual is able to produce many more proteins than the number of genes encoding them (up to 10^{14} for antibody genes in vertebrates). This can be achieved at the DNA level by somatic hyper mutation, by somatic gene conversion, or by somatic combinatorial associations of gene segments such as the rearrangement of the genes encoding the lymphocyte receptors of the vertebrate immune system (Flajnik and Du Pasquier 2012). This can also be achieved at the RNA level and the most common mechanism is by alternative splicing of pre-messenger RNA molecules (Maniatis and Tasic 2002). Diversification by alternative splicing and exon duplication have been suggested as major mechanisms generating protein diversity and expanding gene functions in multicellular organisms (Kondrashov and Koonin 2001). An extreme example of such diversification is that of the human Down syndrome cell adhesion molecule (Dscam) homolog of insects and crustaceans (pancrustaceans). The *Dscam* locus of pancrustaceans has three arrays of tandem duplicated exons that undergo mutually exclusive alternative splicing, producing 10^3 to 10^4 Dscam isoforms (Schmucker et al. 2000; Watson et al. 2005; Brites et al. 2008). Homologs of Dscam are present in many different metazoa (Shapiro et al. 2007) but the exon duplications and the alternative splicing mechanism found in pancrustaceans is not yet known in any other taxa. The human DSCAM and the hypervariable Dscam of pancrustaceans are both involved in similar developmental processes controlling neural wiring (reviewed in Hattori et al. 2008). In addition, the diversity of Dscam isoforms in pancrustaceans plays a role in the immune system providing potentially an increased specific recognition capacity of different pathogens (Watson et al. 2005; Dong et al. 2006; Watthanasurorot et al. 2011; Dong et al. 2012). Thus, duplication and alternative splicing of Dscam are possibly associated with the evolution of new functions of Dscam in pancrustaceans. How this situation evolved is largely unknown (Brites et al. 2008; Armitage et al. 2012). The evolutionary history of this unique gene in arthropods is in the center of this study and it revolves around the following questions: What is the origin of the complex mechanism of mutually exclusive alternative splicing that seems unique to arthropods? And is the immune function of Dscam a derived character of pancrustaceans?

To address them we looked at basal arthropods; the tick *Ixodes scapularis*, a chelicerate, and the centipede *Strigamia maritima*, a myriapode. Recent phylogenetic reconstructions indicate that myriapods are possibly a sister group to pancrustacea and are hence particularly useful in understanding both arthropod and pancrustacea evolution (e.g., Regier et al. 2010; Rota-Stabelli et al. 2010).

We found that Dscam convergently diversified in all arthropods; basal groups obtain their diversity from large Dscam multigene families (unlike pancrustaceans) and by mutually exclusive alternative splicing (like pancrustaceans) but by an apparently primitive mechanism.

Materials and Methods

THE DSCAM MOLECULE

Dcams consist in immunoglobulin (Ig) and fibronectin (FN) domains: 9(Ig)-4(FN)-1(Ig)-2(FN), followed by a transmembrane domain and a cytoplasmic tail. The exons encoding half of the second (Ig2) and third (Ig3) immunoglobulin domains, and the complete domain Ig7 of the extracellular segment have duplicated extensively forming three arrays which undergo mutually exclusive alternative splicing ensuring that only one exon per array is included in the mature mRNA (Schmucker et al. 2000; Graveley 2005; Kreahling and Graveley 2005; Olson et al. 2007; Fig. 1). In this way, the *Drosophila melanogaster Dscam* gene has the potential to generate 19,008 different extracellular Dscam isoforms (Fig. 1). Combined with two alternative transmembrane domains and the regular alternative splicing affecting four exons encoding the cytoplasmic tail, a total of 152,064 different isoforms can be encoded by the same gene in a single fly (Yu et al. 2009). Its crustacean homolog in *Daphnia magna* has the potential to express more than 13,000 different transcripts (Brites et al. 2008). To differentiate this highly diversified gene from other nondiversified *Dscam* genes (of which *D. melanogaster* has three paralogs), we designate it hereafter as "Dscam-hv."

GENE RECOVERY

The program tblastn was used to search several metazoan genomes for Dscam related genes (Table S1). We did a general search using the whole Dscam-hv of *D. melanogaster* as a "probe" and selected the most related genes based on amino acid similarity and domain architecture. Several architectural criteria were used: the Ig1 motif GxxxxC (where x stands for any amino acid and C refers to the canonical cysteine in the β strand of the Ig1 domain) which is a distinctive signature of Dscam (in regular Ig domains G is at position −8 in relation to the referred cysteine); the presence of Ig1 to Ig4, which form a horse-shoe structure typical of Dscam and other related CAMs (Meijers et al. 2007); the presence of Ig10 between the FNIII domains. Finally, we searched for the transmembrane domains and cytoplasmic tails sequence similarities. In all Dscam-related genes found we searched for duplicated exons using the Dscam-hv variable regions of Ig2, Ig3, and Ig7. All homologs were annotated by hand using a prediction of the protein structure obtained with SMART (http://smart.embl-heidelberg.de; Schultz et al. 1998; Letunic et al. 2009) and PHYRE2 http://www.sbg.bio.ic.ac.uk/~phyre/ (Kelley and Sternberg 2009).

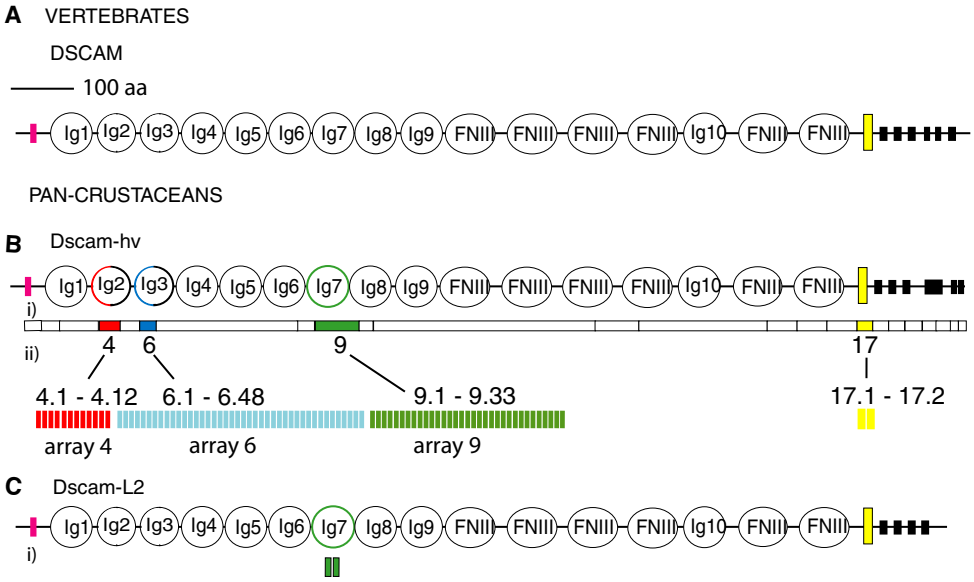

Figure 1. **Dscam domain representation; Ig, immunoglobulin domains; FNIII, fibronectin III domains. The yellow and black boxes represent the transmembrane and cytoplasmic domains. The magenta box represent the leader peptide (A) DSCAM of vertebrates represented by the homolog in *Homo sapiens*; (B) Dscam-hv of pancrustaceans represented by the homolog in *Drosophila melanogaster*; (i) mRNA, each box corresponds to a constitutive exons and the colored boxes 4, 6, and 9, correspond to exons that are the result of mutually exclusive alternative splicing within arrays of duplicated exons which are present in three arrays, as indicated in (ii); (C) Dscam-L2 of pancrustaceans represented by the *D. melanogaster* homolog; (i) two Ig7 exons submitted to mutually exclusive alternative splicing.**

IDENTIFICATION AND ANNOTATION OF THE DSCAM OF MYRIAPODES (*S. MARITIMA*) AND CHELICERATA (*I. SCAPULARIS*)

The genomes of *I. scapularis* (http://www.vectorbase.org/index.php) and that of *S. maritima* 24× scaffolding (http://www.hgsc.bcm.tmc.edu/blast.hgsc) were investigated as referred earlier. Each gene was named after the species name followed by a number (Figs. S2, S5). In each species, a further tblastn was performed using intraspecific Dscam sequences to ensure that no genes were missed. We have also scrutinized the EST database for *I. scapularis* to look for Dscam expression (http://iscapularis.vectorbase.org/SequenceData/EST/).

PHYLOGENETIC RECONSTRUCTIONS

Multiple amino acid alignments were built using CLUSTALW and edited through Jalview (Waterhouse et al. 2009). The positions of some amino acids (G, W, C, D, Y) and their properties (hydrophobic or not) are distinct features of Ig domains (Lefranc and Lefranc 2001) and were used to correct the alignments manually. Phylogenetically conflicting regions were eliminated following Gblocks selected blocks (Castresana 2000; Talavera and Castresana 2007). The program ProTest 1.4 was used to estimate substitution models and related the parameters (Drummond and Strimmer 2001; Guindon and Gascuel 2003; Abascal et al. 2005). Phylogenies were built with both Bayesian and maximum likelihood (ML) methods, using MrBayes 3.1.2 and RAxML (Stamatakis 2006), respectively. For Bayesian analysis we used a γ 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 the two runs. Tree length, log-likelihood score, and α value of the γ distribution were examined prior to the termination of MrBayes to ensure that all parameters had reached stationarity. Convergence was checked using AWTY (Nylander et al. 2008) by plotting the posterior probabilities of all splits for the two runs and the number of generations when necessary was increased. For the ML analysis we run RAxML through the Cipres Portal (Miller et al. 2009) with at least 1000 bootstrap replicates.

To determine the homology of Dscam-related genes found in basal metazoans, we estimated phylogenies of 42 proteins including Dscam and other CAMs whose Ig1 to Ig4 domains form a horse-shoe structure (Table S1). This phylogeny was rooted using the sequence of the human neural CAM (NCAM), which does not form a horse-shoe structure.

The phylogeny of Dscam homologs representative of major metazoan clades was reconstructed based on alignments comprising Ig2 to FNIII-2 domains given that Ig1 was not found in many cases. To include incomplete homologs of *Ixodes* with multiple exons coding for Ig7 and Ig8, we used aligments containing Ig8 to FNIII-2 domains. To trace the origins of Ig7, phylogenetic

trees of Ig7 coding exons of Dscam orthologs and paralogs of representatives of arthropods and deuterostomes were produced (number of exons included = 177).

STRIGAMIA MARITIMA DISSECTIONS, RNA EXTRACTION, AND CDNA SYNTHESIS

Adult individuals of *S. maritima* were sampled near Brora, Scotland, and kept alive at 4°C starved. RNA was extracted from whole-body, hemocytes, and heads using Trizol (Invitrogen, Carlsbad, California) following manufacturer instructions. For hemocytes and heads, to increase RNA yield, RNA samples were precipitated overnight in isopropanol at −80°C with 5 μg of RNAse-free glycogen added (Invitrogen). Hemocytes were obtained by cutting three individuals in several sections and withdrawing the hemolymph by capillary action using microcapillary glass tubes (Harvard apparatus GC100TF-10). To check Dscam expression in the nervous system, the heads from the same individuals were used for RNA extraction. All material was immediately stored in RNA*later* (Ambion) solution.

To obtain the 5′ leader region of the *Sm35* gene of *S. maritima*, we used SMART technology (SMART^{tm} RACE cDNA Amplification Kit; Clontech, Sant-Germain-en-Laye, France) on whole-body mRNA following the instruction of the manufacturer and a specific reverse primer annealing to Ig3.

The expression of the duplicated exons of *Sm35* was investigated by sequencing reverse transcriptase-polymerase chain reaction (RT-PCR) amplicons obtained with primers specific to Ig6 and Ig8 coding exons. The RNA yield of hemocytes was low and a One Step PCR kit (Qiagen) was used to perform a multiplex PCR with *Sm35*-specific primers and primers specific to actin (positive control). All PCR products were cloned using pCR 2.1- TOPO vector (Invitrogen) and sequenced with Sanger sequencing.

Results

THE DSCAM FAMILY WITHIN THE IMMUNOGLOBULIN SUPERFAMILY CAMS OF BILATERIA

We searched for Dscam-related genes in demosponges (*Amphimedon queenslandica*), cnidarians (*Nematostella vectensis*), and placozoans (*Tricoplax adhaerens*; Table S1). To test whether they belong to the Dscam family we built a phylogeny including those, other metazoan Dscams, and other CAMs from the immunoglobulin superfamily whose first four Ig domains form a horse-shoe structure (roundabout, axonin, L1CAM, and hemolin). Most Dscam genes formed relatively well-supported clades and most likely have a monophyletic origin although the latter could not be recovered with strong statistical support (Fig. 2). The same is true for roundabout and axonin, CAMs which are used by the nervous system and to which the gene of *T. adhaerens* is most closely related. The genes from *A. queenslandica* and three of the genes in *N. vectensis* genes blasted significantly to Dscam but did not form any well-supported clade (Fig. 2). The position of *N. vectensis* gene *Nv_1* is unclear (Fig. 2), yet if based on the Ig8 to FNIII-2 domains, Nv_1 forms a well-supported clade with the human DSCAMs (Fig. S1), reflecting their 30% similarity (*E* values below e^{-171}). Furthermore, their cytoplasmic tails also share similar SH2, ITIM, and polyproline motifs (data not shown) indicating that they probably use similar signaling pathways. In subsequent analysis *Nv_1* was used as outgroup.

DIVERSIFICATION OF DSCAM IN CHELICERATES AND MYRIAPODES

Extracellular domain diversification by gene duplication

We found many Dscam-related genes in both *I. scapularis* and *S. maritima* genomes. Our purpose was not an exhaustive description of all those genes but an analysis of relevant comparative aspects of Dscam genes from two basal arthropods. Because of the multiplicity of domains most automatic gene predictions needed manual correction, hence we have annotated and analyzed only a fraction of the Dscam genes of those organisms. Although all statements about absence of genes or domains have to be taken carefully, especially in the case of *I. scapularis* genome which has many undetermined sequences, we can claim that there are no arrays of duplicated exons coding for Ig2 and Ig3 like in the Dscam-hv of pancrustaceans.

In the myriapod *S. maritima* genome, we found over 100 Dscam-related sequences (with $E > 10^{-4}$) including incomplete or noncanonical genes. We estimate that 60 to 80 complete Dscam genes might be present in the genome. Arrays of exons coding for half of Ig2 and Ig3 domains like in pancrustaceans were not found. In contrast, approximately 70% of the Dscam genes contain arrays of duplicated Ig7 coding exons (varying from four to five copies; Figs. S2, S10). In the 12 genes that we annotated in detail, the exon duplications were probably already present before the genes duplicated as they are more similar between genes than within each gene (Fig. S3). One would then expect that those Ig7 domains have similar amino acid divergence compared to the remaining ectodomains of those paralogous Dscam genes. Interestingly, the duplicated Ig7 domains are less divergent than the remaining ectodomains (Fig. S4), which might be due to gene conversion or recombination occurring in this region.

In the tick *I. scapularis*, we found 27 genes with strong similarity to Dscam in contrast to four genes reported by Armitage et al. 2012. Fifteen almost complete homologs were reconstructed (Fig. S5). None exhibits an exact canonical configuration generally lacking the third and fourth FNIII domains and the tenth Ig domain (Fig. S5). This could explain the discrepancy of our results and those of Armitage et al. We did not find exon duplications coding for half of Ig2 and Ig3 like in Dscam-hv. Instead

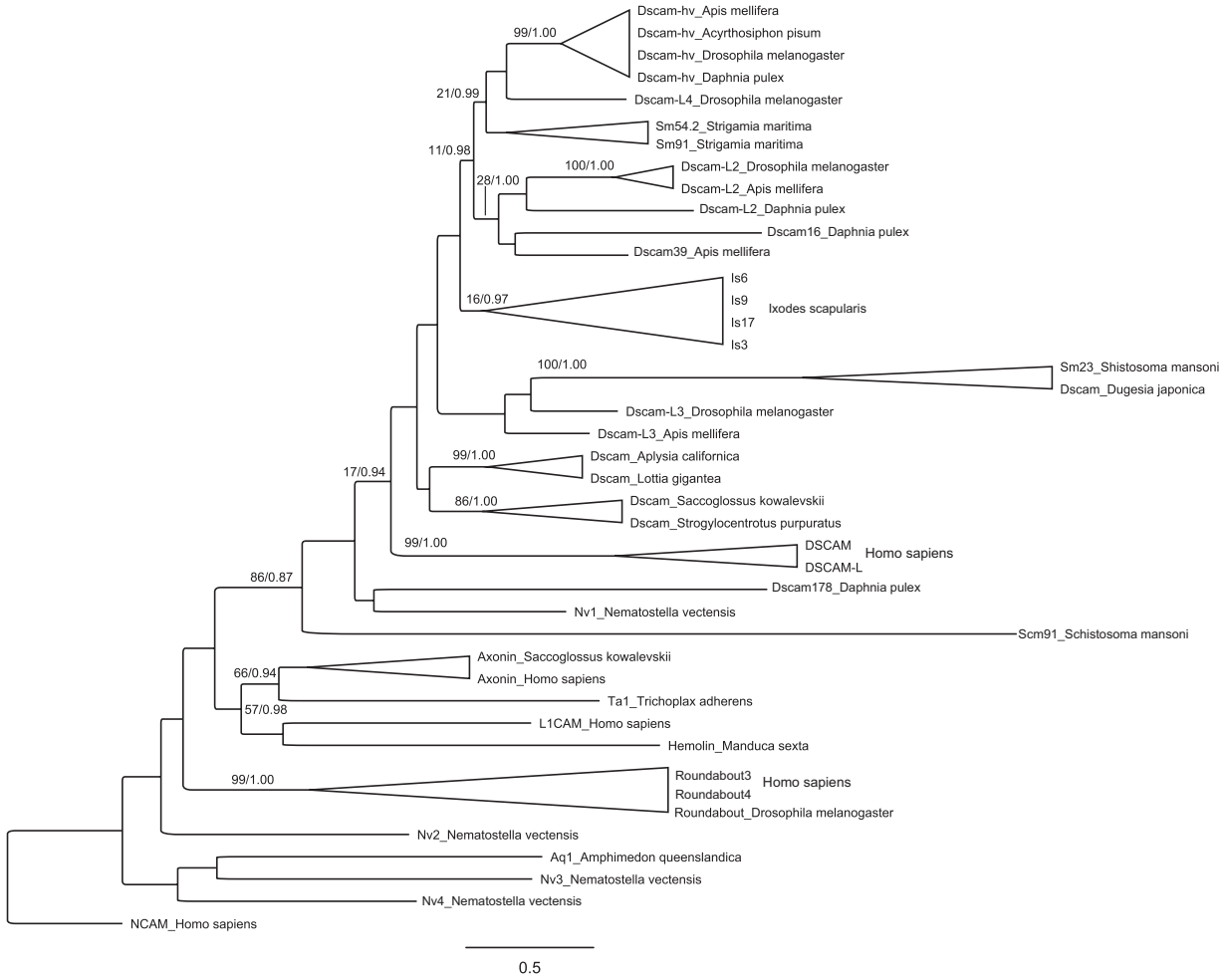

Figure 2. Maximum likelihood topology 42 CAMs whose first four Ig domains form a horse-shoe tertiary structure. Support values at nodes are bootstrap values relative to 1000 replicates (left value) and posterior probabilities (right value) when higher than 60% and/or than 0.95, respectively. The tree is rooted with the human NCAM, a CAM which does not form a horse-shoe structure. The alignment was produced with only the first four Ig domains of the molecules. The amino acid substitution model WAG and the observed amino acid frequencies were used.

we found four genes *Is27*, *Is28*, *Is29*, and *Is53*, each with several duplications of exons coding for Ig7 and Ig8 (Fig. 3A). The multiple Ig7 and Ig8 coding exons are in alternate positions in the genome, a feature not observed in any other *Dscam* (Fig. 3). The exon and intron structure of these genes suggests that they could be alternatively spliced but no related ESTs were found. Genes *Is27*, *Is28*, and *Is29* are located in the same contig separated approximately by 1900 bp. Genes *Is28* and *Is29* are duplicates of each other, whereas the origin of *Is27* remains elusive (Fig. S1). The Ig7 and Ig8 duplicated exons diverged extensively (Fig. 3B). Some duplications might have occurred in a gene specific manner (e.g., *Is28* 8.1 and 8.2) or they were under intragene conversion, whereas others are more similar between different genes (*Is27* Ig8.5 and *Is28* Ig8.11). *Is53* is a chimaera due duplications of partial regions of genes *Is27* and *Is28* (Fig. 3A, B). The amino acid conservation is very strong between *Is53* and *Is27* and *Is28* but not the nucleotide sequences, excluding artifacts of the assembly. Lacking pseudoexons, all these genes look functional and suggest an important role of gene conversion and recombination. Genes *Is15*, *Is4*, *Is9*, *Is10*, and *Is3* are physically close in the genome and all are transcribed in the same direction, except *Is3* (Fig. S6). The phylogenetic relationships among them are mostly unresolved except for *Is3*, which is most closely related to *Is26*, a gene present on a different contig (Figs. S1 and 5).

Dscam diversification by alternative splicing in myriapods

To investigate whether mutually exclusive alternative splicing was present in one of the Dscam genes of *S. maritima*, we obtained RNA from whole single animals, cloned, and sequenced RT-PCR fragments containing duplicated exons coding for Ig7 from the gene *Sm35*. We found Ig7 duplicated exons expressed in several

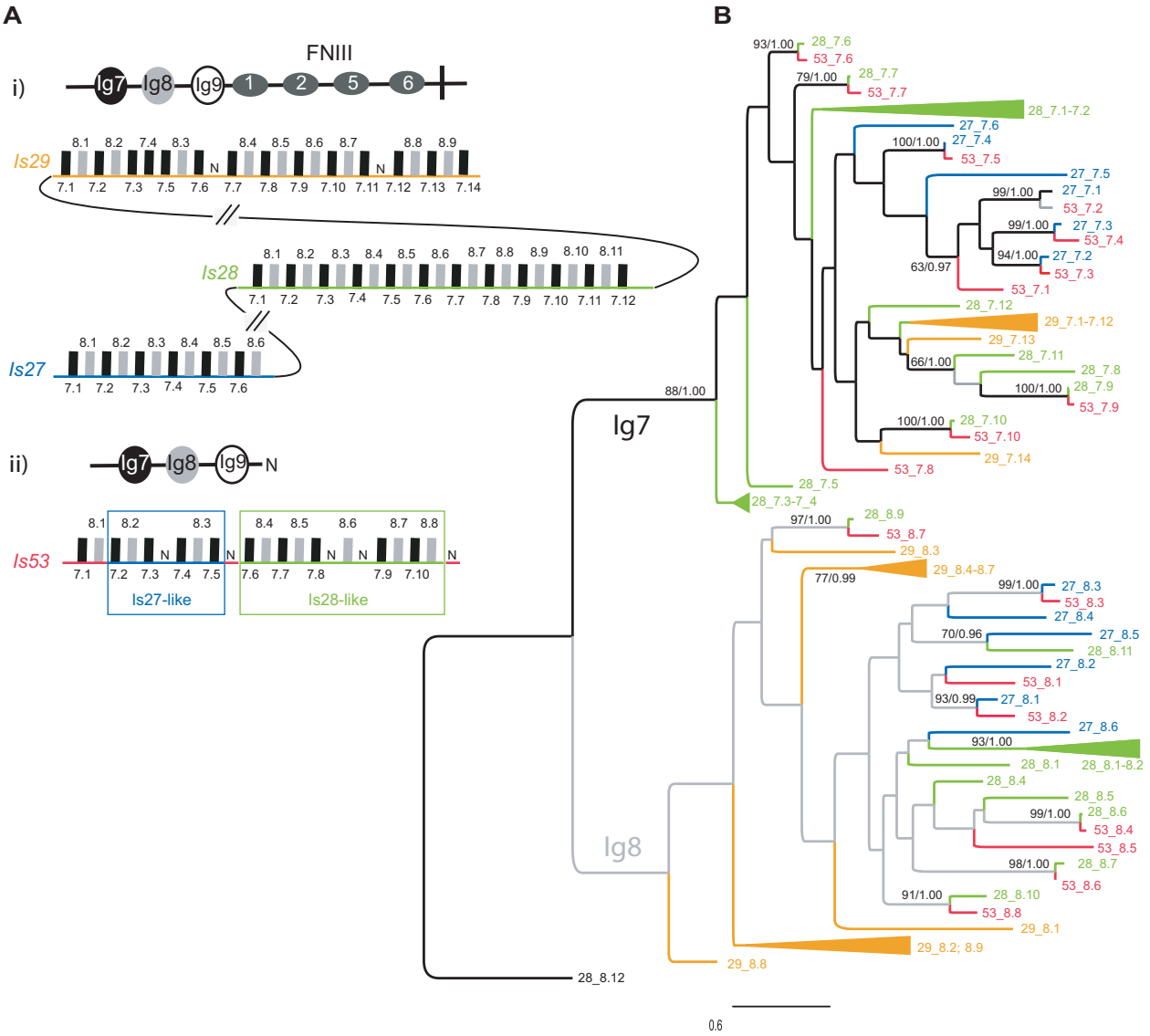

Figure 3. **(A)** ***Ixodes scapularis*** **Dscam homologs with duplicated exons coding for Ig7 and Ig8. Below each protein reconstruction (i) and (ii), is the representation of how alternative exons of each gene coding for Ig7 (black boxes) and Ig8 (gray boxes) are organized in the DNA.** ***N*** **represents undetermined sequence (i) protein reconstruction coded by genes** ***Is27***, ***Is28***, **and** ***Is29***, **which are all adjacent in the same contig. (ii) Protein reconstruction coded by** ***Is53***. **The boxes Is27-like and Is28-like indicate the exons that are homologous to the exons of** ***Is27*** **and** ***Is28*** **genes, respectively. (B) Maximum likelihood topology of the 76 duplicated exons coding for Ig7 (black branches) and Ig8 (gray branches) in** ***I. scapularis*** **Dscam homologs** ***Is27*** **(blue branches),** ***Is28*** **(green branches),** ***Is29*** **(orange branches), and** ***Is53*** **(magenta branches). Support values at nodes are bootstrap values relative to 1000 replicates (left value) and posterior probabilities (right value) when higher than 60% and/or than 0.95, respectively. The tree is rooted for convenience with exon 8.12 from gene** ***Is28*** **because this exon has the lowest amino acid similarity relative to all other exons in the tree. Statistically confident monophyletic clades of exons were collapsed for convenience. The amino acid substitution model WAG and the observed amino acid frequencies were used.**

possible ways. The four duplicated Ig7 coding exons can be expressed in a mutually exclusive alternative splicing fashion like in Dscam-hv (Fig. 4 WB, lane b, band 3). In some cases, two or more alternative exons can be expressed together (Fig. 4 WB, lane b, bands 4, GenBank accession numbers: KC479660, KC479661). Moreover, Ig7 coding exons can be skipped all together (Fig. 4 WB, lane b, band 1, GenBank accession number: KC479656). The latter could represent an RNA "artifact" or splicing intermediates that may not be translated. The dominance of this band could also be explained by a more efficient PCR amplification. Alternatively, it could reveal the possibility of a major variation in domain composition in this part of the Dscam molecule. However, we have no protein data to confirm the existence of a shorter DSCAM molecule (or with multiple Ig7). Bands 2 revealed that

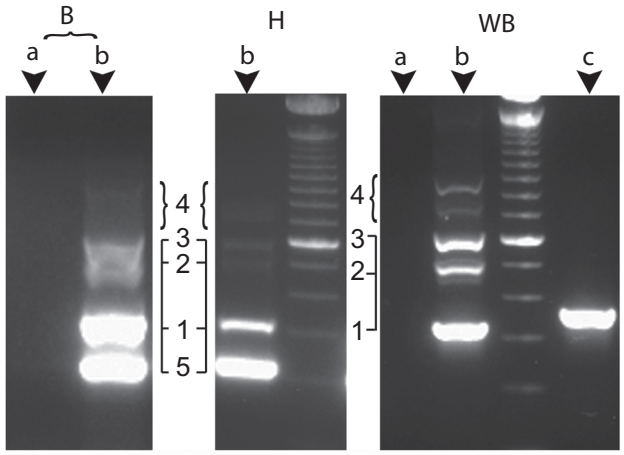

Figure 4. ***Strigamia maritima*** **expression of the** ***Sm35*** **region encompassing duplicated exons coding for Ig7. Whole body (WB), hemocytes of three individuals (H), and heads of three individuals (B). a, negative control of b; b, region encompassing Ig7 coding duplicated exons of** ***Sm35***; **c, expression of** ***Sm 35*** **constitutive exons coding for Ig9; d, 100bp ladder. All bands were cloned and sequenced; 1, transcripts with exons coding for Ig6 and Ig8, missing Ig7 coding exons altogether; 2, transcripts for which Ig7 coding exons were spliced using a premature splicing site, scaffold nucleotide position 11107; 3, transcripts for which Ig7 coding exons were submitted to mutually exclusive alternative splicing. 4, larger bands correspond to transcripts with more than one Ig7 coding exon; 5, expression of actin (positive control) which for preparations H and B was obtained by multiplex PCR.**

in certain instances a premature splicing site is used (GenBank accession number: KC479655, see details in the figure legend; Fig. 4, WB, lane b).

This suggests that the mechanism of mutually exclusive alternative splicing of the Dscam-hv gene has evolved initially in the array of exon duplications coding for Ig7 and it was already present in the ancestor of the pancrustaceans and myriapods, being possibly retained in the pancrustacean gene Dscam-L2.

Alternatively spliced Dscam of myriapodes is expressed by hemocytes and the nervous system

In pancrustaceans Dscam diversity can be used both in the nervous and the immune systems. We investigated whether *Sm35* is expressed both by hemocytes and by nervous system cells of *S. maritima* by RT-PCR. The sequences of cloned RT-PCR fragments show that this gene is expressed and mutually exclusive alternatively spliced by both hemocytes (GenBank accession numbers: KC479657–KC479659) and likely nervous system cells (Fig. 4H, B, respectively, lanes b). Again several transcripts containing none or more than one Ig7 coding exons were obtained in the nervous system enriched preparation and to a lesser extent in hemocytes (Fig. 4 lanes b, bands 1 and 4). This result indicates that the expression of Dscam by hemocytes is not a derived character evolved in pancrustaceans but a character most likely already present in their ancestors.

Transmembrane domains and cytoplasmic tails of Ixodes and Strigamia Dscams

One member of the *I. scapularis* Dscam family has two exons coding for transmembrane domains, suggesting that it might use alternative transmembrane domains through alternative splicing (*Is9*; Figs. S5, S11) like insects (Watson et al. 2005). Indeed, we found one EST corresponding to the expression of *Is9* with only one of the transmembrane forms (Fig. S11). *Is13* does not code for a transmembrane domain and possibly generates a Dscam soluble form (supported by one *Is13* EST; Fig. S11).

The sequence conservation between the cytoplasmic domains of *S. maritima* and of *I. scapularis* with those of pancrustaceans is low (data not show). Nevertheless, a few motifs are conserved, in particular CC1 motifs (PTPYATT; Prasad et al. 2007; Andrews et al. 2008; Fig. S7). The comparison of several Dscam cytoplasmic tails of arthropods revealed that the residues GxxDEICPYATFHLLGFREEMD are a good predictor of Dscam genes containing domains diversified by alternative splicing (Fig. S7). Interestingly, these motifs are present in *Sm35* for which alternative splicing and the expression by hemocytes was demonstrated.

EVOLUTION OF THE DSCAM GENE FAMILY

Our data suggest that the Dscam gene with arrays of exons coding for Ig2, Ig3, and Ig7 evolved early in the diversification of pancrustaceans (Fig. 5). Nevertheless, diversification of Dscam homologs occurred in all arthropod groups either by internal duplication of Ig domains or by duplications of complete genes. The genealogy of all Dscam gene reconstructions of *S. maritima* and *I. scapularis* confirmed that Dscam homologs diversified within each taxa independently. Despite their differences, arthropod's Dscams seem indeed to be more strongly related to each other than to any other Dscam genes, forming a monophyletic group (Fig. 5). Noteworthy, Dscam-L2 of pancrustaceans and all the genes of *S. maritima* with Ig7 coding exon duplications do not group together. The insects represented do not share the same four Dscam paralogs, for example *A. mellifera* and the louse *P. humanus Dscam5* and *Dscam6* and *Dscam49* of the pea aphid *Acyrthosiphon pisum* (Fig. 5). Contrarily to previously reported (Brites et al. 2008; Armitage et al. 2012), *D. pulex* has two other paralogs besides Dscam-hv and Dscam-L2. These do not group confidently with any of the other insect Dscam paralogs (Fig. 5). The *S. maritima* Dscam homologs are more closely related to each other than to any other Dscam and the same is true for

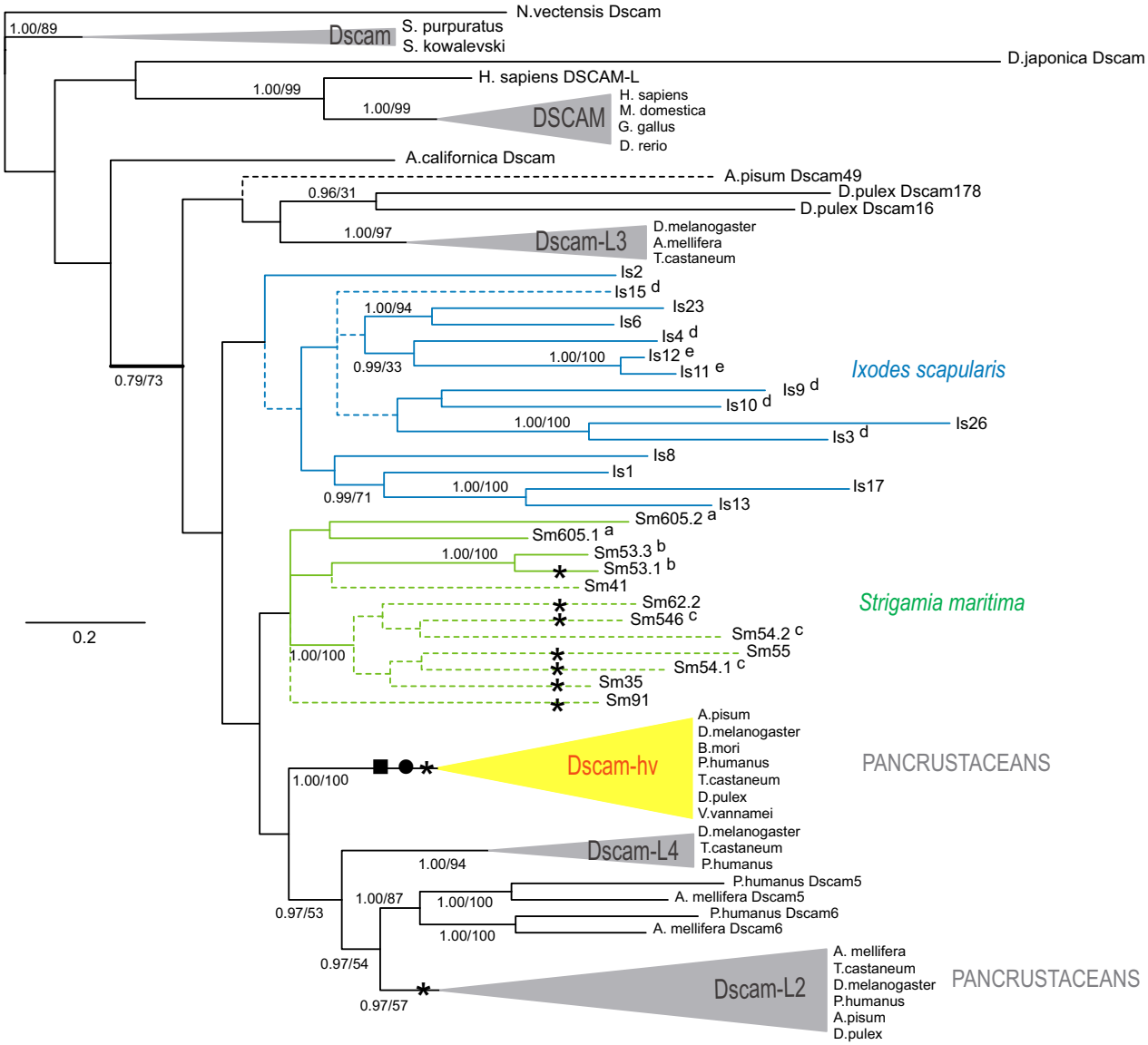

Figure 5. **Maximum likelihood topology of Dscam-related genes in representatives of metazoa. The tree is based in the Dscam regions encompassing Ig2 to FNIII-2 domains and is rooted using the Dscam sequence of the cnidarian *Nematostella vectensis*. Support values at nodes are bootstrap values relative to 1000 replicates (left value) and posterior probabilities (right value) when higher than 60% and/or than 0.95, respectively. Monophyletic clades of orthologs were collapsed for convenience. Genes with internal exon duplications coding for and are indicated with: ■, for Ig2; ●, for Ig3; and *, for Ig7. Genes located in the same genomic scaffold are indicated with the same superscript (a, b, c, d). The dashed branches represent incongruent branches obtained by the maximum likelihood and the Bayesian methods. The possible monophyletic origin of all arthropod Dscams is marked by a thicker internal branch. The amino acid substitution model WAG and the observed amino acid frequencies were used.**

I. scapularis. In both taxa, gene duplication was followed by quick divergence, which makes it difficult to establish confident phylogenetic relationships among paralogs (Fig. 5). The paralogs with Ig7 duplications do not form a monophlyletic group within the *S. maritima* paralogs. Our genealogies are compatible with both a scenario of independent gain of exons in Sm53.1 and Sm91 either by new exon duplications or by recombination followed by divergence or alternatively, some Dscam paralogs might have lost Ig7 duplications.

THE ORIGINS OF THE DUPLICATED EXONS CODING FOR IG7

Present in all arthropods so far studied the Ig7 exons duplications might have evolved in the most common ancestor to all groups. Alternatively, as suggested by the presence of Ig7 and Ig8 duplications in *Ixodes*, Ig7 duplications could have evolved in the common ancestor to myriapods and pancrustaceans. In blast searches, *I. scapularis* and *S. maritima* Ig7 coding exons were more similar to the Ig7 exons of pancrustaceans than to any other

Dscam but a monophyletic origin for all arthropod Ig7 exons could not be found (Fig. S8). This is not unexpected given that those exons are short and except for a few landmark amino acids diverged extensively. We obtained confident monophyletic groups only within species (Fig. S8), the exceptions were the Ig7 exons 11.16 of *D. pulex* and 9.33 of *D. melanogaster* (also found by Lee et al. 2009), and exons 7.6 of *D. pulex* and 7.16 of *Apis mellifera*, indicating that these exons were probably present in the ancestors of pancrustaceans.

The alignment of all Ig7 coding exons revealed interesting differences between Dscam-hv and all the other Dscams. Between the conserved tryptophan 38 and glycine 42, all Ig7 exons except those of Dscam-hv, have a variable nonpolar aminoacid, followed by arginine or lysine and aspartic acid (Fig. S9). This is not observed in any of the Dscam-hv Ig7 coding exons, which have a variable amino acid composition between tryptophan 38 and glycine 42 and invariably, arginine or lysine at position 58, which was never observed outside of the Dscam-hv (Fig. S9). Curiously, exons 11.16 of *D. pulex* and 9.32 and 9.33 of *D. melanogaster* exhibit an intermediary composition at these positions, with aspartic acid before glycine 42 and no charged amino acid at position 58. In both species, these exons are located at the end of the array. Possibly they did not diverge as much as the exons more internally located in the arrays and still retained ancestral features (Brites et al. 2008; Lee et al. 2009). Considering the protein structure models of the *D. melanogaster* Dscam-hv (Sawaya et al. 2008), the significance of these amino acid changes is not clear, but given the prominent differences between Dscam–hv and other Dscams they are likely to be important functionally.

Discussion

THE EVOLUTION OF THE DSCAM FAMILY

Throughout the evolution of metazoans, CAMs were recruited for multiple purposes: cell proliferation and differentiation, apoptosis, and parasite recognition (Buckley et al. 1998; Humphries and Newham 1998). In some CAMs, the first four Ig domains form of a horse-shoe tertiary conformation which creates singular adhesive properties by allowing homophilic and heterophilic adhesion to similar and different proteins, respectively. The appearance of this structure might have allowed the expansion of a subfamily of CAMs used by nervous cells of different metazoans such as axonin, roundabout, contactin, and Dscam among others, and by immune system cells such as hemolin and Dscam. Our results suggest that precursors of Dscam could be already present before the evolution of the Bilateria. Certain regions of the cnidarian NV_1 protein are quite conserved between *Nemastostella vectensis* and humans. Furthermore, Nv_1 shares cytoplasmic motifs with human Dscams (but not with any of the protostome Dscam homologs), denoting the usage of similar signaling pathways. In contrast to the extracellular domains of Dscam of distant taxonomic groups, beyond short motifs, the sequences of the cytoplasmic domains and the number of exon encoding them are little conserved. This suggests that the extracellular and intracellular part of the Dscam family molecules evolved by exon shuffling at different rates and illustrates how the properties of receptor molecules can be accommodated to multiple signaling pathways in different organisms.

THE DSCAM GENES OF ARTHROPODS

All extant arthropods have expanded the *Dscam* gene family. It occurred both by massive duplication of entire genes in chelicerates and myriapodes, and by extensive tandem exon duplications in the pancrustacean *Dscam-hv*, and to a lesser extent in *Dscam-L2* and in some *Dscam* of *I. scapularis* and *S. maritima*. Our analysis suggests a monophyletic origin for the *Dscam* genes in arthropods with expansions in different groups, indicating that selection has acted independently to diversify Dscam. In the remaining metazoans, no Dscam paralogs are known, with the exception of vertebrates in which two paralogs (DSCAM and DSCAM-L in humans) have arisen independently of the arthropod duplications (Crayton et al. 2006; Brites et al. 2008). Why the history of this gene family is so different between arthropods and other metazoans cannot be easily answered. In vertebrates, in the flat worm *Dugesia japonica* and in *Drosophila*, Dscam is essential for the correct development of the nervous system (Yamakawa et al. 1998; Fusaoka et al. 2006; Millard et al. 2007). In arthropods, a functional diversification followed the genetic diversification of Dscam; that is most evident in the pancrustacean Dscam-hv which expresses diverse splicing repertoires both in nervous cells and hemocytes (the immune cells of both insects and crustaceans; Watson et al. 2005; Dong et al. 2006; Brites et al. 2008; Watthanasurorot et al. 2011). Here we show for the first time that the expression of Dscam diversity by hemocytes is not a derived character of pancrustaceans, the hemocytes of the myriapod *S. maritima* also express Dscam variants created by mutually exclusive alternative splicing of Ig7 coding exons.

Concerning the signaling of the molecule we have found CC1 motifs (PYATT; Prasad et al. 2007; Andrews et al. 2008) in all arthropod Dscams and in the Dscam of other invertebrates but not of vertebrates. Interestingly in vertebrates these motifs are also shared by CAMs loosely related to Dscam, such as roundabout, where they are involved in leukocyte mobility control among other functions (Wong et al. 2002; Prasad et al. 2007). Hence, the expression of Dscam diversity by arthropod hemocytes could be related to hemocyte mobility which in turn could have consequences both for immunity (Watthanasurorot et al. 2011) and organogenesis. This suggests that precursor functions of the extant Dscams could already be present in the ancestor of all arthropods. In such

cases it has been proposed that duplications evolve most likely by subfunctionalization, that is partition of the different functions of the ancestral gene among the paralogs (Ohno 1970; Force et al. 1999; Lynch and Conery 2000). However, the possibility that the acquisition of immune functions has evolved only after the genetic diversification by neofunctionalization of new paralogs remains open as the role of Dscam as a singleton gene in nonarthropods has not been studied beyond the nervous system involvement, neither the functions of the many *Ixodes* and *Strigamia* paralogs have been interrogated.

Unlike the results obtained by Armitage et al. (2012), we do not recover a monophyletic origin for all the Dscam genes of pancrustaceans. Nevertheless, the statistical confidence of the more basal nodes of the phylogeny is low (Fig. 5). The fact that different groups of pancrustaceans have different numbers of Dscam paralogs (Fig. 5) suggests that their most recent common ancestor had a large diversity of Dscam genes from which different paralogs were retained in different pancrustacean groups. Alternatively, new duplications and gene losses might have occurred in each specific pancrustacean group as discussed later.

THE ORIGIN OF DSCAM-HV IN PANCRUSTACEANS

Our data confirm that the Dscam gene with arrays of exons coding for Ig2, Ig3, and Ig7 evolved uniquely in the ancestor of pancrustaceans (Brites et al. 2008; Armitage et al. 2012). The diversity of Dscams found in *S. maritima* and *I. scapularis* recapitulates the pancrustacean Dscam-hv, that is Dscam molecules with mutually exclusive alternative splicing of internal duplications, Dscam molecules with alternative transmembrane domains such as in insects, Dscam soluble forms like in pancrustaceans (in decapode crustaceans a Dscam soluble form is encoded in the genome whereas in insects it is produced by proteolytic cleavage of membrane bound forms; Schmucker et al. 2000; Chou et al. 2009). We propose that the ancestors of pancrustaceans had a large number of such Dscam genes and that Dscam-hv through genetic processes such as tandem exon duplication, gene conversion, recombination, and domain lost. The intriguing question is why only those exons that confer diversity to Ig2, Ig3, and Ig7 domains duplicated and not others. Structural aspects of Dscam-hv and the molecular basis of its role in the nervous system provide insights into this question.

The formation of Dscam dimers through homophilic binding of identical isoforms leads to a self-avoidance behavior of nervous cells and is essential for neural wiring in *D. melanogaster* (Hughes et al. 2007; Matthews et al. 2007; Soba et al. 2007; Wojtowicz et al. 2007). Remarkably, the Dscam regions involved in dimer formation are fractions of Ig2, Ig3, and Ig7 domains coded by the duplicated exons (Meijers et al. 2007; Sawaya et al. 2008). In this way the genetic diversification caused by the duplications, coupled with the strong specificity of Dscam's homophilic binding, provide a huge repertoire of highly specific "key-locks" which nervous cells exploit extensively (Hughes et al. 2007; Matthews et al. 2007; Meijers et al. 2007; Soba et al. 2007; Wojtowicz et al. 2007; Sawaya et al. 2008). We propose that the homophilic binding between Dscam molecules having internal duplications coding for Ig2, Ig3, and Ig7 was the mechanism that drove selection on all those duplications because that increased the number of possible Dscam dimers. Duplications that conferred direct functional diversity would be selected whereas others would be lost by drift or by purifying selection in the pancrustacean ancestors.

Another possible explanation is that the duplicated regions are more prone to duplication, because they reside on recombination hot spots. A third possibility still, suggested by the existence in *Strigamia* and *Ixodes* of contiguous Dscam genes separated in some cases by relatively short genomic sequences, is that the transcription of such contiguous genes is not totally independent, producing a step-wise expression. Under this scenario, again based on the selection imposed by the specificity acquired via dimers formation, the composition of the ectodomains could have been shaped mainly by domain lost.

THE ORIGIN OF THE MUTUALLY EXCLUSIVE ALTERNATIVE SPLICING OF THE DUPLICATED EXONS

The mutually exclusive alternative splicing of the internal exon duplications coding for half of Ig2 and Ig3 domains and the complete Ig7 ensures that only one exon per array of duplications is present in the mature RNA. The Ig2 and Ig3 exon duplications encode only half domains, thus any duplicated exon transcribed constitutively would generate a nonfunctional protein and be deleterious. In the case of Ig7, given that it is encoded by a single exon, duplicates constitutively expressed could potentially code for a functional protein with several Ig7 domains. A plausible scenario is that the regulators of the alternative splicing mechanism of Ig7 were used in the ancestors of pancrustaceans to splice exons coding for Ig2 and Ig3 domains. In that case we would predict that the three arrays of duplications have common regulating features but this is still subject of controversy (Graveley 2005; Olson et al. 2007; Lee et al. 2009; Yang et al. 2011). The *S. maritima* Sm35 Ig7 transcripts revealed several possible combinations between the alternative exons and the flanking regions (Fig. 4). This has been rarely found in the transcripts of *Daphnia magna* (Brites et al. 2008; D. Brites, unpubl. data), suggesting that in *S. maritima* the mechanism is in a transitory state toward the greater efficiency of diversification seen in pancrustaceans. Therefore, further investigating *S. maritima* will possibly contribute to the understanding of the evolution of both the mechanism of generating Dscam diversity and more generally of the mechanism of mutually exclusive alternative splicing.

THE EVOLUTION OF SOMATIC VERSUS GERMLINE DIVERSIFICATION

A great diversity of cell surface receptors is a necessity in several biological systems, the two best examples being the nervous and the immune systems where individualization and specialization of many cells are at the basis of their function. The way to reach molecular diversity can be as simple as the creation of multigene families or as sophisticated as multiple somatic adaptations creating randomly huge repertoires of receptors, far larger than the number of genes present in the organism. Our results show that both modes of diversification can exist within a single phylum and within a single gene family. Pancrustaceans show a remarkable somatic diversification of Dscam due to mutually exclusive alternative splicing, making obsolete the conservation of a large number of duplicates. However, in *Strigamia* and *Ixodes* evolution has selected the creation and the maintenance of a large multigene family and alternative splicing is simpler and produces a smaller number of isoforms. This is one striking example occurring in one gene family of the following recognized phenomenon across different animals and genes: large gene families tend to have less alternative splicing occurrences than smaller gene families or singleton genes, and therefore the roles of gene duplication and alternative splicing in protein diversification may not be selected independently of each other during evolution (Su et al. 2006; Jin et al. 2008; Su and Gu 2012). One hypothesis for this negative correlation between alternative splicing and gene duplication (the "function-sharing" hypothesis) proposes that after gene duplication of a pleiotropic ancestral gene which is already alternatively spliced, different spliced exons with different functions are fixed over time in each of the duplicated genes and alternative splicing is lost (Yu et al. 2003; Kopelman et al. 2005; Su et al. 2006; Talavera and Castresana 2007). In *Strigamia* one possible scenario is that the exon duplications encoding Ig7 are being lost because their functional role is taken over by other paralogs. In pancrustaceans, the contrary might have happened, different paralogs may have disappeared as its functions became redundant once Dscam-hv evolved. The later might reflect the evolution toward a finer tuning capacity of regulating the expression of Dscam diversity.

In the context of the immunological role of Dscam it is clear that the creation within a single genome of multiple copies of slightly different genes, together with alternative splicing possibilities, corresponds to a frequent "evolutionary behavior" of immune genes under pressure to create diversified repertoires of recognition (Flajnik and Du Pasquier 2012). An analogous example of the Dscam situation is the immunoglobulin heavy chain locus in vertebrates where the number of genes is higher in species such as frogs where the somatic adaptation are not optimally selected, compared to mammals where selection of somatic variants is optimal (Flajnik and Du Pasquier 2012) .

The convergent evolution that led different arthropod groups to generate Dscam diversity suggests that diversification evolved as an adaptation driven by a common selection. The expression of Dscam in hemocytes of *S. maritima* leaves open the possibility of an important immunological role. However, we cannot answer whether the main driving diversifying selective pressure was exerted by the nervous or the immune system. Interestingly, selection tests based on species divergence and population genetic analysis, indicated that exon duplicates of Dscam-hv in pancrustaceans seem to have diverged mainly under neutral evolution (Brites et al. 2011). This suggests an evolutionary scenario in which accumulating amino acid diversity was more important than the exact amino acid sequences created.

ACKNOWLEDGMENTS

The authors are very thankful to M. Akam for providing us with access to the genome of *Strigamia maritima*. DE was supported by the Swiss National Science Foundation. The authors declare no conflicts of interest.

LITERATURE CITED

Abascal, F., R. Zardoya, and D. Posada. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104–2105.

Andrews, G. L., S. Tanglao, W. T. Farmer, S. Morin, S. Brotman, M. A. Berberoglu, H. Price, G. C. Fernandez, G. S. Mastick, F. Charron, and T. Kidd. 2008. Dscam guides embryonic axons by Netrin-dependent and -independent functions. Development 135:3839–3848.

Armitage, S. A., R. Y. Freiburg, J. Kurtz, and I. G. Bravo. 2012. The evolution of Dscam genes across the arthropods. BMC Evol. Biol. 12:53. doi: 10.1186/1471-2148-12-53.

Brites, D., S. McTaggart, K. Morris, J. Anderson, K. Thomas, I. Colson, T. Fabbro, T. J. Little, D. Ebert, and L. Du Pasquier. 2008. The Dscam homologue of the crustacean Daphnia is diversified by alternative splicing like in insects. Mol. Biol. Evol. 25:1429–1439.

Brites, D., F. Encinas-Viso, D. Ebert, L. Du Pasquier, and C. R. Haag. 2011. Population genetics of duplicated alternatively spliced exons of Dscam in *Daphnia* and *Drosophila*. PLoS ONE 6: e27947. doi:910.1371/journal.pone.0027947.

Buckley, C. D., G. E. Rainger, P. F. Bradfield, G. B. Nash, and D. L. Simmons. 1998. Cell adhesion: more than just glue (Review). Mol. Membr. Biol. 15:167–176.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17:540–552.

Chou, P. H., H. S. Chang, I. T. Chen, H. Y. Lin, Y. M. Chen, H. L. Yang, and K. C. H. C. Wang. 2009. The putative invertebrate adaptive immune protein Litopenaeus vannamei Dscam (LvDscam) is the first reported Dscam to lack a transmembrane domain and cytoplasmic tail. Develop. Comp. Immunol. 33:1258–1267.

Crayton, M. E., 3rd, B. C. Powell, T. J. Vision, and M. C. Giddings. 2006. Tracking the evolution of alternatively spliced exons within the Dscam family. BMC Evol. Biol. 6:16. doi:10.1186/1471-2148-6-16.

Dong, Y., C. M. Cirimotich, A. Pike, R. Chandra, and G. Dimopoulos. 2012. Anopheles NF-kappaB-regulated splicing factors direct pathogen-specific repertoires of the hypervariable pattern recognition receptor AgDscam. Cell Host Microbe 12:521–530.

Dong, Y., H. E. Taylor, and G. Dimopoulos. 2006. AgDdscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. PLoS Biol 4. doi:10.1371/journal.pbio.0040229.

Drummond, A., and K. Strimmer. 2001. PAL: an object-oriented programming library for molecular evolution and phylogenetics. Bioinformatics 17:662–663.

Endo, T., K. Ikeo, and T. Gojobori. 1996. Large-scale search for genes on which positive selection may operate. Mol. Biol. Evol. 13:685–690.

Flajnik, M. F., and L. Du Pasquier. 2012. Evolution of the immune system. Pp. 67–128 *in* W. E. Paul, ed. Fundamental immunology. Lippincott Williams & Wilkins, Philadelphia.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545.

Fusaoka, E., T. Inoue, K. Mineta, K. Agata, and K. Takeuchi. 2006. Structure and function of primitive immunoglobulin superfamily neural cell adhesion molecules: a lesson from studies on planarian. Genes Cells 11:541–555.

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

Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696–704.

Hattori, D., S. S. Millard, W. M. Wojtowicz, and S. L. Zipursky. 2008. Dscam-mediated cell recognition regulates neural circuit formation. Annu Rev Cell Dev Biol. 24:597–620.

Hughes, M. E., R. Bortnick, A. Tsubouchi, P. Baumer, M. Kondo, T. Uemura, and D. Schmucker. 2007. Homophilic Dscam interactions control complex dendrite morphogenesis. Neuron 54:417–427.

Hulpiau, P., and F. van Roy. 2009. Molecular evolution of the cadherin superfamily. Intl. J. Biochem. Cell Biol. 41:349–369.

Humphries, M. J., and P. Newham. 1998. The structure of cell-adhesion molecules. Trends Cell Biol. 8:78–83.

Jin, L., K. Kryukov, J. C. Clemente, T. Komiyama, Y. Suzuki, T. Imanishi, K. Ikeo, and T. Gojobori. 2008. The evolutionary relationship between gene duplication and alternative splicing. Gene. 427:19–31.

Kelley, L. A., and M. J. Sternberg. 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protocols 4:363–371.

Kondrashov, F. A., and E. V. Koonin. 2001. Origin of alternative splicing by tandem exon duplication. Human Mol. Genet. 10:2661–2669.

Kopelman, N. M., D. Lancet, and I. Yanai. 2005. Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. Nat. Genet. 37:588–589.

Kreahling, J. M., and B. Graveley. 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.

Lee, C., N. Kim, M. Roy, and B. R. Graveley. 2009. Massive expansions of Dscam splicing diversity via staggered homologous recombination during arthropod evolution. RNA 16:91–105.

Lefranc, M.-P., and G. Lefranc. 2001. The immunoglobulin facts book. Academic Press, London.

Letunic, I., T. Doerks, and P. Bork. 2009. SMART 6: recent updates and new developments. Nucleic Acids Res. 37:D229–D232.

Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. Science (New York, NY) 290:1151–1155.

Malnic, B., P. A. Godfrey, and L. B. Buck. 2004. The human olfactory receptor gene family. Proc. Natl. Acad. Sci. USA 101:2584–2589.

Maniatis, T., and B. Tasic. 2002. Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature 418:236–243.

Matthews, B. J., M. E. Kim, J. J. Flanagan, D. Hattori, J. C. Clemens, S. L. Zipursky, and W. B. Grueber. 2007. Dendrite self-avoidance is controlled by Dscam. Cell. 129:593–604.

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

Millard, S. S., J. J. Flanagan, K. S. Pappu, W. Wu, and S. L. Zipursky. 2007. Dscam2 mediates axonal tiling in the *Drosophila* visual system. Nature 447:720-U714.

Miller, M., M. Holder, R. Vos, P. Midford, T. Liebowitz, L. Chan, P. Hoover, and T. Warnow. 2009. The CIPRES portals. CIPRES. Available at

Nylander, J. A. A., J. C. Wilgenbusch, D. L. Warren, and D. L. Swofford. 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24: 581–583.

Ohno, S. 1970. Evolution by gene duplication. Springer, Heidelberg, Germany.

Olson, S., M. Blanchette, J. Park, Y. Savva, G. W. Yeo, J. M. Yeakley, D. C. Rio, and B. R. Graveley. 2007. A regulator of Dscam mutually exclusive splicing fidelity. Nat. Struct. Mol. Biol. 14:1134–1140.

Pal, C., B. Papp, and M. J. Lercher. 2006. An integrated view of protein evolution. Nat. Rev. Genet. 7:337–348.

Prasad, A., Z. Qamri, J. Wu, and R. K. Ganju. 2007. Pivotal advance: slit-2/Robo-1 modulates the CXCL12/CXCR4-induced chemotaxis of T cells. J. Leuk. Biol. 82:465–476.

Rast, J. P., L. C. Smith, M. Loza-Coll, T. Hibino, and G. W. Litman. 2006. Genomic insights into the immune system of the sea urchin. Science (New York, NY) 314:952–956.

Regier, J. C., J. W. Shultz, A. Zwick, A. Hussey, B. Ball, R. Wetzer, J. W. Martin, and C. W. Cunningham. 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature 463:1079–1098.

Rota-Stabelli, O., E. Kayal, D. Gleeson, J. Daub, J. L. Boore, M. J. Telford, D. Pisani, M. Blaxter, and D. V. Lavrov. 2010. Ecdysozoan mitogenomics: evidence for a common origin of the legged invertebrates, the Panarthropoda. Genome Biol. Evol. 2:425–440.

Sawaya, M. R., W. M. Wojtowicz, I. Andre, B. Qian, W. Wu, D. Baker, D. Eisenberg, and S. L. Zipursky. 2008. A double S shape provides the structural basis for the extraordinary binding specificity of Dscam isoforms. Cell. 134:1007–1018.

Schmucker, D., J. C. Clemens, H. Shu, C. A. Worby, J. Xiao, M. Muda, J. E. Dixon, and S. l. Zypursky. 2000. Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell 101: 671–684.

Schultz, J., F. Milpetz, P. Bork, and C. P. Ponting. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. USA 95:5857–5864.

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

Soba, P., S. Zhu, K. Emoto, S. Younger, S. J. Yang, H. H. Yu, T. Lee, L. Y. Jan, and Y. N. Jan. 2007. Drosophila sensory neurons require Dscam for dendritic self-avoidance and proper dendritic field organization. Neuron 54:403–416.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Su, Z., and X. Gu. 2012. Revisit on the evolutionary relationship between alternative splicing and gene duplication. Gene. 504:102–106.

Su, Z., J. Wang, J. Yu, X. Huang, and X. Gu. 2006. Evolution of alternative splicing after gene duplication. Genome Res. 16:182–189.

Talavera, G., and J. Castresana. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56:564–577.

Trowsdale, J., and P. Parham. 2004. Mini-review: defense strategies and immunity-related genes. Eur. J. Immunol. 34:7–17.

Waterhouse, A. M., J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton. 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191.

Watson, L. F., F. T. Püttmann-Holgado, F. Thomas, D. L. Lamar, M. Hughes, M. Kondo, V. I. Rebel, and D. Schmucker. 2005. Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science (New York, NY) 309:1874–1878

Watthanasurorot, A., P. Jiravanichpaisal, H. P. Liu, I. Soderhall, and K. Soderhall. 2011. Bacteria-induced Dscam isoforms of the Crustacean, *Pacifastacus leniusculus*. PLoS Pathog. 7(6):e1002062. doi:10.1371/journal.ppat.1002062.

Wojtowicz, W. M., W. Wu, I. Andre, B. Qian, D. Baker, and S. L. Zipursky. 2007. A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. Cell 130:1134–1145.

Wong, K., H. T. Park, J. Y. Wu, and Y. Rao. 2002. Slit proteins: molecular guidance cues for cells ranging from neurons to leukocytes. Curr. Opin. Genet. Develop. 12:583–591.

Woolhouse, M., J. Webster, E. Domingo, B. Charlesworth, and B. Levin. 2002. Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nature genetics 32:569–577.

Yamakawa, K., Y.-K. Huo, M. A. Haendel, R. Hubert, X.-N. Chen, G. E. Lyons, and J. R. Korenberg. 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. Human Mol. Genet. 7:227–237.

Yang, Y., L. Zhan, W. Zhang, F. Sun, W. Wang, N. Tian, J. Bi, H. Wang, D. Shi, Y. Jiang, et al. 2011. RNA secondary structure in mutually exclusive splicing. Nat. Struct. Mol. Biol. 18:159–168.

Yu, H. H., J. S. Yang, J. Wang, Y. Huang, and T. Lee. 2009. Endodoamin diversity in the *Drosophila* Dscam and its roles in neuronal morphogenesis. J. Neurosci. 29:1904–1914.

Yu, W. P., S. Brenner, and B. Venkatesh. 2003. Duplication, degeneration and subfunctionalization of the nested synapsin-Timp genes in Fugu. Trends Genet. 19:180–183.

Associate Editor: A. Navarro

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table S1. Accession numbers of Dscam homologs and other cell adhesion molecule proteins from selected metazoan representatives.

Figure S1. Bayesian topology of a partial region (Ig8 to FNII-2) of Dscam related genes in representatives of metazoa.

Figure S2. *Strigamia maritima* reconstructions of Dscam homologs.

Figure S3. Maximum likelihood topology of the nucleotide sequences of the duplicated exons coding for Ig7 in the different *S. maritima* Dscam homologs.

Figure S4. Number of amino acid substitutions per site calculated with a pair-wise analysis of the Poisson corrected distance among different Dscam domains of paralogs.

Figure S5. *Ixodes scapularis* reconstructions of Dscam homologs.

Figure S6. *Ixodes scapularis* reconstructions of Dscam homologs present in contig 92235.

Figure S7. Conservation of CC1 motif PTPYATT between Human Roundabout and DSCAM family molecules from invertebrates.

Figure S8. Maximum likelihood topology depicting the phylogenetic relationship between Ig7 coding exons ($n = 177$) for different Dscam from different species.

Figure S9. Representation of the amino acid conservation of exons coding for Ig7 of Dscam-hv of six pancrustacea species and of all other Dscam homologs in the remaining species (Table S1).

Figure S10. Aminoacid sequences of the *Strigamia maritima* reconstructions.

Figure S11. Amino acid sequences of the *Ixodes scapularis* reconstructions.