

Repeated Evolution of Identical Domain Architecture in Metazoan Netrin Domain-Containing Proteins

Lucas Leclère* and Fabian Rentzsch

Sars International Centre for Marine Molecular Biology, University of Bergen, Norway

*Corresponding author: E-mail: lucas.leclere@sars.uib.no.

Accepted: July 11, 2012

Abstract

The majority of proteins in eukaryotes are composed of multiple domains, and the number and order of these domains is an important determinant of protein function. Although multidomain proteins with a particular domain architecture were initially considered to have a common evolutionary origin, recent comparative studies of protein families or whole genomes have reported that a minority of multidomain proteins could have appeared multiple times independently. Here, we test this scenario in detail for the signaling molecules netrin and secreted frizzled-related proteins (sFRPs), two groups of netrin domain-containing proteins with essential roles in animal development. Our primary phylogenetic analyses suggest that the particular domain architectures of each of these proteins were present in the eumetazoan ancestor and evolved a second time independently within the metazoan lineage from laminin and frizzled proteins, respectively. Using an array of phylogenetic methods, statistical tests, and character sorting analyses, we show that the polyphyly of netrin and sFRP is well supported and cannot be explained by classical phylogenetic reconstruction artifacts. Despite their independent origins, the two groups of netrins and of sFRPs have the same protein interaction partners (Deleted in Colorectal Cancer/neogenin and Unc5 for netrins and Wnts for sFRPs) and similar developmental functions. Thus, these cases of convergent evolution emphasize the importance of domain architecture for protein function by uncoupling shared domain architecture from shared evolutionary history. Therefore, we propose the terms merology to describe the repeated evolution of proteins with similar domain architecture and discuss the potential of merologous proteins to help understanding protein evolution.

Key words: multidomain protein, domain architecture, convergence, netrin, sFRP, phylogeny.

Introduction

Protein domains are distinct units that can fold autonomously into a particular, stable, three-dimensional structure and are often conserved during evolution. In eukaryotes, the majority of proteins are multidomain proteins, and the particular number and order of these domains defines the domain architecture of the protein (Koonin et al. 2002). During evolution, domains can be recombined into different arrangements to create proteins with new functions, a process that contributes significantly to the expansion of protein repertoires despite a limited number of domains. The generation of multidomain proteins occurs through gene fusion and fission events that can lead to gain, rearrangement, or loss of domains, whereas gene duplication can generate protein families with shared domain architecture (Weiner et al. 2006; Moore et al. 2008). Thus, the unique domain architecture of each protein family is classically considered to have originated only once during evolution (Vogel et al. 2005). However, recent work

proposed that a subset of domain architectures could have appeared several times independently. In an analysis of 62 genomes, Gough (2005) estimated that between 0.4% and 4% of domain architectures could be the result of convergent evolution. In a study based on gene trees instead of species trees, Forslund et al. (2008) argued that even between 5.6% and 12.4% of domain architecture could have originated several times independently, although approximately two-third of the cases involve only the loss of domains. These findings suggest that in contrast to traditional concepts, convergent evolution of domain architecture might significantly contribute to the expansion of proteomes.

Although the strength of the above-mentioned studies lies in their broad sampling of complete genomes, the huge size of these datasets necessarily limits the depth of analysis and the type of approaches that can be used. In the case of the analyses based on species trees (Gough 2005; Kummerfeld

and Teichmann 2005), the different domain architectures are plotted on species trees to reconstruct the most parsimonious scenario for the origin of these multidomain proteins. This type of approach cannot take into account cases of horizontal gene transfer or independent evolution of the same domain architecture in the same lineage. Approaches based on domain phylogeny (Forslund et al. 2008) take into account these two possibilities (Yanai et al. 2002; Jordan et al. 2003; Zhang et al. 2010; Wu et al. 2011); however, they rely on the phylogenies of short protein domains that often lead to only moderate support for the obtained tree topologies and that can be subject to various tree reconstruction artifacts. Therefore, detailed studies of individual cases are necessary to validate or reject the possibility of independent evolution of the same domain architecture.

Here, we focus on two families of secreted developmental regulators, netrin and secreted frizzled-related proteins (sFRPs), because complex phylogenetic patterns for these two protein families have been noticed before but have led to contradictory interpretations (see below). Netrins and sFRPs are essential regulators of embryonic development (Serafini et al. 1996; Satoh et al. 2006). Netrins regulate axon guidance and other developmental processes by binding to the transmembrane receptors Deleted in Colorectal Cancer (DCC)/Neogenin and Unc5 (Moore et al. 2007; Rajasekharan and Kennedy 2009; Lai Wing Sun et al. 2011), whereas sFRPs act as modulators of the Wnt signaling pathway, which plays a prominent role in axial patterning (Bovolenta et al. 2008; Petersen and Reddien 2009; Mii and Taira 2011). Both protein groups contain a so-called netrin domain (Banyai and Patthy 1999; Chong et al. 2002), which is characterized by an enrichment of basic residues and a particular spacing pattern of cysteines, that cause it to fold into one β -barrel and two α -helices, which are located at the N- and C-termini of the domain (Banyai and Patthy 1999; Liepinsh et al. 2003; Bramham et al. 2005). Netrin domains are present in various multidomain proteins that have diverse overall structure and function, for example, complement components C3–C5 or WFIKKN (WAP, Follistatin/kazal, immunoglobulin [IG], Kunitz, and netrin domain-containing protein). They are also found in the single domain protein tissue inhibitor of metalloproteases (TIMPs), present in metazoans and Eubacteria (Brew and Nagase 2010).

Netrin and sFRP multidomain proteins are widespread in metazoans and constitute multigene families. Netrin proteins are composed of two parts: the N-terminal part is a supra-domain (Vogel et al. 2004) that consists of one LamininNT domain plus three epidermal growth factor (EGF) domains, homologous to domains VI and V of laminins, and the C-terminal part contains one netrin domain (Banyai and Patthy 1999; Koch et al. 2000; Qin et al. 2007; Rajasekharan and Kennedy 2009). In contrast, netrin-G1 and netrin-G2 proteins are composed of the LamininNT-3EGF supra-domain plus a particular C-terminal domain that is not related to the

netrin domain (Nakashiba et al. 2000; Yin et al. 2002; Rajasekharan and Kennedy 2009). sFRPs are composed of two domains: the N-terminal part is a frizzled-cysteine-rich domain (CRD) domain and the C-terminal part is a netrin domain (Banyai and Patthy 1999; Chong et al. 2002; Bovolenta et al. 2008).

sFRP polyphyly among frizzled-type proteins has been noticed previously (Adamska et al. 2010), but the shared domain architecture was considered as evidence for an artifact of the phylogenetic reconstruction. For netrins, grouping of the LamininNT-3EGF supra-domain of netrin-1/2/3/5, netrin-4, and netrin-G with different Laminin groups has been observed (Koch et al. 2000; Nakashiba et al. 2000; Yin et al. 2002; Moore et al. 2007; Rajasekharan and Kennedy 2009) but based on their common domain architecture, the former two groups of netrins have been considered to come from a single ancestor. Recently, Fahey and Degnan (2012) reinterpreted this phylogenetic pattern as an indication of independent evolution of the same domain architecture. However, the phylogenetic support for an independent origin of different netrin groups was weak, and none of the above-mentioned studies have tested this possibility thoroughly. Therefore, we have analyzed the evolution of netrin domain-containing proteins in detail and have recovered a phylogenetic pattern supporting a convergence of domain architecture for both netrin and sFRP proteins. We assessed the strength of these hypotheses of convergence using a broad range of reconstruction methods and tests. We show that an independent origin of netrin-1/2/3/5 and netrin-4 is strongly supported and cannot be explained by known tree reconstruction artifacts (functional convergence, gene conversion, mutational saturation, heterogeneity of base composition, long-branch attraction, and heterotachy). Polyphyly of sFRPs was clearly favored by phylogenetic analyses and was not caused by the tested reconstruction artifacts. However, statistical tests did not reject monophyletic tree topologies for sFRPs, probably because of the short size of the domains contained in these proteins. These findings strongly suggest that the protein architecture shared by the two groups of netrins and the two groups of sFRPs does not reflect common evolutionary ancestry but instead is the result of independent events of domain rearrangement. The similar molecular interactions and functions of the two groups of netrins and sFRPs provide a striking example for the importance of domain architecture for protein function, independently of shared evolutionary history.

Materials and Methods

Whole Genomes Analyses

We searched for frizzled-CRD, netrin, and LamininNT domains and frizzled, netrin-1/2/3/5, netrin-4, netrin-G, Laminin, sFRP-1/2/5, and sFRP-3/4 proteins, and the netrin receptors

Neogenin/DCC and Unc5 in 388 complete genomes belonging to all major eukaryote clades. This included all draft and finished eukaryote genomes available from the Joint Genome Institute (JGI) and NCBI (see [supplementary table S1, Supplementary Material](#) online for databases, genome, and sequence information) on 1 September 2011. Gene searches were performed using BLAST (blastp and tblastn) with *Nematostella*, *Strongylocentrotus*, *Drosophila*, and *Mus* proteins as query sequences, against protein and genome databases with the default BLAST parameters and an e-value threshold of 0.1. We then used various validated sequences as queries for a second round of BLAST search on complete genomes. In addition, NCBI Conserved Domain Database (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?>), PFAM (<http://pfam.sanger.ac.uk/>—domain references: netrin, PF01759; CRD-frizzled, PF01392; frizzled-7TM, PF01534; LamininNT, PF00055; EGF, PF00053), Interpro (<http://www.ebi.ac.uk/interpro/>), UniProt (<http://www.uniprot.org/uniprot/>), and Superfamily (<http://supfam.cs.bris.ac.uk/SUPERFAMILY/>) were searched for the different proteins and domains in complete Eukaryota, Eubacteria, and Archea genomes.

Data Sampling and Assembly for Phylogenetic Analyses

To reconstruct the phylogeny of netrin domains, frizzled-CRD domains, and LamininNT-3EGF supra-domains, we recovered all genes containing at least one of these domains or supra-domains by BLAST search (tblastn and blastp) against a selection of metazoan and choanoflagellate complete genomes: *Mus musculus* (NCBI), *Danio rerio* (NCBI), *Branchiostoma floridae* (JGI), *Strongylocentrotus purpuratus* (NCBI + SpBase), *Caenorhabditis elegans* (NCBI + WormBase), *Drosophila melanogaster* (NCBI + FlyBase), *Capitella teleta* (JGI), *Lottia gigantea* (JGI), *Nematostella vectensis* (JGI), *Trichoplax adhaerens* (JGI), *Amphimedon queenslandica* (spongezome), and *Monosiga brevicollis* (JGI). As outgroups we included Usherin, TIMP, and Smoothened sequences for the LamininNT-3EGF, netrin, and frizzled-CRD domain, respectively. The 11 metazoan genomes include 304 frizzled-CRD domain-containing sequences. As a result of massive tandem domain duplication, more than 80 of these domains are encoded in the *Branchiostoma floridae* genome. Initial maximum-likelihood (ML) analysis (data not shown) with all 304 sequences showed that one-third of them fall in a clade containing almost only frizzled receptors and sFRPs. We excluded sequences that did not fall into this clade from further analysis.

Each domain dataset was aligned independently using the software Muscle (Edgar 2004) under default parameters and adjusted manually in BioEdit (Hall 1999). Partial sequences, positions ambiguously aligned or containing more than 70% of gaps and/or missing data were deleted (see [supplementary table S2, Supplementary Material](#) online for details

about sequences). Alignments of nucleotide sequences were generated based on the amino acid alignments using BioEdit.

To exclude fast-evolving sequences and test the impact of the outgroup on the topology, we removed *D. melanogaster*, *C. elegans*, *S. purpuratus*, *T. adhaerens*, *A. queenslandica*, and *M. brevicollis* and outgroup sequences (Usherin, TIMP, and Smoothened), and some particularly unstable sequences between nonparametric bootstrap replicates of the complete amino acid dataset (see [supplementary figs. S7–S9, Supplementary Material](#) online for more details). We refer to these matrices as “reduced datasets” in comparison with the “complete datasets” containing all sequences.

Amino acid alignments for neogenin/DCC and Unc5 were generated as described above with protogenin/hope/punc and ankyrin as outgroups for neogenin/DCC and Unc5, respectively (Salbaum and Kappen 2000; Toyoda et al. 2005; Wang et al. 2009). Neogenin and DCC are both composed of four IG and five fibronectin 3 (FN3) domains, whereas the outgroup sequences are composed of four IG and three to five FN3 domains. To produce an accurate domain alignment, we performed phylogenetic analyses of the individual FN3 and IG domains present in the neogenin–protogenin proteins of the mouse genome (see [Supplementary Material](#) online for more details).

Phylogenetic Analyses

Bayesian phylogenetic inferences were performed using MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003) under the amino acid substitution models WAG + $\Gamma(4) + I$. For the nucleotide dataset, partitioned Bayesian analyses were carried out using a GTR + $\Gamma(4) + I$ model for each codon position with parameter settings optimized independently for each of the three codon partitions. We ran twice 2 searches of four chains for 2 million generations each, sampled every 100 generations, for all datasets except for the complete frizzled-CRD data for which the same number of chains was run for 5 million generations. All other settings were kept as default. Convergence was estimated for each search by using the standard deviation of split frequencies and potential scale reduction factors reported by the software and by checking stasis of the likelihood values using the command “sump” in MrBayes. Posterior probabilities (PP) were estimated by constructing a majority rule consensus of trees, sampled every 100 generations from 1,500,000 to 2,000,000 generations. For the complete frizzled-CRD domain phylogeny, we sampled from 4,500,000 to 5,000,000. Finally, the consensus trees of the two independent searches were compared to confirm convergence on the same topology.

ML analyses were performed using PhyML 3.0 (Guindon and Gascuel 2003) with BioNJ starting tree and NNI branch swapping. The best-fitting model of amino acid substitution for each dataset was estimated using ProtTest v2.4 (Abascal et al. 2005) under the Akaike information criterion (starting

tree: tree from a preliminary ML analysis using PhyML and a LG + Γ + I model with 8 rate categories for the γ distribution). The selected models were WAG + Γ + I for netrin and LamininNT-EGF domain datasets, and LG + Γ + I for frizzled-CRD, neogenin/DCC, and Unc5 datasets. We used the GTR + Γ + I model for all ML analyses of the nucleotide datasets. We considered eight rate categories for the γ distribution in all ML analyses.

Maximum parsimony (MP) analyses were performed using PAUP* 4.0b10 (Swofford 2002). All characters were treated as equally weighted, and unordered and gaps were treated as missing data. Heuristic analyses were performed with 500 random addition of sequences and the TBR algorithm for branch swapping.

Branch robustness in the MP and ML analyses was estimated using nonparametric bootstrap (BP) (Felsenstein 1985) with 100 or 500 replicates depending on the analyses (10 random addition sequences for each MP bootstrap replicate). Trees were visualised using Mega 5.0 (Tamura et al. 2011).

Approximately Unbiased and Parametric Bootstrap Tests

For the approximately unbiased (AU) test (Shimodaira 2002), different topologies were generated by rearranging in TreeView (Page 1996) the branching order of the ML trees having the best likelihood after PhyML and RaxML (7.2.6) (Stamatakis 2006) analyses using the complete amino acid domain datasets and the WAG + $\Gamma(8)$ + I model. In addition, for each hypothesis of monophyly, we also included the best topology obtained with RaxML (model WAG + $\Gamma(8)$ + I) under the appropriate constraint and various rearrangements of this topology. Branch lengths were estimated using Tree Puzzle 5.1 under the WAG + $\Gamma(8)$ model. Likelihoods of these different test topologies were compared with each other and with the likelihood of the best PhyML tree by the AU test using CONSEL (Shimodaira and Hasegawa 2001). For a given tested hypothesis, the selected P value corresponded to the highest P value obtained among the topology displaying the tested clade. Monophyly was then rejected if the clade was not found in any of the nonrejected tested topologies.

For the parametric bootstrap test (SOWH test of Goldman et al. 2000), the null hypotheses were that sFRP and/or netrin had a single ancestor and that the polyphyly obtained in domain phylogenies were the result of reconstruction errors. We used the ML tree generated with RaxML under the constraint that sFRP and/or netrin were monophyletic (model WAG + $\Gamma(8)$ + I) as the null hypothesis topology T_0 (list of tested hypothesis in table 2). We generated 100 simulated datasets using Seq-Gen version 1.3.2 (Rambaut and Grassly 1997) of the same size as the original one, taking into account the T_0 topology (including branch length), proportion of amino acid determined empirically, and parameter of the model from the constrained ML analysis including shape

of the γ distribution and proportion of invariable sites. This procedure was repeated for each tested null hypothesis. Unconstrained and constrained ML searches (constrained under the topology T_0 obtained from the constrained ML analysis of the real dataset) under the WAG + $\Gamma(8)$ + I model were conducted on each simulated dataset using PhyML with re-estimation of all free parameters in each case. We computed the difference in ML scores between these two optimizations ($\delta^{(i)} = L_{ML}^{(i)} - L_{T_0}^{(i)}$) for each simulated dataset to generate a frequency distribution. This provided an empirical estimate of the null distribution and allowed us to generate a critical value δ^* corresponding to the 5% tail of the null distribution where the null hypothesis was statistically rejected. This δ distribution was then compared with the $\delta^{(RD)} = L_{ML} - L_{T_0}$ obtained from the real dataset corresponding to the difference in likelihood values between the unconstrained and the constrained (netrin and/or sFRP monophyly) analyses.

Saturation Analysis

The nucleotide and amino acid substitution saturation of the different domains was evaluated by plotting, for each pair of sequences, the total number of differences against the number of substitutions inferred from the ML trees (as the sum of the length of all branches linking these two sequences). Observed and inferred distances were obtained in PAUP* 4.0b10.

Analysis of the Effect of Individual Sites on Netrin and sFRP Polyphyly

We assessed the influence of particular sites on the polyphyly of netrin and sFRP in the laminin-EGF, netrin, and frizzled-CRD domain amino acid matrices by comparing the log-likelihood values ($\ln L$) for each site under unconstrained (polyphyly) and constrained (monophyly) analyses. Constrained analyses included netrin-1/netrin-4, netrin-1/netrin-G, and netrin-4/netrin-G for the LamininNT-EGF supra-domain; sFRP-1/2/5/sFRP-3/4 for the frizzled-CRD domain; and sFRP-1/2/5/sFRP-3/4, netrin-1/2/3/5/netrin-4, and netrin-4/Cnidaria-Bilatera-netrin-1 for the netrin domain. Per-site log likelihood (psln L) were recovered after constrained and unconstrained analyses in RaxML, and the difference in per-site log likelihood (Δ psln L) between competing hypotheses was calculated. To assess the effects of the highest and lowest Δ psln L on polyphyly, we extracted the corresponding sites and reanalyzed the culled matrix under the same conditions.

Identification of Slow-Evolving and Heterotachous Positions in the LamininNT-EGF and Frizzled-CRD Datasets

A simple method to sort sites according to their rate variation was derived from the slow-fast method (Brinkmann and Philippe 1999). Aligned sequences were divided into seven

groups from the complete amino acid alignments: laminin- α , β , γ and netrin-1, 4, G and Usherin for LamininNT-EGF (laminin- β / γ -like, *Monosiga* and *Amphimedon* sequences were not considered); and frizzled-1/2/7-3/6, frizzled-4, frizzled-5/8, frizzled-9/10, sFRP-1/2/5, sFRP-3/4, and Smoothened for frizzled-CRD. We calculated the number of substitutions per amino acid position within each group using PAUP*. The evolutionary rate of a given position was estimated as the sum of the numbers of steps for this position within the seven groups. Positions were then sorted according to their total number of steps (those having the same number of steps were sorted randomly), to produce a list of amino acid positions from the slowest to fastest evolving.

To identify the level of heterotachy per amino acid position, we computed the absolute difference between the total number of steps per site per clades. For the LamininNT-EGF supra-domain, this was done between netrin-1-Laminin- γ and netrin-4-Laminin- β clades, and for the frizzled-CRD domain between the groups frizzled-5/8-frizzled-1/2/7-3/6-sFRP-3/4 and frizzled-4-frizzled-9/10-sFRP-1/2/5. We then sorted positions according to their absolute difference in steps between clades and sites displaying the same value being sorted randomly. Using a chi-square test, we tested for each “ Δ steps per site” category whether the heterogeneity inferred between the subgroups was significant.

From both “fast-evolving” and “heterotachy” site lists, we generated nine matrices containing from 10% to 90% positions. This allowed us to study the evolution of nodal support for netrin and sFRP polyphyly (100 ML bootstrap replicates in PhyML) as increasingly fast-evolving or heterotachous positions were removed. We also plotted these two values per sites with the Δ psln L values from the comparative netrin-4-netrin-1 and sFRP-1/2/5-sFRP-3/4 monophyly–polyphyly analyses described above.

Results

Phylogenetic Analyses Suggest Polyphyly of Netrins and sFRPs

The netrin and sFRP protein families share a C-terminal domain enriched in basic residues and a particular spacing pattern of cysteines, the netrin domain. This domain is also present in several other proteins, such as complement components C3–C5, WIFKKN, and TIMP (Banyai and Patthy 1999), and it can be found in all metazoan genomes. Outside metazoa, according to our genome survey (388 different eukaryote genomes analyzed—table 1 and supplementary table S1, Supplementary Material online), this domain is only found as a single domain protein, TIMP, in a few eukaryotes (*Sphaeroforma arctica* and *Ectocarpus siliculosus*) and various Eubacteria (Brew and Nagase 2010). We found netrin-1/2/3/5 and sFRP-1/2/5 and sFRP-3/4 proteins in bilaterian and nonbilaterian genomes (but no sFRP-1/2/5 in insects and no

sFRP-3/4 in protostomes) and netrin-4 proteins only in deuterostomes (but not in the tunicate *Oikopleura dioica* and the echinoderm *S. purpuratus*). In reconstructing the netrin domain phylogeny using Maximum Likelihood and Bayesian analysis, with amino acid sequences from 11 metazoan genomes (see Materials and Methods), we recovered that netrin and sFRP protein families were each divided into two distantly related clades: netrin-1/2/3/5 and netrin-4, and sFRP-1/2/5 and sFRP-3/4, respectively (fig. 1A). ML analyses of the netrin domain showed monophyly for all the other gene families containing a netrin domain (TIMP, complement component C3–C5, WIFKKN, ADAMTSL5, and PColCE—fig. 1A and supplementary fig. S1, Supplementary Material online). This phylogenetic pattern suggests the possibility of an independent origin of different groups of sFRP and netrin proteins.

To investigate this hypothesis, we next analyzed the N-terminal supra-domain of the netrin proteins, LamininNT-EGF (one LamininNT + 3 EGF domains), which is also present in Laminin proteins. LamininNT domains alone, or coupled with three EGF domains, are present only in metazoan and choanoflagellate genomes (table 1). Phylogenetic analyses of the LamininNT-EGF supra-domain lead to a strongly supported topology (fig. 1B). As in the netrin domain phylogeny, analyses of the LamininNT-EGF supra-domain suggest polyphyly of netrin proteins. ML and Bayesian analyses are highly congruent and place netrin-1/2/3/5 as the sister group of eumetazoan Laminin- γ , netrin-4 as the sister group of Laminin- β , and netrin-G (a chordate specific group of netrin proteins that actually lack a netrin domain, see Nakashiba et al. 2000) inside the clade composed of Laminin- β / γ -like (fig. 1B and supplementary fig. S3, Supplementary Material online), a recently described group of Laminins (Fahey and Degnan 2012). According to this phylogeny in which netrin clades are nested within paraphyletic groups of laminin, and laminin proteins of Porifera are sister groups to eumetazoan netrin–laminin protein subgroups, the N-terminal supra-domains of netrins are unambiguously derived from laminin proteins.

In sFRP proteins, the Netrin domain is combined with another CRD, the frizzled domain. This domain is also present in the frizzled family of G protein-coupled receptors (Frizzled and Smoothened), where it is coupled with the frizzled-7 transmembrane (frizzled-7TM) domain, and in several other proteins (e.g., the receptor tyrosine kinases Musk and ROR). The domain is present in many eukaryote genomes, but the combination with a Netrin domain, as in sFRPs, is restricted to metazoans, whereas the combination with a frizzled-7TM domain occurs in metazoans, “non-Dikarya” fungi and amoebozoans (Dictyosteliida) (table 1). sFRPs are unambiguously derived from frizzled-type proteins as the two sFRP subgroups cluster inside clades of frizzled sequences (fig. 1C and supplementary fig. S5, Supplementary Material online), and the phylogeny of the second domain present in Frizzled receptors (frizzled-7TM domain) clearly groups all frizzled-like sequences together within the superfamily of G protein-coupled

Table 1

Distribution of Frizzled-CRD, LamininNT, and TIMP/Netrin Domains and sFRP and Netrin Proteins in Sequenced Genomes

	Domains			Proteins				
	Fzd-CRD	LamininNT	TIMP/Netrin	sFRP-1/2/5	sFRP-3/4	Netrin-1	Netrin-4	Netrin-G
Vertebrata	+	+	+	+	+	+	+	+
Urochordata	+	+	+	+	+	+	+/- ^a	+/- ^a
Cephalochordata	+	+	+	+	+	+	+	+
Hemichordata	+	+	+	+	+	+	+	-
Echinodermata	+	+	+	+	+	+	-	-
Arthropoda	+	+	+	+/- ^b	-	+	-	-
Nematoda	+	+	+	+	-	+	-	-
Mollusca	+	+	+	+	-	+	-	-
Annelida	+	+	+	+	-	+	-	-
Platyhelminthes	+	+	+	+	-	+	-	-
Cnidaria	+	+	+	+	+	+/- ^c	-	-
Placozoa	+	+	+	-	+	+	-	-
Porifera	+	+	+	? ^d	? ^d	-	-	-
Choanoflagellata	-	+	-	-	-	-	-	-
Filasterea	-	-	+/-	-	-	-	-	-
Fungi	+/-	-	-	-	-	-	-	-
Amoebozoa	+/-	-	-	-	-	-	-	-
Apusozoa	+	-	-	-	-	-	-	-
Chromalveolata	+/-	-	+/-	-	-	-	-	-
Haptophyta	+	-	-	-	-	-	-	-
Cryptophyta	-	-	-	-	-	-	-	-
Rhizaria	+/-	-	-	-	-	-	-	-
Archaeplastida	+/-	-	-	-	-	-	-	-
Excavata	+/-	-	-	-	-	-	-	-
Eubacteria	-	-	+/-	-	-	-	-	-
Archaea	-	-	-	-	-	-	-	-

Note.— +, domain or protein present in all genomes checked; +/-, domain or protein present in some genomes checked; -, domain or protein not present in genomes checked.

^aNot present in the genome of *Oikopleura dioica*.

^bNot present in insect genomes.

^cNot present in the genome of *Hydra magnipapillata*.

^dNo true sFRP (frizzled-CRD + netrin domains) in the complete genome of *Amphimedon queenslandica* but presence of a very divergent sequence in the sponge *Lobomirski baicalensis* (Adell et al. 2007) of unclear homology.

receptors (data not shown). Consistent with the analysis of the Netrin domain, ML analyses of the frizzled domain support polyphyly of the two distinct groups of sFRPs, sFRP-1/2/5 and sFRP-3/4 (fig. 1C). Thus, the combined phylogenetic analyses of the domains present in netrins and in sFRPs suggest that the domain architecture of both protein groups evolved two times independently.

Assessing the Strength of Netrin and sFRP Polyphyly Hypotheses

Independent evolution of the domain architecture in different Netrin and sFRP protein groups is not a parsimonious scenario and, therefore, needs to be thoroughly tested. We rigorously assessed the support for the polyphyly hypothesis using three approaches. First, we analyzed each domain at the amino acid and nucleotide levels using different reconstruction methods. This allowed us to use different types of models and data for

the same alignment and to detect reconstruction artifacts such as those caused by convergence at the amino acid level due to functional constraint (Li et al. 2010). Second, we reduced the number of sequences by removing outgroups, long branches, and particularly unstable sequences in bootstrap replicates. Removal of distant and divergent sequences can lead to more stable and accurate phylogenies (Gates et al. 2007). Third, we performed nonparametric (AU test) and parametric (parametric bootstrap test—SOWH test) likelihood-based statistical tests to assess the strength of the signal supporting sFRP and Netrin polyphyly.

When analyzing the netrin domain, polyphyly for sFRP and Netrin proteins was obtained in ML and Bayesian analyses of the amino acid and nucleotide datasets (fig. 1A and supplementary figs. S1 and S2, Supplementary Material online). In most of these analyses netrin-4 was the sister group of sFRP-3/4. However, the ML bootstrap or Bayesian posterior probability values for this topology were low. Removal of

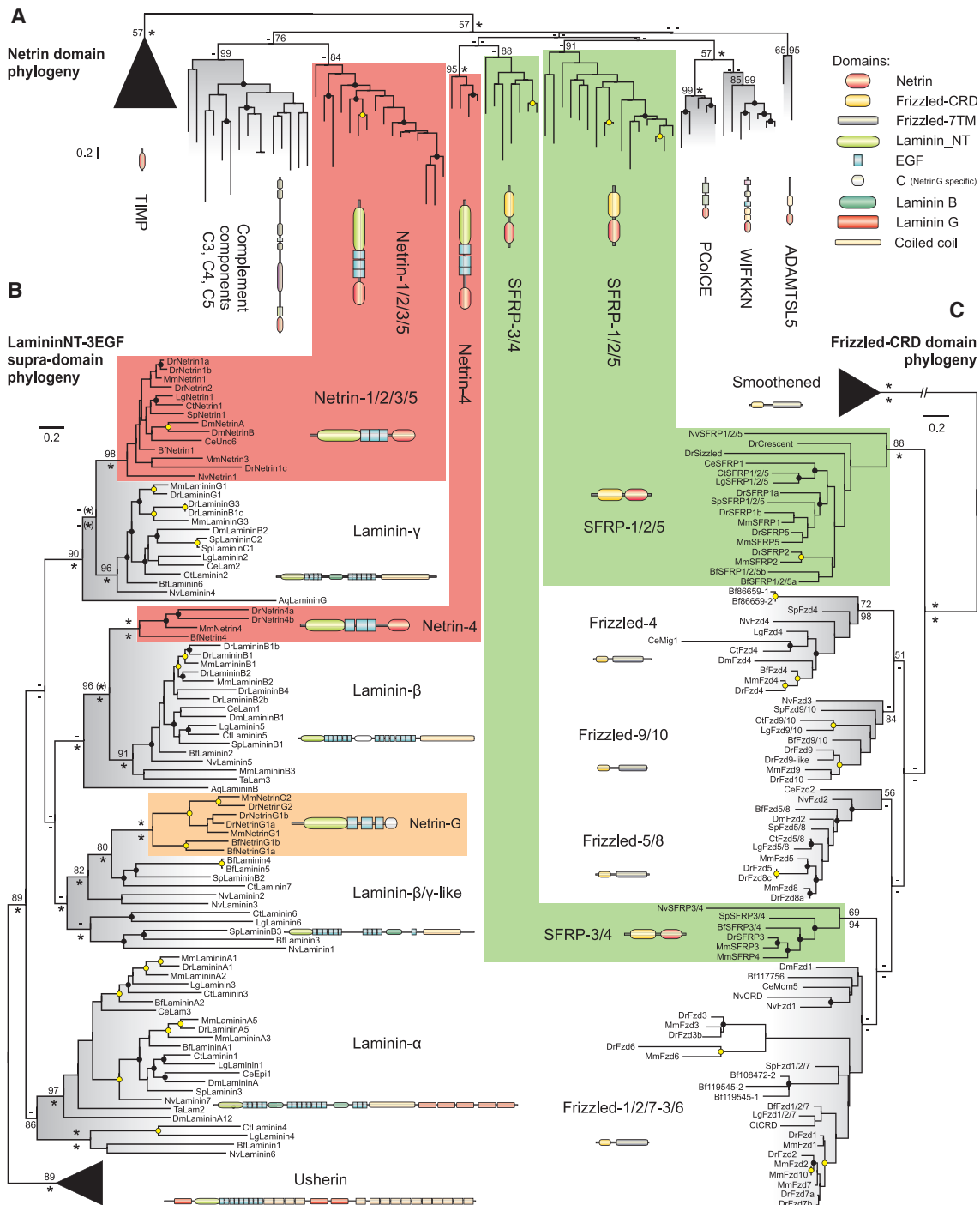


FIG. 1.—Phylogenetic analyses of the complete amino acid domain datasets support polyphyly of netrins and sFRPs. (A) Netrin domain maximum likelihood (ML) analysis under a WAG + $I(8)$ + I model (111 aa, 101 sequences, $-\ln L$ 21377.15); (B) LamininNT-3EGF supra-domain ML analysis under a model WAG + $I(8)$ + I (363 aa, 99 sequences, $-\ln L$ 50081.22); (C) Frizzled-CRD domain ML analysis under a model LG + $I(8)$ + I (112 aa, 87 sequences, $-\ln L$ 10857.68). For deep branches, nonparametric bootstrap values BP (ML)—500 replicates—are indicated on the left (A) or above the branches (B and C), and Bayesian posterior probability (PP) are indicated on the right or below the branches. Asterisks indicate branches with maximum support for both BP (ML) and PP. A dash indicates branches with BP (ML) < 50% and PP < 70%. (B) Values in parenthesis correspond to BP (ML) and PP values from analyses without *Amphimedon* and *Monosiga* sequences. For other branches, black dot indicates PP \geq 90%, yellow dot indicates PP \geq 95% and BP (ML) \geq 90%. The scale bar indicates the estimated number of substitution per site. Consistent grouping of netrin and sFRP subfamilies in individual domain phylogenies are highlighted in red and green, respectively. (A–C) Domain composition of proteins are sketched next to each subgroup and are oriented N- to C-terminal from top to bottom in A and from left to right in B and C. Size of netrin and sFRP protein sketches are double that for the other proteins. The two first letters of gene names in B and C correspond to the first letters of genus and species names (see Materials and Methods).

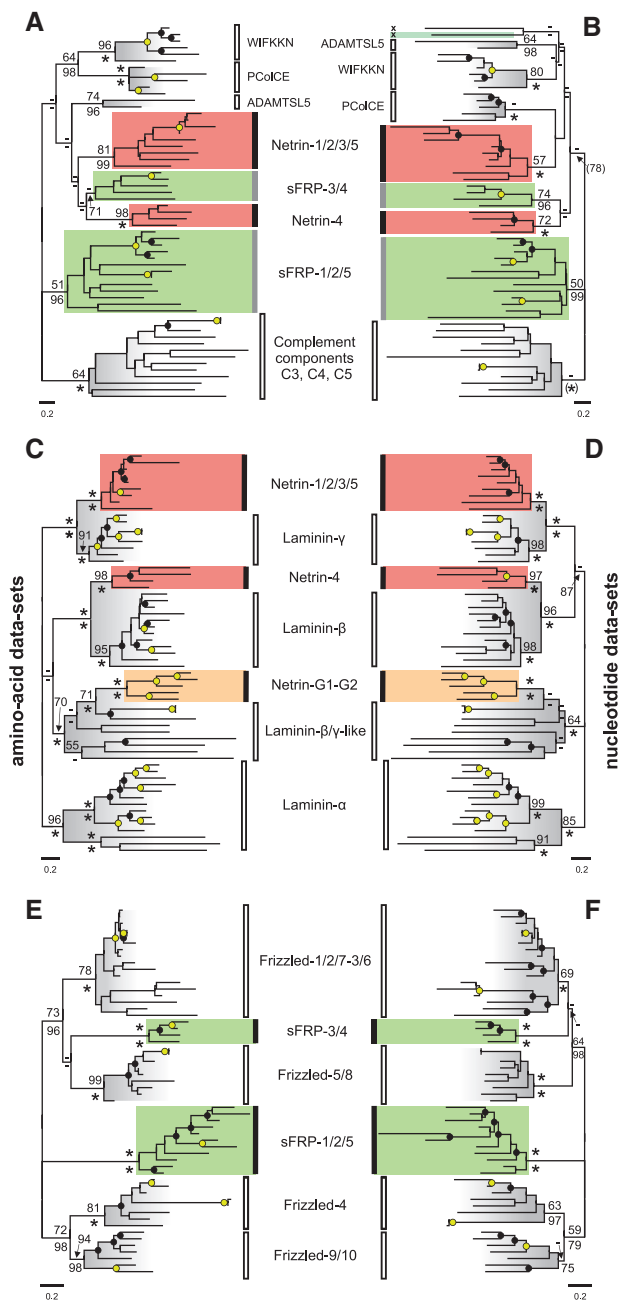


FIG. 2.—Polyphyly of netrins and sFRPs is confirmed in reduced amino acid (A, C, and E) and nucleotide (B, D, and F) datasets. Unstable, fast-evolving and outgroup sequences were excluded from the datasets before re-analyses. (A and B) Netrin domain ML analysis under WAG + $\Gamma(8) + I$ (111 aa, 57 sequences, $-\ln L$ 11711.10) and GTR + $\Gamma(8) + I$ (333 nt, 57 sequences, $-\ln L$ 19511.50) models; (C and D) LamininNT-3EGF supra-domain ML analysis under WAG + $\Gamma(8) + I$ (363 aa, 61 sequences, $-\ln L$ 30275.53) and GTR + $\Gamma(8) + I$ (1089 nt, 61 sequences, $-\ln L$ 58079.49) models; (E and F) frizzled-CRD domain ML analysis under LG + $\Gamma(8) + I$ (112 aa, 56 sequences, $-\ln L$ 5969.88) and GTR + $\Gamma(8) + I$ (1089 nt, 56 sequences, $-\ln L$ 13272.39) models. For deep branches, nonparametric bootstrap values BP (ML)—500 replicates—and Bayesian PP are indicated above and below the branches, respectively. Asterisks indicate branches with maximum support for both BP (ML) and

fast-evolving and outgroup sequences led to similar polyphyletic topologies for sFRPs and netrins, but it did not increase the bootstrap and Bayesian support values for the deep nodes (fig. 2A and B and [supplementary fig. S7, Supplementary Material](#) online). AU and parametric bootstrap tests both failed to reject a topology where netrins and/or sFRPs are monophyletic (table 2), but they excluded grouping of netrin-4 and deuterostome netrin-1/2/3/5, suggesting that the netrin domain of netrin-4 is not derived from the netrin domain of netrin-1/2/3/5.

Analyses of the LamininNT-EGF supra-domain led to a strongly supported topology (fig. 1B), with congruence between phylogenetic analyses of amino acid and nucleotide datasets ([supplementary figs. S3 and S4, Supplementary Material](#) online). Reduction of the sequence and species sampling did not modify the general relationships between laminin and netrin subgroups and led to maximal support in MP and ML bootstraps and Bayesian analyses for grouping the laminin and netrin subgroups (fig. 2C and D and [supplementary fig. S8, Supplementary Material](#) online). We also analyzed the LamininNT domain and the three EGF domains in two separate datasets and obtained in both cases the same netrin–laminin subgrouping (data not shown). Furthermore, when only synonymous substitutions (third codon positions—reduced dataset) of the LamininNT–EGF supra-domain were analyzed in ML, netrin-1/2/3/5, netrin-4, and netrin-G were still sister groups of laminin- γ , laminin- β , and laminin- β/γ -like, respectively (data not shown). In this analysis, some netrin-1/2/3/5 sequences of *Danio* and *Lottia* and Netrin-4 of *Danio* were not clustering together or with any group of Netrin, highlighting the poor conservation of the third codon positions. Nevertheless, this result ruled out the possibility of a reconstruction artifact due to convergent selection pressure at the amino acid level between netrin and laminin subgroups. Both the AU and parametric bootstrap tests strongly rejected grouping of the three netrin subfamilies or the grouping of any of the three possible pairs (see table 2).

These results show that polyphyly of netrin proteins, in LamininNT-EGF and Netrin domain phylogenies, is consistently obtained in complete and reduced amino acid and nucleotide datasets and with different reconstruction methods. However, statistical tests rejected monophyly of netrins only for the LamininNT-EGF supradomain and not for the netrin domain.

Polyphyly of sFRPs in the analyses of the frizzled-CRD domain was recovered in ML analyses of the complete datasets but with low support values (BP and PP < 50%, fig. 1C). In ML analyses of both nucleotide and amino acid datasets, we obtained sFRP-1/2/3/5 as the sister group to a clade containing

PP. A dash indicates branches with BP (ML) < 50% and PP < 70%. For other branches, PP \geq 90% are indicated by a black dot, and PP \geq 95% + BP (ML) \geq 90% are indicated by a yellow dot. The scale bar indicates the estimated number of substitution per site.

Table 2

Results of the Approximately Unbiased and Parametric Bootstrap Tests for Comparison of Alternative Phylogenetic Hypotheses

Hypotheses	ln L	δ ln L	AU Test (P)	PB Test (P)	Rejected
ML netrin domain	-21377.15	-	-	-	
sFRP-1/2/5 as sister to sFRP-3/4	-21378.98	1.83	0.696	0.97	No
Netrin-1/2/3/5 as sister to netrin-4	-21381.31	4.16	0.574	0.81	No
Both netrin and sFRP monophyletic	-21383.31	6.16	0.443	0.62	No
Netrin-4 as sister to Cnidaria/Bilateria netrin-1/2/3/5	-21390.72	13.57	0.219	0.10	No
Netrin-4 as sister to Deuterostomia netrin-1/2/3/5	-21411.41	34.26	0.041	<0.01	Yes
ML LamininNT-EGF supra-domain	-50081.22	-	-	-	
Netrin-1/2/3/5 as sister to NetrinG	-50157.58	76.36	0.002	<0.01	Yes
Netrin-4 as sister to netrin-G	-50186.93	105.71	0.001	<0.01	Yes
Netrin-1/2/3/5 as sister to Netrin4	-50202.58	121.36	0.0004	<0.01	Yes
Netrin-4, netrin-G, netrin-4 grouping together	-50271.00	189.78	0.0002	<0.01	Yes
ML frizzled-CRD domain	-10872.47	-	-	-	
sFRP1/2/5 as sister to sFRP3/4	-10876.87	4.40	0.601	0.59	No

Note.—ML analyses on complete amino acid datasets under WAG+Γ(8)+I model by PhyML. Constrained analyses performed by RaxML. Log-likelihood values recalculated by PhyML using model, topology, and free parameters from RaxML analyses. See [supplementary fig. S11, Supplementary Material](#) online for details about the parametric bootstrap analyses. Bold values indicate significant results at the 5% level.

sFRP-3/4 sequences and four subgroups of frizzled: frizzled-1/2/7-3/6, frizzled-4, frizzled-5/8, and frizzled-9/10 ([supplementary figs. S5 and S6, Supplementary Material](#) online). In the reduced and unrooted datasets, sFRPs were still polyphyletic with a bipartition into sFRP3/4-frizzled5/8-frizzled1/2/7-3/6 and sFRP1/2/5-frizzled4-frizzled9/10. This topology was supported in both amino acid and nucleotide reduced datasets ([fig. 2E and F, supplementary fig. S9, Supplementary Material](#) online) with high Bayesian probability (amino acid [aa] dataset: 96%; nucleotide [nt]: 98%) and moderate ML bootstrap values (aa: 73%, nt: 64%). Parsimony analyses provided a resolved topology only with the reduced dataset, showing the same bipartition but with very low bootstrap values (<50% for both aa and nt datasets). ML analysis of the third codon position also led to a topology where sFRP was polyphyletic, but this analysis failed to recover monophyletic frizzled subgroups. The difference in log likelihood between the two competing hypotheses was small ([table 2](#)), and AU and parametric bootstrap tests did not reject the hypothesis of monophyly of sFRPs in the complete ([table 2](#)) or in the reduced dataset (AU test $P=0.376$). These analyses show that the polyphyly of sFRP in netrin and frizzled-CRD domains is consistent across methods and sampling but that monophyly cannot be ruled out statistically using the current phylogenetic methods. To assess the significance of the observed polyphyly, we therefore tested whether different types of known tree reconstruction artifacts might affect the topology of the obtained trees.

Analysis of Substitution Saturation of the Domains

Accumulation of multiple substitutions at the same position over time erases the true phylogenetic signal and can cause tree reconstruction artifacts. When multiple substitutions affect most of the positions, the dataset can become

mutationally saturated (Jeffroy et al. 2006). We performed a saturation analysis of the different domains at the amino acid ([fig. 3](#)) and nucleotide level on the ML trees ([supplementary fig. S10, Supplementary Material](#) online). The slope of the linear regression between the numbers of observed differences (y axis) and inferred substitutions (x axis) is proportional to the quantity of homoplasy present in the data. Saturation can be detected when the number of inferred substitutions increased, whereas the number of observed differences remains constant (plateau shape and slope close to zero).

Saturation in the complete dataset of the netrin domain appeared high at both the amino acid (slope=0.2505; [fig. 3A](#) and [supplementary fig. S10, Supplementary Material](#) online) and nucleotide (slope=0.2687; [supplementary fig. S10, Supplementary Material](#) online) levels with most of the pairwise comparison located on a plateau (slope=0.0391). This pattern indicates that saturation is reached even for comparisons between relatively closely related proteins, and this is probably causing the difficulties in reconstructing the phylogeny of this domain with accuracy. The LamininNT-EGF supra-domain (slope=0.2992, [fig. 3B](#) and [supplementary fig. S10, Supplementary Material](#) online) and frizzled-CRD (slope=0.3783, [fig. 3C](#) and [supplementary fig. S10, Supplementary Material](#) online) domain complete datasets were also saturated, but less so than the netrin domain. Saturation appeared to be slightly lower in the amino acid datasets than in the nucleotide datasets and lower in the reduced than in the complete datasets (slope for the reduced amino-acid domain datasets: netrin: 0.2798, LamininNT-EGF: 0.3779, and frizzled-CRD: 0.4739; [supplementary fig. S10, Supplementary Material](#) online). This indicates that although the saturation observed in these domains is partially due to fast-evolving sequences and distant outgroups, it is mainly due to the great evolutionary distance separating the proteins containing each of these domains.

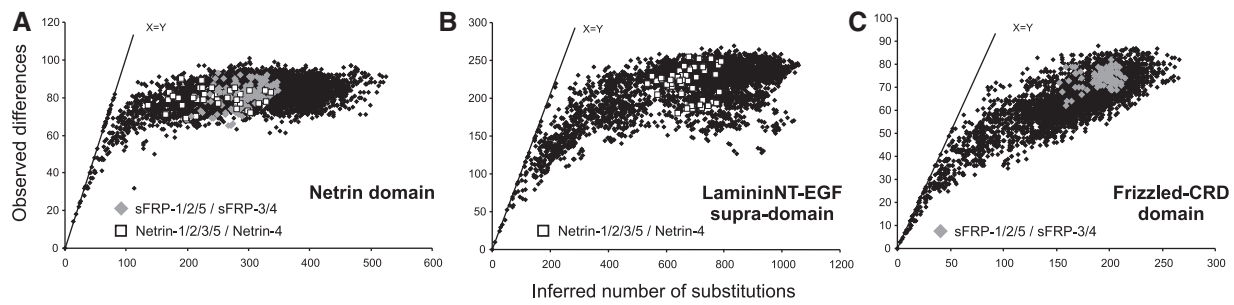


Fig. 3.—Netrin, LamininNT-EGF, and frizzled-CRD domains display a significant level of substitution saturation. Estimation of the substitution saturation of the domains netrin (A), LamininNT-EGF (B), and frizzled-CRD (C) at the amino acid level (complete datasets) as a ratio between inferred (x axis) and observed (y axis) differences for each pair of sequences. Inferred number of substitutions between pairs of sequences were determined using parsimony on the best ML trees. White squares and grey diamonds represent netrin-1/2/3/5-netrin-4 and sFRP-1/2/5-sFRP-3/4 pairwise comparison, respectively. Data points on the straight line $X = Y$ correspond to completely unsaturated comparisons.

Identification of Sites Most Influencing Netrin and sFRP Polyphyly

To investigate from which sites the signal for polyphyly originates for each domain, and test for conflicting signal (phylogenetic vs. nonphylogenetic), we computed the difference in log likelihood per-site ($\Delta\text{psln } L$) between ML analyses with or without monophyletic constraint for netrin and sFRP on the complete amino acid datasets. In figure 4, sites with positive y -axis values have a higher likelihood for the unconstrained topology in which netrin or sFRP is polyphyletic, whereas sites with negative y -axis values have a higher likelihood for the constrained topology in which netrin or sFRP is monophyletic.

When analyzing the LamininNT-EGF supra-domain, we obtained a clear majority of site supporting the polyphyly of the three netrin protein groups (netrin-1/2/3/5 + netrin-4 + netrin-G, fig. 4A) and polyphyly of the netrins sensu stricto (netrin-1/2/3/5 + netrin-4, fig. 4B), with the signal for polyphyly being stronger in the LamininNT domain than in the three EGF domains. Most of the sites were in favor of polyphyly (64%) but some conflicting sites with strong signal against polyphyly were found distributed throughout the protein sequence. For the netrin domain of netrins, the majority of sites (55%) were in favor of netrin polyphyly, and these sites with positive values were not clustered on the gene sequence, arguing against conflict due to gene conversion (fig. 4B).

In sFRP proteins, a narrow majority of sites in both the frizzled-CRD and the netrin domain supported polyphyly (54% for the frizzled-CRD and 55% for the netrin domain, fig. 4C). As for netrin proteins, the sites in favor of sFRP polyphyly in frizzled-CRD and netrin domain were not spatially restricted.

To exclude that the polyphyly of netrins and sFRPs is due to few sites with very high $\Delta\text{psln } L$ values, we progressively removed sites with the highest and lowest $\Delta\text{psln } L$ values from the top 5% to 25% (10%–50% removed sites in total) and performed ML analyses on these reduced datasets. In the domain datasets, the 25% highest plus 25% lowest $\Delta\text{psln } L$

sites represented most of the total sum of the absolute $\Delta\text{psln } L$ between the competing hypotheses (82% for LamininNT-EGF, 91.5% for frizzled-CRD, and 84% for netrin domain). For the LamininNT-EGF domain, removing these sites had no influence on the relationships between laminin and netrin subgroups. Only a few sequences with long branches had variable positions in the different replicates, in particular, the laminin sequences of the poriferan *A. queenslandica*. For the netrin domain, removing the most influential sites did not affect sFRP or netrin polyphyly but had some impact on the relationships between the different protein groups. However, in all analyses, we obtained netrin-4 and sFRP-3/4 as closely related. For the frizzled-CRD domain, removing the most influential sites did not lead to monophyly of sFRP. However, after removing 50% of the most influential sites, the rooting changed from being the sister of sFRP-1/2/5 to the sister of frizzled-3/6. This was probably the result of a long-branch attraction (LBA) artifact since the frizzled-3/6 clade contained only vertebrate sequences and has the longest branch of the ingroup. In all the resulting topologies, ingroup sequences were subdivided into the same two groups as in the complete analysis: sFRP-3/4, frizzled-5/8, frizzled-1/2/7-3/6 and sFRP-1/2/5, frizzled-4, and frizzled-9/10.

All together, these analyses show that the observed polyphyly of Netrin and sFRP proteins are not caused by a dominating influence of few sites with exceptionally high $\Delta\text{psln } L$ values and do not originate from restricted clusters of sites, and therefore, the observed polyphyletic groupings were not caused by gene conversion.

Netrin and sFRP Polyphyly Are Not Caused by Classical Tree Reconstruction Artifacts

For both LamininNT-EGF and frizzled-CRD domain datasets, we could detect a certain amount of substitution saturation and conflict between sites (figs. 3 and 4), possible indications of systematic bias affecting the topology. Three major biases that strongly affect phylogenetic reconstruction have been

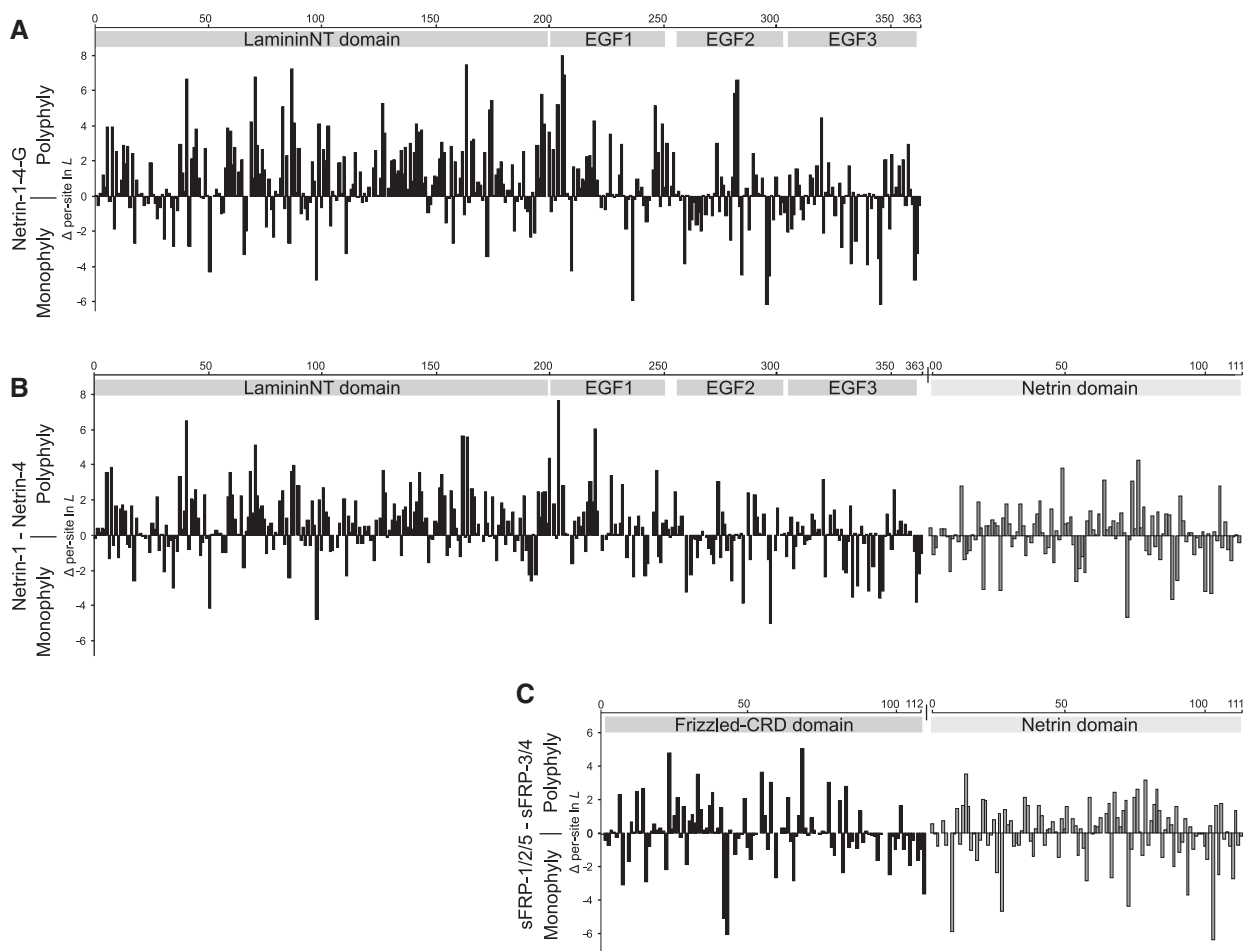


Fig. 4.—Distribution of the polyphyly versus monophyly signal for netrins and sFRPs. Differences in log likelihood per-site ($\Delta_{\text{persite}} \ln L$) between unconstrained and constrained maximum likelihood analyses of (A) LamininNT-EGF supra-domain, with netrin-1/2/3/5 + netrin-4 + netrin-G constrained as monophyletic; (B) LamininNT-EGF and netrin domains, with netrin-1/2/3/5 + netrin-4 constrained as monophyletic; (C) frizzled-CRD and netrin domain, with sFRP-1/2/5 + sFRP-3/4 constrained as monophyletic. The x axes correspond to the alignment columns along the complete amino acid matrices and the y axes correspond to the $\Delta_{\text{persite}} \ln L$ between unconstrained and constrained ML analyses. The sites with positive y axis values have a higher likelihood for the unconstrained topology in which netrin or sFRP is polyphyletic, whereas the sites with negative y axis values have a higher likelihood for the constrained topology in which netrin or sFRP is monophyletic.

described: 1) heterogeneity in base composition (Foster and Hickey 1999; Delsuc et al. 2005; Sheffield et al. 2009); 2) LBA, artificially grouping sequences that share high evolutionary rates (Bergsten 2005; Delsuc et al. 2005); and 3) heterotachy that refers to shifts in site-specific evolutionary rates over time and can lead to the grouping of sequences that share covariant sites (Lopez et al. 2002; Philippe et al. 2005). To clarify the nature of the signal supporting netrin and sFRP polyphylics, we assessed to which extent the phylogenetic reconstruction of LamininNT-EGF and frizzled-CRD datasets were influenced by these tree reconstruction artifacts.

First, we analyzed heterogeneity in base composition of the three domain datasets: netrin, frizzled-CRD and LamininNT-EGF. We could not detect significant variation for base composition among protein groups in the different nucleotide and

amino acid datasets using chi-square test, rejecting this possible source of tree reconstruction artifacts.

To address the possibility that LBA artifacts cause the sFRP and netrin protein polyphylics in the LamininNT-EGF and frizzled-CRD datasets, we selectively analyzed slow evolving sites. These sites are known to retain better phylogenetic signal and are less subject to LBA (Brinkmann et al. 2005). We used a method derived from the slow-fast method (Brinkmann and Philippe 1999) to sort the characters according to the sum of their evolutionary rate in monophyletic groups (see Materials and Methods). This method does not consider the deep nodes under study and thus avoids problems of circularity. In both LamininNT-EGF and frizzled-CRD datasets, the slowest evolving 20% of sites had almost no signal for or against netrin and sFRP polyphyly. Most of the

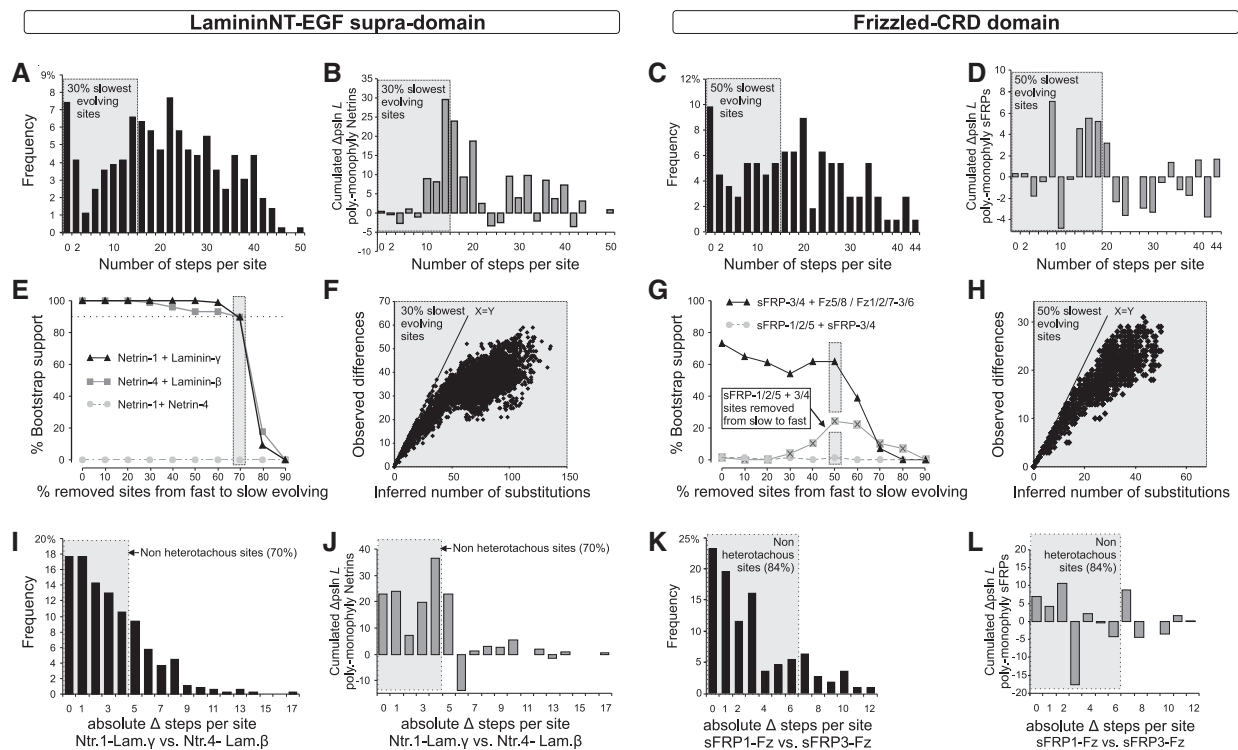


FIG. 5.—Polyphylies of netrins and sFRPs are supported by slow-evolving sites and are not caused by heterotachy in the ML analyses of the LamininNT-EGF (A, B, E, F, I, and J) and frizzled-CRD (C, D, G, H, K, and L) amino acid datasets. (A and C) Proportion of sites for each rate category, corresponding to the calculated number of steps in seven monophyletic groups using parsimony. For displaying purpose, each category contains two merged sequential values. (B and D) Cumulated difference in log likelihood per-site between unconstrained and constrained (B: netrin-1-4 monophyletic; D: sFRP-1/2/5-3/4 monophyletic) ML analysis for all sites within each rate category. (E and G) “Evolution” of the ML bootstrap support values (100 replicates) as fast-evolving sites are progressively removed from the original dataset; (E) 90% of bootstrap support is figured by a dotted line; (G) the “evolution” of BP-ML support value for sFRP monophyly is also indicated as slow-evolving sites are progressively removed from the original dataset. (F and H) Estimation of the mutational saturation as a ratio between inferred (x axis) and observed differences (y axis) for each pair of sequences in the LamininNT-EGF (F) and frizzled-CRD (H) datasets containing, respectively, the 30% and 50% slowest evolving sites. Data points on the straight line $X = Y$ correspond to completely unsaturated comparisons. Data coming from the analyses of the 30% slowest evolving sites of the LamininNT-EGF dataset (in A, B, E, and F) and of the 50% slowest evolving sites of the frizzled-CRD dataset (in C, D, G, and H) are shaded. (I and K) Histogram of the absolute difference of steps per site calculated between the netrin-1-laminin- γ and netrin-4-laminin- β clades for the LamininNT-EGF dataset (I) and between the frizzled-5/8-frizzled-1/2/7-3/6-sFRP-3/4 and frizzled-4-frizzled-9/10-sFRP-1/2/5 clades for the frizzled-CRD dataset (K). (J and L) Cumulated difference in log likelihood per-site between unconstrained and constrained (netrin-1-4 monophyletic in J; sFRP-1/2/5-3/4 monophyletic in L) ML analysis for all sites within each “ Δ steps per site” category. Data coming from the analyses of the 70% nonheterotachous sites of the LamininNT-EGF dataset (I and J) and of the 84% nonheterotachous sites of the frizzled-CRD dataset (K and L) are shaded.

signal in favor of polyphyly came from moderately slow evolving sites (fig. 5A–D). Interestingly, in both cases, most of the signal in favor of monophyly was found to come from the fast-evolving sites (fig. 5B and D). When removing 10%–70% of sites in the LamininNT-EGF dataset, starting from the fastest evolving, branch support for Netrin-1-Laminin- γ and Netrin-4-Laminin- β stayed $\geq 90\%$ (fig. 5E). In none of the bootstrap replicates was monophyly of Netrin obtained. Furthermore, the difference between observed and inferred differences in the slowest evolving 30% of sites was reasonably low (fig. 5F, slope = 0.4652, to compare with 0.2992 for the dataset without site deletion), showing that sites with a low level of saturation also supported netrin polyphyly. Inversely, when removing the character in the opposite

order (starting from the slowest to the fastest), the ML bootstrap support was below 90% for laminin- γ -netrin-1 after removing only 50% of sites (data not shown). In the frizzled-CRD datasets, removing up to 50% of the fastest evolving sites of the “reduced” dataset did not significantly affect the topology or the support values, with moderate support for sFRP polyphyly (fig. 5G, BP-ML: 62%; PP: 92%) and no support for sFRP monophyly still retained (BP-ML: 0%–1%). However, the level of saturation strongly decreased (fig. 5H, slope = 0.6402, to compare with 0.4739 for the dataset without site deletion). Conversely, removing the slow-evolving sites led to a strong increase in support for sFRP monophyly (fig. 5G, BP-ML: 24%–22% after removing 50%–60% of the slowest evolving sites) and of the saturation

level (slope of the plateau = 0.1719, data not shown). These analyses show that most of the signal in favor of sFRP and netrin polyphyly comes from slowly evolving sites and that most of the signal in favor of monophyly comes from fast-evolving and saturated sites, clearly arguing against an LBA artifact as the cause for the polyphyletic tree topologies.

Finally, we assessed the level of heterotachy in the LamininNT-EGF and frizzled-CRD amino acid sites and could also exclude its influence on the topologies. For the LamininNT-EGF supra-domain, we compared the difference in the number of substitution steps per site between the laminin- β -netrin-4 and laminin- γ -netrin-1 clades and sorted sites according to their level of heterotachy between the laminin and netrin subgroups (fig. 5I). Using a chi-square test, we identified sites with a difference in number of steps below five between two groups, as nonheterotachous. They account for approximately 70% of sites. The nonheterotachous sites showed a clear signal in favor of netrin polyphyly, contrary to heterotachous positions (fig. 5J). Furthermore, we found that the laminin-netrin groupings were still strongly supported (BP-ML: 98%–100%; PP: 100%) after removing all the heterotachous positions. Similarly, sorting sites according to their level of heterotachy between the frizzled-5/8-frizzled-1/2/7-3/6-sFRP-3/4 and frizzled-4-frizzled-9/10-sFRP-1/2/5, we could define that 84% of sites in the frizzled-CRD domain were homotachous (sites with difference in number of steps ≤ 6 —fig. 5K). Both homotachous and heterotachous positions provided signal for and against sFRP polyphyly (fig. 5L). However, the same frizzled-sFRP grouping was recovered after removing all heterotachous positions (BP-ML: 57%; PP: 97%). These analyses exclude the possibility that sFRP and netrin polyphyly are due to a reconstruction artifact caused by heterotachy.

Netrin Receptors Phylogeny

Phylogenetic analyses of the neogenin and Unc5 receptors reveal a more classical evolutionary history, with a unique origin in the Cnidaria-Bilateria ancestor and diversification through gene duplication (supplementary figs. S12 and S13, Supplementary Material online). We did not find these receptors in genomes of nonmetazoans, poriferans, or placozoans. For both proteins, diversification occurred in the vertebrates, probably caused by the two genome duplications at the base of vertebrates (reviewed in Kasahara 2007). These events led to the formation of neogenin and DCC and Unc5A, B, C, and D.

Discussion

Repeated Evolution of Domain Architecture of Netrin and sFRP Proteins

The results of the different phylogenetic analyses and statistical tests on the LamininNT-3EGF supra-domain strongly

support a scenario where the N-terminal supra-domain (one LamininNT + 3 EGF domains) of netrins evolved independently three times from the C-terminal part of different laminins: netrin-1/2/3/5 from laminin- γ , netrin-4 from laminin- β (fig. 6A) and netrin-G from laminin- β/γ -like. Laminin- β/γ -like is a newly described group of laminins that shares structural similarities with both laminin- β and laminin- γ and is present in eumetazoans with the exception of ecdysozoans, urochordates, and vertebrates (Fahey and Degnan 2012; supplementary fig. S3, Supplementary Material online). For netrin-1/2/3/5 and netrin-4, the N-terminal part of laminin fused C-terminally to a netrin domain, whereas netrin-G acquired a short and unique CRD (C domain) (Yin et al. 2002). The relatively poor resolution of the netrin domain phylogeny does not allow us to unambiguously determine the origin of the netrin domains found in netrin-1/2/3/5 and netrin-4. However, we could exclude that the netrin domain in netrin-4 is derived from the netrin domain of the older netrin-1/2/3/5. Thus, the domains of these two groups of netrins have completely independent origins.

The phylogenetic distribution of the different netrin groups suggests that netrin-1/2/3/5 was present in the ancestor of eumetazoans (Bilateria, Cnidaria, and Placozoa), and the phylogeny of the LamininNT-EGF supra-domain further shows that netrins did not originate before the common ancestor of Eumetazoa. The netrin-1/2/3/5 group expanded at the base of the vertebrates by gene duplication. Netrin-4 was most likely present in the ancestor of deuterostomes (fig. 6A), although an earlier occurrence followed by multiple losses cannot be ruled out. Netrin-G1 was present in the ancestor of chordates.

For sFRPs, our phylogenetic reconstruction of both Netrin and Frizzled-CRD domains suggest an independent origin for sFRP-1/2/5 and sFRP-3/4 before the last common ancestor of eumetazoans (fig. 6B). For both domains, we could not find obvious reconstruction bias, but statistical tests were not able to reject monophyletic tree topologies. The weak phylogenetic signal is probably due to the short size of these two domains and the ancestry of the domain recombination events. Furthermore, we could show that most of the signal in favor of an independent origin of sFRP-1/2/5 and sFRP-3/4 in the frizzled-CRD domain came from slow-evolving, nonsaturated sites that are more likely to retain genuine phylogenetic signal (Jeffroy et al. 2006), whereas signal in favor of a single origin of all sFRPs was mostly provided by fast-evolving and mutationally saturated sites. These analyses provide evidence in favor of repeated evolution for sFRPs and highlight the relevance of detailed phylogenetic analyses, in addition to statistical tests, for the identification of independent domain architecture evolution.

Our phylogenetic analyses support a scenario with four frizzled (frizzled-4, frizzled-9/10, frizzled-5/8, and frizzled-1/2/3/6/7) and two sFRPs (sFRP-3/4 and sFRP-1/2/5) genes in the ancestor of cnidarians and bilaterians. sFRP-3/4, but not

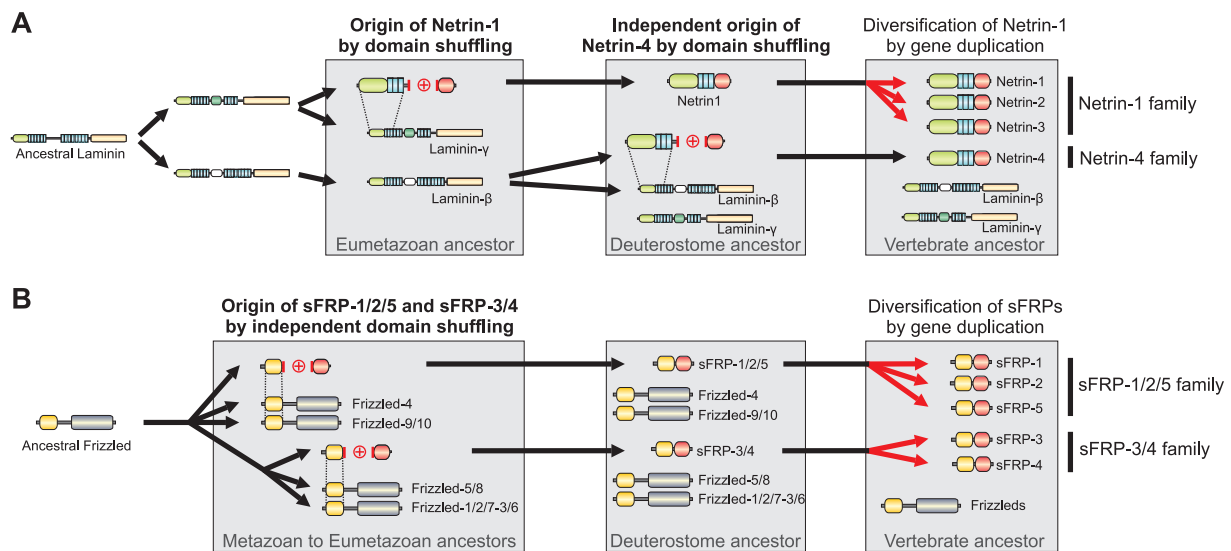


FIG. 6.—Evolutionary scenario for the origin and evolution of netrins and sFRPs. Schematic representation of expansion of (A) netrins and (B) sFRP within one evolutionary lineage by both convergent domain shuffling and gene duplication. Note that diversification of laminin and frizzled proteins in vertebrates and origin and diversification of laminin- α , β/γ -like and netrin-G have been omitted.

sFRP-1/2/5, is also present in the placozoan *T. adhaerens* (with a truncated frizzled-CRD domain not included in the phylogenetic analyses), while both groups of sFRPs are absent from the sequenced ctenophore (*Mnemiopsis leidyi*; Pang et al. 2010) and sponge genomes (Adamska et al. 2010; Srivastava et al. 2010). The *A. queenslandica* proteins annotated as sFRP (ADO16571-16574) do not contain a netrin domain but only a single CRD domain and are thus not genuine sFRPs. One poorly conserved sFRP sequence from the freshwater sponge *Lubomirski baicalensis* has been reported, composed of a highly divergent frizzled-CRD domain and a putative netrin domain (Adell et al. 2007). We were unable to assign this sequence to any group in the netrin domain or frizzled-CRD domain phylogenies (not shown), thus we could not determine its origin. Therefore, it remains possible that sFRPs originated earlier, in the ancestor of metazoans, but sequencing of additional poriferan genomes or transcriptomes is required to answer this question.

It is important to note that the hypothesis of repeated evolution of netrin and sFRP domain architectures does not rely on the phylogeny of the netrin domain. We have shown that this domain is saturated and does not provide a reliable phylogenetic signal contrary to the other domains analyzed, which are clearly in favor of the polyphyly hypothesis (see results). However, even if the netrin domains of netrins and sFRPs were monophyletic, the LamininNT-EGF domain of netrins and the frizzled-CRD domain of sFRP would still have been combined twice independently with the same netrin domain. Consequently, even in this scenario, the identical domain architecture of the different netrin subgroups arose by convergent evolution and not by gene duplication.

Possible Mechanism for the Evolution of the Netrin and sFRP Domain Architectures

Recently, it has been shown that the inclusion of coding exons of neighboring genes is the prevalent mode for the gain of domains in metazoan proteins (Buljan et al. 2010). These events of gene fusion are typically preceded by the duplication of the “donor” domain and its recombination to a position adjacent to the “host” protein (Buljan et al. 2010). Our data suggest that the domain architectures of netrins and sFRP may have evolved by this mechanism. The N-terminal addition of a laminin-derived LamininNT-EGF supra-domain or a frizzled-CRD domain to a single netrin domain is the most parsimonious explanation, because the C-terminal addition of a netrin domain to a LamininNT-EGF supra-domain or a frizzled-CRD domain would require an additional loss of the C-terminal domains of the “hosts” laminins and frizzled. However, as we were unable to establish the exact relationships between the netrin domains of netrins/sFRPs and the netrin-domain-only proteins TIMP (the potential “host” protein), unambiguous support for this scenario is not available. Finally, the presence of conserved introns in both the Netrin and the LamininNT-EGF domains clearly argues against the involvement of retrotransposition as a possible mechanism for the origin of netrins.

Functional Convergences of Netrin and sFRP Proteins Result From the Convergences of Domain Architecture

Domain architecture is thought to be a determining factor for the functional properties of a protein, and thus, multidomain protein with the same domain architecture is expected to have

similar functions (Bashton and Chothia 2007). This is what is indeed observed for many paralogous proteins; however, in paralogs, the shared domain architecture is a consequence of a shared evolutionary history. Thus, the described independent origin of netrins provides an intriguing confirmation of the importance of domain architecture for protein function. In fact, members of both the netrin-1/2/3/5 and netrin-4 subgroups are secreted molecules that bind to DCC/neogenin and Unc5 transmembrane receptors (Koch et al. 2000; Qin et al. 2007; Lejmi et al. 2008; Staquicini et al. 2009) and function in netrin signaling-mediated axon guidance and angiogenesis (Koch et al. 2000; Qin et al. 2007; Rajasekharan and Kennedy 2009). Strikingly, neither Laminin, Netrin-G (lacking the netrin domain and binding to specific netrin-G ligands, see Seiradake et al. 2011) nor TIMP (proteins composed of the netrin domain only) have been shown to bind DCC/neogenin and Unc5 proteins or to function in this signaling pathway (Rajasekharan and Kennedy 2009; Brew and Nagase 2010). Furthermore, we could show that contrary to netrin ligands, the primary netrin receptors, neogenin and Unc5, have both a unique origin in the ancestor of Eumetazoa with diversification through gene duplication in the ancestor of vertebrates (supplementary fig. S13, Supplementary Material online). Because netrin-1/2/3/5 and netrin-4 subgroups originated independently, this shared binding property cannot be explained by a conserved function present in the ancestor of these proteins, but most probably is the consequence of the convergent domain architecture.

In addition to DCC/neogenin and Unc5, netrins from both subgroups have been shown to bind to Integrin $\alpha 3\beta 1$ (Yebra et al. 2003; Stanco et al. 2009; Yebra et al. 2011), where, at least in the case of netrin-1, this interaction is mediated by the netrin domain. However, because the netrin domain-only protein TIMP2 has also been shown to bind to Integrin $\alpha 3\beta 1$ (Seo et al. 2003), this shared function of netrins might not reflect a consequence of their shared domain architecture but rather an ancestral property of netrin domains.

Accepting the polyphyletic origin for sFRPs, they constitute a similar example for functional convergence based on convergence of domain architecture. sFRP proteins have been extensively described as inhibitors of the Wnt signaling pathway, and they bind to secreted Wnt proteins and thereby prevent the interaction of Wnts with frizzled transmembrane receptors (Bovolenta et al. 2008; Mii and Taira 2011). This mechanism appears to have a conserved function in axial patterning in Metazoa (Petersen and Reddien 2009). Both sFRP-1/2/5 and sFRP-3/4 proteins bind to and antagonize signaling molecules of the Wnt family (reviewed in Bovolenta et al. 2008), and recent studies show that both the frizzled-CRD and netrin domains of sFRP-1/2/5 and sFRP-3/4 proteins are necessary for optimal Wnt inhibition (Lin et al. 1997; Bhat et al. 2007; Lopez-Rios et al. 2008).

General Considerations on the Convergence of Domain Architecture

Identical domain architecture of multidomain proteins is frequently considered as evidence for paralogy when occurring in one genome, and for orthology when occurring in the genomes of different taxa (e.g., for sFRPs, Adamska et al. 2010). These simplistic assignments may confound the inference of the evolutionary origin of multidomain proteins and their associated cellular functions. To the best of our knowledge, current terminology does not cover the independent evolution of identical domain architecture (Koonin 2005). As the different parts (domains) of these proteins have different evolutionary histories, we propose the concept of “merology” (derived from the Greek word “méros” meaning part and “logos” meaning relation) to describe the repeated evolution of similar domain architecture and “merologous proteins” to refer to nonhomologous proteins that display the same domain organization.

A study using phylogenetic trees of domains from 96 genomes of Bacteria, Archaea, and Eukaryota has suggested that convergent evolution of domain architecture may occur more frequently than previously suspected (Forslund et al. 2008). Depending on the criteria used for the generation of protein datasets, between 5.6% and 12.4% of domain architectures were identified as candidates for convergent evolution. The cases of netrins and sFRPs described in detail here belong to a particular subset of these events for two reasons. First, only one-third of the documented cases included the independent gain of domains, as is the case for netrins and sFRPs. Second, the repeated evolution of netrins and sFRPs occurred within the same genomic background, that is, the netrin-4 group evolved in a genome in which the netrin-1/2/3/5 group was already present. This is contrary to most cases described by Forslund et al. (2008), in which the same domain architecture evolved in different taxa.

The phylogenetic distribution of merologous proteins identified by Forslund et al. (2008) suggests that many of them originated relatively recently. The cases in which merologs evolved recently could help to understand the genomic mechanisms that promote this type of convergence, for example, whether particular features of “host” and “donor” genes predispose them to recombine with each other. In addition, studies of merologous proteins, in particular those displaying functional convergence, could add a new perspective to the understanding of the relationship between domain architecture and protein function. Currently, research on proteins with shared domain architecture focuses on duplicated paralogs undergoing structural and functional divergence. In the case of merologs, the situation is reversed: proteins originate from ancestral sequences with different domain architecture and probably different functions and converge to similar structures and potentially similar functions. Thus, merologs are particularly interesting cases that may

help to explain why only a fraction of all possible domain combinations exists and why some domains are more frequently found in multidomain proteins than others (Basu et al. 2008).

Supplementary Material

Supplementary tables S1 and S2, supplementary figs. S1–S13, and supplementary material including detailed phylogenies and analyzed data sets of the netrin, frizzled-CRD, and LamininNT-EGF domains are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org>).

Acknowledgments

We thank Michael Manuel, Gemma Richards, and Chiara Sinigaglia for critical reading and Rein Aasland and Daniel Chourrout for comments on an earlier version of the manuscript. Funding of this work was provided by the Sars Centre core budget.

Literature Cited

- Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105.
- Adamska M, et al. 2010. Structure and expression of conserved Wnt pathway components in the demosponge *Amphimedon queenslandica*. *Evol Dev*. 12:494–518.
- Adell T, Thakur AN, Muller WE. 2007. Isolation and characterization of Wnt pathway-related genes from Porifera. *Cell Biol Int*. 31:939–949.
- Banyai L, Patthy L. 1999. The NTR module: domains of netrins, secreted frizzled related proteins, and type I procollagen C-proteinase enhancer protein are homologous with tissue inhibitors of metalloproteinases. *Protein Sci*. 8:1636–1642.
- Bashton M, Chothia C. 2007. The generation of new protein functions by the combination of domains. *Structure* 15:85–99.
- Basu MK, Carmel L, Rogozin IB, Koonin. 2008. Evolution of protein domain promiscuity in eukaryotes. *Genome Res*. 18:449–461.
- Bergsten J. 2005. A review of long-branch attraction. *Cladistics* 21:163–193.
- Bhat RA, Stauffer B, Komm BS, Bodine PV. 2007. Structure-function analysis of secreted frizzled-related protein-1 for its Wnt antagonist function. *J Cell Biochem*. 102:1519–1528.
- Bovolenta P, Esteve P, Ruiz JM, Cisneros E, Lopez-Rios J. 2008. Beyond Wnt inhibition: new functions of secreted frizzled-related proteins in development and disease. *J Cell Sci*. 121:737–746.
- Bramham J, et al. 2005. Functional insights from the structure of the multifunctional C345C domain of C5 of complement. *J Biol Chem*. 280:10636–10645.
- Brew K, Nagase H. 2010. The tissue inhibitors of metalloproteinases (TIMPs): an ancient family with structural and functional diversity. *Biochim Biophys Acta*. 1803:55–71.
- Brinkmann H, Philippe H. 1999. Archaea sister group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. *Mol Biol Evol*. 16:817–825.
- Brinkmann H, van der Giezen M, Zhou Y, Poncelin de Raucourt G, Philippe H. 2005. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. *Syst Biol*. 54:743–757.
- Buljan M, Frankish A, Bateman A. 2010. Quantifying the mechanisms of domain gain in animal proteins. *Genome Biol*. 11:R74.
- Chong JM, Uren A, Rubin JS, Speicher DW. 2002. Disulfide bond assignments of secreted frizzled-related protein-1 provide insights about frizzled homology and netrin modules. *J Biol Chem*. 277:5134–5144.
- Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet*. 6:361–375.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32:1792–1797.
- Fahey B, Degnan BM. 2012. Origin and evolution of laminin gene family diversity. *Mol Biol Evol*. 29:1823–1836.
- Felsenstein J. 1985. Confidence-limits on phylogenies—an approach using the bootstrap. *Evolution* 39:783–791.
- Forslund K, Henricson A, Hollich V, Sonnhammer EL. 2008. Domain tree-based analysis of protein architecture evolution. *Mol Biol Evol*. 25:254–264.
- Foster PG, Hickey DA. 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *J Mol Evol*. 48:284–290.
- Gatesy J, DeSalle R, Wahlberg N. 2007. How many genes should a systematist sample? Conflicting insights from a phylogenomic matrix characterized by replicated incongruence. *Syst Biol*. 56:355–363.
- Goldman N, Anderson JP, Rodrigo AG. 2000. Likelihood-based tests of topologies in phylogenetics. *Syst Biol*. 49:652–670.
- Gough J. 2005. Convergent evolution of domain architectures (is rare). *Bioinformatics* 21:1464–1471.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 52:696–704.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp*. 41:95–98.
- Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence? *Trends Genet*. 22:225–231.
- Jordan IK, Henze K, Fedorova ND, Koonin EV, Galperin MY. 2003. Phylogenomic analysis of the *Giardia* intestinalis transcarboxylase reveals multiple instances of domain fusion and fission in the evolution of biotin-dependent enzymes. *J Mol Microbiol Biotechnol*. 5:172–189.
- Kasahara M. 2007. The 2R hypothesis: an update. *Curr Opin Immunol*. 19:547–552.
- Koch M, et al. 2000. A novel member of the netrin family, beta-netrin, shares homology with the beta chain of laminin: identification, expression, and functional characterization. *J Cell Biol*. 151:221–234.
- Koonin EV. 2005. Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet*. 39:309–338.
- Koonin EV, Wolf YI, Karev GP. 2002. The structure of the protein universe and genome evolution. *Nature* 420:218–223.
- Kummerfeld SK, Teichmann SA. 2005. Relative rates of gene fusion and fission in multi-domain proteins. *Trends Genet*. 21:25–30.
- Lai Wing Sun K, Correia JP, Kennedy TE. 2011. Netrins: versatile extracellular cues with diverse functions. *Development* 138:2153–2169.
- Lejmi E, et al. 2008. Netrin-4 inhibits angiogenesis via binding to neogenin and recruitment of Unc5B. *Proc Natl Acad Sci U S A*. 105:12491–12496.
- Li Y, Liu Z, Shi P, Zhang J. 2010. The hearing gene Prestin unites echolocating bats and whales. *Curr Biol*. 20:R55–R56.
- Liepinsh E, Banyai L, Pintacuda G, Trexler M, Patthy L, Otting G. 2003. NMR structure of the netrin-like domain (NTR) of human type I procollagen C-proteinase enhancer defines structural consensus of NTR domains and assesses potential proteinase inhibitory activity and ligand binding. *J Biol Chem*. 278:25982–25989.
- Lin K, et al. 1997. The cysteine-rich frizzled domain of Frzb-1 is required and sufficient for modulation of Wnt signaling. *Proc Natl Acad Sci U S A*. 94:11196–11200.
- Lopez-Rios J, Esteve P, Ruiz JM, Bovolenta P. 2008. The netrin-related domain of Sfrp1 interacts with Wnt ligands and antagonizes their activity in the anterior neural plate. *Neural Dev*. 3:19.

- Lopez P, Casane D, Philippe H. 2002. Heterotachy, an important process of protein evolution. *Mol Biol Evol.* 19:1–7.
- Mii Y, Taira M. 2011. Secreted Wnt "inhibitors" are not just inhibitors: regulation of extracellular Wnt by secreted frizzled-related proteins. *Dev Growth Differ.* 53:911–923.
- Moore AD, Bjorklund AK, Ekman D, Bornberg-Bauer E, Elofsson A. 2008. Arrangements in the modular evolution of proteins. *Trends Biochem Sci.* 33:444–451.
- Moore SW, Tessier-Lavigne M, Kennedy TE. 2007. Netrins and their receptors. *Adv Exp Med Biol.* 621:17–31.
- Nakashiba T, et al. 2000. Netrin-G1: a novel glycosyl phosphatidylinositol-linked mammalian netrin that is functionally divergent from classical netrins. *J Neurosci.* 20:6540–6550.
- Page RDM. 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci.* 12:357–358.
- Pang K, Ryan JF, Mullikin JC, Baxeveanis AD, Martindale MQ. 2010. Genomic insights into Wnt signaling in an early diverging metazoan, the ctenophore *Mnemiopsis leidyi*. *Evodevo* 1:10.
- Petersen CP, Reddien PW. 2009. Wnt signaling and the polarity of the primary body axis. *Cell* 139:1056–1068.
- Philippe H, Zhou Y, Brinkmann H, Rodrigue N, Delsuc F. 2005. Heterotachy and long-branch attraction in phylogenetics. *BMC Evol Biol.* 5:50.
- Qin S, Yu L, Gao Y, Zhou R, Zhang C. 2007. Characterization of the receptors for axon guidance factor netrin-4 and identification of the binding domains. *Mol Cell Neurosci.* 34:243–250.
- Rajasekharan S, Kennedy TE. 2009. The netrin protein family. *Genome Biol.* 10:239.
- Rambaut A, Grassly NC. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput Appl Biosci.* 13:235–238.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
- Salbaum JM, Kappen C. 2000. Cloning and expression of nope, a new mouse gene of the immunoglobulin superfamily related to guidance receptors. *Genomics* 64:15–23.
- Satoh W, Gotoh T, Tsunematsu Y, Aizawa S, Shimono A. 2006. Sfrp1 and Sfrp2 regulate anteroposterior axis elongation and somite segmentation during mouse embryogenesis. *Development* 133:989–999.
- Seiradake E, et al. 2011. Structural basis for cell surface patterning through NetrinG-NGL interactions. *EMBO J.* 30:4479–4488.
- Seo DW, et al. 2003. TIMP-2 mediated inhibition of angiogenesis: an MMP-independent mechanism. *Cell* 114:171–180.
- Serafini T, et al. 1996. Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system. *Cell* 87:1001–1014.
- Sheffield NC, Song H, Cameron SL, Whiting MF. 2009. Nonstationary evolution and compositional heterogeneity in beetle mitochondrial phylogenomics. *Syst Biol.* 58:381–394.
- Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst Biol.* 51:492–508.
- Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17:1246–1247.
- Srivastava M, et al. 2010. The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* 466:720–726.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Stanco A, et al. 2009. Netrin-1-alpha3beta1 integrin interactions regulate the migration of interneurons through the cortical marginal zone. *Proc Natl Acad Sci U S A.* 106:7595–7600.
- Staquicini FI, et al. 2009. Discovery of a functional protein complex of netrin-4, laminin gamma1 chain, and integrin alpha6beta1 in mouse neural stem cells. *Proc Natl Acad Sci U S A.* 106:2903–2908.
- Swofford DL. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 4.0b10 (AIVitec). Sunderland (MA): Sinauer Associates.
- Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.
- Toyoda R, Nakamura H, Watanabe Y. 2005. Identification of protogenin, a novel immunoglobulin superfamily gene expressed during early chick embryogenesis. *Gene Expr Patterns.* 5:778–785.
- Vogel C, Berzuini C, Bashton M, Gough J, Teichmann SA. 2004. Supra-domains: evolutionary units larger than single protein domains. *J Mol Biol.* 336:809–823.
- Vogel C, Teichmann SA, Pereira-Leal J. 2005. The relationship between domain duplication and recombination. *J Mol Biol.* 346:355–365.
- Wang R, et al. 2009. Autoinhibition of UNC5b revealed by the cytoplasmic domain structure of the receptor. *Mol Cell.* 33:692–703.
- Weiner J III, Beaussart F, Bornberg-Bauer E. 2006. Domain deletions and substitutions in the modular protein evolution. *FEBS J.* 273:2037–2047.
- Wu B, Huan T, Gong J, Zhou P, Bai Z. 2011. Domain combination of the vertebrate-like TLR gene family: implications for their origin and evolution. *J Genet.* 90:401–408.
- Yanai I, Wolf YI, Koonin EV. 2002. Evolution of gene fusions: horizontal transfer versus independent events. *Genome Biol.* 3:research0024.
- Yebara M, et al. 2003. Recognition of the neural chemoattractant Netrin-1 by integrins alpha6beta4 and alpha3beta1 regulates epithelial cell adhesion and migration. *Dev Cell.* 5:695–707.
- Yebara M, et al. 2011. Endothelium-derived Netrin-4 supports pancreatic epithelial cell adhesion and differentiation through integrins alpha2-beta1 and alpha3beta1. *PLoS One.* 6:e22750.
- Yin Y, Miner JH, Sanes JR. 2002. Laminins: laminin- and netrin-related genes expressed in distinct neuronal subsets. *Mol Cell Neurosci.* 19:344–358.
- Zhang Q, Zmasek CM, Godzik A. 2010. Domain architecture evolution of pattern-recognition receptors. *Immunogenetics* 62:263–272.

Associate editor: Eugene Koonin