

Research article

Open Access

Chitinase family GH18: evolutionary insights from the genomic history of a diverse protein family

Jane D Funkhouser* and Nathan N Aronson Jr

Address: Department of Biochemistry and Molecular Biology, College of Medicine, University of South Alabama, Mobile, Alabama 36688, USA

Email: Jane D Funkhouser* - jfunkhouser@usouthal.edu; Nathan N Aronson - naronson@jaguar1.usouthal.edu

* Corresponding author

Published: 26 June 2007

Received: 6 November 2006

BMC Evolutionary Biology 2007, **7**:96 doi:10.1186/1471-2148-7-96

Accepted: 26 June 2007

This article is available from: <http://www.biomedcentral.com/1471-2148/7/96>

© 2007 Funkhouser and Aronson; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Chitinases (EC.3.2.1.14) hydrolyze the β -1,4-linkages in chitin, an abundant N-acetyl- β -D-glucosamine polysaccharide that is a structural component of protective biological matrices such as insect exoskeletons and fungal cell walls. The glycoside hydrolase 18 (GH18) family of chitinases is an ancient gene family widely expressed in archaea, prokaryotes and eukaryotes. Mammals are not known to synthesize chitin or metabolize it as a nutrient, yet the human genome encodes eight GH18 family members. Some GH18 proteins lack an essential catalytic glutamic acid and are likely to act as lectins rather than as enzymes. This study used comparative genomic analysis to address the evolutionary history of the GH18 multiprotein family, from early eukaryotes to mammals, in an effort to understand the forces that shaped the human genome content of chitinase related proteins.

Results: Gene duplication and loss according to a birth-and-death model of evolution is a feature of the evolutionary history of the GH18 family. The current human family likely originated from ancient genes present at the time of the bilaterian expansion (approx. 550 mya). The family expanded in the chitinous protostomes *C. elegans* and *D. melanogaster*, declined in early deuterostomes as chitin synthesis disappeared, and expanded again in late deuterostomes with a significant increase in gene number after the avian/mammalian split.

Conclusion: This comprehensive genomic study of animal GH18 proteins reveals three major phylogenetic groups in the family: chitobiasis, chitinases/chitolectins, and stabilin-I interacting chitolectins. Only the chitinase/chitolectin group is associated with expansion in late deuterostomes. Finding that the human GH18 gene family is closely linked to the human major histocompatibility complex paralogon on chromosome 1, together with the recent association of GH18 chitinase activity with Th2 cell inflammation, suggests that its late expansion could be related to an emerging interface of innate and adaptive immunity during early vertebrate history.

Background

The recent increase in genome projects has provided DNA sequence data useful for understanding the evolutionary dynamics that resulted in the conservation of ancient proteins as modern protein families. One such family of pro-

teins is the glycoside hydrolase family 18 chitinases (GH18), widely expressed in archaea, prokaryotes and eukaryotes. Definition of this family is based on amino acid sequence similarity [1-3]. In eukaryotes the proteins are mainly expressed by fungi, arthropods and nematodes

(for review see [4]), but they are also expressed in mammals, with members recently shown to be involved in important physiological processes that include T-cell mediated inflammation and allergy [5,6]. The chitinases hydrolyze chitin, an abundantly produced homopolymer of N-acetyl- β -D-glucosamine that provides architectural reinforcement of biological structures such as insect exoskeletons, fungal cell walls, nematode egg shells, and other biological matrices involved in protection and self defense. Chitin has not been detected in higher plants and vertebrates, where cellulose and hyaluronan, respectively, may replace certain chitin functions [7,8]. In chitinous species, the chitinases along with chitin synthetases are essential for remodeling chitin-containing structures during growth and development. Certain bacterial genera and plants, which do not produce chitin, use chitinases, respectively, for the conversion of insoluble chitin into metabolizable nutrient sources and for defense from chitinous pathogens.

Mammals are not known to synthesize chitin or to metabolize it as a nutrient, yet the human genome encodes eight well-documented genes for proteins now classified as glycoside hydrolase family 18 members. Members of this family are known to adopt the TIM (triosephosphate isomerase) fold consisting of a strongly conserved $(\beta/\alpha)_8$ -barrel structure [1-3]. Often, separate chitin-binding domains (CBM14) [2,3] are present in the carboxyl terminal region of the proteins (additional file 1: GH18 family domain structure). The protein family includes chitinases as well as homologous proteins termed chitolectins. The latter lack the key active-site glutamate residue that donates a proton required for hydrolytic enzyme activity, but retain highly conserved residues involved in oligosaccharide binding and overall three-dimensional structure. Traditionally, chitinases are classified in two glycoside hydrolase families, GH18 and GH19, with different structures and catalytic mechanisms [9]. Family GH18 includes the chitinases from viruses, bacteria, fungi and animals as well as classes III and V from plants [10]. The GH19 chitinases are identified mostly in plants (classes I, II and IV), nematodes, and some bacteria [11]. Recent data indicate chitinase activity is also present in protein families GH48 and GH20 [12,13]. N-acetyl- β -D-glucosaminidases such as those in family GH20 also can participate in chitin degradation by hydrolyzing GlcNAc from the non-reducing end of chito-oligosaccharides [14].

Three of the GH18 proteins encoded by the human genome have demonstrated enzymatic activity. They are: di-N-acetylchitobiase, active in the lysosomal degradation of asparagine-linked glycoproteins [15]; chitotriosidase, a vesicular and secreted protein produced by activated macrophages and highly elevated in serum from patients with Gaucher disease [16]; and acidic mammalian chitinase

(AMCase), expressed mostly in the gastrointestinal tract and lung [17], and recently shown to be induced by Th2-helper cells in an aeroallergen model of asthma [5]. A gene similar to the AMCase gene encodes a fourth hypothetical protein with a truncated GH18 domain. Three human genes (OVGP1, CHI3L1 and CHI3L2) encode chitolectins likely to be involved in tissue remodeling during inflammation and/or development. The OVGP1 gene encodes a large glycoprotein secreted by oviduct epithelial cells in response to estrogen [18,19]. The glycoprotein includes an extended C-terminal repeat region similar to sequences found in mucins. A number of cell types, including macrophages, articular chondrocytes and synovial cells, secrete proteins encoded by the CHI3L1 and CHI3L2 genes under inflammatory conditions [20-22]. A recently identified CHID1 gene encodes a lysosomal GH18 chitolectin that interacts with stabilin-1, an endocytic/sorting receptor of macrophages that serves as a marker of alternative activation [23]. The function of the CHID1 protein is not yet established, and the sequence similarity to other GH18 proteins is low.

Although the biological functions of most human GH18 proteins are not yet delineated, the available protein characterizations indicate that a process of gene duplication and diversification has resulted in multiple functions that are not related to nutrient utilization or growth-related turnover of chitinous structures. Instead, recent data pertaining to the mammalian proteins point to their prominent roles in defense against fungal or other pathogens and in inflammation and remodeling processes [6]. In this communication, we addressed the evolutionary history of the GH18 multiprotein family from early eukaryotes to mammals in an effort to understand the forces that shaped the human genome content of chitinase-related proteins. The present study indicates that over evolutionary time the GH18 family evolved by decline and expansion according to selective forces associated with speciation. An expansion of chitinase genes occurred as chitinous species appeared early in metazoan evolution, a decline followed as chitin disappeared as an important structural and protective feature of early deuterostomes, and a second expansion began in non-chitinous vertebrates where chitinases and chitolectins may have evolved to assume increasingly important roles associated with pathogen recognition, processing and antigen presentation.

Results

Human GH18 Family Genes and Their Relationship to Other Mammalian GH18 Genes

Seven of the eight human GH18 family members are located on chromosome 1. The chitobiase gene is located at 1p22, 26 Mb upstream of the regions containing six other chitinase/chitolectin genes, which cluster into two

groups, at 1p13 and at 1q32 (Table 1): Genes CHI3L2, CHIA, RP11-165H20.1 and OVGPI cluster at 1p13; and CHIT1 and CHI3L1 at 1q32. Only the recently annotated and weakly homologous CHID1 gene is not located on chromosome 1, but at chromosome 11p15.5. A similar but unique pattern of homologues is present in the mouse genome, which includes ten genes encoding GH18 family members. The genes are located on mouse chromosomes 1, 3 and 7. Genes Ctbs, Chia, Chi3l3, Chi3l4, Ovgp1, Bclp2 and BC051070 are located on chromosome 3, and Chit1 and Chi3l1 are on chromosome 1. The mouse CHID1 homolog, designated 3110023E09Rik [Entrez Gene ID 68038], is located at chromosome 7F5. The rat genome is not complete, but homologous rat genes are located on chromosomes 2, 13 and 18. There is conservation of synteny in the chromosomes of these three mammals that contain GH18 family members.

From the genome sequences currently available, it is evident based on protein sequence similarities that the mammalian genomes include orthologous clades of GH18 family members, and that the chitinase/chitolectin clade (Figure 1) includes orthologous groups of paralogous family members. Construction of a phylogenetic tree from alignment of GH18 domains from 36 mammalian proteins (additional file 2: Mammalian Sequences) illustrates these relationships. Figure 1 shows the tree, constructed using the minimum evolution method. The lysosomal chitobiases and stabilin-1 interacting chitolectins form outlying clades of orthologous proteins distinct from the true chitinases and chitolectins, which form four separate subgroups. Subgroup I represents the acidic chitinases and related chitolectins; subgroup II, the chitotriosidase enzymes; subgroup III, the non-enzymatic proteins associated with injury, repair and remodeling; and subgroup IV, the oviduct glycoproteins and related sequences. Each of the groups is not represented in all the genomes. This may be explained by genome projects that are not complete; or gene duplication or gene loss may have occurred

relatively late in vertebrate evolution. Among GH18 enzymes, chitobiase is biochemically unique as the only member that splits off a monosaccharide from the reducing end of chito-oligosaccharides [15].

Chitinase/Chitolectin Genes in Early Eukaryotes (*D. discoideum* and *H. echinata*)

To assess the family GH18 genes in early eukaryotes and their relationship to the human genes, we examined the genome of the social amoeba *Dictyostelium discoideum*. *D. discoideum* diverged from the animal-fungal lineage after the plant-animal split, and has retained a significant amount of the diversity of the ancestral eukaryotic genome [24]. Thus, chitinases present in the *D. discoideum* genome should provide evolutionary insight into the ancestral metazoan GH18 proteins present before an apparent expansion of chitinases associated with the biological importance of chitin in protostomes (discussed in the next section). A search of the *D. discoideum* genome database at NCBI (Build 1.1; November 22, 2005) revealed six genes encoding chitinase-related proteins. Five of the genes cluster on two chromosomes: a 2.6 Mbp region of chromosome 2 encodes Entrez genes 3393954, 3395463, and 3394351; and a 3.5 Kb region of chromosome 5 encodes Entrez genes 3388787 and 3388788. A sixth gene (Entrez Gene 3387295) also located on chromosome 5 encodes a hypothetical protein [GenBank:XP_635730] that aligned with the human stabilin-1 interacting protein (Expect = 2e-48). Phylogenetic analysis (Figure 2) indicated that the three chromosome 2 genes are co-orthologs of the human chitobiase. Two chromosome 5 genes are more distant homologs, distinct from the human chitinases/chitolectins. The third chromosome 5 gene is a human CHID1 ortholog. If the eukaryotic ancestral genome included orthologs of the mammalian chitinase/chitolectin group, and the presence of chitinases in plants [11] indicates it did, they likely have been lost in the *D. discoideum* lineage. As the cells of *Dictyostelium* are professional phagocytes that consume

Table 1: Human Genes Encoding GH18 Family Members

Entrez Gene ID	Gene (Aliases)	Chr ^a Location	Protein Accession	Description
27159	CHIA	1p13.2	NP_970615 & NP_068569	acidic chitinase (AMCase, eosinophil chemotactic cytokine)
1118	CHIT1	1q32.1	NP_003456	chitinase I (chitotriosidase)
1116	CHI3L1 (YKL 40, GP39, TSA1902)	1q32.1	NP_001267	chitinase 3-like 1 (cartilage protein 39)
1117	CHI3L2 (YKL 39)	1p13.3	NP_003991	chitinase 3-like 2 (chondrocyte protein 39)
RP11-165H20.1	(LOC149620)	1p13.2	NP_001013643	NCBI Ref Seq: NP_001013643 was permanently suppressed because it is a nonsense mediated mRNA decay (NMC) candidate
5016	OVGPI	1p13.2	NP_002548	oviductal glycoprotein I; oviductin
1486	CTBS	1p22	NP_004379	chitobiase, di-N-acetyl
66005	CHID1	11p15.5	NP_076436	stabilin-1 interacting chitinase-like protein

^a Chromosome

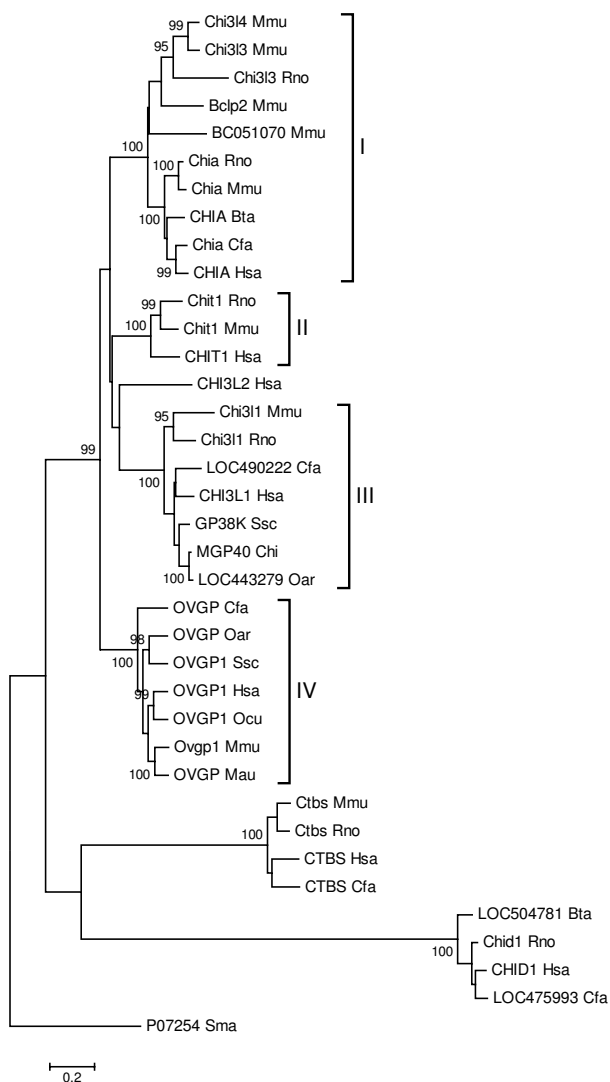


Figure 1
Phylogenetic tree of GH18 domains from 36 mammalian proteins. The phylogenetic trees presented in this paper were constructed with Mega version 3.1 software [86]. Minimum evolution, neighbor-joining, and maximum parsimony methods were used with Poisson correction for multiple amino acid substitution and with 1000 random bootstrap replicates. The figure shows the minimum evolution tree (the three methods produced very similar topologies). Alignment of the GH18 domain amino acid sequences used ClustalX (1.83) [84]. Gene symbols (or protein symbols where genes symbols are not available) and binomen species abbreviations of the form "Gsp" (Genus, species) are used. The GH18 proteins comprise three major phylogenetic groups, with the chitinase/chitolectin group forming four subgroups denoted I, II, III and IV (discussed in the text). The scale at the bottom left is in units of amino acid substitutions per site. Bootstrap values $\geq 95\%$ are shown. The tree is rooted with the *Serratia marcescens* family GH18 chitinase A [GenBank:P07254].

bacteria as a primary nutritional source [25], the chitobiases and the related proteins are likely involved in the uptake, internalization, killing and subsequent digestion of bacterial cell wall components. Indeed, chitobiase will hydrolyze GlcNAc β -D-(1-4)MurNAc, the repeating disaccharide unit of the bacterial cell wall peptidoglycan [26]. However, the proteins also may contribute to the control of microbial growth within the digestive phagosomes [27].

Further information about the early evolution of chitin-degrading proteins comes from the recent cloning of a cDNA encoding a chitinase from a chitinous basal metazoan, the cnidarian *Hydractinia echinata* [28]. The 1467 bp cDNA encodes a typical GH18 glycoside hydrolase of 464 amino acids [GenBank:CAG25409]. The protein sequence includes a secretory signal at the N-terminus and a highly conserved catalytic motif including its hallmark proton-donating glutamic acid. The highest scoring human sequence aligned by NCBI BLAST with the *H. echinata* protein sequence (Expect = $3e-90$) is chitotriosidase [GenBank:NP_003456] (44% identity and 59% similarity over 414 residues, including the GH18 domain). However, all the human chitinases/chitolectins produced high scoring alignments (Expect < $2e-57$). In contrast to the proteins from *D. discoideum*, the phylogenetic analysis (Figure 2) indicated orthology of the CAG25409 protein with the human chitinases/chitolectins.

Although *H. echinata* feeds on small crustaceans, other unidentified chitinase(s) may participate in their diges-

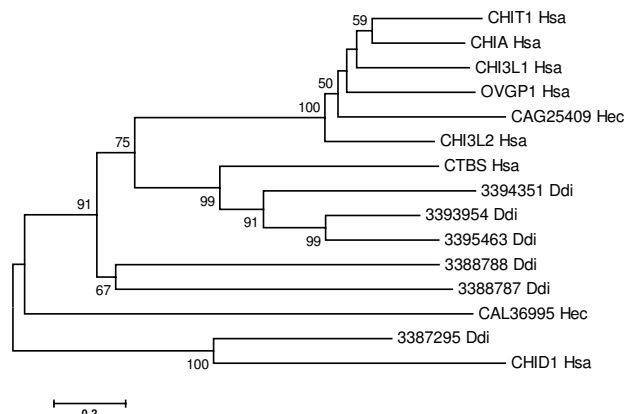


Figure 2
Phylogenetic relationship of *D. discoideum*, *H. echinata*, and human GH18 proteins. The figure shows a minimum evolution tree from a phylogenetic analysis conducted as described for Figure 1. Gene (or protein symbols when gene symbols are not available) and binomen species abbreviations are used. The tree is rooted at midpoint. Bootstrap values $\geq 50\%$ are shown.

tion [28]. (A second more recently identified *H. echinata* protein [NCBI accession CAL36995] with annotated glycosyl-hydrolase and chitin-binding domains is not closely related to the human or *Dictyostelium* proteins, Figure 2; but NCBI BLAST identifies high scoring alignments with fungal chitinases.) The cDNA expression patterns determined by in situ hybridization indicated that the CAG25409 protein has a role in body pattern formation and immunity. *H. echinata* proteins involved in immunity are of particular interest because *H. echinata* has well developed allo-recognition responses that may relate to the early evolution of vertebrate and nonvertebrate immune systems [29]. Thus, functions of this early metazoan chitinase suggest an ancestral pattern of chitinase/chitolectin involvement in differentiation, development and immunity that continued from the earliest metazoans to present mammals. The growth and developmental function is evident in imaginal disc growth factors and chitinase-like proteins from the protostomes *D. melanogaster* and *C. gigas* (see next section) and in human chitolectins including OVGPI, CHI3L1, and CHI3L2, which are secreted proteins reported to be involved in recognition processes associated with fertilization, tissue inflammation, and wound repair [19,30-33].

Evolutionary Relationship between Chitinase-Related Genes from Human and Protostomes (*D. melanogaster* and *C. elegans*)

As expected, comparison of family GH18 genes from genera of animals from which genome sequences are complete indicated that the highly chitin-metabolizing protostomes *C. elegans* and *D. melanogaster* have more chitinase-related genes. A search of the *C. elegans* database at the NCBI and other resources such as Wormbase [34] and ACEView [35] identified 37 genes or predicted genes encoding chitinases and related GH18 family members (additional file 3: *C. elegans* Sequences). Searches of *D. melanogaster* genome resources at NCBI and FlyBase [36] identified 17 recognized or predicted genes encoding GH18 family proteins (additional file 4: *D. melanogaster* Sequences). Many of these protostome genes were identified in earlier publications [37,38].

A phylogenetic tree created from multiple alignment of GH18 domain sequences from *C. elegans*, *D. melanogaster* and human proteins indicated a general lack of orthology among the members of the families of GH18 proteins (Figure 3). The human chitinases/chitolectins grouped with high significance with only a single protein from each protostome. In the case of *C. elegans*, all the human proteins grouped with the worm chitinase, cht-1 [GenBank:NP_508588], encoded on the X chromosome. The gene encodes a predicted 66.9 kD protein with one GH18 domain and two N-terminal CBM14 (chitin binding) domains. This domain structure differs from the verte-

brate enzymes, where a single CBM14 domain is present in each of the enzymatically active chitinases (CHIT1 and AMCase). The *C. elegans* cht-1 protein aligned with the human CHIT1 protein with 40% identity and 56% similarity over 477 residues, including the GH18 domain and a CBM14 domain. For *D. melanogaster*, domains from a single protein [GenBank: NP_647768] located on chromosome 3L at 3.07 Mb, clustered with the human GH18 domains. The NP_647768 protein comprises two GH18 domains but does not include a CBM14 domain. The two domains aligned with the human CHIT1 protein over 367–375 residues with 45–52% identity and 65% similarity. The results from the phylogenetic analysis of human and protostomes sequences (Figure 3) indicate a single clade encompasses these two proteostome proteins and all the human GH18 family proteins except the two outlying groups (Figure 1). The separation of this clade (bootstrap value 89%) suggests that the human GH18 chitinases/chitolectins are likely co-orthologs of the cht-1 protein from *C. elegans* and the NP_647768 protein from *D. melanogaster*. Interestingly, an analysis that included the *H. echinata* chitinase placed it in the same clade (bootstrap value 85%) (additional file 6: Supplemental phylogenetic tree). From these data, we infer that the present day *H. echinata* CAG25409 protein, the *C. elegans* NP_508588 protein, and the *D. melanogaster* NP_647768 protein are descended from an ancient common ancestor; and the human chitinases/chitolectins appear to be lineal descendents of this ancestral protein.

Further, the phylogenetic analysis of the protostome sequences indicated that *D. melanogaster* and *C. elegans* genomes each included a one-to-one ortholog for the human stabilin-1 interacting chitolectin. The GH18 domains from the *C. elegans* protein [GenBank:NP_508185], the *D. melanogaster* protein [GenBank:NP_609190], and the human CHID1 gene product formed a monophyletic group supported by a bootstrap value of 100% (Figure 3). In contrast, sequence alignments and phylogenetic analyses did not identify unequivocal protostome orthologs for the human chitobiase (CTBS).

So far, the *C. elegans* GH18 family members are only predicted from sequence relationships, and their physiological functions are not yet documented. Many of the *C. elegans* proteins include only the GH18 domain, but some sequence annotations include conserved domains known to bind chitin, oligosaccharides and peptidoglycans. The functions of some of the *D. melanogaster* proteins, however, are well established. They include enzymatic functions involved in chitin turnover associated with molting [4] as well as a more recently characterized cell proliferation role that promotes growth in imaginal discs [39,40]. These growth-promoting factors (IDGFs) comprise a fam-

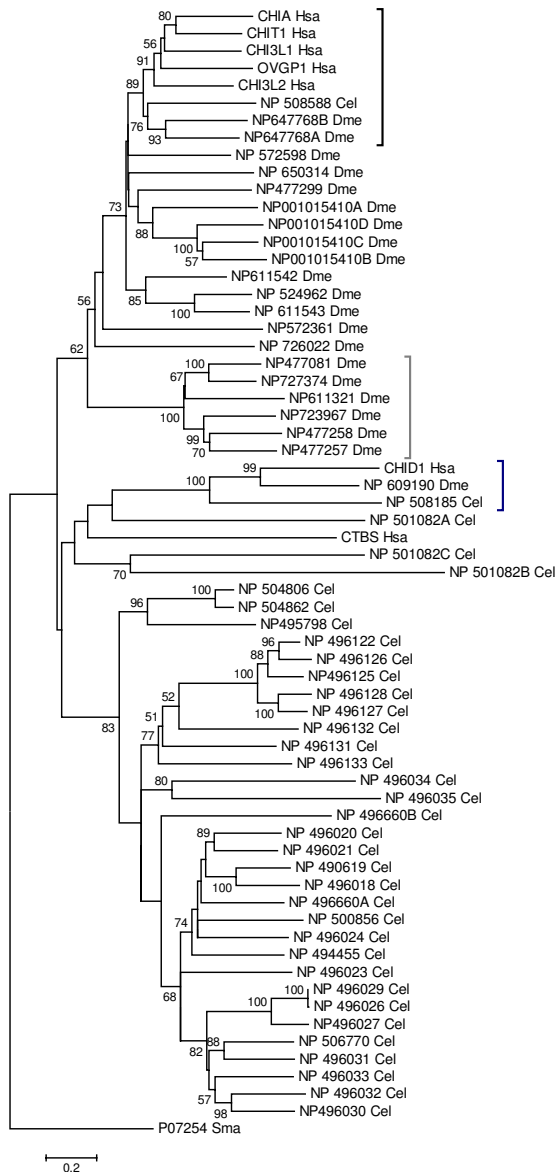


Figure 3
Phylogenetic analysis of GH18 domains of chitinase/chitolectins from human, *D. melanogaster* and *C. elegans*. The figure shows a minimum evolution tree from a phylogenetic analysis conducted as described in the legend for Figure 1. Binomen abbreviations are used for species, gene symbols for human proteins, and accession numbers for *D. melanogaster* and *C. elegans* proteins. A black bracket at the right indicates the clade encompassing all the human chitinases/chitolectins, two domains from a single protein of *D. melanogaster*, and a single protein of *C. elegans*; a dark gray bracket indicates the clade encompassing the *D. melanogaster* imaginal disc growth factors (chitolectins); and a dark blue bracket the stabilin-I interacting chitolectin group. Scale, bootstrapping and rooting are as described for Figure 1. Bootstrap values $\geq 50\%$ are shown.

ily of five soluble proteins. They differ from the chitinases in that they lack the C-terminal chitin-binding domain, and glutamine replaces the critical active-site glutamate residue involved in the catalytic mechanism of all known GH18 chitinase enzymes. The GenBank accession numbers for the IDGFs are [NP 477258](#), [NP 477257](#), [NP 723967](#), [NP 727374](#) and [NP 611321](#). They form a monophyletic group supported by a bootstrap value of 100% (Figure 3). One protein [GenBank:[NP 477081](#)] included in the group is not annotated as IDGF; however, BLASTP identifies its alignments with IDGFs as the best hits in the *D. melanogaster* genome. Interestingly, although some mammalian chitolectins of subgroups I, III and IV (Figure 1) are associated with growth, differentiation and remodeling, they did not group with the imaginal disc growth factors in the phylogenetic analysis (Figure 3). In contrast to the polar glutamine in *D. melanogaster* imaginal disc factors, the substitutions for the catalytic glutamate residue in the mammalian subgroups III and IV proteins OVG1, CHI3L1 and CHI3L2 are the non-polar residues leucine or isoleucine. The mouse Chi3l3 (Ym1) and Chi3l4 (Ym2) proteins of subgroup I (Figure 1), however, have glutamines substituted for the active-site glutamates.

The three-dimensional structures of some of the chitolectins, including IDGF2, are available [40-42]. Analysis of the structure-function properties of IDGF2 suggests the interesting possibility that the *D. melanogaster* IDGFs may have evolved to acquire new functions as growth factors by an evolutionary progression from carbohydrate-binding to protein-binding molecules [40]. The 3D structures of the binding sites of two of the mammalian chitolectins, human cartilage glycoprotein-39 (CHI3L1) [GenBank:[NP 001267](#)] and mouse Ym1 (Chi3l3) [GenBank:[NP 034022](#)], support carbohydrate binding in a manner similar to that observed for active chitinases, with no indication of the evolutionary progression to protein binding as proposed for IDGF2 [41,42]. This is consistent with our phylogenetic analysis, which shows a much closer evolutionary relationship of the human chitolectin proteins to a putative chitinase [GenBank:[NP 647768](#)] and several other *D. melanogaster* proteins than to the IDGFs (Figure 3). However, based on a recent higher resolution structure of the mouse Ym1 protein, the carbohydrate-binding groove in this mammalian chitolectin may be somewhat different from that of the human cartilage glycoprotein-39 (CHI3L1) [43].

Recently two family GH18 proteins were cloned from the oyster *Crassostrea gigas*, providing the first information about the protein family from the Lophotrochozoan branch of protostomes [44,45]. The proteins were *Cg*-Clp1 [GenBank:[CA196028](#)], a chitinase-like protein homologous to the human CHI3L1 (GP-39), and *Cg*-Chit

[GenBank:CAI96026], an atypical family 18 chitinase. Data from gene expression profiles and biochemical characterization indicated the Cg-Clp1 protein is involved in control of growth and development, as reported for the *D. melanogaster* IDGF proteins. The Cg-Chit protein characterization, however, suggests that it plays an important role as an immunity effector in molluscs. Analysis of phylogenetic relationships of the Cg-Clp1 protein with proteins from other branches of Bilateria supported a closer relationship of the Cg-Clp-1 protein to mammalian chitinases/chitolectins than to the *D. melanogaster* IDGFs [44]. Currently, six GH18 domain protein sequences from *Crassostrea gigas* are available from the NCBI protein database [GenBank: CAI96025, CAI96026, CAI96027, CAI96028, CAI96023, CAI96024]. Three of the sequences, including Cg-Chit (Chit3) [GenBank:CAI96026], include the proton-donating glutamate residue in the catalytic motif, indicating that they are active chitinases. The other three sequences have a glutamine residue substituted for the glutamate, as seen in *D. melanogaster* IDGF proteins, and are likely chitolectins. We created phylogenetic trees from sequence alignments of the oyster GH18 domains with domains from *C. elegans*, *D. melanogaster*, and from selected mammalian species and, as reported for the Cg-Clp1 protein [44,45], found support for an ancestral relationship of *C. gigas* proteins with the mammalian chitinase/chitolectins. Our analysis supported separate monophyletic groups for the human chitolectins, *D. melanogaster* IDGF proteins, and the *C. gigas* chitinase-like proteins.

The phylogenetic relationships between the human chitinase/chitolectin sequences and the *C. elegans* and *D. melanogaster* sequences shown in Figure 3 suggest that the present multigene chitinase/chitolectin families in each individual species reflect unique innovations that accompanied the evolutionary radiation of bilaterians into the protostome and deuterostome divisions. The limited orthology of the vertebrate chitinase/chitolectin family with the protostome chitinases/chitolectins is consistent in both deuterostomes and protostomes with extensive gene duplication, followed by retention and diversification as well as gene loss after the radiation. In *C. elegans* and *D. melanogaster*, expansion and diversification of chitinases likely were driven by the appearance and importance of chitinous structures in nematodes and insects and the need for proteins associated with chitin turnover in development, protection and metabolism. The factors important in the human genome, however, are more difficult to understand. Clearly, the human chitinase/chitolectin proteins are more similar to each other than to homologs in *C. elegans* and *D. melanogaster*. On this basis, we infer that most of the human genes are relatively recent products of gene duplication events. The genes produced

by the duplications therefore have evolved in vertebrates in the absence of chitin biosynthesis.

Chitinases/Chitolectins in Early Deuterostomes (*C. intestinalis* and *S. purpuratus*)

The expansion and diversification of chitinase-related genes observed in the protostomes *C. elegans* and *D. melanogaster* apparently did not occur in the earliest deuterostomes. Available data from the urochordate *Ciona intestinalis* (sea squirt) draft genome as well as from echinodermata *Strongylocentrotus purpuratus* (sea urchin) indicate far fewer chitinase-related genes in non-chitinous organisms than already observed for *C. elegans* and *D. melanogaster*. The availability of these genomes provides an opportunity to examine chitinase-related genes from genomes highly relevant for elucidation of the ancestral state prior to two rounds of genome duplication postulated to have occurred in a lineage leading to vertebrates [46-48].

Searches of the data available from the JGI (Joint Genome Institute) *C. intestinalis* draft genome v2.0 (March 2005) [49] and from *S. purpuratus* (genome assembly build 1.1; August 2005) identified three sequences from the *C. intestinalis* genome and four from the *S. purpuratus* genome that produced high scoring alignments with human GH18 family members (Accession numbers listed in additional file 5: Other deuterostome sequences). Figure 4 shows a phylogenetic tree created from an alignment of the *C. intestinalis* and *S. purpuratus* sequences with mammalian GH18 protein sequences. The analysis indicates that the mammalian chitinases/chitolectins are co-orthologs of *C. intestinalis* JGI gene model protein ID229566. The mammalian chitobiases and stabilin-1 interacting proteins are orthologs, respectively, of the other two *C. intestinalis* proteins JGI model ID 239218 and JGI model ID 223016. Further searches of the *C. intestinalis* genome and EST databases available at JGI did not identify other candidate genes for chitinase/chitolectin proteins. Thus, each of the major mammalian phylogenetic groups (Figure 1) must be represented in the *C. intestinalis* genome by a single gene. Three of the *S. purpuratus* proteins [GenBank:XP_798621, XP_783710, and XP_798628] are orthologs of the mammalian chitobiases. The protein XP_783668 is an ortholog of mammalian stabilin-1 interacting chitin-binding protein. As described previously for the *D. discoideum* genome, searches of the *S. purpuratus* genome did not identify genes with homology to the mammalian chitinases/chitolectins.

Collectively, the *C. intestinalis* and *S. purpuratus* data, which identify a single chitinase/chitolectin gene in *C. intestinalis* and none in *S. purpuratus*, indicated that a selective expansion of this group likely occurred later in deuterostome evolution to produce the current mamma-

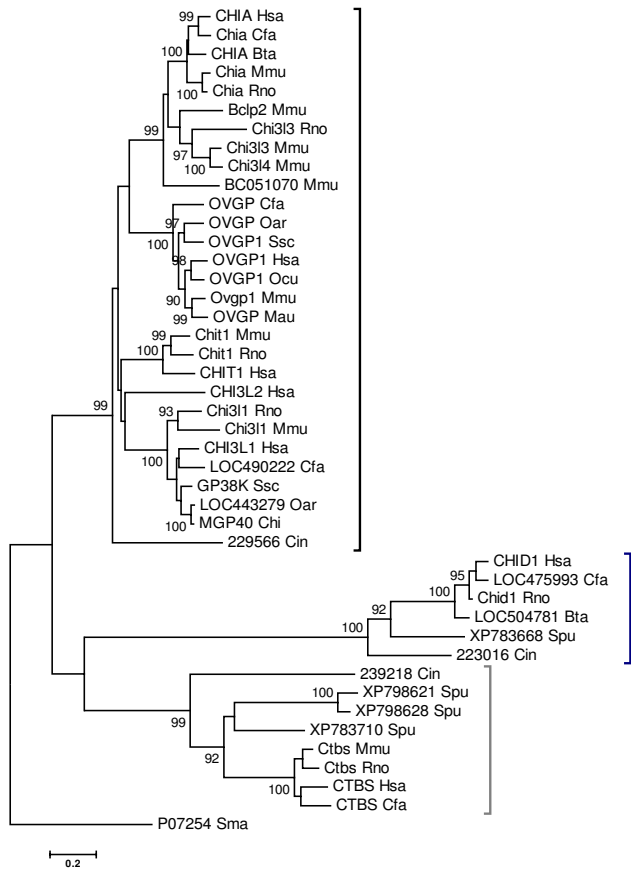


Figure 4
Phylogenetic analysis of GH18 family members from mammals, *C. intestinalis* and *S. purpuratus*. The figure shows the minimum evolution tree from a phylogenetic analysis conducted as described in the legend for Figure 1. JGI model protein ID numbers are used for *C. intestinalis* proteins, accession numbers for the *S. purpuratus* proteins, and gene symbols (or protein symbols when gene symbols are not available) for mammalian proteins. A black bracket indicates the chitinases/chitolectins, a dark blue bracket the stabilin-1 interacting chitolectins, and a dark gray bracket the chitobiases.

lian chitinase/chitolectin phylogenetic group (Figure 1). The two outlying groups, chitobiases and stabilin-1 chitolectins, did not undergo a similar expansion. Notably, the *C. intestinalis* genome assembly v2.0 (March 2005) represents the third assembly release, and is essentially complete with some scaffolds (179 Mb) not yet mapped to chromosomes [50]. The *S. purpuratus* genome sequencing is in progress; therefore, the preliminary assembly (build 1.1) [51] may not include all GH18 family genes.

Vertebrate Chitinases/Chitolectins (*X. tropicalis*, *G. gallus*, *M. domestica*)

Searches of the *X. tropicalis* genome (August 2005 Assembly) and the *G. gallus* genome (May 2006 Assembly) for

GH18 family genes indicated that each of these non-mammalian vertebrate genomes includes a single cluster of chitinase homologs as well as one chitobiase and one stabilin-1 interacting protein homolog (additional file 5: Other deuterostome sequences). In the case of *X. tropicalis*, a 95 kb region of scaffold 41 encodes three chitinase genes [Entrez Genes 548945, 548404 and 448265], each including a GH18 domain and a CBM14 domain. An 18 kb region of *G. gallus* chromosome 26 encodes two chitinase genes with GH18 and CBM14 domains [Entrez Genes 41993, 395072] and an open reading frame encoding a partial GH18 domain, possibly a pseudogene. The presence of the proton-donating glutamate in the active site of the GH18 domains predicts that the *X. tropicalis* and *G. gallus* gene products are active chitinases. The data showing close clustering of the *X. tropicalis* and *G. gallus* chitinase genes indicate that tandem duplications at a single locus occurred in the evolution of the gene families of these species. A possibility based on observation of conserved synteny is that the *X. tropicalis* scaffold 41 and the *G. gallus* chromosome 26 may represent current versions of an ancestral GH18 locus that expanded and rearranged in mammals after the mammalian-avian split (approx. 310 mya) [52]. The present overall arrangement of the mammalian genes indicates both an expansion and separation of the chitinase/chitolectin genes after the last common ancestor of birds and mammals

Monodelphis domestica (the gray short-tailed opossum) was interesting to examine as a representative of an intermediate between the existing mammalian and non-mammalian vertebrate models. As a result of common ancestry, the opossum GH18 family genes may reveal changes that reflect innovations that accompanied mammalian evolution after the avian split. The metatherian ("marsupial") mammals are phylogenetically distinct from current placental mammals having diverged approximately 150–180 mya; however, they are more closely related to one another than to any other vertebrates (birds, amphibians, fishes) [53,54].

The opossum genome includes, on chromosome 2, seven Genescan Gene Prediction models [55] encoding GH18 family proteins (additional file 5: Other deuterostome sequences). Six of the models (c_307.104 through c_307.108 and c_214.85) predict chitinase/chitolectin homologs and the seventh model (c_199.2) predicts a human chitobiase homolog. Three of the gene models (c_307.104, c307.107, and c_214.85) include two GH18 domains. The models encoding chitinase/chitolectins form clusters at the 107 Mb and the 542 Mb regions of chromosome 2, and the chitobiase model is located at 44 Mb. This arrangement resembles that of human GH18 family genes on chromosome 1.

Whole genome alignments of the human genome with the opossum genome (Jan 2006 Assembly) indicated that human genes CHI3L2, RP11-165H20.1, CHIA and OVGP1 align, respectively, with Genescan models c_307.104 (first GH18 domain), c_307.105, c_307.106, and c_307.108. Human genes CHI3L1 and CHIT1 align with the two domains of c_214.85. Thus, the opossum genome includes in the same orientation and order an orthologous gene for each human chitinase/chitolectin, and the human family GH18 family gene arrangement extends at least as far back as marsupials. The opossum genome includes in the same region an additional gene model (c_307.107) with two GH18 domains, which is not included in the human orthology. Further investigation indicated that these two domains (denoted c_307.107A/B in Figure 7, discussed below) are co-orthologs of the *X. tropicalis* putative chitinase [Entrez Gene 448265]. Congruency in number and physical arrangement of GH18 family genes in the opossum and human genomes indicates that an expansion and separation of chitinase/chitolectin genes occurred before the divergence of the metatharian and placental mammals (150–180 mya).

Chitinase/Chitolectin Encoding Regions Are Adjacent to the Human MHC Paralogon Region at Chromosome 1

An interesting observation in view of the recent discovery of an association of the chitinase proteins with Th2 cell mediated immunity (discussed earlier [5,6]) is that the human GH18 family genes are located adjacent to the MHC paralogon genes on the two arms of human chromosome 1 (Figure 5). The human MHC consists of a large number of genes essential to antigen presentation and immune function, including the MHC class I and class II genes that serve pivotal roles in adaptive immunity. In addition to the MHC locus at chromosome 6p21.3, the human genome includes three paralogous genomic regions (paralogons) located on different chromosomes (chromosome 1p11-p32/1q21-q25, chromosome 9q32-q34, and chromosome 19 p13.1-p13.3) [48]. The GH18 cluster including CHI3L2, RP11-165H20.1, CHIA, and OVGP1 is located at 1p13.3-p13.2 and the cluster including CHI3L1 and CHIT1 is located at 1q32.1.

Furthermore, regions of human chromosomes 1 and 6 adjacent to and including the human MHC, MHC paralogon, and GH18 family genes align with chitinase-containing regions of *G. gallus* chromosome 26, *X. tropicalis* scaffold 41, and *M. domestica* chromosome 2 (Figure 6). Taken as a whole, as much as 82% of the 5.1 Mb *G. gallus* chromosome 26, which encodes the chitinase genes, aligns with regions adjacent to the human MHC and the MHC paralogon at chromosome 1. The *G. gallus* MHC, however, is located separately in a 92 kb region of chromosome 16 (a microchromosome of 433 Kb). This region

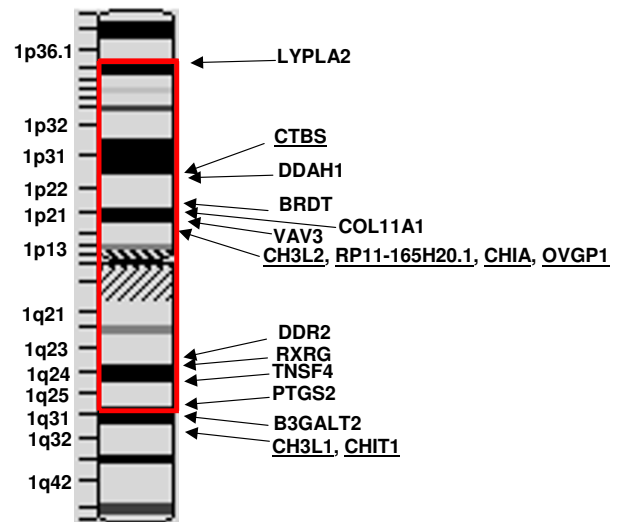


Figure 5
Proximity of the human family GH18 genes to the human MHC paralogon on chromosome 1. An ideogram from NCBI Map Viewer shows human chromosome 1 with region encoding MHC paralogon genes boxed in red. Arrows point to the locations of the human family GH18 genes (symbols underlined to distinguish them from MHC related genes) and representative genes from the MHC paralogous region. The choice of genes is based on a recent comparative analysis of the MHC [48]. The MHC paralog symbols are: LYPLA2, lysophospholipase II (1p36.12-p35.1); DDAH1, dimethylarginine dimethylaminohydrolase I (1p22); BRDT, bromodomain, testis-specific (1p22.1); COL11A1, collagen, type XI alpha 1 (1p21); VAV3, vav 3 oncogene (1p13.3), DDR2, discoidin domain receptor family, member 2 (1q12-q23); RXRG, retinoid X receptor, gamma, 1q22-q23); TNFSF4, tumor necrosis factor (ligand) superfamily, member 4 (1q25); PTGS2, prostaglandin-endoperoxide synthase 2 (1q25.2-q25.3); and B3GALT2, UDP-Gal:beta GlcNAc beta 1,3-galactosyltransferase, polypeptide 2 (1q31).

is generally considered to represent a compact version of the more complex human MHC as it consists of only 19 genes, virtually all having counterparts in the human genome [56]. The *X. tropicalis* scaffold 41 (3.9 Mb) is almost the same size as *G. gallus* chromosome 26 and shows similar synteny with human chromosomes 1 and 6 (Figure 6).

The opossum MHC, which is similar in size and complexity to MHCs of eutherian mammals [57], is located on chromosome 2 (260 Mb region). Opossum chromosome 2 not only contains the GH18 family genes, but also its alignments with the human genome encompassed extensive regions of human chromosomes 1 and 6 including, respectively, the MHC paralogon and the MHC (Figure 6). Approximately 40% of opossum chromosome 2 aligned

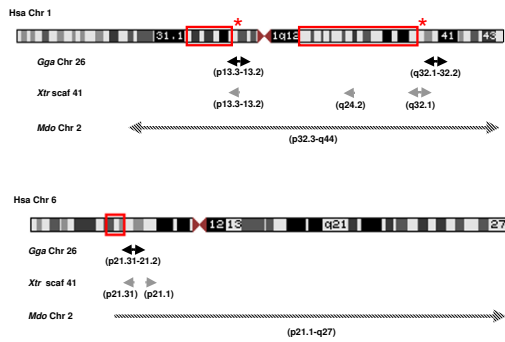


Figure 6
Syntenic relationships of *X. tropicalis*, *G. gallus*, *M. domestica* and human genomic regions involving MHC. The red rectangles on ideograms of human chromosomes show the locations of the MHC (chr 6) and the two MHC paralogous regions (chr 1). Red asterisks indicate the loci for the human chitinase/chitolectin genes on chr1 (p13.3-p13.2) and chr1 (q32.1). The labeled horizontal arrows below the two ideograms designate the regions and orientations of human chromosomes 1 and 6 that correspondingly align with *G. gallus* chromosome 26 (black); *X. tropicalis* scaffold 41 (gray), and *M. domestica* chromosome 2 (diagonal-pattern).

with the human chromosome 1 and another 40% with human chromosome 6. From the perspective of the human genome, as much as 80% of the 247 Mb human chromosome 1 and 85% of the 170 Mb human chromosome 6 aligned with opossum chromosome 2. Collectively, these data from alignment of the human genome with the *X. tropicalis*, *G. gallus*, and opossum genomes demonstrate conserved general patterns of synteny between family GH18 genes and regions including or adjacent to the MHC and MHC paralogon 1. This conserved synteny suggests an ancient organizational relationship between the MHC and GH18 genes.

Speculations Regarding the Chitinase/Chitolectin Genes

Host defense associated with innate immunity is one of the important established functions of chitinases, extending as far back in evolution as the common ancestor of plant and animal kingdoms [11]. However, the recent descriptions of the involvement of chitinases/chitolectins in human and mouse T cell effector mechanisms [5,6] suggest an additional role concerned with the adaptive Th2 immune response in mammals. The evolution of an adaptive immune system with somatic rearranging genes is a relatively recent occurrence. Current evidence indicates that genes critical for adaptive immunity first appeared in jawed vertebrates (gnathostomes) about 500

mya [58]. Comparative genomics of the MHC indicate that the paralogous regions on chromosomes 1, 6, 9, and 19 may have originated by block or whole genome duplications from a common ancestral proto-MHC region present in protochordates (*Ciona* and *Branchiostoma*) [48,59,60]. Previous studies indicate that two rounds of whole genome or large block duplication took place before the emergence of jawed vertebrates, but after the emergence of cephalochordates. A recent review discusses the duplications and the architecture of such a proto-MHC in protochordates, and suggests that this region may have included genes that evolved long before the emergence of the adaptive immune system [48]. Immune-related genes may have been among early genes that were co-opted or recruited for additional roles related to adaptive immunity [47,61]. Earlier references [5,6] as well as a recently published review describing chitinases and chitolectins as "critical mediators" of Th2 inflammatory responses [62] are consistent with the view that chitinase/chitolectin genes were among those recruited.

Considering these new observations related to the evolutionary history of the chitinases, involvement of mammalian GH18 proteins in both innate and adaptive immunity may be related to their expression in phagocytic cells, particularly macrophages [20,21,63] and dendritic cells [64], which have roles in both types of immunity [65]. The association of GH18 proteins with phagocytosis is evident early in animal evolution, as we identified five homologs of the human lysosomal chitobiase in the *D. discoideum* genome. Since *D. discoideum* cells are professional phagocytes that engulf and digest bacteria as a nutrient source, these enzymes are likely among the hydrolases in their phagosomes and lysosomes involved in digestion of bacterial cell walls [25]. During metazoan evolution, the primary digestive function of phagosomes has expanded to include roles in development, tissue remodeling, apoptosis, and in innate and adaptive immunity [66]. In mammalian phagocytic cells, engulfment, and degradation of pathogens as well as mechanisms related to MHC II antigen presentation are associated with endosomal/lysosomal compartments, where several members of the GH18 family are located [20,23,67]. Therefore, the chitinase/chitolectin genes may have evolved and expanded during vertebrate evolution in association with evolutionary changes in phagocytosis, particularly changes associated with vesicular antigen processing pathways.

Many observations are now remarkably consistent with chitinase family GH18 proteins having roles in vesicular antigen processing and MHC II presentation: the requirement of AMCase (CHIA) enzymatic activity for its mediation of Th2 responses in inflammation and asthma [5]; an acidic pH optimum of AMCase catalytic activity corre-

sponding to the acidic endosomal/lysosomal compartments involved in antigen processing [17]; the presence of AMCCase (CHIA) and other chitinase/chitolectins, particularly mouse Ym1 (Chi3l3) and human chitinases, in macrophages and other antigen presenting cells [5,63,68,69]; and regulation of both chitinase and chitolectin gene expression by inflammatory cytokines such as interferon γ , IL-4 and IL-13 [5,63,68-70]. An interesting speculation is that chitinase-related genes were present in an ancestral proto-MHC region and were among those recruited for additional roles in immunity. The patterns of conserved synteny of GH18 proteins with the MHC and the chromosome 1 paralogon are consistent with this possibility. Recently the organization of the proto-MHC region of the cephalochordate amphioxus *Branchiostoma floridae* (traditionally considered the closest living invertebrate to the vertebrates, a view recently challenged by new phylogenetic data [71]) became available and was used to reconstruct chordate and gnathostome ancestor proto-MHC regions [72]. Interestingly, a homolog for the human NEK7 gene, located at 1q31.3, approx. 5Mb centromeric to the CHIT1 gene, is present in the reconstructed proto-MHC. In *X. tropicalis*, an NEK7 homolog is located on scaffold 41 approx. 4 Mb upstream of the chitinase gene cluster. The opossum chromosome 2 also encodes a NEK7 homolog (located at 99 Mb, 8 Mb upstream of the CHIT1 homolog located at 107 Mb). Thus, synteny between ancient proto-MHC genes and chitinase family members might have existed at least since the emergence of amphibians (approx. 350 mya) [73].

Evolution by a birth-and-death process

Our comparative genomic analysis of the GH18 protein family indicates that considerable loss, duplication and diversification of these genes occurred in vertebrates. The diversification and expansion occurred in the chitinase/chitolectin phylogenetic group with no indication of similar changes in the chitobiases and stabilin-1 interacting groups. A phylogenetic tree, Figure 7, shows the relationship of homologous chitinase/chitolectin genes from several vertebrate species including six chitinase genes from *D. rerio* (zebrafish), which shared a common ancestor with mammals approximately 420 mya [53]. The tree also includes the *C. intestinalis* chitinase [JGI protein model ID 229566] and is rooted with the bacterial GH18 protein *Serratia marcescens* [GenBank:P07254]. As observed for earlier metazoan species, Figure 7 shows that distantly related vertebrates such as humans, *M. domestica*, *G. gallus*, *X. tropicalis*, and *D. rerio* have significantly different sets of chitinase/chitolectin genes. According to the tree, the genes form four major clusters (A, B, C, D). Cluster A corresponds to the chitinase/chitolectin subgroup I (Figure 1); cluster C, to subgroup IV; cluster D, to subgroup II; and cluster E to subgroup III. Cluster B does not include genes from placental mammals and is not included in Figure 1.

The dynamic evolutionary process for these genes must reflect the morphological and other changes that accompanied the stages of vertebrate evolution.

The process of gene expansion, deletion and diversification must have occurred more than once to explain the long evolutionary history of the GH18 family of chitinase genes. Although the literature indicates multiple chitinases are present in plants and bacteria, we clearly identified chitobiase genes, but no chitinase/chitolectin genes in the nearly complete *D. discoideum* genome. An important expansion of chitinase/chitolectin genes must have occurred in protostomes *C. elegans* and *D. melanogaster*, as chitinous structures in these species became important protective and architectural features. Chitin synthase (biosynthesis) and chitinases lost importance in early deuterostomes, and examination of chitinase/chitolectin genes in *C. intestinalis* and *S. purpuratus* genomes indicated that the expansion and diversification of chitinases/chitolectins did not continue in these early non-chitinous deuterostomes (Figure 4). However, comparison of the genes from these early genomes with higher vertebrate and mammalian genomes indicated a significant expansion of genes during late vertebrate evolution. Comparison of GH18 family genes from *X. tropicalis*, *G. gallus* and *M. domestica* revealed an increase of chitinase/chitolectin genes after the avian-mammalian split (Figure 7).

Figure 8 provides a summary of the contraction and expansion of these genes during eukaryotic evolution. Overall, the data presented show a general lack of orthology between the protostome, early deuterostome, and mammalian chitinases/chitolectins and support the view that the mammalian genes evolved new functions associated with innovations that accompanied evolution of the mammalian species. The phylogenetic data for the GH18 family are consistent with the model of birth-and-death evolution described by Nei and colleagues [74].

The evolutionary history of the GH18 family includes other examples of recruitment of family members for novel functions. The emergence of a subfamily of xylanase inhibitors within the GH18 family recently reported in plants represents one such example [75]. A factor central to the evolution of novel functions for the GH18 family may be their very common protein fold, the TIM barrel conformation, known to form a stable core structure with specific binding and catalytic functions residing in the loop regions separated from the alpha helices and beta sheets. This type of structure/function relationship in proteins may be particularly well suited to divergence of new functions from an ancestral protein by the process of gene duplication and subsequent modification by mutation, since changes in the loop regions would not produce major changes in protein stability [75,76]. McMahon et

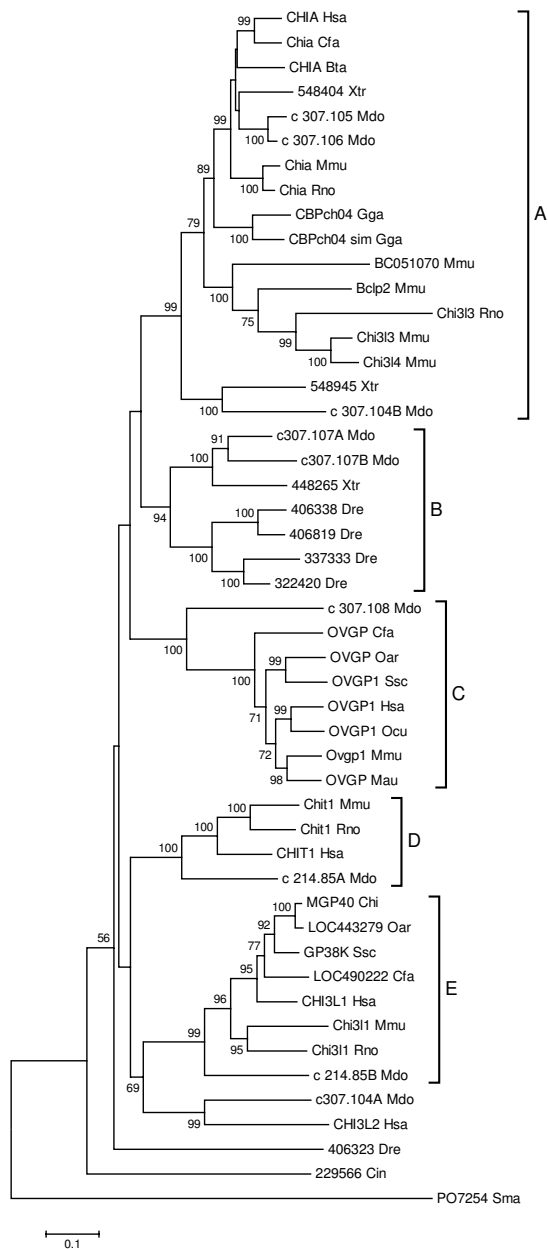


Figure 7
Phylogenetic tree of chordate chitinases/chitolectins.
 The figure shows the minimum evolution tree from a phylogenetic analysis conducted as described in the legend for Figure 1. Besides mammals, the species are *X. tropicalis* (Xtr), *D. rerio* (Dre), *G. gallus* (Gga), *M. domestica* (Mdo) and *C. intestinalis* (Cin). Gene or protein symbols are used for mammals; Gene ID numbers for *X. tropicalis*, *D. rerio* and *G. gallus*; Genscan Gene Prediction symbols for *M. domestica*; and the JGI Model protein ID number for *C. intestinalis*. Groups A-E and their correspondence to subgroups I-IV in Fig. 1 are described in the text. Scale, rooting and bootstrapping are as described for Figure 1. Bootstrap values $\geq 50\%$ are shown.

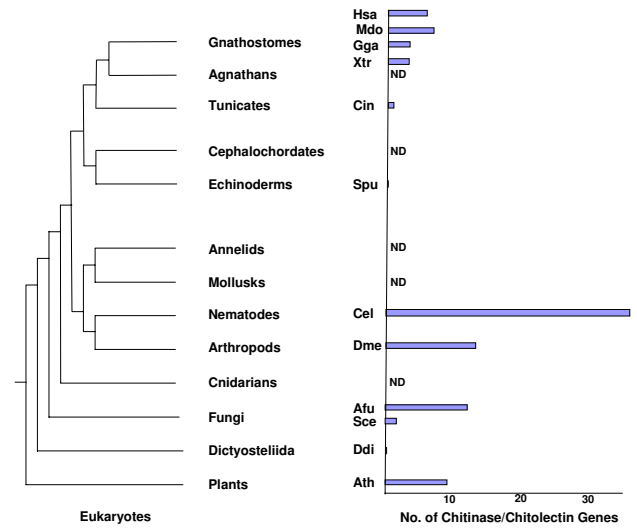


Figure 8
Summary diagram indicating contraction and expansion of the chitinase/chitolectin protein family during eukaryotic evolution. A schematic of eukaryote phylogeny is shown alongside a bar graph indicating the number of chitinase/chitolectin genes identified in the genomes of selected species corresponding to the tree classifications. The schematic is according to the recent animal phylogeny suggested by multigene analyses of bilaterian animals [71]. The species for which the number of genes is shown are the gnathostomes (jawed-vertebrates) human (Hsa), opossum (Mdo), chicken (Gga) and *Xenopus* (Xtr); the chordate *C. intestinalis* (Cin); the echinoderm *S. purpuratus* (Spu); the nematode *C. elegans* (Cel); the arthropod *D. melanogaster* (Dme); and the Dictyosteliida *D. discoideum* (Ddi), all discussed in the text. Also included in the graph are data for the fungi *Aspergillus fumigatus* and *Saccharomyces cerevisiae* and the plant *Arabidopsis thaliana* (Additional file 7: Other chitinase sequences). The *A. fumigatus* and *S. cerevisiae* genes were obtained by querying Fungal Genomes Central at NCBI.

al. (2005) recently emphasized the importance of scaffold stability in the evolution of proteins displaying sequence variation related to binding diversity [77]. To our knowledge, recruitment for development or host defense purposes has not been reported for other GH families.

Conclusion

Our comprehensive genomic study of animal GH18 family proteins provides valuable new insights into phylogenetic relationships of the proteins and the evolutionary history of the protein family. This represents the first such detailed study of animal GH18 family proteins. The family consists of three major phylogenetic groups: chitobiases, chitinases/chitolectins, and stabilin-1 interacting chitolectins. Only the chitinase/chitolectin group is associated with expansion in late deuterostomes. The phylogeny of animal GH18 genes is consistent with evolution of

the family by a birth-and-death process according to the model described for multigene families of the vertebrate immune system [74]. The finding that the human GH18 gene family is closely associated with the human MHC paralogon on chromosome 1, together with published data indicating an association of GH18 chitinase activity with Th2 cell inflammation, suggests the possibility that the late expansion could be related to an emerging interface of innate and adaptive immunity during early vertebrate history.

Methods

Genome searches

For *C. elegans* and *D. melanogaster* genomes, BLASTP was used to query the NCBI Entrez Genome nucleotide and protein databases with the complete human chitinase sequences [GenBank:NP_970615 and NP_003456] and the GH18 domain sequences. The Entrez protein link to the NCBI's Conserved Domain Database search service (CDD) [78] identified the GH18 domain family members (GH18 is designated CDD|24313 which corresponds to pfam00704, smart60636, and Interpro IPR001223 in the respective domain-based databases). The BLAST protein searches used the Entrez protein graphical viewer BLink [79] to identify closely related sequence neighbors. Other resources were Ensembl v37 – February 2006 [80] used to search for IPR001223 domains, Flybase [36] and Aceview [35].

Query of *Ciona intestinalis* genome sequence v2.0 (March 2005) available from JGI Joint Genome Institute, (JGI) [49] using human protein sequences and GH18 domains [GenBank:NP_970615 and NP_003456] along with searches of the *C. intestinalis* genome using the domain GH18 [Interpro IPR001223] identified the *C. intestinalis* chitinase-related genes.

The human chromosome 1 regions encoding chitinase/chitolectin genes were compared to whole genome assemblies from *Monodelphis domestica*, *Gallus gallus*, *Xenopus tropicalis* and *Danio rerio* using the UCSC browser [81]. The UCSC Genome Browser databases [82] used in this study were: NCBI Build 35, May 2004 Assembly (hg17) produced by the International Human Genome Sequencing Consortium; Opossum Jan 2006 (UCSC version monDom4) (produced by The Broad Institute, Cambridge, MA, USA; the February 2004 chicken (*Gallus gallus*) draft assembly, produced by the Genome Sequencing Center at the Washington University School of Medicine in St. Louis, Mo. USA; the August 2005 Assembly frog (*Xenopus tropicalis*) whole genome shotgun (WGS) assembly version 3.0, DOE Joint Genome Institute (JGI) Walnut Creek, CA, USA; the May 2005 zebrafish (*Danio rerio*) Zv5 assembly produced by The Wellcome Trust Sanger Institute in collaboration with the Max Planck Institute for Developmental Biology in Tübingen, Germany, and the

Netherlands Institute for Developmental Biology (Hubrecht Laboratory), Utrecht, The Netherlands.

Other genome resources were: NCBI Entrez genome databases for *Strongylocentrotus purpuratus* (sea urchin) Build 1.1, August 2005; *Dictyostelium discoideum* Build 1.1, November 2005, and *Danio rerio* (zebrafish) Build Zv4, July 2005 and Ensembl genomes *Danio rerio* Zv5, May 2005 Assembly; *Xenopus tropicalis* JGI 4.0 June 2005 Assembly, and *Monodelphis domestica* MonDom 4.0, January 2006 Assembly. In some cases, NCBI Map Viewer [83] was used for chromosomal locations.

Phylogenetic Analysis

For multiple sequence alignments, the Clustal X (version 1.83) Program [84] aligned the GH18 domain amino acid sequences. Boundaries of GH18 domains were obtained from the multiple sequence alignment by editing multiple alignments to conform to the GH 18 domain of human chitotriosidase (CHIT1), for which the 3D structure is available [PDB: 1HKK]. The alignment parameters used were those suggested by Hall [85]. Values used for pairwise alignments were gap opening penalty 35 and gap extension penalty 0.75. Values for multiple alignment were gap opening penalty 15, gap extension penalty 0.3, and delay divergent sequences 25%. Protein weight matrix chosen from the multiple alignment parameters menu was Gonnet series. Duplicate and alternatively spliced protein isoforms were excluded so that only one gene product was used for analysis. Phylogenetic analyses were conducted with MEGA version 3.1 software [86] using maximum parsimony, minimum evolution, and neighbor-joining methods with the Poisson-correction for multiple amino acid substitutions. In all cases, the three methods produced very similar topologies. A summary composite tree that includes the early eukaryote, protostome, early deuterostome and vertebrate sequences discussed in the text is included as additional file 8: Composite tree.

Authors' contributions

JDF & NNA conceived the study. JDF collected and analyzed the data and wrote the paper. NNA provided substantial editorial advice and assistance and provided critical insights into GH18 family biochemistry and biology. Both authors have read and approved the final manuscript.

Additional material

Additional file 1

GH18 Family Domain Structure.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-7-96-S1.doc>]

Additional file 2*Mammalian Sequences.*

Click here for file

[\[http://www.biomedcentral.com/content/supplementary/1471-2148-7-96-S2.doc\]](http://www.biomedcentral.com/content/supplementary/1471-2148-7-96-S2.doc)**Additional file 3***C. elegans Sequences.*

Click here for file

[\[http://www.biomedcentral.com/content/supplementary/1471-2148-7-96-S3.doc\]](http://www.biomedcentral.com/content/supplementary/1471-2148-7-96-S3.doc)**Additional file 4***D. melanogaster Sequences.*

Click here for file

[\[http://www.biomedcentral.com/content/supplementary/1471-2148-7-96-S4.doc\]](http://www.biomedcentral.com/content/supplementary/1471-2148-7-96-S4.doc)**Additional file 5***Other Deuterostome Sequences.*

Click here for file

[\[http://www.biomedcentral.com/content/supplementary/1471-2148-7-96-S5.doc\]](http://www.biomedcentral.com/content/supplementary/1471-2148-7-96-S5.doc)**Additional file 6***Supplementary Phylogenetic Tree.*

Click here for file

[\[http://www.biomedcentral.com/content/supplementary/1471-2148-7-96-S6.doc\]](http://www.biomedcentral.com/content/supplementary/1471-2148-7-96-S6.doc)**Additional file 7***Other Chitinase Sequences.*

Click here for file

[\[http://www.biomedcentral.com/content/supplementary/1471-2148-7-96-S7.doc\]](http://www.biomedcentral.com/content/supplementary/1471-2148-7-96-S7.doc)**Additional file 8***Composite tree.*

Click here for file

[\[http://www.biomedcentral.com/content/supplementary/1471-2148-7-96-S8.doc\]](http://www.biomedcentral.com/content/supplementary/1471-2148-7-96-S8.doc)**Acknowledgements**

This research was supported by NIH grant GM-59471 (N.A.) and NSF-EPS-CoR funds to the Alabama Structural Biology Consortium. We acknowledge the contribution of Ms. Jie Li, USA Biomedical Library, who generously assisted with manuscript preparation.

References

- Henrissat B: **Classification of chitinases modules.** *Exs* 1999, **87**:137-156.
- Stam M, Blanc E, Coutinho P, Henrissat B: **Evolutionary and mechanistic relationships between glycosidases acting on alpha- and beta-bonds.** *Carbohydr Res* 2005, **340(18)**:2728-2734.
- CAZy** [<http://www.cazy.org/>]
- Merzendorfer H, Zimoch L: **Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases.** *J Exp Biol* 2003, **206(Pt 24)**:4393-4412.
- Zhu Z, Zheng T, Homer RJ, Kim YK, Chen NY, Cohn L, Hamid Q, Elias JA: **Acidic mammalian chitinase in asthmatic Th2 inflammation and IL-13 pathway activation.** *Science* 2004, **304(5677)**:1678-1682.
- Elias JA, Homer RJ, Hamid Q, Lee CG: **Chitinases and chitinase-like proteins in T(H)2 inflammation and asthma.** *J Allergy Clin Immunol* 2005, **116(3)**:497-500.
- Takeo S, Fujise M, Akiyama T, Habuchi H, Itano N, Matsuo T, Aigaki T, Kimata K, Nakato H: **In vivo hyaluronan synthesis upon expression of the mammalian hyaluronan synthase gene in Drosophila.** *J Biol Chem* 2004, **279(18)**:18920-18925.
- Shen Z, Jacobs-Lorena M: **Evolution of chitin-binding proteins in invertebrates.** *J Mol Evol* 1999, **48(3)**:341-347.
- Fukamizo T: **Chitinolytic enzymes: catalysis, substrate binding, and their application.** *Curr Protein Pept Sci* 2000, **1(1)**:105-124.
- Ohno T, Armand S, Hata T, Nikaidou N, Henrissat B, Mitsutomi M, Watanabe T: **A modular family 19 chitinase found in the prokaryotic organism Streptomyces griseus HUT 6037.** *J Bacteriol* 1996, **178(17)**:5065-5070.
- Kasprzewska A: **Plant chitinases – regulation and function.** *Cell Mol Biol Lett* 2003, **8(3)**:809-824.
- Fujita K, Shimomura K, Yamamoto K, Yamashita T, Suzuki K: **A chitinase structurally related to the glycoside hydrolase family 48 is indispensable for the hormonally induced diapause termination in a beetle.** *Biochem Biophys Res Commun* 2006, **345(1)**:502-507.
- Kubota T, Miyamoto K, Yasuda M, Inamori Y, Tsujibo H: **Molecular characterization of an intracellular beta-N-acetylglucosaminidase involved in the chitin degradation system of Streptomyces thermoviolaceus OPC-520.** *Biosci Biotechnol Biochem* 2004, **68(6)**:1306-1314.
- Mark B, Vocadlo D, Knapp S, Triggs-Raine B, Withers S, James M: **Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase.** *J Biol Chem* 2001, **276(13)**:10330-10337.
- Kuranda MJ, Aronson NN Jr: **A di-N-acetylchitinase activity is involved in the lysosomal catabolism of asparagine-linked glycoproteins in rat liver.** *J Biol Chem* 1986, **261(13)**:5803-5809.
- Hollak CE, van Weely S, van Oers MH, Aerts JM: **Marked elevation of plasma chitotriosidase activity. A novel hallmark of Gaucher disease.** *J Clin Invest* 1994, **93(3)**:1288-1292.
- Boot RG, Blommaert EF, Swart E, Ghauharali-van der Vlugt K, Bijl N, Moe C, Place A, Aerts JM: **Identification of a novel acidic mammalian chitinase distinct from chitotriosidase.** *J Biol Chem* 2001, **276(9)**:6770-6778.
- Takahashi K, Sendai Y, Matsuda Y, Hoshi H, Hiroi M, Araki Y: **Mouse oviduct-specific glycoprotein gene : genome organization and structure of the 5'-flanking regulatory region.** *Biol Reprod* 2000, **62**:217-226.
- Agarwal A, Yeung WS, Lee KF: **Cloning and characterization of the human oviduct-specific glycoprotein (HuOGP) gene promoter.** *Mol Hum Reprod* 2002, **8(2)**:167-175.
- van Eijk M, van Roomen CP, Renkema GH, Bussink AP, Andrews L, Blommaert EF, Sugar A, Verhoeven AJ, Boot RG, Aerts JM: **Characterization of human phagocyte-derived chitotriosidase, a component of innate immunity.** *Int Immunol* 2005, **17(11)**:1505-1512.
- Rehli M, Niller HH, Ammon C, Langmann S, Schwarzfischer L, Andreesen R, Krause SW: **Transcriptional regulation of CHI3L1, a marker gene for late stages of macrophage differentiation.** *J Biol Chem* 2003, **278(45)**:44058-44067.
- Knorr T, Obermayr F, Bartnik E, Zien A, Aigner T: **YKL-39 (chitinase 3-like protein 2), but not YKL-40 (chitinase 3-like protein 1), is up regulated in osteoarthritic chondrocytes.** *Ann Rheum Dis* 2003, **62(10)**:995-998.
- Kzhyshkowska J, Mamidi S, Gratchev A, Kremmer E, Schmutzmaier C, Krusel L, Haus G, Utikal J, Schledzewski K, Scholtze J, Goerdts S: **Novel stabilin-1 interacting chitinase-like protein (SI-CLP) is up-regulated in alternatively activated macrophages and secreted via lysosomal pathway.** *Blood* 2006, **107(8)**:3221-3228.
- Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Sugang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q, Tunggal B, Kummerfeld S, Madera M, Konfortov BA, Rivero F, Bankier AT, Lehmann R, Hamlin N, Davies R, Gaudet P, Fey P, Pilcher K, Chen G, Saunders D, Sodergren E, Davis P, Kerhornou A, Nie X, Hall N, Anjard C, Hempill L, Bason N, Farbrother P, Desany B, Just E, Morio T, Rost R, Churcher C, Cooper J, Haydock S, van Driessche N, Cronin A, Good-

- head I, Muzny D, Mourier T, Pain A, Lu M, Harper D, Lindsay R, Hauser H, James K, Quiles M, Madan Babu M, Saito T, Buchrieser C, Wardroper A, Felder M, Thangavelu M, Johnson D, Knights A, Louseged H, Mungall K, Oliver K, Price C, Quail MA, Urushihara H, Hernandez J, Rabinowitsch E, Steffen D, Sanders M, Ma J, Kohara Y, Sharp S, Simmonds M, Spiegler S, Tivey A, Sugano S, White B, Walker D, Woodward J, Winckler T, Tanaka Y, Shaulsky G, Schleicher M, Weinstock G, Rosenthal A, Cox EC, Chisholm RL, Gibbs R, Loomis WF, Platzer M, Kay RR, Williams J, Dear PH, Noegel AA, Barrell B, Kuspa A: **The genome of the social amoeba Dictyostelium discoideum.** *Nature* 2005, **435(7038)**:43-57.
25. Clarke M, Maddera L: **Phagocytosis meets prey: uptake, internalization, and killing of bacteria by Dictyostelium amoebae.** *Eur J Cell Biol* 2006, **85(9-10)**:1001-1010.
 26. Aronson NN Jr, Backes M, Kuranda MJ: **Rat liver chitinase: purification, properties, and role in the lysosomal degradation of Asn-linked glycoproteins.** *Arch Biochem Biophys* 1989, **272(2)**:290-300.
 27. Muller I, Subert N, Otto H, Herbst R, Ruhling H, Maniak M, Leippe M: **A Dictyostelium mutant with reduced lysozyme levels compensates by increased phagocytic activity.** *J Biol Chem* 2005, **280(11)**:10435-10443.
 28. Mali B, Mohrlen F, Frohme M, Frank U: **A putative double role of a chitinase in a cnidarian: pattern formation and immunity.** *Dev Comp Immunol* 2004, **28(10)**:973-981.
 29. Frank U, Leitz T, Muller WA: **The hydroid Hydractinia: a versatile, informative cnidarian representative.** *Bioessays* 2001, **23(10)**:963-971.
 30. Hu B, Trinh K, Figueira WF, Price PA: **Isolation and sequence of a novel human chondrocyte protein related to mammalian members of the chitinase protein family.** *J Biol Chem* 1996, **271(32)**:19415-19420.
 31. Lok IH, Britton-Jones CM, Yuen PM, Haines CJ: **Variable expression of oviductin mRNA at different stages of human reproductive cycle.** *J Assist Reprod Genet* 2002, **19(12)**:569-576.
 32. Recklies AD, White C, Ling H: **The chitinase 3-like protein human cartilage glycoprotein 39 (HC-gp39) stimulates proliferation of human connective-tissue cells and activates both extracellular signal-regulated kinase- and protein kinase B-mediated signalling pathways.** *Biochem J* 2002, **365(Pt 1)**:119-126.
 33. Steck E, Breit S, Breusch SJ, Axt M, Richter W: **Enhanced expression of the human chitinase 3-like 2 gene (YKL-39) but not chitinase 3-like 1 gene (YKL-40) in osteoarthritic cartilage.** *Biochem Biophys Res Commun* 2002, **299(1)**:109-115.
 34. **WormBase** [<http://www.wormbase.org/>]
 35. **ACEView** [<http://www.ncbi.nlm.nih.gov/EB/Research/Acembly/>]
 36. **FlyBase** [<http://flybase.bio.indiana.edu/>]
 37. Popovici C, Roubin R, Coulier F, Pontarotti P, Birnbaum D: **The family of Caenorhabditis elegans tyrosine kinase receptors: similarities and differences with mammalian receptors.** *Genome Res* 1999, **9(11)**:1026-1039.
 38. Zhu Q, Deng Y, Vanka P, Brown SJ, Muthukrishnan S, Kramer KJ: **Computational identification of novel chitinase-like proteins in the Drosophila melanogaster genome.** *Bioinformatics* 2004, **20(2)**:161-169.
 39. Kawamura K, Shibata T, Saget O, Peel D, Bryant PJ: **A new family of growth factors produced by the fat body and active on Drosophila imaginal disc cells.** *Development* 1999, **126(2)**:211-219.
 40. Varela PF, Llera AS, Mariuzza RA, Tormo J: **Crystal structure of imaginal disc growth factor-2. A member of a new family of growth-promoting glycoproteins from Drosophila melanogaster.** *J Biol Chem* 2002, **277(15)**:13229-13236.
 41. Sun YJ, Chang NC, Hung SI, Chang AC, Chou CC, Hsiao CD: **The crystal structure of a novel mammalian lectin, Ym1, suggests a saccharide binding site.** *J Biol Chem* 2001, **276(20)**:17507-17514.
 42. Fusetti F, Pijning T, Kalk KH, Bos E, Dijkstra BW: **Crystal structure and carbohydrate-binding properties of the human cartilage glycoprotein-39.** *J Biol Chem* 2003, **278(39)**:37753-37760.
 43. Tsai ML, Liaw SH, Chang NC: **The crystal structure of Ym1 at 1.31 Å resolution.** *J Struct Biol* 2004, **148(3)**:290-296.
 44. Badariotti F, Kypriotou M, Lelong C, Dubos MP, Renard E, Galera P, Favrel P: **The phylogenetically conserved molluscan chitinase-like protein I (Cg-Clp1), homologue of human HC-gp39, stimulates proliferation and regulates synthesis of extracellular matrix components of mammalian chondrocytes.** *J Biol Chem* 2006, **281(40)**:29583-29596.
 45. Badariotti F, Thuau R, Lelong C, Dubos MP, Favrel P: **Characterization of an atypical family 18 chitinase from the oyster Crassostrea gigas: Evidence for a role in early development and immunity.** *Dev Comp Immunol* 2006, **31(6)**:559-70.
 46. Ohno S: **Gene duplication and the uniqueness of vertebrate genomes circa 1970-1999.** *Semin Cell Dev Biol* 1999, **10(5)**:517-522.
 47. Azumi K, De Santis R, De Tomaso A, Rigoutsos I, Yoshizaki F, Pinto MR, Marino R, Shida K, Ikeda M, Ikeda M, et al.: **Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: "waiting for Godot".** *Immunogenetics* 2003, **55(8)**:570-581.
 48. Flajnik MF, Kasahara M: **Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system.** *Immunity* 2001, **15(3)**:351-362.
 49. **JGI** [<http://genome.jgi-psf.org/Cioin2/Cioin2.home.html>]
 50. **JGI info** [<http://genome.jgi-psf.org/Cioin2/Cioin2.info.html>]
 51. **Sea Urchin Genome Resources** [http://www.ncbi.nlm.nih.gov/genome/guide/sea_urchin/]
 52. Burt D: **Chicken genome: current status and future opportunities.** *Genome Res* 2005, **15(12)**:1692-1698.
 53. Kumar S, Hedges SB: **A molecular timescale for vertebrate evolution.** *Nature* 1998, **392(6679)**:917-920.
 54. Woodburne MO, Rich TH, Springer MS: **The evolution of tribospheny and the antiquity of mammalian clades.** *Mol Phylogenet Evol* 2003, **28(2)**:360-385.
 55. Burge C, Karlin S: **Prediction of complete gene structures in human genomic DNA.** *J Mol Biol* 1997, **268(1)**:78-94.
 56. Kaufman J, Milne S, Gobel TW, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S: **The chicken B locus is a minimal essential major histocompatibility complex.** *Nature* 1999, **401(6756)**:923-925.
 57. Belov K, Deakin JE, Papenfuss AT, Baker ML, Melman SD, Siddle HV, Gouin N, Goode DL, Sargeant TJ, Robinson MD, Wakefield MJ, Mahony S, Cross JG, Benos PV, Samollow PB, Speed TP, Graves JA, Miller RD: **Reconstructing an ancestral mammalian immune supercomplex from a marsupial major histocompatibility complex.** *PLoS Biol* 2006, **4(3)**:e46.
 58. Pancer Z, Cooper MD: **The evolution of adaptive immunity.** *Annu Rev Immunol* 2006, **24**:497-518.
 59. Castro LF, Furlong RF, Holland PW: **An antecedent of the MHC-linked genomic region in amphioxus.** *Immunogenetics* 2004, **55(11)**:782-784.
 60. Yu C, Dong M, Wu X, Li S, Huang S, Su J, Wei J, Shen Y, Mou C, Xie X, et al.: **Genes "Waiting" for Recruitment by the Adaptive Immune System: The Insights from Amphioxus.** *J Immunol* 2005, **174**:3493-3500.
 61. Klein J, Nikolaidis N: **The descent of the antibody-based immune system by gradual evolution.** *Proc Natl Acad Sci USA* 2005, **102(1)**:169-174.
 62. Kawada M, Hachiya Y, Arihiro A, Mizoguchi E: **Role of mammalian chitinases in inflammatory conditions.** *Keio J Med* 2007, **56(1)**:21-27.
 63. Nair MG, Gallagher IJ, Taylor MD, Loke P, Coulson PS, Wilson RA, Maizels RM, Allen JE: **Chitinase and Fizz family members are a generalized feature of nematode infection with selective upregulation of Ym1 and Fizz1 by antigen-presenting cells.** *Infect Immun* 2005, **73(1)**:385-394.
 64. Arora M, Chen L, Paglia M, Gallagher I, Allen JE, Vyas YM, Ray A, Ray P: **Simvastatin promotes Th2-type responses through the induction of the chitinase family member Ym1 in dendritic cells.** *Proc Natl Acad Sci USA* 2006, **103(20)**:7777-7782.
 65. Gordon S: **The role of the macrophage in immune regulation.** *Res Immunol* 1998, **149(7-8)**:685-688.
 66. Desjardins M, Houde M, Gagnon E: **Phagocytosis: the convoluted way from nutrition to adaptive immunity.** *Immunol Rev* 2005, **207**:158-165.
 67. Fisher KJ, Aronson NN Jr: **Cloning and expression of the cDNA sequence encoding the lysosomal glycosidase di-N-acetylchitinase.** *J Biol Chem* 1992, **267(27)**:19607-19616.
 68. Malaguarnera L, Musumeci M, Di Rosa M, Scuto A, Musumeci S: **Interferon-gamma, tumor necrosis factor-alpha, and lipopolysaccharide promote chitotriosidase gene expression in human macrophages.** *J Clin Lab Anal* 2005, **19(3)**:128-132.

69. Welch JS, Escoubet-Lozach L, Sykes DB, Liddiard K, Greaves DR, Glass CK: **TH2 cytokines and allergic challenge induce YmI expression in macrophages by a STAT6-dependent mechanism.** *J Biol Chem* 2002, **277(45)**:42821-42829.
70. Webb DC, McKenzie AN, Foster PS: **Expression of the Ym2 lectin-binding protein is dependent on interleukin (IL)-4 and IL-13 signal transduction: identification of a novel allergy-associated protein.** *J Biol Chem* 2001, **276(45)**:41969-41976.
71. Delsuc F, Brinkmann H, Chourrout D, Philippe H: **Tunicates and not cephalochordates are the closest living relatives of vertebrates.** *Nature* 2006, **439(23)**:965-968.
72. Danchin E, Vitiello V, Vienne A, Richard O, Gouret P, McDermott MF, Pontarotti P: **The major histocompatibility complex origin.** *Immunol Rev* 2004, **198**:216-232.
73. Nonaka M, Namikawa C, Kato Y, Sasaki M, Salter-Cid L, Flajnik MF: **Major histocompatibility complex gene mapping in the amphibian *Xenopus* implies a primordial organization.** *Proc Natl Acad Sci USA* 1997, **94(11)**:5789-5791.
74. Nei M, Gu X, Sitnikova T: **Evolution by the birth-and-death process in multigene families of the vertebrate immune system.** *Proc Natl Acad Sci USA* 1997, **94(15)**:7799-7806.
75. Durand A, Hughes R, Roussel A, Flatman R, Henrissat B, Juge N: **Emergence of a subfamily of xylanase inhibitors within glycoside hydrolase family 18.** *FEBS J* 2005, **272(7)**:1745-1755.
76. Brandon C, Tooze J: **Introduction to Protein Structure.** Garland 1999.
77. McMahon SA, Miller JL, Lawton JA, Kerkow DE, Hodes A, Marti-Renom MA, Doulatov S, Narayanan E, Sali A, Miller JF, Ghosh P: **The C-type lectin fold as an evolutionary solution for massive sequence variation.** *Nat Struct Mol Biol* 2005, **12(10)**:886-892.
78. **Conserved Domain Database** [<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>]
79. **NCBI Entrez protein** [<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Protein>]
80. **Ensembl** [<http://www.ensembl.org/index.html>]
81. **UCSC Genome Browser** [<http://genome.ucsc.edu/index.html?org=human>]
82. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ, Weber RJ, Haussler D, Kent WJ: **The UCSC Genome Browser Database.** *Nucleic Acids Res* 2003, **31(1)**:51-54.
83. **NCBI Map View** [<http://www.ncbi.nlm.nih.gov/mapview/>]
84. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22(22)**:4673-4680.
85. Hall BG: **Phylogenetic Trees Made Easy.** 2nd edition. Sinauer; 2004.
86. Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Brief Bioinform* 2004, **5(2)**:150-163.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

