



# Comprehensive analysis of patched domain-containing genes reveals a unique evolutionary pattern

Y. Zhong\*, L.J. Gu\*, X.G. Sun, S.H. Yang and X.H. Zhang

State Key Laboratory of Pharmaceutical Biotechnology,  
School of Life Sciences, Nanjing University, Nanjing, China

\*These authors contributed equally to this study.

Corresponding author: X.H. Zhang

E-mail: xiaohuizhang@nju.edu.cn

Genet. Mol. Res. 13 (3): 7318-7331 (2014)

Received March 26, 2013

Accepted December 18, 2013

Published February 13, 2014

DOI <http://dx.doi.org/10.4238/2014.February.13.11>

**ABSTRACT.** Patched domain-containing genes are members of a small family originally identified in *Drosophila*. A common feature of transmembrane patched domain-containing proteins is their function in the transport of sterols, sterol-modified proteins, and lipids. Recently, an expansion phenomenon of patched domain-containing genes was found in *Caenorhabditis elegans*; the major contributor to this higher number was patched-related (PTR) type genes. However, little is known about their expansion pattern and evolutionary origin. We performed a systematic genome-wide survey of patched domain-containing genes in species ranging from protozoa to vertebrates, as well as some plants. We found that patched domain-containing genes are conserved in plant and animal genomes and their expansion likely occurred in the early stages of nematode speciation. Based on analysis of phylogenetic and reconciled trees and calculation of synonymous substitutions, we discovered that the PTR genes appear to have experienced two expansions within a relatively short period after the speciation of nematodes. We also found that some patched domain-containing genes possessing a relatively

recent evolutionary origin, such as PTR and PTCHD1, had fewer exons and shorter nucleotide coding sequence lengths compared with the older ones. It appears that the different types of patched domain-containing genes have different evolutionary patterns in different species.

**Key words:** Patched domain-containing genes; Gene family; Unique expansion pattern; Molting

## INTRODUCTION

Patched domain-containing genes are proposed to encode a 12 transmembrane (TM) protein, with five of the TMs comprising a 180-amino acid sterol-sensing domain (Carstea et al., 1997; Loftus et al., 1997), and they were initially identified in *Drosophila*, functioning in the patterning of anterior/posterior body segments and limbs (Ingham, 1998; Johnson and Scott, 1998). Subsequently, they have been identified in a number of species, and their roles appear to be conserved from fruit flies to vertebrates. A collective feature shared by patched domain-containing proteins is their function in the transport of sterols, sterol-modified proteins or lipids (Kuwabara and Labouesse, 2002). The TM protein patched is a key member of this family that upon binding of its ligand Hedgehog (Hh) activates the downstream effector Smoothed (Smo) (Hooper and Scott, 1989).

Proteins containing the patched domain in animals can be divided into five types: patched (PTC/PTCH), patched domain-containing (PTCHD), patched-related (PTR), dispatched (DISP), and human Niemann-Pick C disease (NPC1 and NPC1L1) (Kuwabara and Labouesse, 2002). Three PTC (PTC1, PTC2 and PTC3) homologous genes are found in *Caenorhabditis elegans*; it is known that PTC1 is functionally confined to germ line cytokinesis (Kuwabara et al., 2000) and that PTC3 is essential for osmoregulation (Soloviev et al., 2011). Other known functions of patched domain-containing proteins include DISP, which functions in the release of Hh from Hh producing cells (Burke et al., 1999), and NPC1, a protein with sequence similarity to PTC, which is involved in cholesterol trafficking (Carstea et al., 1997). Recently, it was reported that the inherited neurodevelopmental disorder autism is related to the X-chromosome gene PTCHD1 (Noor et al., 2010). The PTR group of patched domain-containing proteins was identified in *C. elegans* and was shown to be closely related to PTC in predicted topology (Kuwabara et al., 2000). Functional analysis of PTC, PTR and Hh-r in *C. elegans* has revealed that they are involved in molting, cytokinesis, growth, and pattern formation (Zugasti et al., 2005).

Much research has previously been done to reveal the diverse functions of patched domain-containing genes. In these studies, Zugasti et al. (2005) identified more patched domain-containing genes caused by expansion of PTR genes in *C. elegans* than in *Drosophila melanogaster*, *Mus musculus*, and *Homo sapiens*. However, no investigation about the expansion pattern, expansion time and evolutionary process of these genes has been reported. It is clear that the patched domain-containing gene family manifests lineage-specific expansion in nematodes, although this family is a small one in plants and animals from our analysis and a previous study (Zugasti et al., 2005). We know that lineage-specific expansion plays a considerable role in the organization and differentiation of the proteome of eukaryotes (Lespinet et al., 2002), and gene duplication is regarded as a principal driving force in adaptive evolution of the genome (Ohno, 1970). Duplicates are considered the raw materials and primary

mechanisms for the genesis of new genes and novel gene functions (Long and Langley, 1993). At least 8-20% of *D. melanogaster*, *C. elegans*, and *Saccharomyces cerevisiae* genomes are considered to be caused by gene duplication, and the rate of gene duplications in *C. elegans* is higher than that in *Drosophila* and in yeast (Lynch and Conery, 2000). Because little has been reported in regard to the expansion pattern and evolutionary pattern of these patched domain-containing genes, this study aimed to conduct a broad survey from protozoa to vertebrates and in several plant species at the genome-wide level. It was shown that patched domain-containing genes are conserved across plants and animals and emerged before the split of plants and animals. Also, the expansion of patched domain-containing genes likely occurred in the early stage of nematode speciation. According to our analysis of the phylogenetic tree and reconciled tree of patched domain-containing genes in 19 species and calculation of synonymous substitutions ( $K_s$ ) of PTR genes, we interpreted our data to indicate that PTR genes might have experienced two expansions in a relatively concentrated period after the speciation of nematodes. We also found that those genes possessing a relatively recent evolutionary origin, such as PTR and PTCHD1, had fewer exons and shorter nucleotide coding sequences (CDS) compared with the evolutionarily older ones. Our study suggests that genes of different types may have different evolutionary patterns in different species.

## MATERIAL AND METHODS

### Identification of patched domain-containing genes

The genome sequences of *Ostreococcus tauri*, *Oryza sativa*, *Arabidopsis thaliana*, *Amphimedon queenslandica*, *C. elegans*, *Drosophila melanogaster*, *Ciona intestinalis*, *Gallus gallus*, *Danio rerio*, *Monodelphis domestica*, *M. musculus*, *Macaca mulatta*, and *H. sapiens* were downloaded from the FTP site (<ftp://ftp.ncbi.nih.gov/genomes/>) of the National Center for Biotechnology Information (NCBI) and *Physcomitrella patens*, *Paramecium tetraurelia*, *Hydra magnipapillata*, *Nematostella vectensis*, *Strongylocentrotus purpuratus*, and *Xenopus tropicalis* were obtained from the DOE Joint Genome Institute (JGI, <http://genome.jgi.doe.gov/>). The four worm genomes of *Ascaris suum*, *Meloidogyne hapla*, *C. briggsae*, and *C. remanei* were retrieved from WormBase (<ftp://ftp.wormbase.org/pub/wormbase/>). The patched domain-containing genes in these genomes were identified using a standard patched domain obtained from the Pfam website (Punta et al., 2012) (<http://pfam.sanger.ac.uk/>) to BLAST against the genome sequences with an e value  $\leq 10$ . Only genes possessing a standard CDS, which meant that CDS lengths were integer multiples of three, could be recognized as candidates. These candidate proteins were then checked by Pfam analysis to predict the presence of a patched domain (PF02460); if this exact domain was confirmed in candidate proteins, it was then considered to belong to the patched domain-containing gene family.

### Sequence alignment and phylogenetic analysis

The amino acid sequences of the patched domain-containing proteins in 19 genomes were first aligned using the MUSCLE program with default options and then manually corrected using MEGA v5.0 (Tamura et al., 2011). The alignments were used to construct a neighbor-joining tree using ClustalW 2.0 (Larkin et al., 2007) with 1000 bootstrap replicates. The

protein sequences of patched domains from the three *Caenorhabditis* were aligned and a phylogenetic tree was constructed on the basis of the neighbor-joining method (Saitou and Nei, 1987) with a Kimura 2-parameter model by MEGA v5.0 with 1000 replicates. Phylogenetic trees were all displayed using MEGA v5.0.

### Reconciliation analysis

We constructed the reconciled tree on the PrIMETV website (<http://prime.sbc.su.se/primetv/>). Two kinds of trees, including a species tree and a phylogenetic tree of the gene family were needed for reconciliation analysis. The species tree was obtained from the NCBI-Taxonomy Common Tree (<http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi>) including the 19 species studied, while the phylogenetic tree for the gene family was ready-made by ClustalW 2.0. A reconciled tree could be easily attained from the illustrations of the PrIMETV website.

### Calculation of $K_s$

To understand the process of duplication in PTR genes,  $K_s$ , which are used to reflect the duplication time, were calculated by MEGA v5.0. CDSs of PTR genes in *C. elegans* were aligned using the MUSCLE program with default options in MEGA v5.0 according to alignments of protein sequences. The  $K_s$  were then computed pairwise on the basis of the multiple sequence alignments by MEGA v5.0.

### Physical location analysis

The exon numbers and CDS lengths of all patched domain-containing genes in our research were obtained on the basis of the genome annotations for each species by Practical Extraction and Report Language (Perl) scripts. The information regarding the physical position for the patched domain-containing genes in *C. elegans* was obtained according to its genome annotation, and then a figure for physical location was made using the GenomePixelizer software (Kozik et al., 2002).

## RESULTS

### Identification of the patched domain-containing gene family

To explore the origin of the expansion event, we identified patched domain-containing genes across a wide range of animal phyla, including lower animals (e.g., *P. tetraurelia* and *A. queenslandica*) and higher animals (e.g., *M. domestica*). Furthermore, the four plants *O. tauri*, *P. patens*, *O. sativa*, and *A. thaliana* were also analyzed. On the basis of the conserved patched domain (PF02460), we performed BLAST searches and Pfam analysis to investigate the patched domain-containing gene family at the genome-wide level. A total of 174 patched domain-containing genes, averaging 10.24 per genome, were detected in 17 representative animals, and 10 patched domain-containing genes were also identified in the four plant genomes (Table 1). Narrow ranges of 1 to 11 copies for the species, except *Caenorhabditis*, indicated

that the relatively low copies were sufficient for their survival requirements. Furthermore, unlike the known large number of patched domain-containing genes in *Caenorhabditis*, the genes of higher animals had even distributions for the seven protein types. For example, the protein type of DISp was absent in *Caenorhabditis* genomes, while it was present at low copy numbers in other species (1 to 2 per species). These results indicated that although patched domain-containing genes had different copy numbers in each species, they were still conserved across the plant and animal kingdoms and emerged before the split of plants and animals.

**Table 1.** Number of each protein type for patched-containing genes in eukaryotic genomes.

Species	Protein type							Total
	PTC/PTCH	PTCHD	PTR	DISp	NCR/NPC/NPC1L1	Hypothetical protein	Unclassified	
<i>Paramecium tetraurelia</i>	-*	-	-	-	-	-	3	3
<i>Amphimedon queenslandica</i>	0	0	0	0	1	0	-	1
<i>Hydra magnipapillata</i>	-	-	-	-	-	-	3	3
<i>Nematostella vectensis</i>	-	-	-	-	-	-	11	11
<i>Caenorhabditis elegans</i>	3	0	24	0	2	1	-	30
<i>Caenorhabditis briggsae</i>	2	0	24	0	2	1	-	29
<i>Caenorhabditis remanei</i>	2	0	24	0	2	0	-	28
<i>Drosophila melanogaster</i>	1	0	1	1	2	0	-	5
<i>Strongylocentrotus purpuratus</i>	0	0	1	2	2	0	-	5
<i>Ciona intestinalis</i>	1	0	0	1	3	1	-	6
<i>Danio rerio</i>	0	1	1	2	2	1	-	7
<i>Xenopus tropicalis</i>	2	0	1	1	1	0	-	5
<i>Gallus gallus</i>	0	2	1	2	1	1	-	7
<i>Monodelphis domestica</i>	2	2	1	1	2	1	-	9
<i>Mus musculus</i>	2	2	1	1	2	1	-	9
<i>Macaca mulatta</i>	1	1	1	1	2	1	-	7
<i>Homo sapiens</i>	2	2	1	1	2	1	-	9
<i>Ostreococcus tauri</i>	-	-	-	-	-	-	4	4
<i>Physcomitrella patens</i>	-	-	-	-	-	-	3	3
<i>Oryza sativa</i>	-	-	-	-	-	-	1	1
<i>Arabidopsis thaliana</i>	-	-	-	-	-	-	2	2

Blue branches represent animals, while green branches represent plants. \*No gene could be classified in the corresponding types.

We also identified 55 and 27 patched domain-containing genes in another two non-rhabditid nematodes, *A. suum* and *M. hapla*, respectively (data not shown). Thus, the results indicated that the expansion of patched domain-containing genes likely occurred in the early stage of nematode speciation and nematodes might have retained them to fulfill basic survival requirements.

### Phylogenetic analysis of the patched domain-containing family

To further investigate the evolutionary history and process of patched domain-containing genes, we constructed an unrooted phylogenetic tree based on the amino acid sequences of the core patched domain from the 19 species studied (Figure 1 and Figure S1). In Table 1, 27 patched domain-containing genes could not be classified into distinct types because there was no related literature. The genes from lower animals and plants may represent a more ancient situation of patched domain-containing genes. However, some of them could be easily distinguished according to the phylogenetic relationship in our phylogenetic tree. The phylogenetic tree could be divided into 7 clades representing the corresponding types of patched domain-



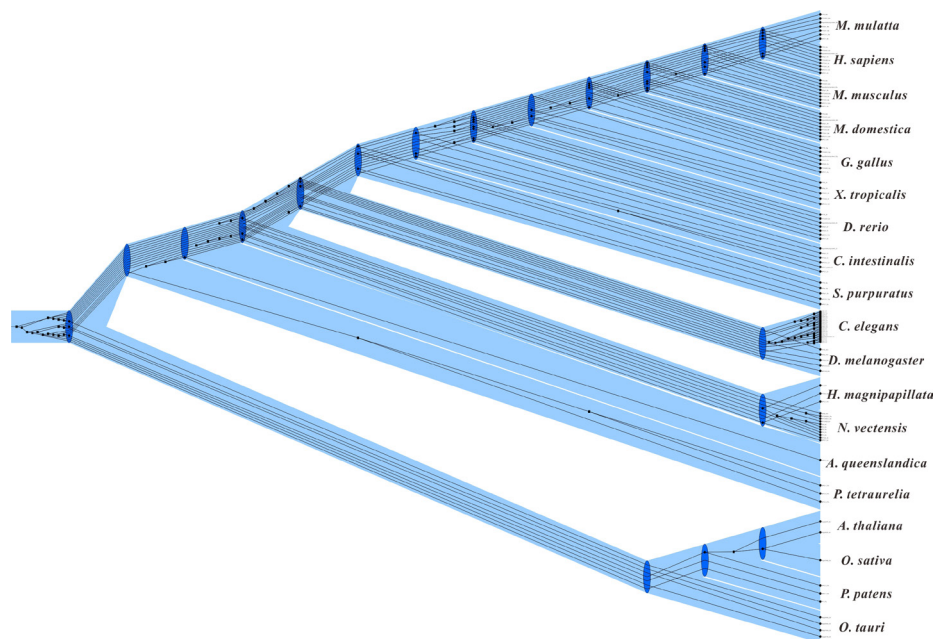


*elegans* and 9 from other animals, while subclade 2 was *C. elegans* specific, including 15 members. From this, we inferred that the PTR gene family might have experienced two expansions after the speciation of nematodes necessary for adaption to their environment or some other physiological need, while other higher animals only reserved one copy in their own genomes.

### Construction of the reconciled tree

To further confirm our inference, we constructed a reconciled tree, which could be viewed as a gene-family tree evolving inside another species tree (Sennblad et al., 2007).

At the bottom of the reconciled tree (Figure 2), 19 independent blue branches represented the 19 species in our study, and the black dots indicated patched domain-containing genes. After the first speciation node in the tree, two distinct clades, plants and animals, were recognized. It was very clear that there were many more genes in the branch of *C. elegans* than other species. This tree illustrated the evolutionary details of patched domain-containing genes concerning gene duplication and distribution in *C. elegans*. The two types, PTC/PTCH and NCR/NPC1/NPC1L1, were present as a single copy before the occurrence of *C. elegans*, while during the speciation of *C. elegans*, they duplicated respectively and engendered three PTC/PTCH genes (PTC1/PTC2/PTC3) and two NCR/NPC1/NPC1L1 genes (NCR1/NCR2). Conversely, two ancestral genes duplicated into six copies before speciation of *C. elegans*, and then, all of them were kept as ancestors of the PTR gene in the *C. elegans* genome, except one which was the ancestral gene for DISp-type genes.



**Figure 2.** Reconciled tree of patched-containing genes in 19 species. The species tree is presented by the blue profile and the gene-family tree is shown by the black lines. Blue circles mean the nodes of speciation, and the black dots indicate gene-family tree vertices corresponding to speciation, while the black squares denote gene duplication events.

Three of the six copies were conserved during the appearance of higher species and were the incarnation of DISp, PTCHD1/Hypothetical protein and PTR genes found in higher animals. Within the period of *C. elegans* speciation, we could easily distinguish that five of the six copies duplicated 20 times (20 black squares) at different times and finally led to the PTR expansion. Furthermore, it was clear that these duplication squares distinctly separated into two groups, which corresponded to the two subclades described in the previous section (Figure 1). Therefore, few gene duplications occurred before speciation of *C. elegans*, and then after other species split, there were two expansion events of PTR genes, which resulted in the expansion of patched domain-containing genes in *C. elegans*, and in nematodes in general.

To explore the expansion time of PTR genes, we calculated the *Ks* value of PTR genes in *C. elegans*. It was clear that there were two continuous peaks in the *Ks* range of 2 to 3 (Figure S2), indicating that the PTR genes in *C. elegans* experienced two duplications in a relatively concentrated period. We also calculated the *Ks* values of PTR genes of the two subclades in *C. elegans* (Figure 1, clade G), 2.43 and 2.87 were got from sub-clade 1 and sub-clade 2, respectively, which demonstrated that the duplication of genes in subclade 2 might have occurred earlier than in subclade 1 containing the PTR genes from other animals. In addition, the two *Ks* values were consistent with the two duplication peaks (Figure S2), further clarifying that PTR genes had experienced two duplications, which accorded with the conclusion gotten from the phylogenetic and reconciled trees.

### Patterns of exon evolution

To better understand the evolutionary pattern of patched domain-containing genes, we surveyed the exon number and CDS length of all patched domain-containing genes identified in this study (Table 2). The exon number ranged from 4 to 20 in *C. elegans* (Table S1). PTC2 had the smallest exon number and the shortest CDS length, maybe because it is a pseudogene and was also recently duplicated (Kuwabara et al., 2000). For the more ancient genes, such as NCR1, NCR2 and PTC3, we found more exons and longer CDS lengths in *C. elegans*. The phenomenon that younger genes had fewer exons than older ones was more evident in the higher animals. On the basis of the phylogenetic tree (Figure 1 and Figure S1), we noted that PTC/PTCH, NPC/NPC1/NPC1L1, and PTCHD2 had an ancient evolutionary origin, whereas PTCHD1, Hypothetical protein, PTR, and DISp had a relatively recent origin. To determine the distribution of exon numbers and CDS lengths, we calculated their median values across each gene type (Table 2 and Table S2). The median values of the exon numbers for those older genes, PTC/PTCH (PTCH1/PTCH2), PTCHD2, and NCR/NPC1/NPC1L1, fell in the range of 19.5 to 22.5. The lower exon numbers of these genes in *C. elegans* and *D. melanogaster* could be explained by prior research, which described that nematodes and fruit flies have significant intron loss (Kiontke et al., 2004; Putnam et al., 2007). In contrast to the median exon number of the more ancient genes, the distribution of exon numbers for the younger genes, PTCHD1, Hypothetical protein, PTR, and DISp, was significantly lower (*t*-test,  $P < 0.01$ ). Furthermore, the distribution of CDS lengths for younger genes (~2600 bp) was significantly lower (*t*-test,  $P < 0.01$ ) when compared with the distribution for older genes (3834 to 4104 bp), except for DISp genes. We interpret these data as indicating that the evolutionary pattern may be different between young and old genes, which is reflected in the different exon numbers and CDS lengths of patched domain-containing genes.



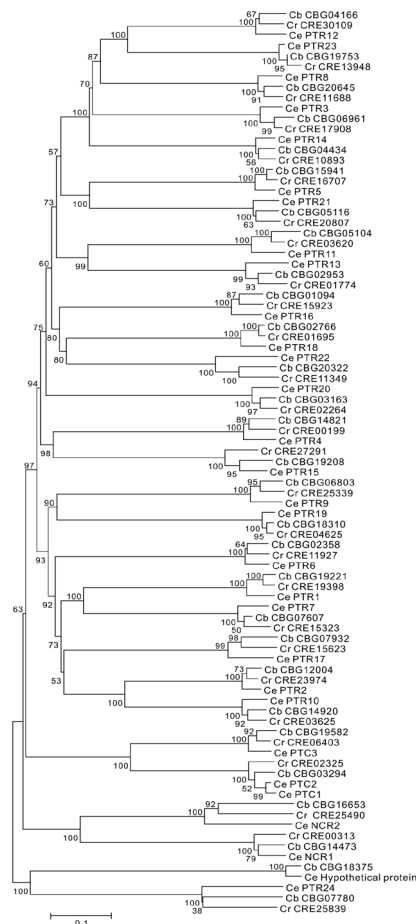
**Table 2.** Exon number and CDS length for each type of patched-containing gene.

Species	PTC/PTCH		PTCHD				PTR		DISp		NCR/NPC1/ NPC1L1		Hypothetical protein	
	Number*	Length†	PTCHDI		PTCHD2		Number	Length	Number	Length	Number	Length	Number	Length
			Number	Length	Number	Length								
<i>O. tauri</i>	-	-	-	-	3	4350	-	-	-	-	-	-	-	-
<i>P. patens</i>	-	-	-	-	-	-	-	-	-	-	27‡	3609‡	-	-
<i>O. sativa</i>	-	-	-	-	-	-	-	-	-	-	38	3519	-	-
<i>A. thaliana</i>	-	-	-	-	-	-	-	-	-	-	37‡	3762‡	-	-
<i>P. tetraurelia</i>	-	-	-	-	-	-	-	-	-	-	7‡	2739‡	-	-
<i>A. queenslandica</i>	-	-	-	-	-	-	-	-	-	-	16	4134	-	-
<i>H. magnipapillata</i>	7	1173	-	-	-	-	-	-	-	10	2763	-	-	-
<i>N. vectensis</i>	25	3360	-	-	12‡	3295‡	-	3	2433	15	3834	-	-	-
<i>C. elegans</i>	9‡	4104‡	-	-	-	-	12.5‡	-	-	18‡	3988.5‡	-	19	3681
<i>D. melanogaster</i>	6	3861	-	-	-	-	9	3510	7	3657	8.5‡	3862.5‡	-	-
<i>S. purpuratus</i>	-	-	-	-	-	-	3	2529	5‡	3654‡	18.5‡	3058.5‡	-	-
<i>C. intestinalis</i>	27	4359	-	-	-	-	-	6	3147	16.5‡	2379.5‡	-	6	1131
<i>X. tropicalis</i>	23.5‡	4332‡	-	-	-	-	-	-	7	1872	19	4263	-	-
<i>G. gallus</i>	-	-	3	2349	22	4374	4	2400	7	1872	7.5‡	4471.5‡	-	-
<i>D. rerio</i>	-	-	3	2664	-	-	5	1872	8‡	4413‡	25	3786	3	2664
<i>M. domestica</i>	21.5‡	3895.5‡	4	2982	20	4158	4	2610	7	4617	22‡	3823.5‡	3	2694
<i>M. musculus</i>	22.5‡	3927‡	3	2667	20	4044	6	2898	7	4566	22‡	3919.5‡	3	2853
<i>M. mulatta</i>	27	4656	3	2667	-	-	4	2721	7	4566	22‡	3918‡	3	2664
<i>H. sapiens</i>	23‡	3976.5‡	3	2667	20	4179	5	2799	7	4572	22.5‡	3886.5‡	4	2478
Median	22.5	4104	3	2667	20	4044	4	2304	7	4575	22.5‡	3958.5‡	3	2490
					20	4044	10	2703	7	4354.5	19.5	3834	3	2664

\*Exon number. †CDS length (bp). ‡Medians of the genes for the corresponding types in each species.

### Analysis of the patched domain-containing genes in three *Caenorhabditis* species

To analyze the particular evolutionary pattern of patched domain-containing genes in worms, we constructed another phylogenetic tree using the patched domain-containing genes only in *C. elegans*, *C. briggsae* and *C. remanei*. This phylogenetic tree was constructed with MEGA v5.0 using the amino acid sequences of the core patched domains. Apart from PTC2 and the Hypothetical protein genes, all other 28 genes had clearly orthologous relationships between the three *Caenorhabditis* species (Figure 3), indicating that the different types of patched domain-containing genes were conserved and generated at least before the three *Caenorhabditis* species split. It was clear from this tree that genes of *C. briggsae* were apparently more closely related to *C. remanei* than to *C. elegans*. Indeed, the two species *C. briggsae* and *C. remanei* are known to have a closer genetic relationship and are thought to be sister species (Kiontke et al., 2004). In addition, two clades with high bootstrap support which were composed of NCR and PTR24, respectively, were located outside with longer branch lengths, which suggested that they were more ancient than other genes.



**Figure 3.** Phylogenetic tree based on the core patched domains in three *Caenorhabditis* species. Ce = *C. elegans*; Cb = *C. briggsae* and Cr = *C. remanei*.

To further analyze the genetic evolutionary events in the genome of worms, the physical positions of patched domain-containing genes were deduced across the six chromosomes and illustrated by the GenomePixelizer software ([Figure S3](#)). This analysis was only performed in *C. elegans* because it has intact chromosome information available. It is clear that patched domain-containing family members were scattered throughout all six chromosomes of the *C. elegans* genome, without any clustering. However, there were significantly more patched domain-containing genes (*t*-test,  $P < 0.05$ ) located on chromosomes I and II (7 and 9) than those on other chromosomes (3, 1, 4, and 5). To test whether segmental duplications contributed to the expansion of patched genes in *C. elegans*, the method used by Schauser et al. (2005) was employed to investigate the evolutionary relationship between duplicated segments; however, no evidence of segmental duplication could be found between the gene pairs analyzed.

## DISCUSSION

### Duplication pattern of patched domain-containing genes in nematodes

Lineage-specific expansion is defined as the proliferation of a protein family in a particular lineage compared with other lineages. Although the patched domain-containing gene family is a small one in plants and animals, it was clear from our analysis that this family expanded in a lineage-specific manner in nematodes. Previous research has found that gene duplications are very widespread and more frequently observed in *C. elegans* than in either *Drosophila* or yeast (Katju and Lynch, 2003). A gene duplication event is an important source of raw materials for the genesis of new genes in eukaryotic genomes, particularly in the origin of multi-gene families (Lespinet et al., 2002). It can occur via three major mechanisms: unequal crossing over, transposition and chromosomal (or genome) duplication, which leads to tandem gene duplication, dispersed duplication and large-scale duplication (segmental duplication or whole genome duplication), respectively (Zhang, 2003).

In our study, physical location and the method of Schauser et al. (2005) were used to investigate the duplication pattern of the patched domain-containing genes. Our results showed that no segmental duplication could be detected by BLAST search in an intra-chromosomal or inter-chromosomal fashion. Large-scale duplications (chromosomal or genome duplications) occur frequently in plants but infrequently in animals (Li et al., 2001). Furthermore, regional duplications were not present in *C. elegans*, except three small regional duplications (Semple and Wolfe, 1999), and *C. elegans* does not appear to have experienced a whole-genome duplication event because only very few block duplication events have been detected (Friedman and Hughes, 2001; Cavalcanti et al., 2003). Our finding of no duplication regions could be explained by the possibility that a high rate of deletions or genomic rearrangements could have eliminated the evidence for regional duplications (Semple and Wolfe, 1999), and thus, another mechanism may be more suitable for explaining our results. The large number of physically separated genes found in *C. elegans* suggested that these genes might have duplicated through transposition rather than unequal crossover. As previously reported, dispersed duplication caused by transposition is defined as duplications with intervening genes or genes located on different chromosomes (Zhou et al., 2008). Another explanation leaning towards dispersed duplication is that tandem duplications could become dispersed during subsequent mutation events (Zhou et al., 2008), and it has also been shown that new constrained functional genes in *Drosophila*

were mostly dispersed duplicates. Taken together, we conclude that dispersed duplication could be the predominant mechanism for generating patched domain-containing genes in *C. elegans*.

### Functional roles for the expanded patched domain-containing genes

Duplication occurs in an individual and can either be fixed or lost in the genome. Previous research has suggested that many duplicated genes will be lost (Kimura, 1979), but for fixed duplicates, their evolutionary fate will be influenced by their own functions. Three alternative theories have been proposed to illustrate their fate: nonfunctionalization/pseudogenization, neofunctionalization and subfunctionalization.

It is known that lineage-specific expansion of gene families played an important part in the growth and differentiation of the proteomes of eukaryotes (Lespinet et al., 2002). A previous study and our data both showed that nematode genomes had an obvious expansion of PTR genes relative to the one single PTR gene found in other animals (Zugasti et al., 2005). Thus, the question arises: why did worms duplicate and conserve so many PTR genes in their genome? Species-specific gene duplication can lead to species-specific gene function, which could facilitate species-specific adaptation (Zhang et al., 2002). Although little cholesterol is needed to form its membrane structure, *C. elegans* obtains sterol from exogenous food, because it is a cholesterol auxotrophic organism and lacks enzymes for synthesizing sterol. It has been suggested that the principal role of cholesterol in *C. elegans* may be in membrane structure and for signaling functions in molting and a specialized non-feeding larval stage. Animals complete their metamorphosis and become imagoes via molting, such as the stable fly, *Stomoxys calcitrans*, fruit fly, silkworm, and crustaceans (King and Siddall, 1969), where their molting processes are all controlled by the steroidal hormone  $\beta$ -ecdysone. While in nematodes, ecdysone does not act as a molting hormone, because no ecdysteroids have been discovered in any free-living nematode (Chitwood, 1999) and no ecdysone receptor homologue has been found in the *C. elegans* genome (Kurzychalia and Ward, 2003). Rather than using ecdysone, in *C. elegans* a sterol-derived hormone is involved in larval formation and molting (Matyash et al., 2004), which has been related to many nematode-specific genes, including PTR4 and PTR23 (Frand et al., 2005). During the process of molting in *C. elegans*, it synthesizes and secretes a new exoskeleton, which is a collagenous extracellular matrix distinct from the exoskeleton of the fruit fly underneath the old one (Frand et al., 2005). It is also known that the PTR and Hh genes of *C. elegans* are involved in cell growth, patterning and molting, depending on sterols (Zugasti et al., 2005). Taken together, these studies may indicate that the expansion of PTR genes facilitate nematode-specific adaptation.

In addition to the PTR genes, NCR1 and NCR2 are also involved in sterol metabolism (Li et al., 2004). Additionally, some patched domain-containing genes might have experienced neofunctionalization, which is thought to be one of the most important outcomes of gene duplication. For example, PTR7/DAF6 is required for lumen formation (Perens and Shaham, 2005) and PTC1 is fundamentally confined to functioning in germ-line cytokinesis (Kuwabara et al., 2000).

### ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (#30970198, #30930008 and #J1103512) and the Qing Lan Project (grant awarded to S.H. Yang).

[Supplementary material](#)**REFERENCES**

- Burke R, Nellen D, Bellotto M, Hafen E, et al. (1999). Dispatched, a novel sterol-sensing domain protein dedicated to the release of cholesterol-modified hedgehog from signaling cells. *Cell* 99: 803-815.
- Carstea ED, Morris JA, Coleman KG, Loftus SK, et al. (1997). Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis. *Science* 277: 228-231.
- Cavalcanti AR, Ferreira R, Gu Z and Li WH (2003). Patterns of gene duplication in *Saccharomyces cerevisiae* and *Caenorhabditis elegans*. *J. Mol. Evol.* 56: 28-37.
- Chitwood DJ (1999). Biochemistry and function of nematode steroids. *Crit. Rev. Biochem. Mol. Biol.* 34: 273-284.
- Frand AR, Russel S and Ruvkun G (2005). Functional genomic analysis of *C. elegans* molting. *PLoS Biol.* 3: e312.
- Friedman R and Hughes AL (2001). Gene duplication and the structure of eukaryotic genomes. *Genome Res.* 11: 373-381.
- Hooper JE and Scott MP (1989). The *Drosophila* patched gene encodes a putative membrane protein required for segmental patterning. *Cell* 59: 751-765.
- Ingham PW (1998). The patched gene in development and cancer. *Curr. Opin. Genet. Dev.* 8: 88-94.
- Johnson RL and Scott MP (1998). New players and puzzles in the Hedgehog signaling pathway. *Curr. Opin. Genet. Dev.* 8: 450-456.
- Katju V and Lynch M (2003). The structure and early evolution of recently arisen gene duplicates in the *Caenorhabditis elegans* genome. *Genetics* 165: 1793-1803.
- Kimura M (1979). The neutral theory of molecular evolution. *Sci. Am.* 241: 98-100, 102, 108.
- King DS and Siddall JB (1969). Conversion of alpha ecdysone to beta ecdysone by crustaceans and insects. *Nature* 221: 955-956.
- Kiontke K, Gavin NP, Raynes Y, Roehrig C, et al. (2004). *Caenorhabditis* phylogeny predicts convergence of hermaphroditism and extensive intron loss. *Proc. Natl. Acad. Sci. U. S. A.* 101: 9003-9008.
- Kozik A, Kochetkova E and Michelmore R (2002). GenomePixelizer - a visualization program for comparative genomics within and between species. *Bioinformatics* 18: 335-336.
- Kurzchalia TV and Ward S (2003). Why do worms need cholesterol? *Nat. Cell. Biol.* 5: 684-688.
- Kuwabara PE and Labouesse M (2002). The sterol-sensing domain: multiple families, a unique role? *Trends Genet.* 18: 193-201.
- Kuwabara PE, Lee MH, Schedl T and Jefferis GS (2000). A *C. elegans* patched gene, *ptc-1*, functions in germ-line cytokinesis. *Genes Dev.* 14: 1933-1944.
- Larkin MA, Blackshields G, Brown NP, Chenna R, et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947-2948.
- Lespinet O, Wolf YI, Koonin EV and Aravind L (2002). The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res.* 12: 1048-1059.
- Li J, Brown G, Ailion M, Lee S, et al. (2004). NCR-1 and NCR-2, the *C. elegans* homologs of the human Niemann-Pick type C1 disease protein, function upstream of DAF-9 in the dauer formation pathways. *Development* 131: 5741-5752.
- Li WH, Gu Z, Wang H and Nekrutenko A (2001). Evolutionary analyses of the human genome. *Nature* 409: 847-849.
- Loftus SK, Morris JA, Carstea ED, Gu JZ, et al. (1997). Murine model of Niemann-Pick C disease: mutation in a cholesterol homeostasis gene. *Science* 277: 232-235.
- Long M and Langley CH (1993). Natural selection and the origin of jingwei, a chimeric processed functional gene in *Drosophila*. *Science* 260: 91-95.
- Lynch M and Conery JS (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151-1155.
- Matyash V, Entchev EV, Mende F, Wilsch-Brauninger M, et al. (2004). Sterol-derived hormone(s) controls entry into diapause in *Caenorhabditis elegans* by consecutive activation of DAF-12 and DAF-16. *PLoS Biol.* 2: e280.
- Noor A, Whibley A, Marshall CR, Gianakopoulos PJ, et al. (2010). Disruption at the PTCHD1 locus on Xp22.11 in autism spectrum disorder and intellectual disability. *Sci. Transl. Med.* 2: 49ra68.
- Ohno S (1970). Evolution by Gene Duplication. Springer-Verlag, London: George Alien & Unwin Ltd. Berlin, Heidelberg and New York.
- Perens EA and Shaham S (2005). *C. elegans* *daf-6* encodes a patched-related protein required for lumen formation. *Dev. Cell.* 8: 893-906.
- Punta M, Cogill PC, Eberhardt RY, Mistry J, et al. (2012). The Pfam protein families database. *Nucleic Acids Res.* 40: D290-D301.
- Putnam NH, Srivastava M, Hellsten U, Dirks B, et al. (2007). Sea anemone genome reveals ancestral eumetazoan gene



- repertoire and genomic organization. *Science* 317: 86-94.
- Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.
- Schauser L, Wieloch W and Stougaard J (2005). Evolution of NIN-like proteins in *Arabidopsis*, rice, and *Lotus japonicus*. *J. Mol. Evol.* 60: 229-237.
- Semple C and Wolfe KH (1999). Gene duplication and gene conversion in the *Caenorhabditis elegans* genome. *J. Mol. Evol.* 48: 555-564.
- Sennblad B, Schreil E, Berglund Sonnhammer AC, Lagergren J, et al. (2007). primetv: a viewer for reconciled trees. *BMC Bioinformatics* 8: 148.
- Soloviev A, Gallagher J, Marnef A and Kuwabara PE (2011). *C. elegans* patched-3 is an essential gene implicated in osmoregulation and requiring an intact permease transporter domain. *Dev. Biol.* 351: 242-253.
- Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.
- Zhang J (2003). Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18: 292-298.
- Zhang J, Zhang YP and Rosenberg HF (2002). Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat. Genet.* 30: 411-415.
- Zhou Q, Zhang G, Zhang Y, Xu S, et al. (2008). On the origin of new genes in *Drosophila*. *Genome Res.* 18: 1446-1455.
- Zugasti O, Rajan J and Kuwabara PE (2005). The function and expansion of the Patched- and Hedgehog-related homologs in *C. elegans*. *Genome Res.* 15: 1402-1410.