The Hedgehog signaling domain was acquired from a prokaryote.

Henk Roelink.
Department of Molecular and Cell Biology, University of California at Berkeley, USA
roelink@berkeley.edu

Abstract

The presence of well-conserved N-terminal Hh domains in several bacterial strains indicates that the metazoan Hedgehog (Hh) gene was acquired via horizontal gene transfer from bacteria. The bilaterian and cnidarian Hh proteins are rooted within the bacterial clade, while the cnidarian hedgling protein forms the outgroup. This relationship indicates that the N-terminal signaling part of the Hh proteins was acquired from a bacterium, and not from the hedgling protein.

Introduction

Hedgehog genes play central roles during development of most bilaterians. Hedgehog proteins are synthesized as pro-proteins, which are characterized by two distinct domains, and N-terminal domain which is retained as the ligand (the "Hedge" domain), and a C-terminal domain which is required for the autoproteolytic processing (the "Hog" domain). The Hedge domain (also referred to as Hh-N) has all characteristics of a Zn^{2+} metalloprotease [1], although no proteolytic target of it has been identified. The Hog domain resembles that of inteins, self-excising proteins that re-join the adjacent peptides [2].

Both the Hedge and Hog domains are modular. The Hedge domain can be found as the extreme N-terminal end of large Cadherin-like membrane proteins in Choanoflagellates, Sponges, and Cnidarians [3], members of the Hedgling family. Cnidarians also have Hh proteins, as well as intein-like Hog sequence, leading to the hypothesis that Hh genes were a rearrangement of the N-domain of Hedgling with a Hog/intein domain from an unknown protein [4].

Results

Here I propose an alternate hypothesis for the origin of the Hh protein. By performing a protein BLAST in prokaryotic genomes using Shh-N as the query sequence, I found that highly conserved Hh-N domains are coded for in several bacterial genomes (Figure 1 A). A BLAST phylogenetic tree (nearest neighbor joining) of the Hh-N containing proteins including bacterial, cnidarian, insect and mammalian hedgehog proteins showed that all metazoan Hh proteins are rooted within the bacterial clade. Cnidarian (Nematostella) hedgling is the outgroup (Figure 1B). In all instances, the Hh-N domain is part of a larger protein. The Hh-N domains in bacteria are located at the C-terminal ends of larger proteins, all with unknown function. In contrast, in metazoans the Hh-N domain is the amino-terminal part of both hedgling and the Hh pro-proteins (Figure 1 A).

Performing the alignment with only the Hh-N domain (the conserved block in figure 1A), did not change the overall interpretation. Again, the metazoan Hh proteins rooted within the bacterial clade, with the cnidarian hedgling N-domain being the outgroup (Figure 2). The hog domain is widely distributed among prokaryotes and eukaryotes including fungi, plants, and algae. Nevertheless, the eukaryotic hog domains are more similar to each other than to the bacterial hog domains. This indicates that a horizontally transferred bacterial *Hh-N* domain was recombined with a eukaryotic *Hog* domain to give rise to the ancestral *Hh* gene.

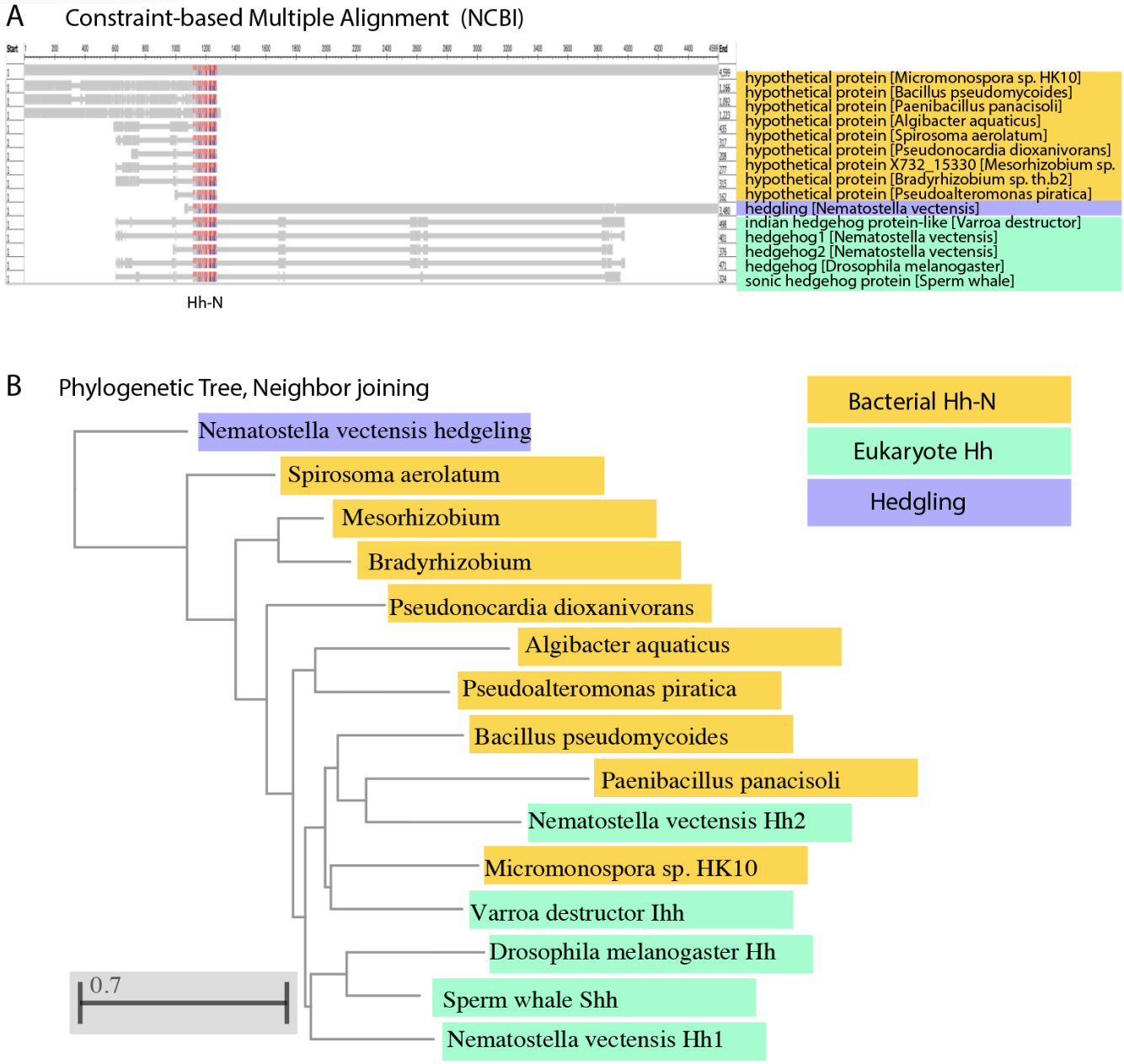

Figure 1. Hh-N containing proteins are present both in prokaryotes and eukaryotes. A: Alignment of several Hh-N containing proteins from bacteria (amber), metazoans (green) and the cnidarian hedgling protein (purple). B: The phylogenetic tree derived from the lineup in A. Note that all eukaryotic Hh-N containing proteins are rooted within the bacterial clade, except Nematostella hedgling.

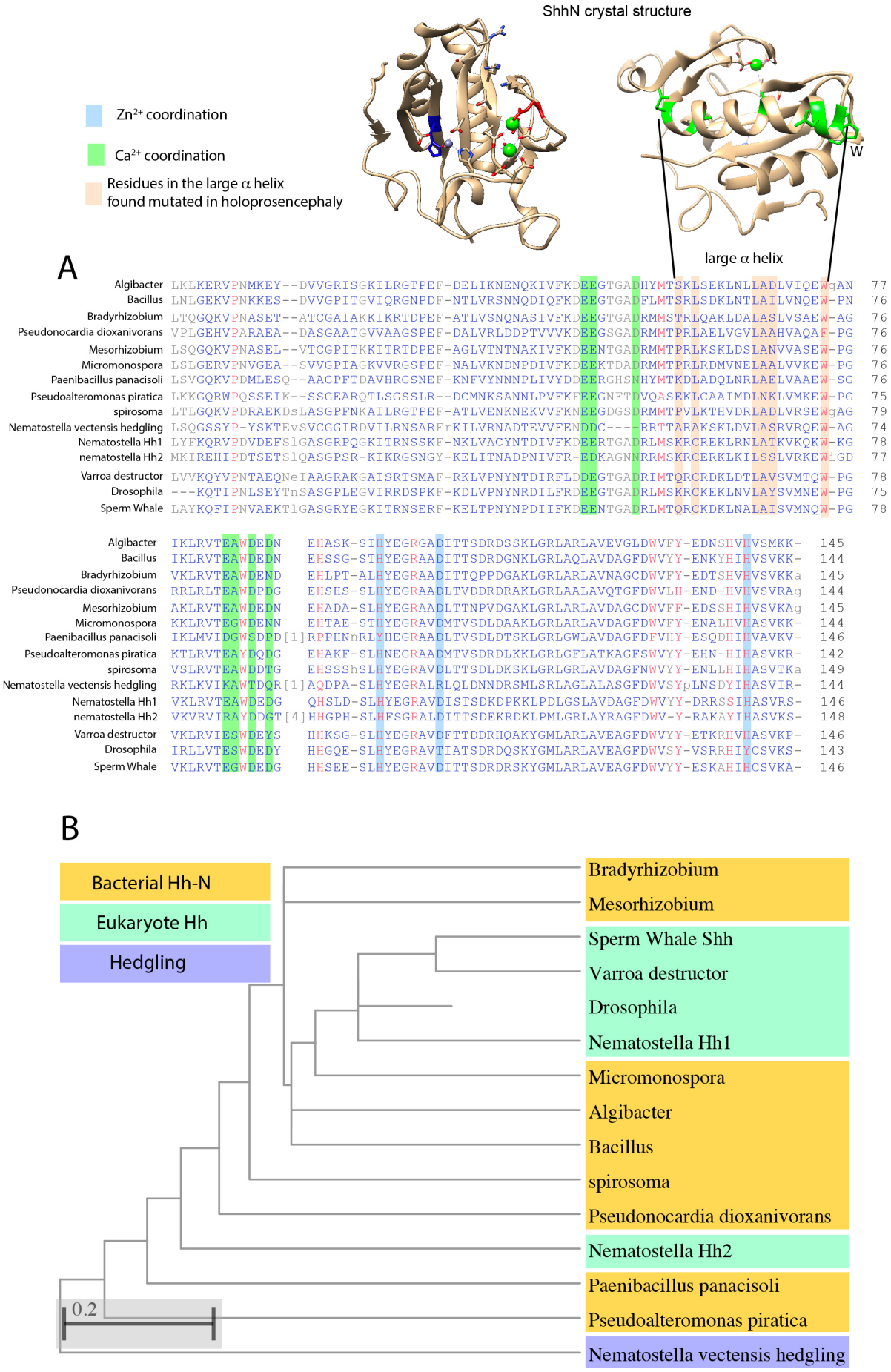
ShhN crystal structure
Zn2+ coordination
Ca2+ coordination
Residues in the large α helix found mutated in holoprosencephaly
W
large α helix
A
Algibacter LKLKERVPNMKEY--DVVGRISGKILRGTPEF-DELIKNENQKIVFKDEEGTGADHYMTSKLSEKLNLLADLVIQEWgAN 77
Bacillus LNLGEKVPNKKES--DVVGPITGVIQRGNPDF-NTLVRSNNQDIQFKDEEGTGADFLMTSRLSDKLNTLAILVNQEW-PN 76
Bradyrhizobium LTQGQKVPNASET--ATCGAIAKKIKRTDPEF-ATLVSNQNASIVFKDEEGTGADRMMSTRLQAKLDALASLVSAEW-AG 76
Pseudonocardia dioxanivorans VPLGEHVPARAEA--DASGAATGVVAAGSPEF-DALVRLDDPTVVVKDEEGSGADRMMTPRLAELVGVLAAHVAQAF-PG 76
Mesorhizobium LSQGQKVPNASEL--VTCGPITKKITRTDPEF-AGLVTNTNAKIVFKDEENTGADRMMTPRLKSKLDSLANVVASEW-PG 76
Micromonospora LSLGERVPNVGEA--SVVGPIAGKVVRGSPEF-NALVKNDNPDIVFKDEEKTDADRMMTPRLRDMVNELAALVVKEW-PG 76
Paenibacillus panacisoli LSVGQKVPDMLESQ--AAGPFTDAVHRGSNEF-KNFVYNNNPLIVYDDEERGHSNHYMTKDLADQLNRLAELVAAEW-SG 76
Pseudoalteromonas piratica LKKGQRWPQSSEIK--SSGEARQTLSGSSLR--DCMNKSANNLPVFKFEEGNFTDVQASEKLCAAIMDLNKLVMKEW-PG 75
spirosoma LTLGQKVPDRAEKDsLASGPFNKAILRGTPEF-ATLVENKNEKVVFKNEEGDGSDRMMTPVLKTHVDRLADLVRSEWgAG 79
Nematostella vectensis hedgling LSQGSSYP-YSKTEvSVCGGIRDVILRNSARFrKILVRNADTEVVFENDDC---RRTTARAKSKLDVLASRVRQEW-AG 74
Nematostella Hh1 LYFKQRVPDVDEFSlGASGRPQGKITRNSSKF-NKLVACYNTDIVFKDEERTGADRLMSKRCREKLRNLATKVKQKW-KG 78
nematostella Hh2 MKIREHIPDTSETSlQASGPSR-KIKRGSNGY-KELITNADPNIVFR-EDKAGNNRRMSKRCERKLKILSSLVRKEWiGD 77
Varroa destructor LVVKQYVPNTAEQNeIAAGRAKGAISRTSMAF-RKLVPNYNTDIRFLDDEGTGADRIMTQRCRDKLDTLAVSVMTQW-PG 78
Drosophila ---KQTIPNLSEYTnSASGPLEGVIRRDSPKF-KDLVPNYNRDILFRDEEGTGADRLMSKRCKEKLNVLAYSVMNEW-PG 75
Sperm Whale LAYKQFIPNVAEKTlGASGRYEGKITRNSERF-KELTPNYNPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQW-PG 78
Algibacter IKLRVTEAWDEDN EHASK-SIHYEGRGADITTSDRDSSKLGRLARLAVEVGLDWVFY-EDNSHVHVSMKK- 145
Bacillus IKLRVTEAWDEDN EHSSG-STHYEGRAADITTSDRDGNKLGRLAQLAVDAGFDWVYY-ENKYHIHVSVKK- 144
Bradyrhizobium VKLRVTEAWDEND EHLPT-ALHYEGRAADITTQPPDGAKLGRLARLAVNAGCDWVFY-EDTSHVHVSVKKa 145
Pseudonocardia dioxanivorans RRLRLTEAWDPDG EHSHS-SLHYEGRAADLTVDDRDRAKLGRLAALAVQTGFDWVLH-END-HVHVSVRAg 144
Mesorhizobium AKLRVTEAWDEDN EHADA-SLHYEGRAADLTTNPVDGAKLGRLARLAVDAGCDWVFF-EDSSHIHVSVKAg 145
Micromonospora KKLRVTEGWDENN EHTAE-STHYEGRAVDMTVSDLDAAKLGRLARLAVDAGFDWVFY-ENALHVHASVKK- 144
Paenibacillus panacisoli IKLMVIDGWSDPD[1]RPPHNnRLYHEGRAADLTVSDLDTSKLGRLGWLAVDAGFDFVHY-ESQDHIHVAVKV- 146
Pseudoalteromonas piratica KTLRVTEAYDQDG EHAKF-SLHNEGRAADMTVSDRDLKKLGRLGFLATKAGFSWVYY-EHN-HIHASVKR- 142
spirosoma VSLRVTEAWDDTG EHSSShSLHYEGRAVDLTTSDLDKSKLGRLGRLAVDAGFNWVYY-ENLLHIHASVTKa 149
Nematostella vectensis hedgling RKLKVIKAWTDQR[1]AQDPA-SLHYEGRALRLQLDNNDRSMLSRLAGLALASGFDWVSYpLNSDYIHASVIR- 144
Nematostella Hh1 VKLRVTEAWDEDG QHSLD-SLHYEGRAVDISTSDKDPKKLPDLGSLAVDAGFDWVYY-DRRSSIHASVRS- 146
nematostella Hh2 VKVRVIRAYDDGT[4]HHGPH-SLHFSGRALDITTSDEKRDKLPMLGRLAYRAGFDWV-Y-RAKAYIHASVKS- 148
Varroa destructor VKLRVIESWDEYS HHKSG-SLHYEGRAVDFTTDDRHQAKYGMLARLAVEAGFDWVYY-ETKRHVHASVKP- 146
Drosophila IRLLVTESWDEDY HHGQE-SLHYEGRAVTIATSDRDQSKYGMLARLAVEAGFDWVSY-VSRRHIYCSVKS- 143
Sperm Whale VKLRVTEGWDEDG HHSEE-SLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYY-ESKAHIHCSVKA- 146
B
Bacterial Hh-N
Eukaryote Hh
Hedgling
Bradyrhizobium
Mesorhizobium
Sperm Whale Shh
Varroa destructor
Drosophila
Nematostella Hh1
Micromonospora
Algibacter
Bacillus
spirosoma
Pseudonocardia dioxanivorans
Nematostella Hh2
Paenibacillus panacisoli
Pseudoalteromonas piratica
Nematostella vectensis hedgling
0.2

Figure 2. The Hh-N domain is highly conserved between metazoans and bacteria. A: lineup of the conserved Hh-N domains. In most Hh-N domains the residues involve in metal ion coordination are conserved. B: The phylogenetic tree of the lineup in A shows clustering of most metazoan Hh proteins within the bacterial clade. Nematostella Hh2 is more diverged but is still rooted within the bacterial clade. Hedgling is the outgroup.

Discussion

It has been proposed that Hh arose in a cnidarian ancestor by the recombination of the *hedgling N-domain* with a eukaryotic *hog* domain before the urbilaterian (a common ancestor of protostomes and deuterostomes) arose, likely in a cnidarian ancestor [4]. This is consistent with the presence of hedgling in Choanoflagellates, sponges and cnidarians, but not in bilaterians, and the presence of Hh in cnidarians and bilaterians. However, the Hedge domain of eukaryotic Hh proteins is more similar to several bacterial proteins than to the cnidarian hedgling N-domain. This indicates that a bacterial *Hh-N* gave rise to the *Hh* gene by the recombination to a cnidarian *hog* domain. This would require horizontal gene transfer of the *Hh-N* domain from a bacterium into the cnidarian genome before the evolution of the ur-bilaterian.

Although all bacterial Hh-N containing proteins are hypothetical, they are likely Zn^{2+} peptidase pro-proteins [5]. The conservation is in this domain is strong, and it is likely that all the Hh-N proteins fold in a similar manner. The near complete conservation of the residues involved in Zn^{2+} and Ca^{2+} binding further support this notion and indicate that most Hh-N domains function as peptidases. The targets of the Hh-N peptidase activity remain unknown. Some of the highly conserved residues are found mutated in the congenital birth defect holoprosencephaly, which is caused by aberrant Shh signaling. These includes several residues in the large a helix (Figure 2A), but also include several of the residues involved in Zn^{2+} coordination [6], and thus peptidase activity. As this demonstrates that these residues are critical for Hh signaling, it is likely that a conserved functional activity is present in bacteria.

Methods

All alignments were performed in protein blast at the NCBI website (www.ncbi.nlm.nih.gov)

Accession numbers for the proteins used for the lineup and tree in Figure 1.

WP_082159544.1	hypothetical protein [Micromonospora sp. HK10]
WP_098188151.1	hypothetical protein [Bacillus pseudomycoides]
WP_028591995.1	hypothetical protein [Paenibacillus panacisoli]
WP_069829738.1	hypothetical protein [Algibacter aquaticus]
WP_080055297.1	hypothetical protein [Spirosoma aerolatum]
WP_103381118.1	hypothetical protein [Pseudonocardia dioxanivorans]
ESZ39647.1	hypothetical protein X732_15330 [Mesorhizobium sp. L2C066B000]
WP_029077494.1	hypothetical protein [Bradyrhizobium sp. th.b2]
WP_040135141.1	hypothetical protein [Pseudoalteromonas piratica]
ABX84114.1	hedgling [Nematostella vectensis]
XP_022667503.1	indian hedgehog protein-like [Varroa destructor]
ABX89897.1	hedgehog1 [Nematostella vectensis]
ABX89898.1	hedgehog2 [Nematostella vectensis]
NP_001034065.1	hedgehog [Drosophila melanogaster]
XP_007124731.1	sonic hedgehog protein [Physeter catodon]

Blast input for Hh-N domain used for the lineup and tree in Figure 2

>Algibacter
LKLKERVPNMKEYDVVGRISGKILRGTPEFDELIKNENQKIVFKDEEGTGADHYMTSKLSEKLNLLADLVIQEWGANIKLRVTEAWD
EDNEHASKSIHYEGRGADITTSDRDSSKLGRLARLAVEVGLDWVFYEDNSHVHVSMKK

>Bacillus
LNLGEKVPNKKESDVVGPITGVIQRGNPDFNTLVRSNNQDIQFKDEEGTGADFLMTSRLSDKLNTLAILVNQEWPNIKLRVTEAWD
EDNEHSSGSTHYEGRAADITTSDRDGNKLGRLAQLAVDAGFDWVYYENKYHIHVSVKK

>Bradyrhizobium
LTQGQKVPNASETATCGAIAKKIKRTDPEFATLVSNQNASIVFKDEEGTGADRMMSTRLQAKLDALASLVSAEWAGVKLRVTEAW
DENDEHLPTALHYEGRAADITTQPPDGAKLGRLARLAVNAGCDWVFYEDTSHVHVSVKKA

>Pseudonocardia dioxanivorans
VPLGEH
VPARAEADASGAATGVVAAGSPEFDALVRLDDPTVVVKDEEGSGADRMMTPRLAELVGVLAAHVAQAFPGRRLRLTEAWDPDG
EHSHSSLHYEGRAADLTVDDRDRAKLGRLAALAVQTGFDWVLHENDHVHVSVRAG

>Mesorhizobium
LSQGQKVPNASELVTCGPITKKITRTDPEFAGLVTNTNAKIVFKDEENTGADRMMTPRLKSKLDSLANVVASEWPGAKLRVTEAW
DEDNEHADASLHYEGRAADLTTNPVDGAKLGRLARLAVDAGCDWVFFEDSSHIHVSVKAG

>Micromonospora
LSLGERVPNVGEASVVGPIAGKVVRGSPEFNALVKNDNPDIVFKDEEKTDADRMMTPRLRDMVNELAALVVKEWPGKKLRVTEG
WDENNEHTAESTHYEGRAVDMTVSDLDAAKLGRLARLAVDAGFDWVFYENALHVHASVKK

>Nematostella Hh1
LYFKQRVPDVDEFSLGASGRPQGKITRNSSKFNKLVACYNTDIVFKDEERTGADRLMSKRCREKLRNLATKVKQKWKGVKLRVTEA
WDEDGQHSLDSLHYEGRAVDISTSDKDPKKLPDLGSLAVDAGFDWVYYDRRSSIHASVRS

>Nematostella Hh2
MKIREHIPDTSETSLQASGPSRKIKRGSNGYKELITNADPNIVFREDKAGNNRRMSKRCERKLKILSSLVRKEWIGDVKVRVIRAYDD
GTSKKRHHGPHSLHFSGRALDITTSDEKRDKLPMLGRLAYRAGFDWVYRAKAYIHASVKS

>Nematostella vectensis hedgling
LSQGSSYPYSKTEVSVCGGIRDVILRNSARFRKILVRNADTEVVFENDDCRRTTARAKSKLDVLASRVRQEWAGRKLKVIKAWTDQR
TAQDPASLHYEGRALRLQLDNNDRSMLSRLAGLALASGFDWVSYPLNSDYIHASVIR

>Paenibacillus panacisoli
LSVGQKVPDMLESQAAGPFTDAVHRGSNEFKNFVYNNNPLIVYDDEERGHSNHYMTKDLADQLNRLAELVAAEWSGIKLMVIDG
WSDPDVRPPHNNRLYHEGRAADLTVSDLDTSKLGRLGWLAVDAGFDFVHYESQDHIHVAVKV

>Pseudoalteromonas piratica
LKKGQRWPQSSEIKSSGEARQTLSGSSLRDCMNKSANNLPVFKFEEGNFTDVQASEKLCAAIMDLNKLVMKEWPGKTLRVTEAYD
QDGEHAKFSLHNEGRAADMTVSDRDLKKLGRLGFLATKAGFSWVYYEHNHIHASVKR

>spirosoma
LTLGQKVPDRAEKDSLASGPFNKAILRGTPEFATLVENKNEKVVFKNEEGDGSDRMMTPVLKTHVDRLADLVRSEWGAGVSLRVT
EAWDDTGEHSSSHSLHYEGRAVDLTTSDLDKSKLGRLGRLAVDAGFNWVYYENLLHIHASVTKA

>Varroa destructor
LVVKQYVPNTAEQNEIAAGRAKGAISRTSMAFRKLVPNYNTDIRFLDDEGTGADRIMTQRCRDKLDTLAVSVMTQWPGVKLRVIE
SWDEYSHHKSGSLHYEGRAVDFTTDDRHQAKYGMLARLAVEAGFDWVYYETKRHVHASVKP

>Drosophila
KQTIPNLSEYTNSASGPLEGVIRRDSPKFKDLVPNYNRDILFRDEEGTGADRLMSKRCKEKLNVLAYSVMNEWPGIRLLVTESWDED
YHHGQESLHYEGRAVTIATSDRDQSKYGMLARLAVEAGFDWVSYVSRRHIYCSVKS

>Sperm Whale
LAYKQFIPNVAEKTLGASGRYEGKITRNSERFKELTPNYNPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG
WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYYESKAHIHCSVKA

References

1. Hall, T. M., Porter, J. A., Beachy, P. A. & Leahy, D. J. A potential catalytic site revealed by the 1.7-A crystal structure of the amino-terminal signaling domain of Sonic hedgehog. *Nature* **378,** 212–216 (1995).
2. Bumcrot, D. A., Takada, R. & McMahon, A. P. Proteolytic processing yields two secreted forms of sonic hedgehog. *Mol. Cell. Biol.* **15,** 2294–2303 (1995).
3. Fairclough, S. R. *et al.* Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate Salpingoeca rosetta. *Genome biology* **14,** R15 (2013).
4. Adamska, M. *et al.* The evolutionary origin of hedgehog proteins. *Curr. Biol.* **17,** R836–7 (2007).
5. Rebollido-Rios, R. *et al.* Signaling domain of Sonic Hedgehog as cannibalistic calcium-regulated zinc-peptidase. *PLoS Comput. Biol.* **10,** e1003707 (2014).
6. Roessler, E. *et al.* The mutational spectrum of holoprosencephaly-associated changes within the SHH gene in humans predicts loss-of-function through either key structural alterations of the ligand or its altered synthesis. *Hum. Mutat.* **30,** E921–35 (2009).