J Mol Evol (2002) 55:790–794
DOI: 10.1007/s00239-002-2361-x

JOURNAL OF MOLECULAR EVOLUTION

Letter to the Editor

The Occurrence of Type S1A Serine Proteases in Sponge and Jellyfish

Ana Rojas, Russell F. Doolittle

Center for Molecular Genetics, University of California, San Diego, La Jolla, CA 92093-0634, USA

Received: 26 February 2002 / Accepted: 7 June 2002

Abstract. Although serine proteases are found in all kinds of cellular organisms and many viruses, the classic "chymotrypsin family" (Group S1A by the 1998 Barrett nomenclature) has an unusual phylogenetic distribution, being especially common in animals, entirely absent from plants and protists, and rare among fungi. The distribution in Bacteria is largely restricted to the genus *Streptomyces*, although a few isolated occurrences in other bacteria have been reported. The family may be entirely absent from Archaea. Although more than a thousand sequences have been reported for enzymes of this type from animals, none of them have been from early diverging phyla like Porifera or Cnidaria. We now report the existence of Group S1A serine proteases in a sponge (phylum Porifera) and a jellyfish (phylum Cnidaria), making it safe to conclude that all animal groups possess these enzymes.

Key words: Serine proteases — "Chymotrypsin family" — Sponge — Jellyfish

The origin of the "chymotrypsin family" of serine proteases [EC 3.4.21; Group S1A (Barrett et al. 1998)] in animals poses a perplexing problem in phylogenetics. There are literally hundreds of different enzymes from this group among invertebrate and vertebrate animals, but members of the family are not found at all among plants or protists. In fungi they are restricted to a single small group of Ascomycetes, and in Bacteria, with only a few exceptions, they are restricted to the genus *Streptomyces*. These enzymes, which are mostly extracellular and contain three or more disulfide bonds, are extremely abundant in both vertebrate and invertebrate animals.

Although these enzymes have been reviewed extensively in the past (e.g., Lesk and Fordham 1996), the issue of where and how the animal cohort of enzymes originated has not been specifically addressed in recent times. The current version of the MEROPS database contains about a thousand entries for such enzymes from 48 different animal species. Of these, about 200 are from two dozen invertebrate species. None of these, however, are from early-diverging phyla such as Porifera and Cnidaria. Given the overall sporadic occurrence of these enzymes and the frequent invocation of horizontal gene transfer in explanation (e.g., Screen et al. 2000), we felt it important to determine whether sponges and jellyfish possess these enzymes, and if so, where their sequences fall in a general phylogeny.

We isolated total RNA (Chomczynski and Sacchi, 1987) from a sponge (*Verongia aurea*) and a jellyfish (*Aurelia aurita*) and prepared cDNA, first in single-stranded form and then double-stranded (Chenchick et al. 1996). We constructed sets of degenerate primers based on extensive alignments of genes coding for this kind of protease among various inverte-

Correspondence to: R.F. Doolittle; *email*: rdoolittle@ucsd.edu

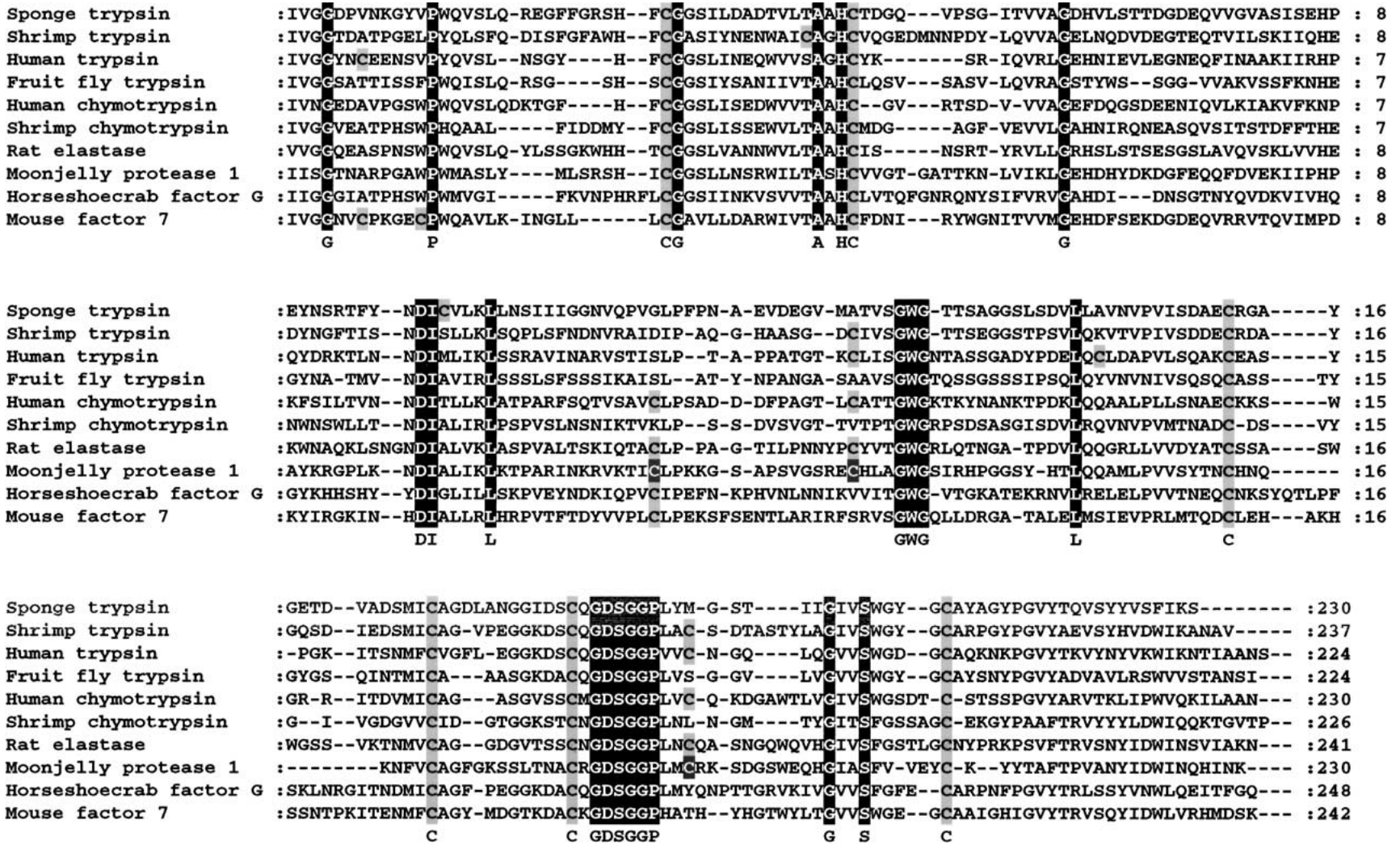
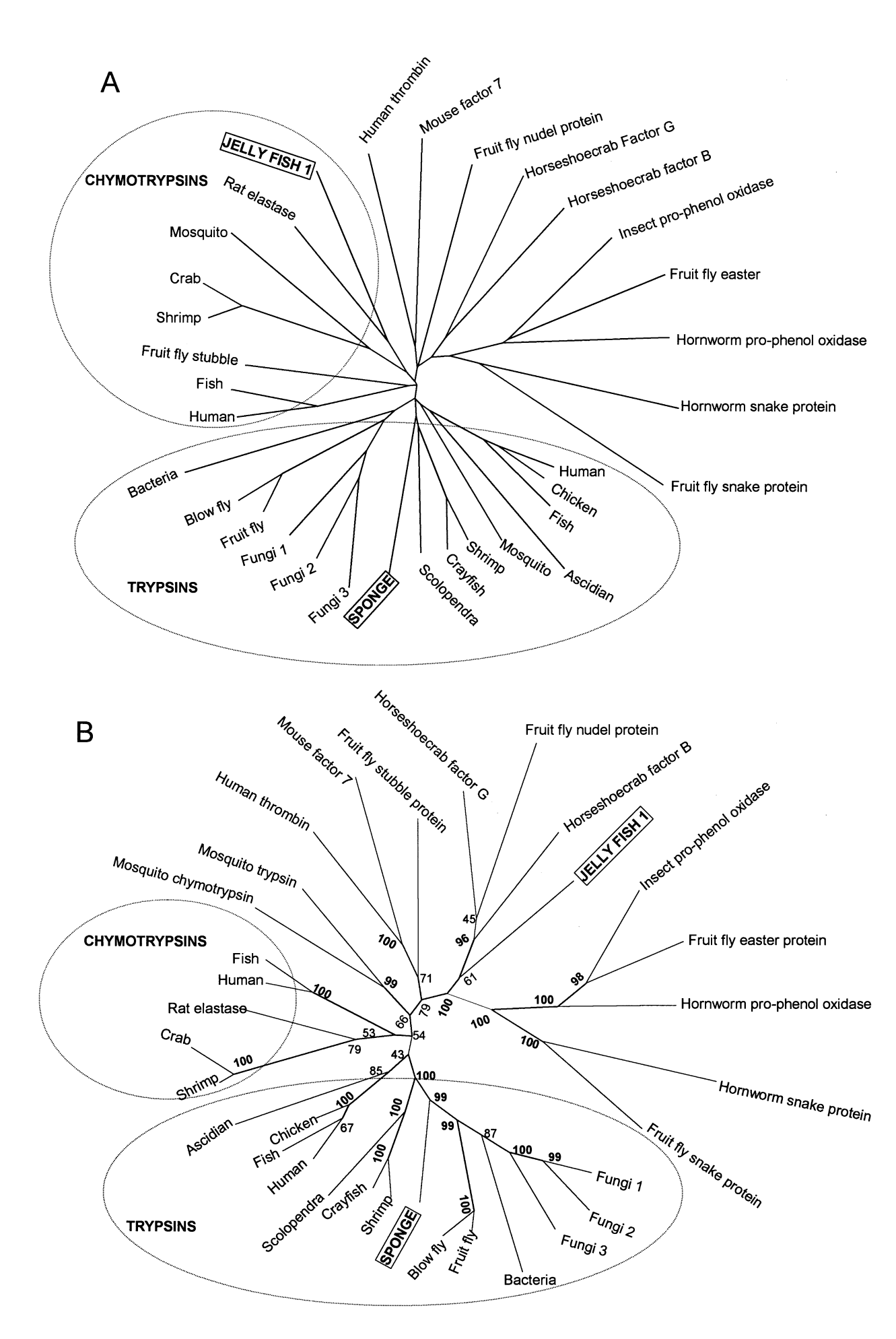


Fig. 1. Alignment of serine protease sequences from the sulfur sponge (*Verongia aurea*) and moon jellyfish (*Aurelia aurita*) with a representative set of S1A sequences from vertebrate and invertebrate animals.


A
Human thrombin
Mouse factor 7
Fruit fly nudel protein
Horseshoecrab Factor G
Horseshoecrab factor B
Insect pro-phenol oxidase
Fruit fly easter
Hornworm pro-phenol oxidase
Hornworm snake protein
Fruit fly snake protein
JELLY FISH 1
CHYMOTRYPSINS
Rat elastase
Mosquito
Crab
Shrimp
Fruit fly stubble
Fish
Human
Bacteria
Blow fly
Fruit fly
Fungi 1
Fungi 2
Fungi 3
SPONGE
Scolopendra
Crayfish
Shrimp
Mosquito
Ascidian
Human
Chicken
Fish
TRYPSINS
B
Horseshoecrab factor G
Fruit fly nudel protein
Mouse factor 7
Fruit fly stubble protein
Horseshoecrab factor B
Human thrombin
JELLY FISH 1
Insect pro-phenol oxidase
Mosquito trypsin
Mosquito chymotrypsin
Fruit fly easter protein
CHYMOTRYPSINS
Hornworm pro-phenol oxidase
Fish
Human
Rat elastase
Hornworm snake protein
Crab
Fruit fly snake protein
Shrimp
Ascidian
Chicken
Fish
Human
Scolopendra
Crayfish
Shrimp
SPONGE
Blow fly
Fruit fly
Bacteria
Fungi 1
Fungi 2
Fungi 3
TRYPSINS

brate animals. These primers were used in conjunction with sets of universal primers with adapters that facilitated the isolation of the appropriate messages (Matz et al. 1999a, b). Extensive use was made of kits from Clonetech, Invitrogen, and Quiagen. PCR was conducted on a Perkin Elmer thermocycler. *Drosophila melanogaster* dsDNA was used as a positive control to test the effectiveness of the primers.

Bands of DNA of the expected lengths (600–700 nt) were extracted from agarose gels after electrophoresis and subcloned into appropriate vectors. Multiple transformants were picked and submitted for DNA sequencing at the UCSD campus sequencing facility. In the end, sequences corresponding to serine proteases were isolated from each of the three kinds of starting DNA, including an already identified serine protease from fruitfly (GenBank Accession AAL 49280), a trypsin-like enzyme from sponge, and two serine proteases from jellyfish. In the cases of the sponge and jellyfish, sequences were confirmed by the use of different sets of primers that yielded overlapping regions. The single sequence identified in sponge encompassed 243 amino acid codons. Of the two isolated from jellyfish, one spanned 232 residues, but in the other the sequence did not extend beyond the 5′ degenerate primers, and only 183 codons were obtained. The two jellyfish sequences differed at 53 of the 185 comparable positions (71% identical at the amino acid level). The sponge and one of the jellyfish sequences were confirmed by constructing exact primers from them, reisolating bands from new cDNA from separate specimens, and determining the sequences again.

The three sequences have been submitted to GenBank; jellyfish serine protease 1, AF486486; jellyfish serine protease 2, AF486487; and sponge trypsin, AF486488. An alignment of the sponge and jellyfish sequences with those from several other animals is shown in Fig. 1.

Phylogenetic trees were constructed with these sequences and a representative set from other animals, both vertebrate and invertebrate, as well as several from bacteria and fungi, by two methods (Fig. 2A and B). Although the two trees are very similar, there are a few discrepancies. In both trees the digestive proteases—mostly trypsins and chymotrypsins—tend to separate from the proteases with nondigestive functions, and in both trees the newly discovered sponge protease clusters with the trypsins. As for the jellyfish enzymes, in one case they cluster with the chymotrypsin–elastase group (Fig. 2A), but in a Bayesian-inferred tree they appear among the nondigestive enzymes (Fig. 2B).

Apart from the phylogenetic trees, there is some ancillary information about what kinds of proteases these may be. For example, the sponge sequence has an aspartic acid at "position 189," a diagnostic residue that occurs in all known trypsins (Hannenshalli and Russell 2000). In contrast, the two jellyfish sequences have asparagine at that position. The sponge and jellyfish enzymes also differ with regard to the codon employed for the serine at the active site, the sponge having a typical trypsin signature with a TCT codon, whereas one jellyfish sequence has the AGT codon and the other AGC. When the new sequences were searched against GenBank with Blast (Altschul et al. 1990), the highest-scoring match for the sponge sequence was an arthropod trypsin (48% identity), whereas the best match for either of the jellyfish sequences was rat elastase (36% identity). The jellyfish enzymes are the first S1A serine proteases from an invertebrate animal with five disulfide bonds, all the cysteines of which match vertebrate elastase and chymotrypsin exactly, an observation that tends to support the phylogeny shown in Fig. 2A.

Although the S1A serine proteases from animals cover a wide variety of functional designations and are involved in physiological processes ranging from blood clotting to molting, the largest number of reported enzymes have to do with feeding and digestion, including about 100 entries labeled "trypsin." Jellyfish (Cnidaria) have genuine organs and tissues, including a mouth and gastrovascular cavity. They have a well-developed digestive apparatus that allows for the capture, swallowing, and digestion of food, and we fully expected that they would have digestive enzymes. Sponges, on the other hand, are loose colonies of cells that are filter feeders. They lack a true body cavity. The surrounding water and small particulate matter pass through pores that perforate special cells called porocytes into an inner chamber, from whence the flow is guided to an exit called the osculum by flagellated cells called choanocytes that line the chamber. Presumably much of the digestion is extracellular, but little is known of the process. The existence of a typical trypsin-like enzyme supports that notion.

The origin of animal serine proteases remains enigmatic, as implied by the positions of three pro-

Fig. 2. Phylogenetic trees were constructed from 33 assorted S1A serine protease sequences, including trypsin from the sulfur sponge (*Verongia aurea*), a serine protease from a moon jellyfish (*Aurelia aurita*), and a representative set of serine protease sequences from vertebrate and invertebrate animals, as well as three from fungi (*Fusarium oxyporum*, *Metarhizium anisopliae*, and *Cochliobolus carbonum*) and one from a bacterium (*Streptomyces griseus* trypsin), by two methods. The upper tree (A) was made by the progressive distance method of Feng and Doolittle (1996); the lower tree (B) was made by the Bayesian method of Huelsenbeck and Ronquist (2001). In the latter case, numbers at nodes denote clade confidence values.

tease sequences from Ascomycetes (fungi) and a bacterial sequence (*Streptomyces griseus* trypsin) in the phylogenetic trees (Fig. 2A and B). If the animal enzymes are the result of vertical descent in a conventional manner, then it must be presumed that protists, plants, and most fungi have lost these genes. With regard to the fungi, good cases have been made for the transfer of genes between ascomyctes and bacteria of the genus *Streptomyces* in both directions. Thus, the ascomycete *Fusarium oxysporum*, which has an animal-like trypsin (Rypniewski et a. 1993), also has a respiratory system which looks to have been acquired from *Streptomyces* (Takaya et al. 1998). But another ascomycete has a chymotrypsin which appears to have been transferred in the other direction, i.e., from the fungus to *Streptomyces* (Screen et al. 2000). We plan to address these apparently contradictory observations in a separate publication devoted strictly to phylogenetic analysis (Rojas and Doolittle, in preparation).

Acknowledgments. Sponge and jellyfish specimens were provided by the Birch Aquarium, UCSD. *Drosophila melanogaster* DNA was a gift from Prof. Stephen Wasserman. We thank Mikhail Matz for providing various protocols for preparing cDNA from invertebrate animals, and we are also grateful to Dirk Buscher for much valuable advice and consultation. A.R. was supported by an NSCORT postdoctoral fellowship.

References

Altshul SF, Gish W, Meyers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Barrett AJ, Rawlings ND, Woessner JF (1998) Handbook of proteolytic enzymes. Academic Press, San Diego, CA

Chenchick A, Diachenko L, Moqadam F, Tarabykin V, Lukyanov S, Siebert PD (1996) Full length cDNA cloning and determination of mRNA 5′ and 3′ ends by amplification of adaptor-ligated cDNA. Biotechniques 21:526–534

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156–159

Feng DF, Doolittle RF (1996) Progressive alignment of protein sequences and construction of phylogenetic trees from them. Methods Enzymol 266:368–382

Hannenhalli SS, Russell RB (2000) Analysis and prediction of functional sub-types from protein sequence alignments. J Mol Biol 303:61–76

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755

Lesk A, Fordham WD (1996) Conservation and variability in the structures of serine proteases of the chymotrypsin family. J Mol Biol 258:501–537

Matz M, Shagin D, Bogdanova E, Britanova O, Lukyanov S, Diatchenko L, Chenchick A (1999a) Amplification of cDNA ends based on template-switching effect and step-out PCR. Nucleic Acids Res 27:1558–1560

Matz MV, Fradkov AF, Labas YA, Savitsky AP, Zaraisky AG, Markelov ML, Lukyanov SA (1999b). Fluorescent proteins from non-bioluminescent Anthozoa species. Nature Biotechnol 17:969–973

Rypniewski WR, Hastrup S, Betzel Ch, et al. (1993) The sequence and X-ray structure of trypsin from *Fusarium oxysporum*. Prot Eng. 6:341–348

Screen SE, St. Leger RJ (2000) Cloning, expression and substrate specificity of a fungal chymotrypsin. Evidence for a horizontal transfer from an actinomycete to a bacterium. J Biol Chem 275:6689–6694

Takaya N, Kobayashi M, Shoun H (1998) Fungal denitrification, a respiratory system possibly acquired by horizontal gene: transfer from prokaryotes. In: Syvanen M, Kado C (eds) Horizontal gene transfer. Chapman and Hall, London, pp 321–3271