

2003

# Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB

Michal A. Kurowski

*Bioinformatics Laboratory, International Institute of Molecular and Cell Biology, michal@genesilico.pl*

Ashok S. Bhagwat

*Wayne State University, axb@chem.wayne.edu*

Grzegorz Papaj

*International Institute of Molecular and Cell Biology, grzesiek@genesilico.pl*

Janusz M. Bujnicki

*Bioinformatics Laboratory, International Institute of Molecular and Cell Biology, iamb@genesilico.pl*

---

## Recommended Citation

Kurowski *et al.* *BMC Genomics* 2003, 4:48

doi:10.1186/1471-2164-4-48

Available at: <http://digitalcommons.wayne.edu/biomedcentral/224>

Research article

Open Access

## Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB

Michal A Kurowski<sup>1</sup>, Ashok S Bhagwat<sup>2</sup>, Grzegorz Papaj<sup>1</sup> and Janusz M Bujnicki\*<sup>1</sup>

Address: <sup>1</sup>Bioinformatics Laboratory, International Institute of Molecular and Cell Biology, Trojdena 4, 02-109 Warsaw, Poland and <sup>2</sup>Department of Chemistry, Wayne State University, Detroit, MI 48202, U.S.A

Email: Michal A Kurowski - [michal@genesilico.pl](mailto:michal@genesilico.pl); Ashok S Bhagwat - [axb@chem.wayne.edu](mailto:axb@chem.wayne.edu); Grzegorz Papaj - [grzesiek@genesilico.pl](mailto:grzesiek@genesilico.pl); Janusz M Bujnicki\* - [iamb@genesilico.pl](mailto:iamb@genesilico.pl)

\* Corresponding author

Published: 10 December 2003

Received: 16 September 2003

*BMC Genomics* 2003, 4:48

Accepted: 10 December 2003

This article is available from: <http://www.biomedcentral.com/1471-2164/4/48>

© 2003 Kurowski et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** Combination of biochemical and bioinformatic analyses led to the discovery of oxidative demethylation – a novel DNA repair mechanism catalyzed by the *Escherichia coli* AlkB protein and its two human homologs, hABH2 and hABH3. This discovery was based on the prediction made by Aravind and Koonin that AlkB is a member of the 2OG-Fe<sup>2+</sup> oxygenase superfamily.

**Results:** In this article, we report identification and sequence analysis of five human members of the (2OG-Fe<sup>2+</sup>) oxygenase superfamily designated here as hABH4 through hABH8. These experimentally uncharacterized and poorly annotated genes were not associated with the AlkB family in any database, but are predicted here to be phylogenetically and functionally related to the AlkB family (and specifically to the lineage that groups together hABH2 and hABH3) rather than to any other oxygenase family. Our analysis reveals the history of ABH gene duplications in the evolution of vertebrate genomes.

**Conclusions:** We hypothesize that hABH 4–8 could either be back-up enzymes for hABH1-3 or may code for novel DNA or RNA repair activities. For example, enzymes that can dealkylate N3-methylpurines or N7-methylpurines in DNA have not been described. Our analysis will guide experimental confirmation of these novel human putative DNA repair enzymes.

### Background

The 2-oxoglutarate-Fe<sup>2+</sup>-dependent (2OG-Fe<sup>2+</sup>) oxygenase superfamily groups together enzymes that catalyze a variety of reactions that typically involve oxidation of an organic substrate using a di-oxygen molecule [1]. Among them are proteins implicated in the hydroxylation of proline and lysine side chains in collagen, the synthesis of

plant hormones, pigments and flavones, enzymes catalyzing the reaction of oxidative ring expansion in biosynthesis of penicillin and cephalosporin antibiotics, proteoglycans associated with basement membrane in chordates (Iprecan and its homologs). X-ray crystallographic studies of isopenicillin N synthase revealed a variant of the "jelly-roll" fold and provided insight into the

architecture of the conserved active site and the catalytic mechanism of 2OG-Fe<sup>2+</sup> oxygenases [2].

Using the protein fold-recognition method (a bioinformatic approach for determination if the sequence of interest is related to known protein structures) Aravind and Koonin found that the *Escherichia coli* AlkB protein implicated in DNA repair is also a member of the 2OG-Fe<sup>2+</sup> oxygenase superfamily [3]. They proposed that AlkB may catalyze oxidative demethylation of alkylated bases in DNA, while the eukaryotic homologs of AlkB maybe involved in RNA demethylation. In accordance with this prediction, it was demonstrated experimentally that bacterial AlkB carries out the oxidative demethylation of N1-methyladenine (m<sup>1</sup>A) or N3-methylcytosine (m<sup>3</sup>C) lesions in single-stranded and double-stranded DNA; the reaction requires non-heme Fe<sup>2+</sup>, O<sub>2</sub>, and α-ketoglutarate and yields succinate, CO<sub>2</sub> and formaldehyde and the restoration of unalkylated bases in DNA [4,5]. To date, three human homologs of AlkB have been reported, hABH1, hABH2 and hABH3. Although hABH1 was initially reported by one group to partially complement the sensitivity of an *E. coli alkB* mutant towards an alkylating agent [6], this result could not be reproduced by others [7]. hABH2 and hABH3 remove m<sup>1</sup>A, m<sup>3</sup>C and 1-ethyladenine from alkylated polynucleotides in an α-ketoglutarate dependent reaction [7]. Interestingly, bacterial AlkB and hABH3 (but not hABH2) were shown to prefer single-stranded nucleic acids and to repair lesions in RNA [8].

The availability of several completed eukaryotic genome sequences [9-13] allowed us to carry out large-scale searches for novel members of previously identified protein families and to perform "phylogenomic" analyses, i.e. functional predictions via phylogenetic analysis. The premise of phylogenomics [14] is that only the analyses carried out for complete sets of proteins obtained from whole genomes allow to make accurate inferences of the duplication/speciation patterns, identification of orthologs and paralogs, and mapping on known functions onto the evolutionary tree [15,16]. Thereby, functions of uncharacterized genes can be predicted by their phylogenetic position relative to the characterized ones, and by analyzing the rates and patterns of gene evolution (reviews: [14,17,18]). Previous analyses of AlkB and its homologs utilized databases of a smaller size (less complete genomes and less correctly predicted genes in the human genome) that did not allow comprehensive comparisons of the eukaryotic protein families on the genome/proteome scale. This prompted us to carry out a genome-wide search for novel eukaryotic members of the (2OG-Fe<sup>2+</sup>) oxygenase superfamily (with a special focus on potential new close homologs of AlkB) and to carry out the evolutionary analysis to predict their molecular function.

## Results

We have identified 24 members of the 2OG-Fe<sup>2+</sup> oxygenase superfamily in the human genome, including 19 previously known members (according to the annotation in the Ensembl database [19]) and 5 novel ones (see Methods for details). According to the phylogenomic analysis, 8 of these 24 human oxygenases grouped together with the AlkB lineage. This group of proteins included previously reported hABH1, hABH2, and hABH3, three proteins annotated only as oxygenase homologs and two proteins which have not been functionally annotated to date. We propose to name these 5 novel putative AlkB family members as ABHs (AlkB homologs) numbered 4 through 8 (Fig. 1). We also were able to find a number of orthologs of the newly identified ABHs in other fully sequenced metazoan genomes (Table 1) and in the EST division of the GenBank database (data not shown, links between the individual genes/contigs/transcripts are available via the Ensembl website using the gene identifiers provided in Table 1).

To substantiate the hypothesis that the new hABHs are genuine members of the 2OG-Fe<sup>2+</sup> oxygenase superfamily, we have confirmed their structural relationship to the jelly-roll fold by secondary structure prediction and protein fold-recognition i.e. the same techniques that contributed to the original identification of AlkB as a member of the 2OG-Fe<sup>2+</sup> oxygenase superfamily [3]. All protein structure analyses were carried out via the GeneSilico protein structure prediction meta server interface [20]. The consensus pattern of secondary structures predicted for all ABHs (Fig. 1) agrees very well with that observed in the known crystal structures of genuine β-helix oxygenases (see the "clavamate synthase-like" superfamily in the SCOP database [21]). The protein fold-recognition analysis has also corroborated our prediction: according to the FFAS 03 algorithm [22], the sequences of all ABHs are compatible with the structures of 2OG-Fe<sup>2+</sup> oxygenases (such as 1mze, 1gp4, 1bk0 in the Protein Data Bank) as well as other proteins that exhibit the same double-stranded β-helix fold (such as 1lr5 or 2arc). This compatibility has statistically significant value (FFAS score lower than -9.00). Taken together, these results can be regarded as strong indication that all five novel ABHs (4-8) are indeed true members of the 2OG-Fe<sup>2+</sup> oxygenase superfamily.

In order to find whether the new human ABHs belong to the AlkB family (and are likely to be genuine "dealkylases" involved in nucleic acid repair) or to some other 2OG-Fe<sup>2+</sup> oxygenase family, phylogenetic analysis was carried out and the multiple sequence alignment of ABHs (Fig. 1) was scrutinized for the presence of conserved residues and other features characteristic for AlkB. We found that all hABHs exhibit the catalytic residues (HxD...H...RxxxxR)



**Figure 1**

The multiple sequence alignment of *E. coli* AlkB and all human members of the AlkB family described previously (ABH1-3) and in this study (ABH4-8). Residues conserved in at least 4 members of the family are highlighted in black, residues with similar physico-chemical character are highlighted in gray. The length of loops omitted for the clarity of the presentation are shown in parentheses. The N- and C-terminal residues are numbered. The predicted catalytic residues are indicated with asterisks (\*).

**Table 1: AlkB homologs (ABHs) identified in complete eukaryotic genomes. For each gene, database identifiers are provided: Ensembl for the human, mouse and Fugu (pufferfish), Flybase for *D. melanogaster*, and Wormbase for *C. elegans*, respectively and, if available, a RefSeq identifier and a Gene Identification (gi) number.**

Name	Human	Mouse	Fugu	<i>D. melanogaster</i>	<i>C. elegans</i>
ABH1	ENSG00000100601 NP_006011 gi5174385	ENSMUSG0000021041 XP_127049 gi23597748	SINFRUG00000121116	-	-
ABH2	- XP_058581 gi17474993	ENSMUSG0000044339 XP_132383 gi20841151	SINFRUG00000145993	-	-
ABH3	ENSG00000166199 NP_631917 gi21040275	ENSMUSG0000040174 XP_130317 gi25056889	SINFRUG00000138252	-	-
ABH4	ENSG00000160993 NP_060091 gi8923019	ENSMUSG0000039754 XP_132400 gi20846125	SINFRUG00000147682	CG4036 NP_609314 gi24583140	Y09F7.7a NP_741141 gi25148697
ABH5	ENSG00000091542 NP_060228 gi22726193	ENSMUSG0000042650 NP_766531 gi31044423	SINFRUG00000127593	-	-
ABH6	ENSG00000130687 NP_116267 gi14249626	ENSMUSG0000042831 Q8K2U2 gi20988250	SINFRUG00000141687	CG6144 NP_609414 gi24583454	B0564.2 NP_502522 gi17538328
ABH7	ENSG00000125652 NP_115682 gi14150066	ENSMUSG0000002661 NP_079814 gi21313470	SINFRUG00000148920	-	-
ABH8	ENSG00000137760 + ENSG00000184170 NP_620130 gi20270315	ENSMUSG0000025899 BAC38223 gi26349167	SINFRUG00000142661	1) CG17807 NP_611690 gi24658267 2) CG14130 NP_648511 gi24662930	C14B1.10 NP_497751 gi17552176

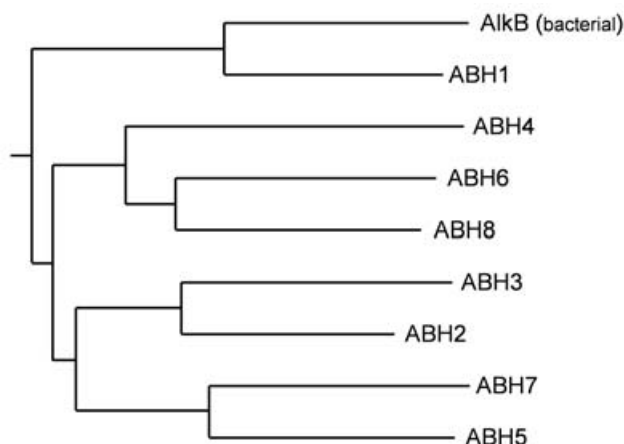
predicted for the AlkB family [3]. It is remarkable that all these proteins share the conserved Arg residue in the C-terminal  $\beta$ -strand (R210 in *E. coli* AlkB), which was found to be characteristic of the AlkB family and not found in the other 2OG-Fe<sup>2+</sup> oxygenases. The latter enzymes typically have a bulky aromatic side chain at this position [3].

Furthermore, in the phylogenetic tree of the 2OG-Fe<sup>2+</sup> oxygenase superfamily all ABHs cluster together with the known AlkB family members from Prokaryota and Eukaryota with strong bootstrap support (>90%; data not shown). Interestingly, the novel ABHs (ABHs 4–8) were found to be more closely related to the hABH2/hABH3 lineage described previously [3] rather than to the lineage comprising *E. coli* AlkB and the ABH1 protein (Fig. 2). This result suggests that eukaryotic ABHs 2–8 have a relatively recent common origin and that they have most likely radiated by a series of duplication events in the metazoan lineage. Our analysis shows that all eight hABHs have orthologs in vertebrate genomes, while the number of ABHs in the genomes of *C. elegans* and *D. melanogaster* is more limited. According to the maximum likelihood phylogenetic tree, ABH2&3, ABH5&7 and ABH4&6&8 are in-paralogous (i.e. originated by duplication following the radiation of the main ABH lineages).

The mouse orthologs of hABH5, hABH6, and hABH7 are within known regions of human-mouse synteny. The *hABH5* gene maps to the 17p11 locus and *hABH6* and *hABH7* map to the 19th chromosome (loci 19q13 and 19p13, respectively). On the other hand, the *hABH4* gene is found outside the region of human-mouse chromosome synteny. It maps to the locus 7q22. There are two separate transcripts in the Ensembl database for the *hABH8* gene (locus 11p11), seemingly corresponding to two protein domains, namely the RRM domain, typically found in RNA-binding proteins and the oxygenase domain homologous to AlkB. However, in the hABH8 orthologs from other metazoan genomes (including *C. elegans* and *D. melanogaster*) these two domains are found fused together to form a single polypeptide. The fusion of the AlkB-like domain with an RNA-binding domain suggests that the ABH8 protein, like previously found for hABH3, may be involved in repair of alkylation damage in RNA. Whether the human hABH8 protein is formed by two independently expressed domains or is encoded by a single polypeptide like its orthologs, remains to be determined experimentally.

## Conclusions

Based on the phylogenomic considerations supported by structure prediction and identification of characteristic conserved residues we predict that five so far uncharacterized human proteins are members of the AlkB family of nucleic acid repair enzymes. hABH4, hABH5, and hABH6



**Figure 2**

The maximum likelihood phylogenetic tree of the AlkB family. Only representative members (from human and *E. coli*), corresponding to the sequences in Fig. 1 are shown for the clarity of presentation. The position of the root was inferred from the large 2OG-Fe<sup>2+</sup> oxygenase superfamily tree (data not shown). The topology of the presented tree received very high p-value support according to the Shimodaira-Hasegawa [31] and "approximately unbiased" [32] tests (0.998 and 0.988, respectively).

have been annotated in Ensembl only as members of the huge 2OG-Fe<sup>2+</sup> oxygenase superfamily without any specification as to the possible substrate or biological function. hABH7 and hABH8 have not been annotated as members of the 2OG-Fe<sup>2+</sup> oxygenase superfamily in any database at the time of this analysis (August 2003). We hypothesize that hABH 4–8 could either be back-up enzymes for hABH1-3 or may code for novel DNA or RNA repair activities. For example, enzymes that can dealkylate N3-methylpurines or N7-methylpurines in DNA have not been described. It is possible that some of the predicted ABHs perform such demethylation reaction. Our prediction has therefore a significant potential to facilitate the experimental analysis aimed at understanding the molecular function of these proteins. However, it must be remembered that for hABH1 no activity could be reproducibly demonstrated and some of the ABHs (4–8) reported in this work may turn out to be inactive or may be difficult to assay. Nonetheless, all ABHs have (at least partial) counterparts in the characterized cDNA or EST databases (Table 1) suggesting that they are all expressed in human cells.

## Methods

Our re-analysis of (2OG-Fe<sup>2+</sup>) oxygenases was initiated with the sequence data obtained from the Pfam and the

Clusters of Orthologous Groups (COG) databases [23,24]. The Pfam database (release 9.0) contains 857 sequences from various species annotated as members of the (2OG-Fe<sup>2+</sup>) oxygenase superfamily. We also used the COG 3145 cluster with 23 prokaryotic members of the AlkB subfamily. The multiple sequence alignments (MSA) from PFAM (oxygenases) and COG (AlkB orthologs) were converted to profiles and used to search for oxygenase/AlkB homologs in the following genomic databases with a variant of the RPS-BLAST program [25]: *Homo sapiens* genome: NCBI assembly 33/ Ensembl 15.33.1; *Mus musculus* genome: MGSC – NCBI assembly 30 / Ensembl 15.30.1; *Takifugu rubripes* genome: ICMB October 2001 release / Ensembl 15.2.1; *Drosophila melanogaster* genome: Flybase 3.1; *Caenorhabditis elegans* genome: Wormbase 102. The RefSeq format and identifiers were used to annotate the sequences.

All bona fide oxygenases/AlkB orthologs and their homologs identified in the genomic databases were aligned using HMMER [26] and subjected to the evolutionary analysis using the SDI/RIO software package [15,16]. Briefly, a maximum likelihood (ML) tree of the bona fide members of the superfamily (sequences from PFAM and COG) was built using Tree-Puzzle [27]. Each newly identified putative oxygenase/AlkB homolog was separately aligned to the genuine oxygenases and its position on the superfamily tree (as well as the statistical support for this position) was determined. The SDI algorithm was used to identify the type of evolutionary relationships (orthology/paralogy) between the tested sequence and the subfamilies of genuine oxygenases, using the established evolutionary tree of species [28] as the reference. Consequently, for each oxygenase candidate it was predicted whether it could have evolved from the known superfamily members by speciation or duplication and what would be the relative timing of its radiation compared to the radiation of other members.

The final evolutionary tree of the oxygenase superfamily was computed using the fast neighbor-joining (NJ) approach [29] (for all sequences analyzed) and the computationally expensive maximum likelihood (ML) approach [30] (for the selected members of each major lineage, including all hABHs). The topologies of major branches of both trees were virtually identical, therefore only the ML tree is shown, for the clarity of the presentation. The final phylogenetic tree was obtained following using the "approximately unbiased" and Shimodaira-Hasegawa tests [31,32].

#### List of abbreviations used

(h)ABH, (human) AlkB homolog

#### Authors' contributions

MAK performed most of database searches, multiple sequence alignment, and phylogenetic analyses and drafted the first version of the manuscript. ASB initiated the study, participated in the interpretation of the results and writing of the manuscript. GP performed database searches and multiple sequence alignment to identify eukaryotic orthologs of all eight ABHs and used this information to identify missing exons in human genes. JMB supervised the study, refined the multiple sequence alignment, participated in phylogenetic analyses, prepared the figures and participated in the discussions of the results and writing of the manuscript. All authors read and approved the final manuscript.

#### Acknowledgements

This work was supported by a grant 1R21CA97899-01A from NCI. M.A.K. and J.M.B. were additionally supported by a grant 3P05A02024 from KBN. J.M.B. is an EMBO/HHMI Young Investigator and a fellow of the Foundation for Polish Science.

#### References

- Hegg EL, Que L Jr: **The 2-His-1-carboxylate facial triad – an emerging structural motif in mononuclear non-heme iron(II) enzymes.** *Eur J Biochem* 1997, **250**:625-629.
- Roach PL, Clifton IJ, Fulop V, Harlos K, Barton GJ, Hajdu J, Andersson I, Schofield CJ, Baldwin JE: **Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes.** *Nature* 1995, **375**:700-704.
- Aravind L, Koonin EV: **The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases.** *Genome Biol* 2001, **2(3)**:RESEARCH0007.
- Trewick SC, Henshaw TF, Hausinger RP, Lindahl T, Sedgwick B: **Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage.** *Nature* 2002, **419**:174-178.
- Falnes PO, Johansen RF, Seeberg E: **AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*.** *Nature* 2002, **419**:178-182.
- Wei YF, Carter KC, Wang RP, Shell BK: **Molecular cloning and functional analysis of a human cDNA encoding an *Escherichia coli* AlkB homolog, a protein involved in DNA alkylation damage repair.** *Nucleic Acids Res* 1996, **24**:931-937.
- Duncan T, Trewick SC, Koivisto P, Bates PA, Lindahl T, Sedgwick B: **Reversal of DNA alkylation damage by two human dioxygenases.** *Proc Natl Acad Sci U S A* 2002, **99**:16660-16665.
- Aas PA, Otterlei M, Falnes PO, Vagbo CB, Skorpen F, Akbari M, Sundheim O, Bjoras M, Slupphaug G, Seeberg E et al.: **Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA.** *Nature* 2003, **421**:859-863.
- Harris TW, Lee R, Schwarz E, Bradnam K, Lawson D, Chen W, Blasier D, Kenny E, Cunningham F, Kishore R et al.: **WormBase: a cross-species database for comparative genomics.** *Nucleic Acids Res* 2003, **31**:133-137.
- Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A et al.: **Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*.** *Science* 2002, **297**:1301-1310.
- The FlyBase Consortium: **The FlyBase database of the *Drosophila* genome projects and community literature.** *Nucleic Acids Res* 2003, **31**:172-175.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W et al.: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
- Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P et al.: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420**:520-562.

14. Eisen JA: **Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis.** *Genome Res* 1998, **8**:163-167.
15. Zmasek CM, Eddy SR: **A simple algorithm to infer gene duplication and speciation events on a gene tree.** *Bioinformatics* 2001, **17**:821-828.
16. Zmasek CM, Eddy SR: **RIO: Analyzing proteomes by automated phylogenomics using resampled inference of orthologs.** *BMC Bioinformatics* 2002, **3**:14.
17. Eisen JA, Wu M: **Phylogenetic analysis and gene functional predictions: phylogenomics in action.** *Theor Popul Biol* 2002, **61**:481-487.
18. Eisen JA, Fraser CM: **Phylogenomics: intersection of evolution and genomics.** *Science* 2003, **300**:1706-1707.
19. Clamp M, Andrews D, Barker D, Bevan P, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V et al.: **Ensembl 2002: accommodating comparative genomics.** *Nucleic Acids Res* 2003, **31**:38-42.
20. Kurowski MA, Bujnicki JM: **GeneSilico protein structure prediction meta-server.** *Nucleic Acids Res* 2003, **31**:3305-3307 [<http://genesilico.pl/meta/>].
21. LoConte L, Brenner SE, Hubbard TJ, Chothia C, Murzin AG: **SCOP database in 2002: refinements accommodate structural genomics.** *Nucleic Acids Res* 2002, **30**:264-267.
22. Rychlewski L, Jaroszewski L, Li W, Godzik A: **Comparison of sequence profiles. Strategies for structural predictions using sequence information.** *Protein Sci* 2000, **9**:232-241.
23. Bateman A, Birney E, Cerruti L, Durbin R, Ewinger L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL: **The Pfam protein families database.** *Nucleic Acids Res* 2002, **30**:276-280.
24. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV: **The COG database: new developments in phylogenetic classification of proteins from complete genomes.** *Nucleic Acids Res* 2001, **29**:22-28.
25. Marchler-Bauer A, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, He S, Hurwitz DI, Jackson JD, Jacobs AR, Lanczycki CJ et al.: **CDD: a curated Entrez database of conserved domain alignments.** *Nucleic Acids Res* 2003, **31**:383-387.
26. Durbin R, Eddy SR, Krogh A, Mitchison G: *Biological sequence analysis: probabilistic models of proteins and nucleic acids* Cambridge: Cambridge University Press; 1998.
27. Schmidt HA, Strimmer K, Vingron M, von Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.** *Bioinformatics* 2002, **18**:502-504.
28. **The Tree of Life web project** [<http://tolweb.org/tree/phylogeny.html>]
29. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4**:406-425.
30. Saitou N: **Property and efficiency of the maximum likelihood method for molecular phylogeny.** *J Mol Evol* 1988, **27**:261-273.
31. Shimodaira H, Hasegawa M: **CONSEL: for assessing the confidence of phylogenetic tree selection.** *Bioinformatics* 2001, **17**:1246-1247.
32. Shimodaira H: **An approximately unbiased test of phylogenetic tree selection.** *Syst Biol* 2002, **51**:492-508.

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](http://www.biomedcentral.com/info/publishing_adv.asp)

