

References

- Abagyan RA, Batalov S (1997) Do aligned sequences share the same fold? *J Mol Biol* 273 (1):355–368. doi:10.1006/jmbi.1997.1287
- Afrasiabi C, Samad B, Dineen D, Meacham C, Sjölander K (2013) The PhyloFacts FAT-CAT web server: ortholog identification and function prediction using fast approximate tree classification. *Nucleic Acids Res* 41(Web Server issue), W242–8. doi:10.1093/nar/gkt399
- Alexeyenko A, Tamas I, Liu G, Sonnhammer ELL (2006) Automatic clustering of orthologs and inparalogs shared by multiple proteomes. *Bioinformatics* (Oxford, England), 22(14), e9–15. doi:10.1093/bioinformatics/btl213
- Altenhoff AM, Dessimoz C (2009) Phylogenetic and functional assessment of orthologs inference projects and methods. *PLoS Comput Biol* 5(1):e1000262. doi:10.1371/journal.pcbi.1000262
- Altenhoff AM, Studer RA, Robinson-Rechavi M, Dessimoz C (2012) Resolving the ortholog conjecture: orthologs tend to be weakly, but significantly, more similar in function than paralogs. *PLoS Comput Biol* 8(5):e1002514. doi:10.1371/journal.pcbi.1002514
- Altenhoff AM, Škunca N, Glover N, Train C-M, Sueki A, Piližota I et al (2015) The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. *Nucleic Acids Res* 43(Database issue), D240–9. doi:10.1093/nar/gku1158
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–3402
- Alva V, Remmert M, Biegert A, Lupas AN, Söding J (2010) A galaxy of folds. *Protein Sci: A Publ Protein Soc* 19(1):124–130. doi:10.1002/pro.297
- Banumathy G, Somaiah N, Zhang R, Tang Y, Hoffmann J, Andrade M et al (2009) Human UBN1 is an ortholog of yeast Hpc2p and has an essential role in the HIRA/ASF1a chromatin-remodeling pathway in senescent cells. *Mol Cell Biol* 29(3):758–770. doi:10.1128/MCB.01047-08
- Barberis M, De Gioia L, Ruzzene M, Sarno S, Coccetti P, Fantucci P et al (2005) The yeast cyclin-dependent kinase inhibitor Sic1 and mammalian p27Kip1 are functional homologues with a structurally conserved inhibitory domain. *Biochem J* 387(Pt 3):639–647. doi:10.1042/BJ20041299
- Bedoya O, Tischer I (2014) Remote homology detection incorporating the context of physicochemical properties. *Comput Biol Med* 45:43–50. doi:10.1016/j.compbiomed.2013.11.012
- Bedoya O, Tischer I (2015) Reducing dimensionality in remote homology detection using predicted contact maps. *Comput Biol Med* 59:64–72. doi:10.1016/j.compbiomed.2015.01.020

- Bernardes JS, Dávila AMR, Costa VS, Zaverucha G (2007) Improving model construction of profile HMMs for remote homology detection through structural alignment. *BMC Bioinform* 8 (1):435. doi:10.1186/1471-2105-8-435
- Bernardes JS, Carbone A, Zaverucha G (2011) A discriminative method for family-based protein remote homology detection that combines inductive logic programming and propositional models. *BMC Bioinform* 12(1):83. doi:10.1186/1471-2105-12-83
- Bhadra R, Sandhya S, Abhinandan KR, Chakrabarti S, Sowdhamini R, Srinivasan N (2006) Cascade PSI-BLAST web server: a remote homology search tool for relating protein domains. *Nucleic Acids Res* 34(Web Server issue), W143–6. doi:10.1093/nar/gkl157
- Bhardwaj G, Ko KD, Hong Y, Zhang Z, Ho NL, Chintapalli SV et al (2012) PHYRN: a robust method for phylogenetic analysis of highly divergent sequences. *PLoS ONE* 7(4):e34261. doi:10.1371/journal.pone.0034261
- Biegert A, Mayer C, Remmert M, Söding J, Lupas AN (2006) The MPI bioinformatics toolkit for protein sequence analysis. *Nucleic Acids Res* 34(Web Server issue), W335–9. doi:10.1093/nar/gkl217
- Blake JD, Cohen FE (2001) Pairwise sequence alignment below the twilight zone. *J Mol Biol* 307 (2):721–735. doi:10.1006/jmbi.2001.4495
- Bork P, Sander C, Valencia A (1993) Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases. *Protein Sci: A Publ Protein Soc* 2(1):31–40. doi:10.1002/pro.5560020104
- Burmester T, Hankeln T (2014) Function and evolution of vertebrate globins. *Acta Physiol (Oxford, England)*, 211(3): 501–514. doi:10.1111/apha.12312
- Chang GS, Hong Y, Ko KD, Bhardwaj G, Holmes EC, Patterson RL, van Rossum DB (2008) Phylogenetic profiles reveal evolutionary relationships within the “twilight zone” of sequence similarity. *Proc Natl Acad Sci USA* 105(36):13474–13479. doi:10.1073/pnas.0803860105
- Comin M, Verzotto D (2011) The irredundant class method for remote homology detection of protein sequences. *J Computat Biol: J Computat Mol Cell Biol* 18(12):1819–1829. doi:10.1089/cmb.2010.0171
- Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet* 9(12):938–950. doi:10.1038/nrg2482
- Dalquen DA, Dessimoz C (2013) Bidirectional best hits miss many orthologs in duplication-rich clades such as plants and animals. *Genome Biol Evol* 5(10):1800–1806. doi:10.1093/gbe/evt132
- Darzentas N, Rigoutsos I, Ouzounis CA (2005) Sensitive detection of sequence similarity using combinatorial pattern discovery: a challenging study of two distantly related protein families. *Proteins* 61(4):926–937. doi:10.1002/prot.20608
- Datta RS, Meacham C, Samad B, Neyer C, Sjölander K (2009) Berkeley PHOG: PhyloFacts orthology group prediction web server. *Nucleic Acids Res* 37(Web Server issue), W84–9. doi:10.1093/nar/gkp373
- Dietmann S, Fernandez-Fuentes N, Holm L (2002) Automated detection of remote homology. *Curr Opin Struct Biol* 12(3):362–367
- Dong Y, Bogdanova A, Habermann B, Zachariae W, Ahringer J (2007) Identification of the *C. elegans* anaphase promoting complex subunit Cdc26 by phenotypic profiling and functional rescue in yeast. *BMC Dev Biol* 7(1):19. doi:10.1186/1471-213X-7-19
- Doolittle RF (1986) Of Urfs and Orfs: a primer on how to analyze derived amino acid sequences. In: University Science Books, Herndon, VA vol 29, pp 1–103. doi:10.1002/jobm.3620290411
- Dufayard J-F, Duret L, Penel S, Gouy M, Rechenmann F, Perrière G (2005) Tree pattern matching in phylogenetic trees: automatic search for orthologs or paralogs in homologous gene sequence databases. *Bioinformatics (Oxford, England)*, 21(11): 2596–2603. doi:10.1093/bioinformatics/bti325
- Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. *Genome Inform Int Conf Genome Inform* 23(1): 205–211
- Eyre TA, Wright MW, Lush MJ, Bruford EA (2007) HCOP: a searchable database of human orthology predictions. *Briefings Bioinform* 8(1):2–5. doi:10.1093/bib/bbl030

- Fariselli P, Rossi I, Capriotti E, Casadio R (2007) The WWWH of remote homolog detection: the state of the art. *Briefings Bioinform* 8(2):78–87. doi:10.1093/bib/bbl032
- Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F et al (2015) HMMER web server: 2015 update. *Nucleic Acids Res* 43(W1):W30–W38. doi:10.1093/nar/gkv397
- Fitch WM (1970) Distinguishing homologous from analogous proteins. *Syst Zool* 19(2):99–113
- Gabaldón T, Koonin EV (2013) Functional and evolutionary implications of gene orthology. *Nat Rev Genet* 14(5):360–366. doi:10.1038/nrg3456
- Galindo A, Hervás-Aguilar A, Rodríguez-Galán O, Vincent O, Arst HN, Tilburn J, Peñalva MA (2007) PalC, one of two Bro1 domain proteins in the fungal pH signalling pathway, localizes to cortical structures and binds Vps32. *Traffic (Copenhagen, Denmark)* 8(10): 1346–1364. doi:10.1111/j.1600-0854.2007.00620.x
- Ginalski K (2003) ORFeus: detection of distant homology using sequence profiles and predicted secondary structure. *Nucleic Acids Res* 31(13):3804–3807. doi:10.1093/nar/gkg504
- Gray GS, Fitch WM (1983) Evolution of antibiotic resistance genes: the DNA sequence of a kanamycin resistance gene from *Staphylococcus aureus*. *Mol Biol Evol* 1(1):57–66
- Grossberger R, Gieffers C, Zachariae W, Podtelejnikov AV, Schleiffer A, Nasmyth K et al (1999) Characterization of the DOC1/APC10 subunit of the yeast and the human anaphase-promoting complex. *J Biol Chem* 274(20):14500–14507
- Gupta MK, Niyogi R, Misra M (2013) An alignment-free method to find similarity among protein sequences via the general form of Chou's pseudo amino acid composition. *SAR QSAR Environ Res* 24(7):597–609. doi:10.1080/1062936X.2013.773378
- Heinicke S, Livstone MS, Lu C, Oughtred R, Kang F, Angiuoli SV et al (2007) The Princeton protein orthology database (P-POD): a comparative genomics analysis tool for biologists. *PLoS ONE* 2(8):e766. doi:10.1371/journal.pone.0000766
- Herrero J, Muffato M, Beal K, Fitzgerald S, Gordon L, Pignatelli M et al (2016) Ensemble comparative genomics resources. Database: *J Biol Databases Curation* 2016, bav096. doi:10.1093/database/bav096
- Höhl M, Ragan MA (2007) Is multiple-sequence alignment required for accurate inference of phylogeny? *Syst Biol* 56(2):206–221. doi:10.1080/10635150701294741
- Höhl M, Rigoutsos I, Ragan MA (2006) Pattern-based phylogenetic distance estimation and tree reconstruction. *Evol Bioinform Online* 2:359–375
- Huerta-Cepas J, Bueno A, Dopazo J, Gabaldon T (2007) PhylomeDB: a database for genome-wide collections of gene phylogenies. *Nucleic Acids Res* 36(Database), D491–D496. doi:10.1093/nar/gkm899
- Hutterer A, Berdnik D, Wirtz-Peitz F, Zigman M, Schleiffer A, Knoblich JA (2006) Mitotic activation of the kinase Aurora-A requires its binding partner Bora. *Dev Cell* 11(2):147–157. doi:10.1016/j.devcel.2006.06.002
- Ivliev AE, Sergeeva MG (2008) OrthoFocus: program for identification of orthologs in multiple genomes in family-focused studies. *Js Bioinform Comput Biol* 6(4):811–824
- Johnson LS, Eddy SR, Portugaly E (2010) Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinform* 11(1):431. doi:10.1186/1471-2105-11-431
- Karwath A, King RD (2002) Homology induction: the use of machine learning to improve sequence similarity searches. *BMC Bioinform* 3(1):11. doi:10.1186/1471-2105-3-11
- Kim S, Kang J, Chung YJ, Li J, Ryu KH (2008) Clustering orthologous proteins across phylogenetically distant species. *Proteins* 71(3):1113–1122. doi:10.1002/prot.21792
- Kim B-H, Cheng H, Grishin NV (2009) HorA web server to infer homology between proteins using sequence and structural similarity. *Nucleic Acids Res* 37(Web Server issue), W532–8. doi:10.1093/nar/gkp328
- Kim J, Ishiguro K-I, Nambu A, Akiyoshi B, Yokobayashi S, Kagami A et al (2015) Meikin is a conserved regulator of meiosis-I-specific kinetochore function. *Nature* 517(7535):466–471. doi:10.1038/nature14097
- Kitajima TS, Kawashima SA, Watanabe Y (2004) The conserved kinetochore protein shugoshin protects centromeric cohesion during meiosis. *Nature* 427(6974):510–517. doi:10.1038/nature02312

- Koonin EV (2005) Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet* 39(1):309–338. doi:10.1146/annurev.genet.39.073003.114725
- Kristensen DM, Wolf YI, Mushegian AR, Koonin EV (2011) Computational methods for Gene Orthology inference. *Briefings Bioinform* 12(5):379–391. doi:10.1093/bib/bbr030
- Kriventseva EV, Rahman N, Espinosa O, Zdobnov EM (2008) OrthoDB: the hierarchical catalog of eukaryotic orthologs. *Nucleic Acids Res* 36(Database issue), D271–5. doi:10.1093/nar/gkm845
- Kueng S, Hegemann B, Peters BH, Lipp JJ, Schleiffer A, Mechtler K, Peters J-M (2006) Wapl controls the dynamic association of cohesin with chromatin. *Cell* 127(5):955–967. doi:10.1016/j.cell.2006.09.040
- Kumar S (2011) Remote homologue identification of *Drosophila* GAGA factor in mouse. *Bioinformatics* 7(1):29–32
- Kumar A, Cowen L (2009) Augmented training of hidden Markov models to recognize remote homologs via simulated evolution. *Bioinformatics (Oxford, England)* 25(13): 1602–1608. doi:10.1093/bioinformatics/btp265
- Kuziemko A, Honig B, Petrey D (2011) Using structure to explore the sequence alignment space of remote homologs. *PLoS Comput Biol* 7(10):e1002175. doi:10.1371/journal.pcbi.1002175
- Lawo S, Bashkurov M, Mullin M, Ferreria MG, Kittler R, Habermann B et al (2009) HAUS, the 8-subunit human Augmin complex, regulates centrosome and spindle integrity. *Current Biol: CB* 19(10):816–826. doi:10.1016/j.cub.2009.04.033
- Lee MM, Bundschuh R, Chan MK (2008) Distant homology detection using a LLength and SStructure-based sequence alignment tool (LESTAT). *Proteins* 71(3):1409–1419. doi:10.1002/prot.21830
- Li L, Stoecckert CJ, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13(9):2178–2189. doi:10.1101/gr.1224503
- Liu K, Raghavan S, Nelesen S, Linder CR, Warnow T (2009) Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees. *Science (New York, N.Y.)* 324 (5934):1561–1564. doi:10.1126/science.1171243
- Liu K, Warnow TJ, Holder MT, Nelesen SM, Yu J, Stamatakis AP, Linder CR (2012) SATe-II: very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. *Syst Biol* 61(1):90–106. doi:10.1093/sysbio/syr095
- Liu B, Zhang D, Xu R, Xu J, Wang X, Chen Q et al (2014) Combining evolutionary information extracted from frequency profiles with sequence-based kernels for protein remote homology detection. *Bioinformatics (Oxford, England)* 30(4): 472–479. doi:10.1093/bioinformatics/btt709
- Liu B, Chen J, Wang X (2015) Protein remote homology detection by combining Chou's distance-pair pseudo amino acid composition and principal component analysis. *Mol Genet Genomics: MGG* 290(5):1919–1931. doi:10.1007/s00438-015-1044-4
- Makarova KS, Koonin EV, Kelman Z (2012) The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. *Biol Direct* 7(1):7. doi:10.1186/1745-6150-7-7
- Maulik U, Sarkar A (2013) Searching remote homology with spectral clustering with symmetry in neighborhood cluster kernels. *PLoS ONE* 8(2):e46468. doi:10.1371/journal.pone.0046468
- Meier A, Söding J (2015) Context similarity scoring improves protein sequence alignments in the midnight zone. *Bioinformatics (Oxford, England)* 31(5): 674–681. doi:10.1093/bioinformatics/btu697
- Mina JG, Okada Y, Wansadhipathi-Kannangara NK, Pratt S, Shams-Eldin H, Schwarz RT et al (2010) Functional analyses of differentially expressed isoforms of the Arabidopsis inositol phosphorylceramide synthase. *Plant Mol Biol* 73(4–5):399–407. doi:10.1007/s11103-010-9626-3
- Mirarab S, Nguyen N, Warnow T (2012) SEPP: SATé-enabled phylogenetic placement. In: Pacific symposium on biocomputing. Pacific symposium on biocomputing, pp. 247–258. doi:10.1142/9789814366496_0024

- Muda HM, Saad P, Othman RM (2011) Remote protein homology detection and fold recognition using two-layer support vector machine classifiers. *Comput Biol Med* 41(8):687–699. doi:10.1016/j.combiomed.2011.06.004
- Mudgal R, Sowdhamini R, Chandra N, Srinivasan N, Sandhya S (2014) Filling-in void and sparse regions in protein sequence space by protein-like artificial sequences enables remarkable enhancement in remote homology detection capability. *J Mol Biol* 426(4):962–979. doi:10.1016/j.jmb.2013.11.026
- Mudgal R, Sandhya S, Kumar G, Sowdhamini R, Chandra NR, Srinivasan N (2015) NrichD database: sequence databases enriched with computationally designed protein-like sequences aid in remote homology detection. *Nucleic Acids Res* 43(Database issue), D300–5. doi:10.1093/nar/gku888
- Murzin AG, Bateman A (1997) Distant homology recognition using structural classification of proteins. *Proteins Suppl* 1:105–112
- Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247(4):536–540. doi:10.1006/jmbi.1995.0159
- NCBI Resource Coordinators (2016) Database resources of the national center for biotechnology information. *Nucleic Acids Res* 44(D1):D7–D19. doi:10.1093/nar/gkv1290
- Nehrt NL, Clark WT, Radivojac P, Hahn MW (2011) Testing the ortholog conjecture with comparative functional genomic data from mammals. *PLoS Comput Biol* 7(6):e1002073. doi:10.1371/journal.pcbi.1002073
- Nelesen S, Liu K, Wang L-S, Linder CR, Warnow T (2012) DACTAL: divide-and-conquer trees (almost) without alignments. *Bioinformatics (Oxford, England)* 28(12): i274–82. doi:10.1093/bioinformatics/bts218
- Nishiyama T, Ladurner R, Schmitz J, Kreidl E, Schleiffer A, Bhaskara V et al (2010) Sororin mediates sister chromatid cohesion by antagonizing Wapl. *Cell* 143(5):737–749. doi:10.1016/j.cell.2010.10.031
- Östlund G, Schmitt T, Forslund K, Köstler T, Messina DN, Roopra S et al (2010) InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res* 38(Database issue), D196–203. doi:10.1093/nar/gkp931
- Ozlu N, Srayko M, Kinoshita K, Habermann B, O’toole ET, Müller-Reichert T et al (2005) An essential function of the *C. elegans* ortholog of TPX2 is to localize activated aurora A kinase to mitotic spindles. *Dev Cell* 9(2): 237–248. doi:10.1016/j.devcel.2005.07.002
- Pelletier L, Ozlu N, Hannak E, Cowan C, Habermann B, Ruer M et al (2004) The *Caenorhabditis elegans* centrosomal protein SPD-2 is required for both pericentriolar material recruitment and centriole duplication. *Current Biol: CB* 14(10):863–873. doi:10.1016/j.cub.2004.04.012
- Penel S, Arigon A-M, Dufayard J-F, Sertier A-S, Daubin V, Duret L et al (2009) Databases of homologous gene families for comparative genomics. *BMC Bioinform* 10 Suppl 6(Suppl 6), S3. doi:10.1186/1471-2105-10-S6-S3
- Penkett CJ, Morris JA, Wood V, Bähler J (2006) YOGY: a web-based, integrated database to retrieve protein orthologs and associated gene ontology terms. *Nucleic Acids Res* 34(Web Server issue), W330–4. doi:10.1093/nar/gkl311
- Perutz MF, ROSSMANN MG, CULLIS AF, MUIRHEAD H, WILL G, NORTH AC (1960) Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. *Nature* 185(4711), 416–422
- Powell S, Szklarczyk D, Trachana K, Roth A, Kuhn M, Muller J et al (2011) eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res* 40(D1):D284–D289. doi:10.1093/nar/gkr1060
- Proost S, Van Bel M, Vaneechoutte D, Van de Peer Y, Inzé D, Mueller-Roeber B, Vandepoele K (2015) PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Res* 43(Database issue), D974–81. doi:10.1093/nar/gku986
- Pryszcz LP, Huerta-Cepas J, Gabaldón T (2011) MetaPhOrs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score. *Nucleic Acids Res* 39(5):e32–e32. doi:10.1093/nar/gkq953

- Rabitsch KP, Gregan J, Schleiffer A, Javerzat J-P, Eisenhaber F, Nasmyth K (2004) Two fission yeast homologs of *Drosophila* Mei-S332 are required for chromosome segregation during meiosis I and II. *Current Biol*: CB 14(4):287–301. doi:10.1016/j.cub.2004.01.051
- Remmert M, Biegert A, Hauser A, Söding J (2012) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* 9(2):173–175. doi:10.1038/nmeth.1818
- Rost B (1999) Twilight zone of protein sequence alignments. *Protein Eng* 12(2):85–94
- Ruan J, Li H, Chen Z, Coghlan A, Coin LJM, Guo Y et al (2008) TreeFam: 2008 Update. *Nucleic Acids Res* 36(Database issue), D735–40. doi:10.1093/nar/gkm1005
- Sánchez-Díaz A, González I, Arellano M, Moreno S (1998) The Cdk inhibitors p25rum1 and p40SIC1 are functional homologues that play similar roles in the regulation of the cell cycle in fission and budding yeast. *J Cell Sci* 111(Pt 6):843–851
- Sandhya S, Mudgal R, Jayadev C, Abhinandan KR, Sowdhamini R, Srinivasan N (2012) Cascaded walks in protein sequence space: use of artificial sequences in remote homology detection between natural proteins. *Mol BioSyst* 8(8):2076–2084. doi:10.1039/c2mb25113b
- Schreiber F, Sonnhammer ELL (2013) Hieranoid: hierarchical orthology inference. *J Mol Biol* 425(11):2072–2081. doi:10.1016/j.jmb.2013.02.018
- Schwickart M, Havlis J, Habermann B, Bogdanova A, Camasses A, Oelschlaegel T et al (2004) Swm1/Apc13 is an evolutionarily conserved subunit of the anaphase-promoting complex stabilizing the association of Cdc16 and Cdc27. *Mol Cell Biol* 24(8):3562–3576. doi:10.1128/MCB.24.8.3562-3576.2004
- Sémon M, Wolfe KH (2007) Consequences of genome duplication. *Curr Opin Genet Dev* 17(6):505–512. doi:10.1016/j.gde.2007.09.007
- Shah AR, Oehmen CS, Webb-Robertson B-J (2008) SVM-HUSTLE—an iterative semi-supervised machine learning approach for pairwise protein remote homology detection. *Bioinformatics (Oxford, England)* 24(6): 783–790. doi:10.1093/bioinformatics/btn028
- Shevchenko A, Roguev A, Schaft D, Buchanan L, Habermann B, Sakalar C et al (2008) Chromatin Central: towards the comparative proteome by accurate mapping of the yeast proteomic environment. *Genome Biol* 9(11):R167. doi:10.1186/gb-2008-9-11-r167
- Shi G, Zhang L, Jiang T (2010) MSOAR 2.0: Incorporating tandem duplications into ortholog assignment based on genome rearrangement. *BMC Bioinform* 11(1):10. doi:10.1186/1471-2105-11-10
- Sinha S, Lynn AM (2014) HMM-Mode: implementation, benchmarking and validation with HMMER3. *BMC Res Notes* 7(1):483. doi:10.1186/1756-0500-7-483
- Söding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33(Web Server issue), W244–8. doi:10.1093/nar/gki408
- Söding J, Remmert M, Biegert A, Lupas AN (2006) HHSenser: exhaustive transitive profile search using HMM-HMM comparison. *Nucleic Acids Res* 34(Web Server issue), W374–8. doi:10.1093/nar/gkl195
- Sonnhammer ELL, Östlund G (2015) InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic acids research* 43(Database issue), D234–9. doi:10.1093/nar/gku1203
- Stingele J, Habermann B, Jentsch S (2015) DNA-protein crosslink repair: proteases as DNA repair enzymes. *Trends Biochem Sci* 40(2):67–71. doi:10.1016/j.tibs.2014.10.012
- Studer RA, Robinson-Rechavi M (2009) How confident can we be that orthologs are similar, but paralogs differ? *Trends Genet*: TIG 25(5):210–216. doi:10.1016/j.tig.2009.03.004
- Szklarczyk R, Wanschers BF, Cuypers TD, Esseling JJ, Riemersma M, van den Brand MA et al (2012) Iterative orthology prediction uncovers new mitochondrial proteins and identifies C12orf62 as the human ortholog of COX14, a protein involved in the assembly of cytochrome c oxidase. *Genome Biol* 13(2):R12. doi:10.1186/gb-2012-13-2-r12
- Szklarczyk R, Wanschers BFJ, Nijtmans LG, Rodenburg RJ, Zschocke J, Dikow N et al (2013) A mutation in the FAM36A gene, the human ortholog of COX20, impairs cytochrome c oxidase

- assembly and is associated with ataxia and muscle hypotonia. *Hum Mol Genet* 22(4):656–667. doi:10.1093/hmg/dd5473
- Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. *Science* (New York, N.Y.) 278(5338):631–637
- Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E (2009) EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* 19(2):327–335. doi:10.1101/gr.073585.107
- Vinga S, Almeida J (2003) Alignment-free sequence comparison—a review. *Bioinformatics* (Oxford, England) 19(4): 513–523
- Vogt G, Etzold T, Argos P (1995) An assessment of amino acid exchange matrices in aligning protein sequences: the twilight zone revisited. *J Mol Biol* 249(4):816–831. doi:10.1006/jmbi.1995.0340
- Wagner I, Volkmer M, Sharan M, Villaveces JM, Oswald F, Surendranath V, Habermann BH (2014) morFeus: a web-based program to detect remotely conserved orthologs using symmetrical best hits and orthology network scoring. *BMC Bioinform* 15(1):263. doi:10.1186/1471-2105-15-263
- Wang Y, Levy DE (2006) *C. elegans* STAT: evolution of a regulatory switch. *FASEB J: Official Publ Fed Am Soc Exp Biol* 20(10):1641–1652. doi:10.1096/fj.06-6051com
- Watson HC, Kendrew JC (1961) The amino-acid sequence of sperm whale myoglobin. Comparison between the amino-acid sequences of sperm whale myoglobin and of human hemoglobin. *Nature* 190:670–672
- Wieser D, Niranjana M (2009) Remote homology detection using a kernel method that combines sequence and secondary-structure similarity scores. *Silico Biol* 9(3):89–103
- Wolf YI, Koonin EV (2012) A tight link between orthologs and bidirectional best hits in bacterial and archaeal genomes. *Genome Biol Evol* 4(12):1286–1294. doi:10.1093/gbe/evs100
- Wu S, Zhang Y (2008) MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins* 72(2):547–556. doi:10.1002/prot.21945
- Yamada K, Tomii K (2014) Revisiting amino acid substitution matrices for identifying distantly related proteins. *Bioinformatics* (Oxford, England) 30(3): 317–325. doi:10.1093/bioinformatics/btt694
- Yang Y, Tantoso E, Li K-B (2008) Remote protein homology detection using recurrence quantification analysis and amino acid physicochemical properties. *J Theor Biol* 252(1):145–154. doi:10.1016/j.jtbi.2008.01.028
- Yona G, Levitt M (2002) Within the twilight zone: a sensitive profile-profile comparison tool based on information theory. *J Mol Biol* 315(5):1257–1275. doi:10.1006/jmbi.2001.5293
- Yu C, Desai V, Cheng L, Reifman J (2012) QuartetS-DB: a large-scale orthology database for prokaryotes and eukaryotes inferred by evolutionary evidence. *BMC Bioinform* 13(1):143. doi:10.1186/1471-2105-13-143
- Zhang Z, Schäffer AA, Miller W, Madden TL, Lipman DJ, Koonin EV, Altschul SF (1998) Protein sequence similarity searches using patterns as seeds. *Nucleic Acids Res* 26(17):3986–3990