

- Altenhoff AM, Boeckmann B, Capella-Gutierrez S, Dalquen DA, DeLuca T, Forslund K, Huerta-Cepas J, Linard B, Pereira C, Pryszcz LP, et al. 2016. Standardized benchmarking in the quest for orthologs. *Nat. Methods* 13:425–430.
- Altenhoff AM, Dessimoz C. 2009. Phylogenetic and functional assessment of orthologs inference projects and methods. *PLoS Comput. Biol.* 5:e1000262.
- Altenhoff AM, Dessimoz C. 2012. Inferring Orthology and Paralogy. In: Anisimova M, editor. *Evolutionary Genomics*. Vol. 855. Totowa, NJ: Humana Press. p. 259–279. Available from: http://link.springer.com/10.1007/978-1-61779-582-4_9
- Altenhoff AM, Schneider A, Gonnet GH, Dessimoz C. 2011. OMA 2011: orthology inference among 1000 complete genomes. *Nucleic Acids Res.* 39:D289–D294.
- Altenhoff AM, Škunca N, Glover N, Train C-M, Sueki A, Piližota I, Gori K, Tomiczek B, Müller S, Redestig H, et al. 2015. The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. *Nucleic Acids Res.* 43:D240-249.
- 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. Eisen JA, editor. *PLoS Comput. Biol.* 8:e1002514.
- Baptiste E, O'Malley MA, Beiko RG, Ereshefsky M, Gogarten JP, Franklin-Hall L, Lapointe F-J, Dupré J, Dagan T, Boucher Y, et al. 2009. Prokaryotic evolution and the tree of life are two different things. *Biol. Direct* 4:34.
- Battenberg K, Lee EK, Chiu JC, Berry AM, Potter D. 2017. OrthoReD: a rapid and accurate orthology prediction tool with low computational requirement. *BMC Bioinformatics [Internet]* 18. Available from: <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-017-1726-5>
- Boeckmann B, Robinson-Rechavi M, Xenarios I, Dessimoz C. 2011. Conceptual framework and pilot study to benchmark phylogenomic databases based on reference gene trees. *Brief. Bioinform.* 12:423–435.
- Chen F, Mackey AJ, Stoeckert CJ, Roos DS. 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res [Internet]* 34. Available from: <https://doi.org/10.1093/nar/gkj123>
- Chen F, Mackey AJ, Vermunt JK, Roos DS. 2007. Assessing Performance of Orthology Detection Strategies Applied to Eukaryotic Genomes. Fairhead C, editor. *PLoS ONE* 2:e383.
- Chen T, Wu TH, Ng WV, Lin W. 2010. DODO: an efficient orthologous genes assignment tool based on domain architectures. Domain based ortholog detection. *BMC Bioinformatics* 11 Suppl 7:S6.
- Chen X, Zhang J. 2012. The Ortholog Conjecture Is Untestable by the Current Gene Ontology but Is Supported by RNA Sequencing Data. Ouzounis CA, editor. *PLoS Comput. Biol.* 8:e1002784.
- Chiu JC, Lee EK, Egan MG, Sarkar IN, Coruzzi GM, DeSalle R. 2006. OrthologID: automation of genome-scale ortholog identification within a parsimony framework. *Bioinforma. Oxf. Engl.* 22:699–707.
- Dagan T, Artzy-Randrup Y, Martin W. 2008. Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc. Natl. Acad. Sci.* 105:10039–10044.

Dalquen DA, Altenhoff AM, Gonnet GH, Dessimoz C. 2013. The Impact of Gene Duplication, Insertion, Deletion, Lateral Gene Transfer and Sequencing Error on Orthology Inference: A Simulation Study. Carmel L, editor. PLoS ONE 8:e56925.

Datta RS, Meacham C, Samad B, Neyer C, Sjölander K. 2009. Berkeley PHOG: PhyloFacts orthology group prediction web server. Nucleic Acids Res. 37:W84-89.

Dessimoz C, Gabaldón T, Roos DS, Sonnhammer ELL, Herrero J, Quest for Orthologs Consortium. 2012. Toward community standards in the quest for orthologs. Bioinforma. Oxf. Engl. 28:900-904.

Doolittle WF. 2009. The practice of classification and the theory of evolution, and what the demise of Charles Darwin's tree of life hypothesis means for both of them. Philos. Trans. R. Soc. B Biol. Sci. 364:2221-2228.

Egan M, Lee EK, Chiu JC, Coruzzi G, Desalle R. 2009. Gene orthology assessment with OrthologID. Methods Mol. Biol. Clifton NJ 537:23-38.

Eisen JA. 1998. Phylogenomics: Improving Functional Predictions for Uncharacterized Genes by Evolutionary Analysis. Genome Res. 8:163-167.

Fitch WM. 1970. Distinguishing Homologous from Analogous Proteins. Syst. Zool. 19:99.

Fitch WM. 2000. Homology a personal view on some of the problems. Trends Genet. TIG 16:227-231.

Fujimoto MS, Suvorov A, Jensen NO, Clement MJ, Snell Q, Bybee SM. 2017. The OGCleaner: filtering false-positive homology clusters. Bioinformatics 33:125-127.

Gabaldón T, Dessimoz C, Huxley-Jones J, Vilella AJ, Sonnhammer EL, Lewis S. 2009. Joining forces in the quest for orthologs. Genome Biol. 10:403.

Gogarten JP, Doolittle WF, Lawrence JG. 2002. Prokaryotic Evolution in Light of Gene Transfer. Mol. Biol. Evol. 19:2226-2238.

Gout J-F, Duret L, Kahn D. 2009. Differential Retention of Metabolic Genes Following Whole-Genome Duplication. Mol. Biol. Evol. 26:1067-1072.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59:307-321.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696-704.

Hachiya T, Osana Y, Pependorf K, Sakakibara Y. 2009. Accurate identification of orthologous segments among multiple genomes. Bioinformatics 25:1474-1474.

Hahn MW. 2007. Bias in phylogenetic tree reconciliation methods: implications for vertebrate genome evolution. Genome Biol. 8:R141.

van der Heijden RT, Snel B, van Noort V, Huynen MA. 2007. Orthology prediction at scalable resolution by phylogenetic tree analysis. BMC Bioinformatics 8:83.

Huerta-Cepas J, Capella-Gutiérrez S, Pryszcz LP, Marcet-Houben M, Gabaldón T. 2014. PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome. Nucleic Acids Res. 42:D897-D902.

- Hughes AL, Friedman R. 2004. Differential loss of ancestral gene families as a source of genomic divergence in animals. *Proc. R. Soc. B Biol. Sci.* 271:S107–S109.
- Hulsen T, Huynen MA, de Vlieg J, Groenen PM. 2006. Benchmarking ortholog identification methods using functional genomics data. *Genome Biol.* 7:R31.
- Huynen MA, Bork P. 1998. Measuring genome evolution. *Proc Natl Acad Sci USA* [Internet] 95. Available from: <https://doi.org/10.1073/pnas.95.11.5849>
- Jothi R, Zotenko E, Tasneem A, Przytycka TM. 2006. COCO-CL: hierarchical clustering of homology relations based on evolutionary correlations. *Bioinforma. Oxf. Engl.* 22:779–788.
- Kaduk M, Sonnhammer E. 2017. Improved orthology inference with Hieranoid 2. *Bioinforma. Oxf. Engl.* 33:1154–1159.
- Koonin EV. 2005. Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* 39:309–338.
- Koonin EV. 2011. Obituary: Walter Fitch and the orthology paradigm. *Brief. Bioinform.* 12:377–378.
- Koonin EV, Wolf YI. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* 36:6688–6719.
- Koonin EV, Wolf YI. 2009. The fundamental units, processes and patterns of evolution, and the Tree of Life conundrum. *Biol. Direct* 4:33.
- Kristensen DM, Wolf YI, Mushegian AR, Koonin EV. 2011. Computational methods for Gene Orthology inference. *Brief. Bioinform.* 12:379–391.
- Kriventseva EV, Rahman N, Espinosa O, Zdobnov EM. 2007. OrthoDB: the hierarchical catalog of eukaryotic orthologs. *Nucleic Acids Res.* 36:D271–D275.
- Kuzniar A, van Ham RCHJ, Pongor S, Leunissen JAM. 2008. The quest for orthologs: finding the corresponding gene across genomes. *Trends Genet. TIG* 24:539–551.
- Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* 62:611–615.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13:2178–2189.
- Liu K, Linder CR, Warnow T. 2011. Multiple sequence alignment: a major challenge to large-scale phylogenetics. *PLoS Curr.* 2:RRN1198.
- Liu K, Raghavan S, Nelesen S, Linder CR, Warnow T. 2009. Rapid and Accurate Large-Scale Coestimation of Sequence Alignments and Phylogenetic Trees. *Science* 324:1561–1564.
- Lynch M, Katju V. 2004. The altered evolutionary trajectories of gene duplicates. *Trends Genet.* 20:544–549.
- Makarova KS. 2005. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res.* 33:4626–4638.
- Mirkin B, Muchnik I, Smith TF. 1995. A Biologically Consistent Model for Comparing Molecular Phylogenies. *J. Comput. Biol.* 2:493–507.

- Mushegian AR, Koonin EV. 1996. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. U. S. A.* 93:10268–10273.
- Nakaya A, Katayama T, Itoh M, Hiranuka K, Kawashima S, Moriya Y, Okuda S, Tanaka M, Tokimatsu T, Yamanishi Y, et al. 2013. KEGG OC: a large-scale automatic construction of taxonomy-based ortholog clusters. *Nucleic Acids Res.* 41:D353–D357.
- Nehrt NL, Clark WT, Radivojac P, Hahn MW. 2011. Testing the Ortholog Conjecture with Comparative Functional Genomic Data from Mammals. Rzhetsky A, editor. *PLoS Comput. Biol.* 7:e1002073.
- O'Connor T, Sundberg K, Carroll H, Clement M, Snell Q. 2010. Analysis of long branch extraction and long branch shortening. *BMC Genomics* 11:S14.
- Page RDM, Charleston MA. 1997. From Gene to Organismal Phylogeny: Reconciled Trees and the Gene Tree/Species Tree Problem. *Mol. Phylogenet. Evol.* 7:231–240.
- Petersen M, Meusemann K, Donath A, Dowling D, Liu S, Peters RS, Podsiadlowski L, Vasilikopoulos A, Zhou X, Misof B, et al. 2017. Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. *BMC Bioinformatics* 18:111.
- Peterson ME, Chen F, Saven JG, Roos DS, Babbitt PC, Sali A. 2009. Evolutionary constraints on structural similarity in orthologs and paralogs. *Protein Sci.* 18:1306–1315.
- Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, Huerta-Cepas J, Gabaldón T, Rattei T, Creevey C, Kuhn M, et al. 2014. eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.* 42:D231-239.
- Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. Poon AFY, editor. *PLoS ONE* 5:e9490.
- Puigbo P, Wolf YI, Koonin EV. 2010. The Tree and Net Components of Prokaryote Evolution. *Genome Biol. Evol.* 2:745–756.
- Remm M, Storm CE, Sonnhammer EL. 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol [Internet]* 314. Available from: <https://doi.org/10.1006/jmbi.2000.5197>
- Ruan J, Li H, Chen Z, Coghlan A, Coin LJM, Guo Y, Hériché J-K, Hu Y, Kristiansen K, Li R, et al. 2008. TreeFam: 2008 Update. *Nucleic Acids Res.* 36:D735-740.
- Salichos L, Rokas A. 2011. Evaluating Ortholog Prediction Algorithms in a Yeast Model Clade. Fairhead C, editor. *PLoS ONE* 6:e18755.
- Schliep K, Lopez P, Lapointe F-J, Baptiste E. 2011. Harvesting Evolutionary Signals in a Forest of Prokaryotic Gene Trees. *Mol. Biol. Evol.* 28:1393–1405.
- Schreiber F, Sonnhammer ELL. 2013. Hieranoid: Hierarchical Orthology Inference. *J. Mol. Biol.* 425:2072–2081.
- Shi G, Peng M-C, Jiang T. 2011. MultiMSOAR 2.0: an accurate tool to identify ortholog groups among multiple genomes. *PloS One* 6:e20892.
- Sjolander K. 2004. Phylogenomic inference of protein molecular function: advances and challenges. *Bioinformatics [Internet]* 20. Available from: <https://doi.org/10.1093/bioinformatics/bth021>

- Sonnhammer ELL, Gabaldón T, Sousa da Silva AW, Martin M, Robinson-Rechavi M, Boeckmann B, Thomas PD, Dessimoz C, Quest for Orthologs consortium. 2014. Big data and other challenges in the quest for orthologs. *Bioinforma. Oxf. Engl.* 30:2993–2998.
- Sonnhammer ELL, Östlund G. 2015. InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res.* 43:D234–239.
- Stamatakis A. 2015. Using RAxML to Infer Phylogenies. *Curr. Protoc. Bioinforma.* 51:6.14.1–14.
- Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science* 278:631–637.
- Tekaia. 2016. Inferring Orthologs: Open Questions and Perspectives. *Genomics Insights*:17.
- Thomas PD, Wood V, Mungall CJ, Lewis SE, Blake JA, on behalf of the Gene Ontology Consortium. 2012. On the Use of Gene Ontology Annotations to Assess Functional Similarity among Orthologs and Paralogs: A Short Report. Bourne PE, editor. *PLoS Comput. Biol.* 8:e1002386.
- Thompson JD, Linard B, Lecompte O, Poch O. 2011. A Comprehensive Benchmark Study of Multiple Sequence Alignment Methods: Current Challenges and Future Perspectives. Badger J, editor. *PLoS ONE* 6:e18093.
- Trachana K, Larsson TA, Powell S, Chen W-H, Doerks T, Muller J, Bork P. 2011. Orthology prediction methods: A quality assessment using curated protein families. *BioEssays* 33:769–780.
- Ullah I, Sjöstrand J, Andersson P, Sennblad B, Lagergren J. 2015. Integrating Sequence Evolution into Probabilistic Orthology Analysis. *Syst. Biol.* 64:969–982.
- 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.* 19:327–335.
- Wall DP, Fraser HB, Hirsh AE. 2003. Detecting putative orthologs. *Bioinforma. Oxf. Engl.* 19:1710–1711.
- Waterhouse RM, Zdobnov EM, Tegenfeldt F, Li J, Kriventseva EV. 2011. OrthoDB: the hierarchical catalog of eukaryotic orthologs in 2011. *Nucleic Acids Res.* 39:D283–D288.
- Wolf YI, Novichkov PS, Karev GP, Koonin EV, Lipman DJ. 2009. The universal distribution of evolutionary rates of genes and distinct characteristics of eukaryotic genes of different apparent ages. *Proc. Natl. Acad. Sci.* 106:7273–7280.
- Yu C, Zavaljevski N, Desai V, Reifman J. 2011. QuartetS: a fast and accurate algorithm for large-scale orthology detection. *Nucleic Acids Res.* 39:e88–e88.
- Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, Seppey M, Loetscher A, Kriventseva EV. 2017. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* 45:D744–D749.
- Zmasek CM, Eddy SR. 2001. A simple algorithm to infer gene duplication and speciation events on a gene tree. *Bioinformatics* 17:821–828.