Model organism database

Model organism databases (**MODs**) are biological databases, or knowledgebases, dedicated to the provision of in-depth biological data for intensively studied model organisms. MODs allow researchers to easily find background information on large sets of genes, plan experiments efficiently, combine their data with existing knowledge, and construct novel hypotheses.[1][2] They allow users to analyse results and interpret datasets, and the data they generate are increasingly used to describe less well studied species.[1] Where possible, MODs share common approaches to collect and represent biological information. For example, all MODs use the Gene Ontology (GO)[3][4] to describe functions, processes and cellular locations of specific gene products. Projects also exist to enable software sharing for curation, visualization and querying between different MODs.[5] Organismal diversity and varying user requirements however mean that MODs are often required to customize capture, display, and provision of data.[1]

Types of data and services

Model organism databases generate, source and collate species-specific information integratively by combining expert knowledge with literature curation and bioinformatics.

Services provided to biological research communities include:

- Genome sequence annotations
  - Location of genes and regulatory regions in the genome
- Functional curation of gene products
  - Discern functions fulfilled by the gene product by looking at a variety of data including Gene Ontology (GO) annotations, phenotypes, gene expression, pathway information
- Protein/RNA sequence annotations
- Anatomical information
- Stock centres

- Orthology

List of model organism databases

Common name	Scientific name	Wikipedia page	Database link-out
Baker's yeast	*Saccharomyces cerevisiae*	Saccharomyces Genome Database	SGD (http://www.yeastgenome.org)[6]
Fission yeast	*Schizosaccharomyces pombe*	PomBase	PomBase (http://www.pombase.org)[7][8][9][10]
Clawed frog	*Xenopus*	Xenbase	Xenbase (http://www.xenbase.org/entry/)[11][12]
Sea urchins, starfish, etc.	Echinodermata	Echinobase	Echinobase (http://www.echinobase.org/entry/)[13]
Fruitfly	*Drosophila melanogaster*	FlyBase	FlyBase (http://flybase.org/)[14]
Bees, wasps, ants	Hymenoptera	Hymenoptera Genome Database	HGD (https://hymenoptera.elsiklab.missouri.edu/)[15]
Mouse	*Mus musculus*	Mouse Genome Informatics	MGI (http://www.informatics.jax.org/)[16]
Nematode	*Caenorhabditis elegans*	WormBase	WormBase (https://wormbase.org/)[17]
Rat	*Rattus norvegicus*	Rat Genome Database	RGD (http://rgd.mcw.edu/)[18]
Social amoeba	*Dictyostelium discoideum*	DictyBase	dictyBase (https://dictycr.org)[19]
Ciliate	*Tetrahymena thermophila*	Tetrahymena Genome Database	TGD (http://ciliate.org/index.php/home/welcome)
Thale cress	*Arabidopsis thaliana*	The Arabidopsis Information Resource	TAIR (https://www.arabidopsis.org/)[20]
Maize	*Zea mays ssp. mays*	-	MaizeGDB (http://www.maizegdb.org/)[21][22]
Soybean	*Glycine soja*	SoyBase	SoyBase (https://soybase.org/)[23]
Zebrafish	*Danio rerio*	Zebrafish Information Network	ZFIN (http://zfin.org/)[24]
-	*Candida albicans*	-	CGD (http://www.candidagenome.org/)[25]
-	*Escherichia coli*	EcoCyc	EcoCyc (http://ecocyc.org/)[26]
Hay bacillus	*Bacillus subtilis*	-	*Subti*Wiki (http://subtiwiki.uni-goettingen.de/)[27]

References

1. Oliver SG, Lock A, Harris MA, Nurse P, Wood V (June 2016). "Model organism databases: essential resources that need the support of both funders and users" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4918006). *BMC Biology*. **14** (1): 49. doi:10.1186/s12915-016-0276-z (https://doi.org/10.1186%2Fs12915-016-0276-z). PMC 4918006 (https://www.ncbi.nlm.nih.g

ov/pmc/articles/PMC4918006). PMID 27334346 (https://pubmed.ncbi.nlm.nih.gov/27334346).

2. Bond M, Holthaus SM, Tammen I, Tear G, Russell C (November 2013). "Use of model organisms for the study of neuronal ceroid lipofuscinosis" (https://doi.org/10.1016%2Fj.bbadis.2013.01.009). *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*. **1832** (11): 1842–65. doi:10.1016/j.bbadis.2013.01.009 (https://doi.org/10.1016%2Fj.bbadis.2013.01.009). PMID 23338040 (https://pubmed.ncbi.nlm.nih.gov/23338040).
3. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. (May 2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3037419). *Nature Genetics*. **25** (1): 25–9. doi:10.1038/75556 (https://doi.org/10.1038%2F75556). PMC 3037419 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3037419). PMID 10802651 (https://pubmed.ncbi.nlm.nih.gov/10802651).
4. Gene Ontology Consortium (January 2015). "Gene Ontology Consortium: going forward" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4383973). *Nucleic Acids Research*. **43** (Database issue): D1049-56. doi:10.1093/nar/gku1179 (https://doi.org/10.1093%2Fnar%2Fgku1179). PMC 4383973 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4383973). PMID 25428369 (https://pubmed.ncbi.nlm.nih.gov/25428369).
5. O'Connor BD, Day A, Cain S, Arnaiz O, Sperling L, Stein LD (2008). "GMODWeb: a web framework for the Generic Model Organism Database" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2481422). *Genome Biology*. **9** (6): R102. doi:10.1186/gb-2008-9-6-r102 (https://doi.org/10.1186%2Fgb-2008-9-6-r102). PMC 2481422 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2481422). PMID 18570664 (https://pubmed.ncbi.nlm.nih.gov/18570664).
6. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, et al. (January 2012). "Saccharomyces Genome Database: the genomics resource of budding yeast" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245034). *Nucleic Acids Research*. **40** (Database issue): D700-5. doi:10.1093/nar/gkr1029 (https://doi.org/10.1093%2Fnar%2Fgkr1029). PMC 3245034 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245034). PMID 22110037 (https://pubmed.ncbi.nlm.nih.gov/22110037).
7. Lock A, Rutherford K, Harris MA, Hayles J, Oliver SG, Bähler J, Wood V (January 2019). "PomBase 2018: user-driven reimplementation of the fission yeast database provides rapid and intuitive access to diverse, interconnected information" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6324063). *Nucleic Acids Research*. **47** (D1): D821–D827. doi:10.1093/nar/gky961 (https://doi.org/10.1093%2Fnar%2Fgky961). PMC 6324063 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6324063). PMID 30321395 (https://pubmed.ncbi.nlm.nih.gov/30321395).
8. Wood V, Harris MA, McDowall MD, Rutherford K, Vaughan BW, Staines DM, et al. (January 2012). "PomBase: a comprehensive online resource for fission yeast" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245111). *Nucleic Acids Research*. **40** (Database issue): D695-9. doi:10.1093/nar/gkr853 (https://doi.org/10.1093%2Fnar%2Fgkr853). PMC 3245111 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245111). PMID 22039153 (https://pubmed.ncbi.nlm.nih.gov/22039153).
9. McDowall MD, Harris MA, Lock A, Rutherford K, Staines DM, Bähler J, et al. (January 2015). "PomBase 2015: updates to the fission yeast database" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4383888). *Nucleic Acids Research*. **43** (Database issue): D656-61. doi:10.1093/nar/gku1040 (https://doi.org/10.1093%2Fnar%2Fgku1040). PMC 4383888 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4383888). PMID 25361970 (https://pubmed.ncbi.nlm.nih.gov/25361970).

10. Lock A, Rutherford K, Harris MA, Wood V (2018). "PomBase: The Scientific Resource for Fission Yeast". *Eukaryotic Genomic Databases*. Methods in Molecular Biology. Vol. 1757. pp. 49–68. doi:10.1007/978-1-4939-7737-6_4 (https://doi.org/10.1007%2F978-1-4939-7737-6_4). ISBN 978-1-4939-7736-9. PMC 6440643 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6440643). PMID 29761456 (https://pubmed.ncbi.nlm.nih.gov/29761456).
11. Karimi K, Fortriede JD, Lotay VS, Burns KA, Wang DZ, Fisher ME, et al. (January 2018). "Xenbase: a genomic, epigenomic and transcriptomic model organism database" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5753396). *Nucleic Acids Research*. **46** (D1): D861–D868. doi:10.1093/nar/gkx936 (https://doi.org/10.1093%2Fnar%2Fgkx936). PMC 5753396 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5753396). PMID 29059324 (https://pubmed.ncbi.nlm.nih.gov/29059324).
12. James-Zorn C, Ponferrada VG, Fisher ME, Burns KA, Fortriede JD, Segerdell E, Karimi K, Lotay VS, Wang DZ, Chu S, Pells TJ, Wang Y, Vize PD, Zorn AM (May 2018). "Navigating Xenbase: An Integrated Xenopus Genomics and Gene Expression Database". *Eukaryotic Genomic Databases*. pp. 251–305. doi:10.1007/978-1-4939-7737-6_10 (https://doi.org/10.1007%2F978-1-4939-7737-6_10). ISBN 978-1-4939-7736-9. PMC 6853059 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6853059). PMID 29761462 (https://pubmed.ncbi.nlm.nih.gov/29761462). {{cite book}}: |journal= ignored (help)
13. Arshinoff, Bradley I.; Cary, Gregory A.; Karimi, Kamran; Foley, Saoirse; Agalakov, Sergei; Delgado, Francisco; Lotay, Vaneet S.; Ku, Carolyn J.; Pells, Troy J.; Beatman, Thomas R.; Kim, Eugene; Cameron, R Andrew; Vize, Peter D.; Telmer, Cheryl A; Croce, Jenifer C.; Ettensohn, Charles A.; Hinman, Veronica F. (2022). "Echinobase: Leveraging an extant model organism database to build a knowledgebase supporting research on the genomics and biology of echinoderms" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8728261). *Nucleic Acids Research*. **50** (D1): D970–D979. doi:10.1093/nar/gkab1005 (https://doi.org/10.1093%2Fnar%2Fgkab1005). PMC 8728261 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8728261). PMID 34791383 (https://pubmed.ncbi.nlm.nih.gov/34791383).
14. Attrill H, Falls K, Goodman JL, Millburn GH, Antonazzo G, Rey AJ, Marygold SJ (January 2016). "FlyBase: establishing a Gene Group resource for Drosophila melanogaster" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4702782). *Nucleic Acids Research*. **44** (D1): D786-92. doi:10.1093/nar/gkv1046 (https://doi.org/10.1093%2Fnar%2Fgkv1046). PMC 4702782 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4702782). PMID 26467478 (https://pubmed.ncbi.nlm.nih.gov/26467478).
15. Elsik CG, Tayal A, Unni DR, Burns GW, Hagen DE (2018). "Hymenoptera Genome Database: Using HymenopteraMine to Enhance Genomic Studies of Hymenopteran Insects". In Kollmar M (ed.). *Eukaryotic Genomic Databases*. Methods in Molecular Biology. Vol. 1757. New York, NY: Springer New York. pp. 513–556. doi:10.1007/978-1-4939-7737-6_17 (https://doi.org/10.1007%2F978-1-4939-7737-6_17). ISBN 978-1-4939-7736-9. PMID 29761469 (https://pubmed.ncbi.nlm.nih.gov/29761469).
16. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE (January 2015). "The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4384027). *Nucleic Acids Research*. **43** (Database issue): D726-36. doi:10.1093/nar/gku967 (https://doi.org/10.1093%2Fnar%2Fgku967). PMC 4384027 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4384027). PMID 25348401 (https://pubmed.ncbi.nlm.nih.gov/25348401).
17. Harris TW, Baran J, Bieri T, Cabunoc A, Chan J, Chen WJ, et al. (January 2014). "WormBase 2014: new views of curated biology" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3965043). *Nucleic Acids Research*. **42** (Database issue): D789-93. doi:10.1093/nar/gkt1063 (https://doi.org/10.1093%2Fnar%2Fgkt1063). PMC 3965043 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3965043). PMID 24194605 (https://pubmed.ncbi.nlm.nih.gov/24194605).

18. Shimoyama M, De Pons J, Hayman GT, Laulederkind SJ, Liu W, Nigam R, et al. (January 2015). "The Rat Genome Database 2015: genomic, phenotypic and environmental variations and disease" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4383884). *Nucleic Acids Research*. **43** (Database issue): D743-50. doi:10.1093/nar/gku1026 (https://doi.org/10.1093%2Fnar%2Fgku1026). PMC 4383884 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4383884). PMID 25355511 (https://pubmed.ncbi.nlm.nih.gov/25355511).
19. Kreppel L, Fey P, Gaudet P, Just E, Kibbe WA, Chisholm RL, Kimmel AR (January 2004). "dictyBase: a new Dictyostelium discoideum genome database" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC308872). *Nucleic Acids Research*. **32** (Database issue): D332-3. doi:10.1093/nar/gkh138 (https://doi.org/10.1093%2Fnar%2Fgkh138). PMC 308872 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC308872). PMID 14681427 (https://pubmed.ncbi.nlm.nih.gov/14681427).
20. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, et al. (January 2012). "The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245047). *Nucleic Acids Research*. **40** (Database issue): D1202-10. doi:10.1093/nar/gkr1090 (https://doi.org/10.1093%2Fnar%2Fgkr1090). PMC 3245047 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245047). PMID 22140109 (https://pubmed.ncbi.nlm.nih.gov/22140109).
21. Lawrence CJ, Dong Q, Polacco ML, Seigfried TE, Brendel V (January 2004). "MaizeGDB, the community database for maize genetics and genomics" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC308746). *Nucleic Acids Research*. **32** (Database issue): D393-7. doi:10.1093/nar/gkh011 (https://doi.org/10.1093%2Fnar%2Fgkh011). PMC 308746 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC308746). PMID 14681441 (https://pubmed.ncbi.nlm.nih.gov/14681441).
22. Andorf CM, Cannon EK, Portwood JL, Gardiner JM, Harper LC, Schaeffer ML, et al. (January 2016). "MaizeGDB update: new tools, data and interface for the maize model organism database" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4702771). *Nucleic Acids Research*. **44** (D1): D1195-201. doi:10.1093/nar/gkv1007 (https://doi.org/10.1093%2Fnar%2Fgkv1007). PMC 4702771 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4702771). PMID 26432828 (https://pubmed.ncbi.nlm.nih.gov/26432828).
23. Grant D, Nelson RT, Cannon SB, Shoemaker RC (January 2010). "SoyBase, the USDA-ARS soybean genetics and genomics database" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2808871). *Nucleic Acids Research*. **38** (Database issue): D843-6. doi:10.1093/nar/gkp798 (https://doi.org/10.1093%2Fnar%2Fgkp798). PMC 2808871 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2808871). PMID 20008513 (https://pubmed.ncbi.nlm.nih.gov/20008513).
24. Howe DG, Bradford YM, Conlin T, Eagle AE, Fashena D, Frazer K, et al. (January 2013). "ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3531097). *Nucleic Acids Research*. **41** (Database issue): D854-60. doi:10.1093/nar/gks938 (https://doi.org/10.1093%2Fnar%2Fgks938). PMC 3531097 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3531097). PMID 23074187 (https://pubmed.ncbi.nlm.nih.gov/23074187).
25. Inglis DO, Arnaud MB, Binkley J, Shah P, Skrzypek MS, Wymore F, et al. (January 2012). "The Candida genome database incorporates multiple Candida species: multispecies search and analysis tools with curated gene and protein information for Candida albicans and Candida glabrata" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245171). *Nucleic Acids Research*. **40** (Database issue): D667-74. doi:10.1093/nar/gkr945 (https://doi.org/10.1093%2Fnar%2Fgkr945). PMC 3245171 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245171). PMID 22064862 (https://pubmed.ncbi.nlm.nih.gov/22064862).

26. Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martínez C, et al. (January 2013). "EcoCyc: fusing model organism databases with systems biology" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3531154). *Nucleic Acids Research*. **41** (Database issue): D605-12. doi:10.1093/nar/gks1027 (https://doi.org/10.1093%2Fnar%2Fgks1027). PMC 3531154 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3531154). PMID 23143106 (https://pubmed.ncbi.nlm.nih.gov/23143106).
27. Zhu B, Stülke J (January 2018). "SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism Bacillus subtilis" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5753275). *Nucleic Acids Research*. **46** (D1): D743–D748. doi:10.1093/nar/gkx908 (https://doi.org/10.1093%2Fnar%2Fgkx908). PMC 5753275 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5753275). PMID 29788229 (https://pubmed.ncbi.nlm.nih.gov/29788229).
