# Supplementary Information

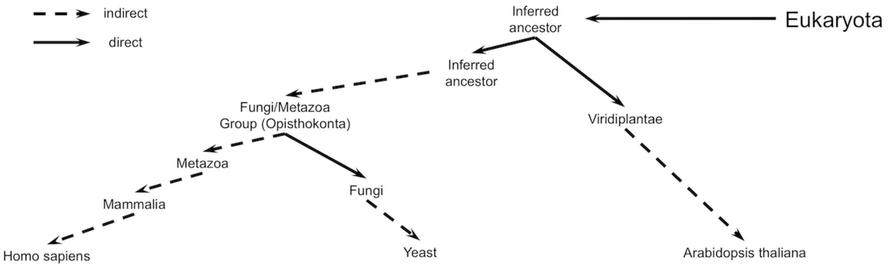
# A daily-updated tree of (sequenced) life as a reference for genome research

Hai Fang[1,*], Matt Oates[1], Ralph Pethica[1], Jenny Greenwood[2], Adam Sardar[1], Owen Rackham[1], Alexandros Stamatakis[3], David A. de Lima Morais[1,4] & Julian Gough[1,*]

[1]Department of Computer Science, University of Bristol, The Merchant Venturers Building, Bristol BS8 1UB, UK; [2]School of Earth Sciences, University of Bristol, Wills Memorial Building, Bristol BS8 1RJ, UK; [3]The Exelixis Lab, Scientific Computing Group, Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnenweg 35, D-69118, Heidelberg, Germany; [4]Centre de Calcul Scientifique, Université de Sherbrooke, Sherbrooke, Quebec J1K 2R1, Canada.

*Correspondence and requests for materials should be addressed to J.G. (gough@cs.bris.ac.uk) or H.F. (hfang@cs.bris.ac.uk).

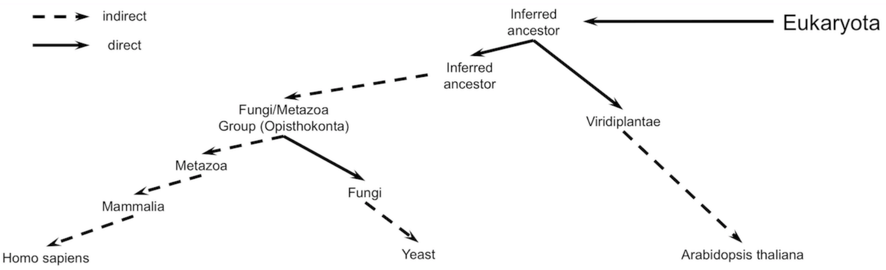
This supplementary information includes two supplementary figures.



| SCOP id[a] (sunid) | SCOP description[a] | Homo sapiens [9606[b]] | Mammalia [40674[b]] | Metazoa [33208[b]] | Fungi/Metazoa group [33154[b]] | Fungi [4751[b]] | Yeast [4932[b]] | Inferred ancestor | Inferred ancestor | Viridiplantae [33090[b]] | Arabidopsis thaliana [3702[b]] | Eukaryota [2759[b]] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Histone-lysine N-methyltransferase (2.1.1.43) from Enzyme Commission (EC)[c]* | | | | | | | | | | | | |
| 48403 | Ankyrin repeat | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 57667 | beta-beta-alpha zinc fingers | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 47370 | Bromodomain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 54160 | Chromo domain-like | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 47676 | Conserved domain common to transcription factors TFIIS, elongin A, CRSP70 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 54171 | DNA-binding domain | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 |
| 57903 | FYVE/PHD zinc finger | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 82185 | Histone H3 K4-specific methyltransferase SET7/9 N-terminal domain | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 |
| 47095 | HMG-box | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 109640 | KRAB domain (Kruppel-associated box) | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 88697 | PUA domain-like | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 54928 | RNA-binding domain, RBD | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 81822 | RuBisCo LSMT C-terminal, substrate-binding domain | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 |
| 82199 | SET domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 48452 | TPR-like | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 63748 | Tudor/PWWP/MBT | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 51045 | WW domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| | **Total:** | **17** | **17** | **16** | **16** | **16** | **13** | **16** | **16** | **16** | **16** | **13** |
| *Oxidation (KW:0558) from the Post-Translational Modificaiton KeyWord (KW) in UniprotKB[d]* | | | | | | | | | | | | |
| 53067 | Actin-like ATPase domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 58113 | Apolipoprotein A-I | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 82936 | Apolipoprotein A-II | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 52317 | Class I glutamine amidotransferase-like | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 57038 | Cyclotides | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 51905 | FAD/NAD(P)-binding domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 55424 | FAD/NAD-linked reductases, dimerisation (C-terminal) domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 49695 | gamma-Crystallin-like | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 |
| 49764 | HSP20-like chaperones | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 56209 | Nitrile hydratase alpha chain | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 |
| | **Total:** | **8** | **8** | **7** | **8** | **8** | **6** | **8** | **8** | **7** | **6** | **8** |
| *Hormone metabolism (UPA00437) from UniPathway (UP)[e]* | | | | | | | | | | | | |
| 46785 | Winged helix DNA-binding domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 55729 | Acyl-CoA N-acyltransferases (Nat) | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 75304 | Amidase signature (AS) enzymes | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 51197 | Clavaminate synthase-like | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 57535 | Complement control module/SCR domain | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 |
| 48264 | Cytochrome P450 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 47473 | EF-hand | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 57196 | EGF/Laminin | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 |
| 54373 | FAD-linked reductases, C-terminal domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 51905 | FAD/NAD(P)-binding domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 52343 | Ferredoxin reductase-like, C-terminal NADP-linked domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 48113 | Heme-dependent peroxidases | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 51735 | NAD(P)-binding Rossmann-fold domains | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 53383 | PLP-dependent transferases | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 63380 | Riboflavin synthase domain-like | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 53335 | S-adenosyl-L-methionine-dependent methyltransferases | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 49879 | SMAD/FHA domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 48239 | Terpenoid cyclases/Protein prenyltransferases | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 48576 | Terpenoid synthases | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 53756 | UDP-Glycosyltransferase/glycogen phosphorylase | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | **Total:** | **20** | **20** | **20** | **20** | **19** | **18** | **20** | **20** | **20** | **19** | **18** |

[a] A list of *superfamilies* annotated by the term as indicated below.
[b] NCBI Taxonomy identifiers.
[c] Available at http://supfam.org/SUPERFAMILY/cgi-bin/dcbo.cgi?type=EC;po=2.1.1.43;focus=SUPRA.
[d] Available at http://supfam.org/SUPERFAMILY/cgi-bin/dcbo.cgi?type=KW;po=KW:0558;focus=SUPRA.
[e] Available at http://supfam.org/SUPERFAMILY/cgi-bin/dcbo.cgi?type=UP;po=UPA00437;focus=SUPRA.

**Figure S1. The exemplified use of sTOL for studying the evolution of enzymes, post-translational modifications and metabolism.** The diagram in the top panel shows the paths covering three kingdoms. The bottom panel lists the details of their presence (1) and absence (0) patterns at the major branching points of eukaryotic evolution.



| SCOP id[a] (sunid) | SCOP Term[a] | Homo sapiens [9606[b]] | Mammalia [40674[b]] | Metazoa [33208[b]] | Fungi/Metazoa group [33154[b]] | Fungi [4751[b]] | Yeast [4932[b]] | Inferred ancestor | Inferred ancestor | Viridiplantae [33090[b]] | Arabidopsis thaliana [3702[b]] | Eukaryota [2759[b]] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Proto-oncogene (KW:0656) from the Disease KeyWord (KW) in UniprotKB[c]* | | | | | | | | | | | | |
| 46785 | Winged helix DNA-binding domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 47454 | A DNA-binding domain in eukaryotic transcription factors | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 103657 | BAR/IMD domain-like | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 81995 | beta-sandwich domain of Sec23/24 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 54277 | CAD & PB1 domains | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 57889 | Cysteine-rich domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 50353 | Cytokine | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 |
| 48065 | DBL homology domain (DH-domain) | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 81296 | E set domains | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 47459 | HLH, helix-loop-helix DNA-binding domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 48726 | Immunoglobulin | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 |
| 57959 | Leucine zipper domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 50729 | PH domain-like | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 103575 | Plexin repeat | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 56112 | Protein kinase-like (PK-like) | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 46955 | Putative DNA-binding domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 90209 | Ran binding protein zinc finger-like | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 47769 | SAM/Pointed domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 63763 | SAND domain-like | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |
| 101912 | Sema domain | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 55550 | SH2 domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 50044 | SH3-domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 54236 | Ubiquitin-like | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 51045 | WW domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| | **Total:** | **24** | **24** | **24** | **23** | **19** | **18** | **22** | **21** | **20** | **19** | **15** |
| *Oncogene (KW:0553 from the Disease KeyWord (KW) in UniprotKB[d]* | | | | | | | | | | | | |
| 46565 | Chaperone J-domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 56672 | DNA/RNA polymerases | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 161229 | E6 C-terminal domain-like | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 161234 | E7 C-terminal domain-like | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 55464 | Origin of replication-binding domain, RBD-like | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 |
| 52540 | P-loop containing nucleoside triphosphate hydrolases | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 56112 | Protein kinase-like (PK-like) | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 55550 | SH2 domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 161240 | T-antigen specific domain-like | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 50494 | Trypsin-like serine proteases | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | **Total:** | **6** | **6** | **6** | **7** | **7** | **6** | **7** | **7** | **7** | **6** | **5** |
| *Embryonic lethality (MP:0008762) from Mouse/Mammalian Phenotype (MP)[e]* | | | | | | | | | | | | |
| 55729 | Acyl-CoA N-acyltransferases (Nat) | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 75632 | Cullin homology domain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 55979 | DNA clamp | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 52029 | GroEL apical domain-like | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 81901 | HCP-like | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 49417 | p53-like transcription factors | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 56366 | SMAD MH1 domain | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | **Total:** | **7** | **7** | **7** | **6** | **6** | **6** | **5** | **5** | **5** | **5** | **5** |

[a] A list of *superfamilies* annotated by the term as indicated below.

[b] NCBI Taxonomy identifiers.

[c] Available at http://supfam.org/SUPERFAMILY/cgi-bin/dcbo.cgi?type=KW;po=KW:0656;focus=SUPRA.

[d] Available at http://supfam.org/SUPERFAMILY/cgi-bin/dcbo.cgi?type=KW;po=KW:0553;focus=SUPRA.

[e] Available at http://supfam.org/SUPERFAMILY/cgi-bin/dcbo.cgi?type=MP;po=MP:0008762;focus=SUPRA.

**Figure S2. The exemplified use of sTOL for studying the evolution of diseases and phenotypes.** The diagram in the top panel shows the paths covering three kingdoms. The bottom panel lists the details of their presence (1) and absence (0) patterns at the major branching points of eukaryotic evolution.