Table S1. Electronic resources related to the structural, sequence, and functional analysis of the Nudix superfamily (DOI: 10.6078/D1CC74).

Resource Number	Title	Type	Explanation	Digital Object Identifier (DOI)
1	mysql_dump_char+nudix_0 4062014.sql	SQL	MySQL database dump file. The database contains experimental data on Nudix proteins with journal references. Refer to Resource 2 for detailed information of its structure.	10.6078/d17p4v
2	The layout of the Nudix MySQL database.pdf	PDF	Schematic schema of MySQL database containing Nudix protein names, relevant experimental data and their source references. Each box indicates a MySQL table. The arrows represent the foreign key linkages between tables. The data entries are stored as kinetic, screening or genetic data, each with distinct table organizations. Each data is linked with a GO term. The references are stored as either PubMed IDs or DOI addresses.	10.6078/d13w2x
3	Nudix data collection.xlsx	Excel	Summary of Nudix enzyme's functions collected from the literature. All experimental data for Nudix proteins were categorized into six different evidence types (represented by six sheets of this file): 1) kcat/Km, 2) kcat, 3) Km, 4) relative activity, 5) other biochemical activity (<i>e.g.</i> gel electrophoresis, HPLC, X-ray crystal structure), and 6) genetic evidence (<i>e.g.</i> knockouts, complementation). All data are compiled in a master sheet "all data" where each row corresponds to one protein activity pair. Additional rows are included where a given activity was investigated more than once. This file also has a sheet "GO terms" for all the Gene Ontology terms associated with at least one activity.	10.6078/d1059d
4	Initial score assignment.xlsx	Excel	To describe confidence in function assignment of Nudix proteins, six evidence types were first classified as either biochemical or genetic, and biochemical evidence were further categorized into four subcategories: 1) kcat/Km, 2) pseudo kcat/Km (explained below), 3) relative activity, and 4) others (represented by five sheets of this file). All data points were assigned a score based on our scoring scheme (Table 4). In each sheets, the data and their corresponding scores are shown side-by-side. There is an additional sheet "kk used for pseudo kk", which contains a set of protein activities that had both their kcat/Km values and relative activities determined. In such cases, the kcat/Km value was used to approximate pseudo kcat/Km values for other relative activities in the same screening group.	10.6078/d1vc7g
5	Final score assignment.xlsx	Excel	To describe confidence in function assignment of Nudix proteins, the two sheets "biochemical scores" and "genetic scores" combine all initial score assignments that belong to the two categories. Within the category, the maximum score of a given protein function was designated as $S_{biochem}$ and $S_{genetic}$ in the sheet "overall scores". The overall score was calculated by the equation: $S_{overall} = 1 - (1 - S_{genetic}) \times (1 - S_{biochem})$. This score was adjusted to	10.6078/d1qp46

6	generate_score_biochem_ge netic.R	R script	take into account the abundance of annotations for a given enzyme using an R script (explained below). The sheet "final scores" shows the score before (Soverall) and after the adjustment (Sfinal). This script combines biochemical and genetic confidence scores to yield and integrated score. Two sheets, "score_biochem.csv" (Resource 7) and "score_genetic.csv" (Resource 8) are used as the input files of the R script "generate_score_biochem_genetic.R", which yields an output file, "score_biochem_genetic.csv" (Resource 9). Soverall is calculated as an additional column in the sheet, and the result is saved as "score_biochem_genetic soverall.csv" (Resource 10).	10.6078/d1kw28
7	score_biochem.csv	csv	Input file for Resource 6; derived from Resource 5.	<u>10.6078/d1g59r</u>
8	score_genetic.csv	csv	Input file for Resource 6; derived from Resource 5.	10.6078/d1bc7t
9	score_biochem_genetic.csv	csv	Output file for Resource 6	10.6078/d16p4j
10	score_biochem_genetic_sov erall.csv	csv	Output file for Resource 6 and input file for Resource 11	10.6078/d1301k
11	generate_score_final.R	R script	The script "generate_score_final.R" is run to compute the final adjusted scores saved as "score_final.csv" from "score biochem genetic soverall.csv"	10.6078/d1z593
12	score_final.csv	csv	Output file for Resource 11, with score S[final]	10.6078/d1tg6t
13	GOA analysis results.xlsx	Excel	Comparison between our literature-based study of Nudix functions and those in GOA. The sheet "GOA_Comparison" shows the final comparison result between our Nudix database and UniProt GOA (release 2013-12-11). The main content in this sheet is reprinted as Table 5 in the main text of the paper. The sheet "Statistics" contains information to derive the GOA-related statistics mentioned in the abstract, introduction, and result part of the main text. The sheet "MySQL_040614" contains the same protein-function-score tuples as in the "Final Score Assignment" Excel file, and in addition provides curated function assignments (Columns D and E) resulting from the GOA comparison. The three sheets "UniProt_GOA_121113-EXP/NONEXP/NONF" contains all GO assignments to Nudix proteins in GOA release 2013-12-11. The sheet "GO_Terms" contains the complete proposed and current GO terms used in this paper to describe the Nudix protein collection; the main content in this sheet is reprinted as Tables 1 and 2. The sheet "GO_Hierarchy_proposed" contains the proposed GO hierarchy, and is reprinted as Figure S2. Finally, the sheet "GO_Hierarchy_010114" is an archive of the existed GO hierarchy for Nudix-related terms as of 2014-01-01.	10.6078/d1pp4w
14	Pipelines to build sequence alignments of the Nudix	PDF	Graphical representation of the steps used to reconstruct sequence alignments of the Nudix superfamily, as described in the Materials and	10.6078/D1K01X

	superfamily.pdf		Methods section. (A) The pipeline to build the 78-PDB structure guided sequence alignment. (B) The pipeline to build the 324-core sequence alignment guided by the 78-PDB sequence alignment. (C) The pipeline to build the alignment of the complete Nudix clan (38,950 sequences). (D) Illustration of how to combine two alignment into one guided by a scaffold alignment.	
15	3DCOMB alignment.py	Python script	The file "3DCOMB alignment.py" is a structural alignment of 78 proteins with PDB entires; can be opened with Chimera v1.6.2 ⁵² .	10.6078/D1F59F
16	46-PDB alignment.fa	FASTA	46-PDB sequence alignment in FASTA format; can be opened by most sequence alignment visualization software such as Jalview ⁵³ .	10.6078/D19G65
17	78-PDB alignment.fa	FASTA	76-PDB sequence alignment in FASTA format; can be opened by most sequence alignment visualization software such as Jalview ⁵³ .	10.6078/D15P47
18	340 select nudix proteins.xlsx	Excel	The Excel file "340 select nudix proteins.xlsx" contains information of 340 select Nudix proteins, the subsets of which comprise of all alignments discussed in the paper except the "Nudix-clan alignment". The information related to the PDB structures is reprinted in Table 3. Columns K, L, M and N indicates the starting and ending residues of domains used in the alignments. For each protein, the best-characterized activity with the highest confidence score, when available, is recorded in Columns O, P and Q; such information was obtained from Resource 4.	10.6078/D12018
19	324-core alignment.fa	FASTA	324-core sequence alignment in FASTA format; can be opened by most sequence alignment visualization software such as Jalview ⁵³ .	10.6078/D1X59S
20	347-select alignment.fa	FASTA	347-select sequence alignment in FASTA format can be opened by most sequence alignment visualization software such as Jalview ⁵³ .	10.6078/D1SG6H
21	Nudix-clan alignment.fa	FASTA	Full Nudix clan alignment in FASTA format; can be opened by most sequence alignment visualization software such as Jalview ⁵³ .	10.6078/D1NP4K
22	347-select tree.newick	Newick	Phylogenetic tree of the 347-select proteins. It can be opened by most phylogeny visualization software such as Dendroscope v3.2.8 ⁶⁶ .	10.6078/D1J01M
23	347-select tree reconciled.dendro	Dendrosc ope	Reconciled tree of proteins in the 347-select set. The files in Dendroscope format have the same tree topologies as the Newick format, and additionally contain rich information such as reconciliation, bootstrap and so on; these files can be opened by earlier version of Dendroscope (v3.2.8 or earlier) ⁶⁶ .	10.6078/D1D594
24	Nudix-clan tree.newick	Newick	Phylogenetic tree of the entire Nudix clan. It can be opened by most phylogeny visualization software such as Dendroscope v3.2.8 ⁶⁶ .	10.6078/D18G6V
25	Nudix-clan tree reconciled.dendro	Dendrosc ope	Reconciled tree of the entire Nudix clan. The files in Dendroscope format have the same tree topologies as the Newick format, and additionally contain rich information such as reconciliation, bootstrap and so on; these files can be opened by earlier version of Dendroscope (v3.2.8 or earlier) ⁶⁶ .	10.6078/D14S3K
26	Figure 5 347-select tree high resolution.pdf	PDF	Original high-resolution PDF used to generate Figure 5 in the main text of the paper, with the 347-select tree. It can be zoomed-in to a high degree to	10.6078/D1101Z

			visualize the details of the trees.	
27	Figure S4 Nudix-clan tree high resolution.pdf	PDF	Original high-resolution PDF used to generate Figure S4 in the main text of the paper, with the entire Nudix clan. It can be zoomed-in to a high degree to visualize the details of the trees.	10.6078/D1W884

Figure S1

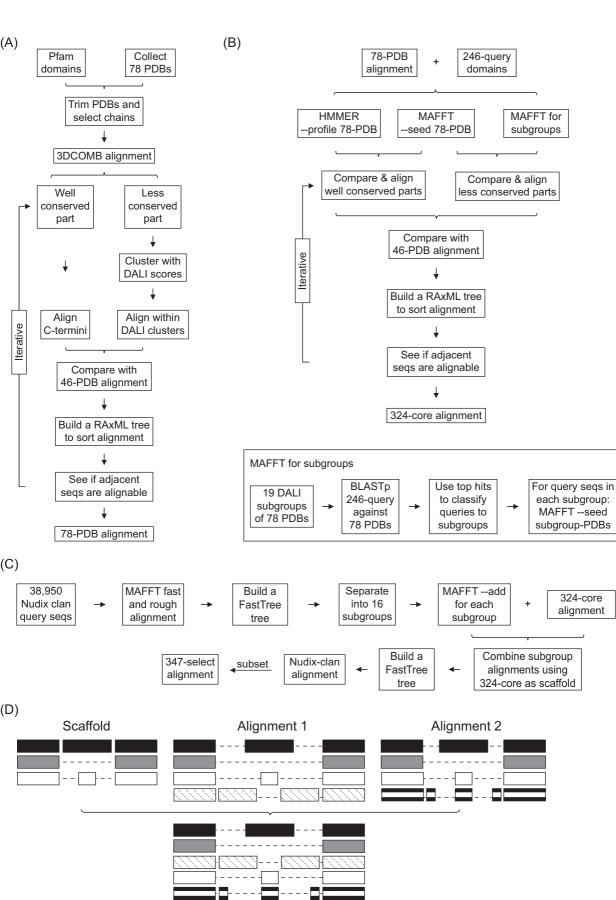


Figure S1. Pipelines to build sequence alignments of the Nudix superfamily. The detailed steps are described in the Materials and Methods section. (A) The pipeline to build the 78-PDB structure guided sequence alignment. (B) The pipeline to build the 324-core sequence alignment guided by the 78-PDB sequence alignment. (C) The pipeline to build the alignment of the complete Nudix clan (38,950 sequences). (D) Illustration of how to combine two alignment into one guided by a scaffold alignment.

534 oxidized purir	ne nucleobase lesion	DNA N-glycosylas	o activity see activity activity (product undefined) hatbase activity (Ps yielding) activity activity activity activity	a name changed
A001 F62 pyrophosphat	2-hydroxy-adeni tase activity	ne DNA N-glycosly	ase activity	a name changed definition changed parent changed parent added
A086 A087	oxythiamine-dipl thiamine triphos	osphatase activity thate phosphatase	activity (product undefined)	* merged f proposed
1 A107	UDP-hexanolam	thiamine-triphosp ine diphosphatase	hafase activity (Pl yielding) activity dishorabelman activity	
A127	CDP-ethanolami CDP-choline dip	ne diphosphatase hosphatase activity	activity	
GO:00002 GO:00008	98 endopolyphosph 10 diacylglycerol di	atase activity shosphate phosph	stase activity	
GO:00021 GO:00043 GO:00044	45 4-amino-5-hydro 109 exopolyphospha 27 inomanic dinhos	xymethyl-2-methyl tase activity nhatase activity	pyrimidine diphosphatase activity	
GO:00002 GO:00021 GO:00043 GO:00047 GO:00087	87 thiamine-pyroph 15 CDP-diacylglyce	osphatase activity rol diphosphatase	diphosphatasa activity stasa activity stasa activity stasa activity stasa activity stasa activity stasa activity	
GO:00096 GO:00191 GO:00431	78 hydrogen-transic 77 dihydroneopterir 35 5-phosphorihosy	cating pyrophosph triphosphate pyro I 1.nyrophosphate	adasa adikity ahasa hadikity phosphohydrolase adikity (PPI yielding) pyrophosphatase adikity diphosphatase adikity y	
GO:00474 GO:00477	30 oligosaccharide- 34 CDP-glycerol dip	diphosphodolichol hosphatase activit	y opiniosphalase activity V	
GO:00478 GO:00503 GO:00503				
' AP001	nucleoside-polyp A004	hosphate phosphi 2-hydroxy-deoxy	by fase activity denosine triphosphate phosphatase activity (product undefine triphosphate phosphatase activity (product undefined) 8-cm or Up even-production activity (P) yielding) 8-cm of UP even-production activity (PP) yielding)	1)
	f A005	8-oxo-guanosine A006	triphosphate phosphatase activity (product undefined) 8-oxo-guanosine triphosphatase activity (Pi yielding)	
	f A008	5-methyl-deoxyc	(tidine triphosphate phosphatase activity (product undefined)	
	f A009 f A010 f A011	5-methyl-uridine 5-hydroxy-cytidin	triphosphate phosphatase activity (product undefined) e triphosphate phosphatase activity (product undefined) (produ	ct undefined)
	f A014		e triphosphate phosphatase activity (product undefined) (produ cytidine triphosphate phosphatase activity (product undefined) adenosine triphosphate phosphatase activity (product undefine	f) lyby (PPI violding)
	' A016	8-oxo-deoxyguar = A015	osine triphosphate phosphatase activity (product undefined) 8-oxo-deoxyguanosine triphosphatase activity (Pi yielding)	
	A029 A030	arabinofuranosyl adenosine tripho	cytosine triphosphate phosphatase activity (product undefined) sphate phosphatase activity (product undefined)	
	f A034		ATPase activity (Pl yielding) ATP diphosphatase activity (PPI yielding) nate phosphatase activity (product undefined)	
			CTP pyrophosphatase activity (PPI yielding) CTPase activity (PI yielding)	
	1 A055		triphosphate phosphatase activity (product undefined) dATP pyrophosphohydrolase activity (PPI yielding) shosphate phosphatase activity (product undefined)	
	AUSI			
	f A060		triphosphate phosphatase activity (product undefined) dGTP diphosphatase activity (PPI yielding) ate phosphatase activity (product undefined)	
	1 A063		UTP diphosphatase activity (PPI yielding) hosphate phosphatase activity (PPI yielding)	
	f A076		deoxyuridine triphosphatase activity (stepwise) sphate phosphatase activity (product undefined)	
	' A080	inosine triphosph == GO:0036220	ate phosphatase activity (product undefined)	
	A082 A088	N4-methyl-decxy adenosine tetrap f Apro	TIP diprocipitates activity (PPI yesting) yildine triphosphates phosphates activity (product undefined) hosphate phosphates activity (product undefined) adenosine testinphosphates activity (AMP yielding) adenosine testinphosphates activity (ATP yielding) adenosine testinphosphates activity (product undefined) phosphate phosphates activity (product undefined) phosphate phosphates activity (product undefined) adenosine pentanbosphate hosphates activity (the product undefined)	
	A090	≃ GO:0047624 guanosine tetrao	adenosine tetraphosphatase activity (ARP yielding) hosphate phosphatase activity (product undefined)	
	1 A090 1 A091 1 A099	adenosine penta f A136	phosphate phosphatase activity (product undefined) adenosine pentaphosphate phosphatase activity (ATP yieldin; phate phosphatase activity (product undefined)	1)
	' A099	A100	phate phosphatase activity (product undefined) thymidine inphosphatase activity (stepwise) TTD-new perfects (50 stellation)	
	^c GO:0017110	nucleoside-dipho	TTP diphosphatase activity (PPI yielding) sphatase activity	
		' A094	TTP diphosphatase activity (PF) yelding) sphatase activity guanosine 3.5°-bis(diphosphate) diphosphate phosphatase ac 1° A095 guanosine 3.5°-bis(diphosphate) diphosphate thymidine-diphosphatase activity cylidine diphosphatase activity cylidine diphosphatase activity adenosine-diphosphatase activity adenosine-diphosphatase activity	tivity (product undefined) lase activity (stepwise)
		GO:0004382 GO:0036384	guanosine-diphosphatase activity cylidine diphosphatase activity	
	≃ GO:0017111	GO:0090450 nucleoside-tripho	undine-diphosphatase activity inosine-diphosphatase activity sohatase activity (PI vieldino)	
		GO:0090450 nucleoside-tripho d A006 A015 d A058 d A064 a A077 d A100 d A113	S-con-GEP phosphatase activity unifine-disposphatase activity unifine-disposphatase activity incrine-diprosphatase activity S-con-decayaments triphosphatase activity (FI yielding) S-con-decayaments triphosphatase activity (FI yielding) decayardine triphosphatase activity (tipspile) decayardine triphosphatase activity (singuise) decayardine triphosphatase activity (singuise) tripmidine triphosphatase activity (singuise) unifine triphosphatase activity (singuise) unifine triphosphatase activity (singuise)	
		" A058 " A064 " A077	deoxycytidine triphosphalase activity (stepwise) deoxyuridine triphosphalase activity (stepwise) numerine triphosphalase activity (stepwise)	
			motor activity GTPase activity (Pi yielding) helicase activity	
		E GO:0008894 E GO:0016887	helicase activity guanosine-5-4 riphosphate. 3-diphosphate diphosphatase activ APPase activity (P) yielding) TiPase activity (P) yielding) TiPase activity (P) yielding) TiPase activity (P) yielding) Siphate diphosphatase regulator activity sophate diphosphatase activity (PP) yielding) Sembly-diseny/glidine (pihosphatase) Sembly-diseny/glidine (pihosphatase)	ity (Pi yielding)
		GO:0036310 ad GO:0043273 ad GO:0050339 GO:0060589	annealing helicase activity CTPase activity (Pi yielding)	
	[∞] GO:0047429	GO:0050339 GO:0060589 nucleoside-tripho	TTPase activity (Pt yielding) nucleoside-triphosphatase regulator activity sohate diphosphatase activity (PPI yielding)	
	00.0041423		5-methyl-deoxycytidine triphosphate pyrophosphatase activity 8-hydroxy-adenosine triphosphate pyrophophatase activity (Pl	(PPi yielding) Pi yielding)
		4 A013 A061 A131	8-hydroxy-adenosine triphosphate pyrophophatase activity (PI 8-hydroxy-deoxyadenosine triphosphate pyrophosphatase act CTP pyrophosphatase activity (PPI yielding) 3'-amino-3'-dATP pyrophosphatase activity (PPI yielding)	ivity (PPI yielding)
		A132 GO:0004170	CTP printingarbase activity (PP) yelden) "Tameso-TTP printingarbase activity (PP) yelden) "Tameso-TTP printingarbase activity (PP) yelden) dTP deplocationase activity (PP) yelden) dTP deplocationase activity (PP) yelden) dTP optionationase activity (PP) yelden) dTP printingarbase activity (PP) yelden) dTP printingarbase activity (PP) yelden) dTP printingarbase activity (PP) yelden) TTP deplocationase activity (PP) yelden) TTP deplocationase activity (PP) yelden) UTP deplocationase activity (PP) yelden) UTP deplocationase activity (PP) yelden) 2-OH-GATP proprincipations activity (PP) yelden) dTP deplocationase activity (PP) yelden) deplocationase activity deplocationase activity deplocationase activity deplocationase activity deplocationase activity descriptional deplocationase activity descriptional deplocationase activity	
		≈ GO:0004636 ad GO:0008413	phosphoribosyl-ATP diphosphatase activity (PPI yielding) 8-oxo-GTP pyrophosphatase activity (PPI yielding)	
		ac GO:0008828 ac GO:0035539 a GO:0035870	dATP pyrophosphohydrolase activity (PPi yielding) 8-oxo-dGTP pyrophosphatase activity (PPi yielding)	
		ad GO:0036217 ad GO:0036218	dTP diphosphatase activity (PPI yielding) dGPP diphosphatase activity (PPI yielding) TTP diphosphatase activity (PPI yielding)	
		ad GO:0036219 ad GO:0036220	GTP diphosphatase activity (PPI yielding) ITP diphosphatase activity (PPI yielding)	
		ad GO:0036221 a GO:0036222	UTP diphosphatase activity (PPi yielding) XTP diphosphatase activity (PPi yielding)	
		© GO:0044714	2-OH-ATP pyroprospiration activity (PP) yielding) ATP diphysphatase activity (PP) yielding)	
	6 GO:0097382	at GO:0047840 deaxynucleoside	dCTP diphosphatase activity (PPI yielding) diphosphatase activity	
		1 A002 1 A035 1 A056	2-hydroxy-deoxyadenosine diphosphatase activity deoxyadenosine diphosphatase activity deoxyadenosine diphosphatase activity	
		¹ A059 ² GO:0044715	deoxyguanosine diphosphatase activity 8-oxo-dGDP phosphatase activity 8-hydroxy-dADP phosphatase activity dIDP diphosphatase activity	
f AP007	CDP-sugar diph	GO:0044717 GO:0097383 sohatase activity	8-hydroxy-dADP phosphatase activity dIDP diphosphatase activity	
	A031	CDP-glucose dip CDP-ribose diph	hosphatase activity osphatase activity	
AP008	GDP-sugar diph 1 A068 1 A069	osphatase activity GDP-beta-fucose	hosphatase activity orphatase activity diphosphatase activity hosphatase activity hosphatase activity rosphatase activity pophatase activity phosphatase activity phosphatase activity	
	' A070 ' A071 ** GO:0052751	GDP-ribciose dip GDP-ribose diph	hosphalase activity sochalase activity	
f AP009				
AP011	1 A116 general coenzyn 1 A036	TDP-glucose dip ne A diphosphatas fatty-acid-acyl-Ci	hosphatase activity a activity	
	A054 A096	choloyl-CoA diph	osphatase activity	
	f A098	trihydroxycopros 3-methyl-3-hydro	lanoyl-CoA diphosphatase activity xyglutaryl-CoA diphosphatase activity	
	A114	3'-dephospho-Co	ocoprolates activity annyl-CoA diphosphalase activity syglutanyl-CoA diphosphalase activity hosphalase activity A diphosphalase activity gliphosphalase activity sphalase activity sphalase activity sphalase activity sphalase activity sphalase activity	
	A114 A118 A120 A122			
	A120 A122 A125 A128			
¹ AP013	f A120 f A122 f A125 f A128 GO:0010945 RNA decapping	CoA pyrophosph	atase activity	
¹ AP013	A120 A122 A125 A128 G0:0010945 RNA decapping A065 A066 A067	CoA pyrophosph	atase activity	
¹ AP013	A120 A122 A125 A125 A128 GO:0010945 RNA decapping A065 A066 A067	CoA pyrophosph activity m7G(5')ppp-mRNA G(5')ppp-mRNA G(5')ppp-mRNA	atase activity	
f AP014	* A120 * A122 * A125 * A125 * A128 * GO:0010945 * RNA decapping * A065 * A067 * A081 * GO:0050072 * GO:0034353	CoA pyrophosph activity m7G(5')ppp-mRNA G(5')ppp-mRNA m7G(5')ppp-snoi m7G(5')ppp-snoi m7G(5')ppp-snoi	taken activity 44. diphosphatase activity (mTGDP yielding) iriphosphatase activity (GMP yielding) iriphosphatase activity (GMP yielding) iriphosphatase activity (GDP yielding) 44. diphosphatase activity (mTGMP yielding) 44. diphosphatase activity (mTGMP yielding) 45. diphosphatase activity (mTGMP yielding)	
	* A120 * A122 * A125 * A125 * A128 * GO:0010945 * RNA decapping * A065 * A067 * A081 * GO:0050072 * GO:0034353	CGA pyrophosph activity m7G(5')ppp-mRNA G(5')ppp-mRNA G(5')ppp-mRNA M7G(5')ppp-mRNA M7G(5')ppp-mRNA RNA pyrophosph sphatase activity IDP-ribose dipho phosphate phospi dinucloside-diplo	atasa activity (A Giphosphatase activity (Int'GDP yielding) piprosphatase activity (CMP yielding) piprosphatase activity (CMP yielding) piprosphatase activity (CMP yielding) (A Giphosphatase activity (CMP yielding) (A Giphosphatase activity (MTCMP yielding) ohydrolase activity polydrolase activity polydrolase activity polydrolase activity polydrolase activity polydrolase activity	
f AP014	* A120 * A122 * A125 * A125 * A128 * GO:0010945 * RNA decapping * A065 * A067 * A081 * GO:0050072 * GO:0034353	CGA pyrophosph activity m7G(5')ppp-mRNA G(5')ppp-mRNA G(5')ppp-mRNA M7G(5')ppp-mRNA M7G(5')ppp-mRNA RNA pyrophosph sphatase activity IDP-ribose dipho phosphate phospi dinucloside-diplo	atters activity Al, diphosphatase activity (anTGDF violiding) proposphatase activity (Completing) proposphatase activity (COMP violiding) proposphatase activity (COMP violiding) PAM triphosphatase activity (COMP violiding) Ald phosphatase activity (COMP violiding) objectives activity proposphatase activity postates activity postates activity postates activity postates activity postates activity	
¹ AP014	F A120 F A122 F A125 F A128 GO:0010945 RNA decapping F A065 A066 A067 GO:0050072 GO:0034353	CoA pyrophosph activity m7G(5')ppp-mRNA G(5')ppp-mRNA m7G(5')ppp-snoi m7G(5')ppp-snoi m7G(5')ppp-snoi	atters activity of Aphyladorus activity (mTODP-yeledino) propositional activity (mTODP-yeledino) propositional activity (mTODP-yeledino) activity (m	
¹ AP014	F A120 F A122 F A125 F A128 GO:0010945 RNA decapping F A065 A066 A067 GO:0050072 GO:0034353	CGA pyrophosph activity m7G(5')ppp-mRNA G(5')ppp-mRNA G(5')ppp-mRNA M7G(5')ppp-mRNA M7G(5')ppp-mRNA RNA pyrophosph sphatase activity IDP-ribose dipho phosphate phospi dinucloside-diplo	attes activity (Indiana) activit	
f AP014	F A120 F A122 F A125 F A128 GO:0010945 RNA decapping F A065 A066 A067 GO:0050072 GO:0034353	CGA pyrophosph activity m7G(5')ppp-mRNA G(5')ppp-mRNA G(5')ppp-mRNA M7G(5')ppp-mRNA M7G(5')ppp-mRNA RNA pyrophosph sphatase activity IDP-ribose dipho phosphate phospi dinucloside-diplo	attes activity (Indiana) activit	
¹ AP014	A 120 A 122 A 122 A 122 A 123 A 123 A 128 GO 0010945 RNA decapping A 066 A 066 A 066 A 067 A 067	CoA pyrophosph activity m7G(5')ppp-mR1A G(5')ppp-mR1A m7G(5')ppp-mR1A m7G(5')ppp-mR1A m7G(5')ppp-mR1A pyrophosph phatase activity IDP-dbose dipho phatase activity IDP-dbose dipho dinucleoside-dipl 1' A017 1' A017 1' A012	attes activity (Indiana) activit	
¹ AP014	1 A120 1 A122 1 A122 1 A122 1 A122 1 A123 1 A128 1 GO-0010945 RNA decapping 1 A066 1 A066 1 A066 1 A066 1 A061 2 GO-005072 1 A070 1 A070 1 A7002	CoA pyrocic Light Coat of the	atters activity ("IOP" yelloting) (A phonositissis as schilly (IOTOP yelloting) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP) (IOP)	divity (product undefined)
¹ AP014	A 120 A 122 A 122 A 122 A 123 A 123 A 128 GO 0010945 RNA decapping A 066 A 066 A 066 A 067 A 067	CoA precise up to the control of the	attes activity of the control of the	
¹ AP014	1 A120 1 A122 1 A122 1 A122 1 A122 1 A123 1 A128 1 GO-0010945 RNA decapping 1 A066 1 A066 1 A066 1 A066 1 A061 2 GO-005072 1 A070 1 A070 1 A7002	CA pyrophosph cachily prophosph cachily pp-m2G(p)pp-m2N, G(f)pp-m2N, G(f)pp-m2	attes activity of the control of the	
¹ AP014	1 A120 1 A122 1 A122 1 A122 1 A122 1 A123 1 A128 1 GO-0010945 RNA decapping 1 A066 1 A066 1 A066 1 A066 1 A061 2 GO-005072 1 A070 1 A070 1 A7002	CoA pyrophosph caching in a cac	states activity (midDP-yeldino) A (phonositiase activity (miDDP-yeldino) phonopiniase activity (miDDP-yeldino) A (phonopiniase activity (miDDP-yeldino) A (phonopiniase activity (miDDP-yeldino) A (phonopiniase activity A (phonopiniase activity postales activity activit	t undefined) ADP yielding) AMP yielding) e activity (product undefin t undefined)
f AP014	1 AP003 1 AP006 1 AP006 1 AP006 1 AP006 1 AP006	CoA pyrophosphacultury of the control of the contro	states activity (miCDP yielding) A physical state activity (miCDP yielding) propognitists activity (miCDP yielding) propognitists activity (miCDP yielding) propognitists activity (miCDP yielding) displantation activity (miCDP yielding) displantation activity proposition activity togic state of the state of t	t undefined) ADP yielding) AMP yielding) e activity (product undefined) t undefined) ADP yielding)
f AP014	1 AP003 1 AP006 1 AP006 1 AP006 1 AP006 1 AP006	CoA pyrophosphacultury of the control of the contro	states activity (miCDP yielding) A physical state activity (miCDP yielding) propognitists activity (miCDP yielding) propognitists activity (miCDP yielding) propognitists activity (miCDP yielding) displantation activity (miCDP yielding) displantation activity proposition activity togic state of the state of t	t undefined) ADP yielding) AMP yielding) e activity (product undefined) t undefined) ADP yielding)
f AP014	1 AP003 1 AP006 1 AP006 1 AP006 1 AP006 1 AP006	CoA pyrophosphic CoA py	states activity ("IOTOP yeledino) composition of the composition of th	t undefined) ADP yielding) AMP yielding) e activity (product undefined) t undefined) ADP yielding) ATP yielding) AMP yielding) a activity (product undefined)
¹ AP014	1 AP003 1 AP006 1 AP006 1 AP006 1 AP006 1 AP006	CoA pyrophosphic CoA py	states activity ("IOTOP yeledino) composition of the composition of th	t undefined) ADP yielding) AMP yielding) e activity (product undefined) t undefined) ADP yielding) ATP yielding) AMP yielding) a activity (product undefined)
f AP014	1 AP003 1 AP006 1 AP006 1 AP006 1 AP006 1 AP006	CoA pyrophosphic CoA py	states activity ("IOTOP yeledino) composition of the composition of th	t undefined) ADP yielding) AMP yielding) e activity (product undefined) t undefined) ADP yielding) ATP yielding) AMP yielding) a activity (product undefined)
¹ AP014	1 AP003 1 AP006 1 AP006 1 AP006 1 AP006 1 AP006	CoA pyrophosphic CoA py	states activity ("IOTOP yeledino) composition of the composition of th	t undefined) ADP yielding) AMP yielding) e activity (product undefined) t undefined) ADP yielding) ATP yielding) AMP yielding) a activity (product undefined)
* GO:00048	A120 A120	CoA pryrephosphate Coal Private Privat	states activity of the control of th	t undefined) ADP yielding) AMP yielding) e activity (product undefined) t undefined) ADP yielding) ATP yielding) AMP yielding) a activity (product undefined)
* GO:00048	A1002	CoA pyrephosphatics of the control o	states activity (InTOP) yielding) A diprocedulates activity (InTOP) yielding) propognitus activity (InTOP) yielding) A diprocedulates activity (IntoP) yielding) (IntoP) yielding	Landelmed ADP yelding) ADP yelding) ADP yelding) a calvely (repolar) ADP yelding) ADP yelding)
* AP014 * GO:00045	A1002	CoA pyrephosphate CoA pyrephosphate Policy CoA	states activity (InTOP) yielding) A diprocedulates activity (InTOP) yielding) propognitus activity (InTOP) yielding) A diprocedulates activity (IntoP) yielding) (IntoP) yielding	Landelmed ADP yelding) ADP yelding) ADP yelding) a calvely (repolar) ADP yelding) ADP yelding)
* AP014 * GC:00045	A1002	CoA pyrephosphate CoA pyrephosphate Policy CoA	states activity (InTOP) yielding) A diprocedulates activity (InTOP) yielding) propognitus activity (InTOP) yielding) A diprocedulates activity (IntoP) yielding) (IntoP) yielding	Landelmed ADP yelding) ADP yelding) ADP yelding) a calvely (repolar) ADP yelding) ADP yelding)
[†] AP014 * GD:00045	A120 A120	CoA pyrephosphate CoA pyrephosphate Policy CoA	states activity (InTOP) yielding) A diprocedulates activity (InTOP) yielding) propognitus activity (InTOP) yielding) A diprocedulates activity (IntoP) yielding) (IntoP) yielding	Landelmed ADP yelding) ADP yelding) ADP yelding) a calvely (repolar) ADP yelding) ADP yelding)
* AP014 * GO:00045	AP003	CAA psymphosphase activity of the control of the co	states activity (InTOP) yielding) A physical state activity (InTOP) yielding) physical states activity (InTOP) yielding) A physical state activity yielding activity yielding activity yielding activity yielding activity yielding activity Into activity	Is medined ADP yelding) ADP yelding) ADP yelding) ADP yelding) a earthyl (product undefine t undefined) ADP yelding) ADP y
¹ AP014 * GG:00045	APO03	CoA prepared to the control of the c	states activity of COP yelloting companies activity (mode) of COP yelloting companies activity (mode) yelloting companies yelloting companies activity (mode) yelloting companies yelloting yelloting companies yelloting yelloting companies yelloting ye	Is medined ADP yelding) ADP yelding) ADP yelding) a calvily (product undefined undefined) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) activity (product undefined product undefined activity (ADP yelding) they (product undefined product undefined activity (ADP yelding) undefined) undefined) undefined)
[†] AP014 * GD:00045	A1203	CoA psymptomegh CoA psymptomeg	states activity (military states) A diphoralisates activity (military states) A diphoralisates activity (military states) A diphoralisates activity (military states) A states activity (militar	Is medined ADP yelding) ADP yelding) ADP yelding) a calvily (product undefined undefined) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) activity (product undefined product undefined activity (ADP yelding) they (product undefined product undefined activity (ADP yelding) undefined) undefined) undefined)
[†] AP014 * GD:00045	AP003	CoA psychopany CoA ps	states activity (military states) A diphoralisates activity (military states) A diphoralisates activity (military states) A diphoralisates activity (military states) A states activity (militar	Is medined ADP yelding) ADP yelding) ADP yelding) a calvily (product undefined undefined) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) activity (product undefined product undefined activity (ADP yelding) they (product undefined product undefined activity (ADP yelding) undefined) undefined) undefined)
[†] AP014 * GD:00045	AP003	CoA psychopany CoA ps	states activity (military states) A diphoralisates activity (military states) A diphoralisates activity (military states) A diphoralisates activity (military states) A states activity (militar	Is medined ADP yelding) ADP yelding) ADP yelding) ADP yelding) a earthyl (product undefine t undefined) ADP yelding) ADP y
[†] AP014 * GG:00045	AP003	CoA psychopany CoA ps	states activity (military states) A diphoralisates activity (military states) A diphoralisates activity (military states) A diphoralisates activity (military states) A states activity (militar	Is medined ADP yelding) ADP yelding) ADP yelding) ADP yelding) a earthyl (product undefine t undefined) ADP yelding) ADP y
GO:00087	APODS	CoA psymptomegical control of the co	states activity (military states) A physical states activity (military states) A physical states activity (military states) A physical states activity (military states) A states activity (mili	Is medined ADP yelding) ADP yelding) ADP yelding) a calvily (product undefined undefined) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) ADP yelding) activity (product undefined product undefined activity (ADP yelding) they (product undefined product undefined activity (ADP yelding) undefined) undefined) undefined)
GO:00087	APODS	CoA psymptomegh CoA psymptomeg	states activity (InTOP) yielding (InTOP)	Is medined ADP yelding) ADP yelding) ADP yelding) ADP yelding) a earthyl (product undefine t undefined) ADP yelding) ADP y

Figure S2. Proposed Gene Ontology hierarchy. This region of the classification is a strict hierarchy, not a DAG, except as indicated below. New terms are begun with either A (terms with experimental support) or AP (parent terms proposed only for the structure of the hierarchy). The relationships between parent and child terms are represented with indents. The proposed changes, when applicable, are indicated by superscripts in front of GO term IDs: a – change name; b – change definition; c – change parent; d – add parent; e – remove; f – newly proposed. When a term has two parents, the term is shown in black under one parent and grey in the other. The removed term is indicated with strikethrough.

Figure S3

CREATS SHOWN 3628 GTU-FT-C ELLEDIUM LARGAM CREAT 2878 CP STANDAM LARGAM CREAT 2878 CP STANDAM LARGAM CREAT CRE				E REU EE GU
CPSS_RMMN 2380 EUN 10 10 10 10 10 10 10 1		Q8EFJ3_SHEON 3GZ8		
P83847_GEOSE 1875 1875 0		, ,		
### AUTH JEMMN 2551 VILLE 15-0 LIGHARDE LIDAKE OWN ED LIEANE ELIPANE OWN ED CONTROL OWN ED OWN ED CONTROL OWN ED OWN ED CONTROL OWN ED OW				
GMM_ECOLT_ JEN/A GNU_N-CD		P83847_GEOSE 1RRS		
OKXOSE ECOLX 21871		MUTYH_HUMAN 1X51	VTWE-PS	eq <mark>lqrkall<mark>qe</mark>lqrwa<mark>gp</mark></mark>
IDL_SALTY 3HYQ		GMM_ECOLI 1RYA		
IDL_SALTY 3HYQ		Q6XQ58 ECOLX 2I8T	GRVQ-KD	ETLEAAFE <mark>r</mark> ltma <mark>elg</mark> l
IOL_ECOLI JPPV				
TOTA_HAMAN 27CK		- '.		
TOT2_HUMAN 2PNV SPEMEMELER				
Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q		_ ·		
QOTTCS_CLOP1 JSCM				
NUDK_EOLI 1/11/10 SUNDS - L. SEPAKEABEERO		_ :		
NUDIA_NUMAN 3091				
OBJEER_BACK 3041				
B9NTJB_STRSU 3085 SAD - NO NATIONALENCESSA				
Less conserved ### August 2.887		Q81EE8_BACCR 3Q4I		
MUDIA_MIMAN 2850 GVU-TRDISCREDING REFELE WIDDIA_MIMAN 2850 GVU-TRDISCREDING REFELE WIDDIA_MIMAN 2857 GVRES - DEAT REAFFLICE REVINERD WIDDIA_CAEEL 1876 GVRES - DEAT REAFFLICE REVINERD APAL_CAEEL 1876 GVRES - DEAT REAFFLICE REVINERD APAL_CAEEL 1876 GVRES - DEAT DEATH REFERENCE APAL_CAEEL 1876 GVRES - DEATH REFERENCE QREVERO_CATERN 3000 GVRES - DEATH REFERENCE ARADM_SYNY3 2000 GVRES - DEATH REFERENCE QREVERO_LISSIN 3180 GVRES - DEATH		B9WTJ0_STRSU 308S		
CONSERVED MUDIC MIMMAN 2850 GVU-TO- TR. SLEGIK RELEBELDE Q2X1A2_THEMA 3857 GVU-TO- TR. Q2X1A3_THEMA 3867 GVU-TO- TR. Q2X1A4_THEMA 3867 GVU-TO- TR.	Less	033199_MYCTU 1MQW		
Q9X1A2_THEMA 3E57	_	NUD16_XENLA 2A8T	GFVD-TR	DI <mark>S</mark> LE <mark>EG</mark> LK <mark>r</mark> ELEEEL <mark>GP</mark>
QSXLA2_THEMA 3ES7 GNA_EG - AA	conserved	NUD16 HUMAN 2XSQ	GFVD-TQ	DR <mark>Sledglnre</mark> lr <mark>eelg</mark> e
Y879_DETRA 205F			GHVR-EG	DGATPREAFLKGLEREVNEEVDV
### APA4_CABEL JRTG				
DAQ082_LISW 350N GGG-DE- ALSTANGESTEEN NUDB_ECOLI 201C SVE-EG- APDAMENVECTIAL QSPM46_DETRA 1.NQZ SUL-AG- PTPAMLEAGEVIAL QSPM46_DETRA 1.NQZ SUL-AG- PTPAMLEAGEVIAL QSPM46_DETRA 1.NQZ SUL-AG- PTPAMLEAGEVIAL QSPM6_FEATT 2.071 SULF-DE- TIVGAMERI, LEAMN QSPM6_FEATT 2.071 SULF-DE- TIVGAMERI, LEAMN QSPM6_FEATT 2.071 SULF-DE- TIVGAMERI, LEAMN QSPM6_FEATT 3.072 SULF-DE- SULFAMC, REPERT SULFAMC,				
NUDB_ECOLI 2012				
Q9RV46_DETRA 1NQZ			CCVC FC	AISEIAKRESTEELINL
QR2VOE_MITEU 3CMG				
QSMHRI_FRATT 2QJJ				
MADM_SYNYS 2QJO		Q82VD6_NITEU 3CNG		
Q2EBB_LISIN 319X GVB-EG SECAME RELEBERS QSLBB_BACFN 36W GVC-EG SECAME RELEBERS QSLBB_BACFN 36W GVC-EG SECAME RELEBERS QSLBB_BFAA 37JY GVC-PG SECAME RELEBERS QSZXR9_BIFAA 37JY GVC-PG SECAME RELEBERS QSZXR9_BIFAA 37JY GVC-PG SECAME RELEBERS QSZXR9_BITEU 288V QFLE-PG SITIOACSRVLEETG QSZXR9_BITEU 288V GVC-KG QSZXR9_BITEU 309S GVC-KG QCK-KG		Q5NHR1_FRATT 2QJT		
QSLBB1_BACFN 3 GWY		NADM_SYNY3 2QJO		
NUDT_9_HUMAN 10VJ		Q92EH0_LISIN 3I9X	GFVD-EN	E <mark>S</mark> AE <mark>Q</mark> AAE <mark>r</mark> ele <mark>eet</mark> sl
NUDT9_HUMAN 1QV3		Q5LBB1 BACFN 3GWY	GKVE-EG	ESLQEALQREIMEEMDY
NUDI_SALTY 3N77				
### ABZZM4_BIFAM 3FJY GKLE-QN-E THRHAMYETGETEG ### NUDT5_HUMAN 2DSC QBSZRS9_NITEU 2BBV QBSCS1_DETRA 2 WAE GGVE-KS-E DIGAAAARELLEEVG QBSCS1_DETRA 2 WAE GGVE-KS-E DIGAAAARELLEEVG CBWVE1_ALTAD 3QSJ GAAD-PS-DAEMAKRAFGRPWCAEDDOPALAVTAL RETAELIGM QBPYE2_METMA 3GRN GKN-PD-E SILEGAVAREVEETGI C3HA76_BACTU 31D9 GRVE-NS-E TLEEAMTREMEETGI WIDC_ECOLI 2 KB GWQ-AD-E QLESARREAEELGE MUDC_ECOLI 1 VKG GFVE-VS-E TLEEAMTREWREETGI APAA_HUMAN 3U53 GHVE-PG-E DULETAL RETOERNEE APAA_HUMAN 3U53 GHVE-PG-E DULETAL RETOERNEE APAA_HUMAN 3 WIS GME-PG-E DULETAL RETOERNEE APAA_HUMAN 3 WIS GME-PG-E DULETAL RETOERNEE APAA_HUMAN 3 WIS GME-PG-E DEGAAVREVVEEGO ANDITA_MOUSE 2 DUK GGW-PG-E DEGAAVREVVEEGO ANDITA_MOUSE 2 DUK GGW-PG-E DEGAAVREVVEEGO ANDITA_MOUSE 2 DUK GGW-PG-E DEGAAVREVVEEGO ANDITA_HUMAN 2 FVV GGW-PG-E DEGAAVREVVEEGO ANDITA_HUMAN 1 XFV GGW-PG-E DEGAAVREVEEGO ANDITA_HUMAN 1 XFV GGW-PG-E DEGAAVREVEEGO				
NUDT5_HUMAN 2DSC GLID-DG PEAAL RELECTOR (82XRS_NITEU 2BBV GVLE-RG STIDACSRIVLEETOR (PRSC1_DETRA 3GRN GVLE-RG STIDACSRIVLEETOR (PRSC1_DETRA 3GRN GVLE-RG STIDACSRIVLEETOR (PRSC1_DETAAR) GVLE-RG				
Q82XR9_NITEU 2B0V				
Q9RSCI_DEIRA 2W4E				
C8WVE1_ALIAD 3QSJ GAD_PS DAEMAKRAFGRPVCAEDDOPALAVTALRETAELIGN 3GRN GKVN PD SLKEGVAREVWEETGI C3H476_BACTU 31D9 GRVE NG SLKEGVAREVWEETGI C3H476_BACTU 31D9 GRVE NG SLKEGVAREVWEETGI VFCD_ECOLI 2FKB GVVQ - AD QLLESAR REAEELEGI NUDC_ECOLI 1VK6 GFVE - VG SULETAVREVWEETGI AP4A_HUMAN 3U53 GHVE - PG SULETAVREVWEETGI AP4A_HUMAN 3U53 GHVE - PG SULETAVREVWEETGI GG6548_AQUAE 317V GTI - PC KPETTAVREVWEETGI VIND10 HUMAN 3MCF GME - PE SULETAVREVWEETGI VIND11 HUMAN 3MCF GME - PE SULETAVREVWEETGI VIND14_MOUSE 2DUK GGME - PE SULETAVREVWEETGI VIND13_HUMAN 2FVV GGME - PE SAEQAYREVYEEAGI VIND13_HUMAN 2FVV GGME - PE SAEQAYREVYEEAGI VIND13_HUMAN 2FVV GGME - PE SAEQAYREVYEEAGI VIND13_HUMAN 2FVV GGME - PE SOVDCAIREVYEETGI VIND14_HUMAN 2FVV GGME - PE SOVDCAIREVEELGI VIND14_HUMAN 2FVV GGME - PE SOVDC				
Q8PYE2_METMA 3GRN		_ :		
C3H476_BACTU 3ID9		_ :		
YFCD_ECOLI 2FKB		_ :		
NUDC_ECOLI 1VK6		C3H476_BACTU 3ID9		
AP4A_HUMAN 3U53		YFCD_ECOLI 2FKB		
O66548_AQUAE 317V		NUDC_ECOLI 1VK6		
Q75UV1_THETH 1VC8		AP4A_HUMAN 3U53	GHVE-PG	EDDLETALRETQEEAGI
Q75UV1_THETH 1VC8		066548_AQUAE 317V	GNIE-PG	EK <mark>P</mark> EETAV <mark>REVWEETG</mark> V
NUD10_HUMAN 3MCF		075UV1 THETH 1VC8	GHPE-PG	ESLEEAAV <mark>RE</mark> VWEETGV
NUDTA_MOUSE 2DUK			GGME-PE	EEPGGAAVREVYEEAGV
NUDT3_HUMAN 2FVV		_ :		
RPPH_ECOLI 2KDV				
O04841_LUPAN 1JKN		′		
DCP2_YEAST 2JVB GKIS-KD-				
DCP2_SCHPO 2QKM				
Q6MPX4_BDEBA 3FFU		_ :	GKIS-KD	ENDIDCCIREVKEEIGF
NUDG_ECOLI 2RRK			GKID-KD	ESDVDCAIREVYEETGF
Q2RXE7_RHORT 3R03				
Q665F4_BARHE 3HHJ		NUDG_ECOLI 2RRK	GKVE-PD	ESQRQALV <mark>R</mark> ELREELGI
MUTT_ECOLI 3A6T				
CONSERVED Q7NNQ3_CHRVO 3F13 GKAN-RG-E		Q6G5F4_BARHE 3HHJ		
Q8ZTD8_PYRAE 1K2E		MUTT_ECOLI 3A6T		
Q8ZTD8_PYRAE 1K2E	Conserved			
Q0TS82_CLOP1 3F6A GHIE-VN-E LPEGACIREAKEEAGL Q836H1_ENTFA 2AZW GEIE-GT-E TKE6AIHREVLEELGI Q97737_STRPN 2PQV GAUQ-VN-E STEDAVNREVKEELGV Q9RVK2_DEIRA 1SU2 GAVE-DG-E NPQDAAVREACEETGL NUDE_ECOLI 1VHZ GLD-PG-E SVYEAANRELKEEVGF ADPP_ECOLI 1KHZ GMIE-EG-E SVEDVARREAIEEAGL Q5SJY9_THET8 2YVP GKVD-EG-E TPEAAARRELREEVGA Q84CU3_THETH 1V8M GLIE-PG-E DPLEARRELAEETGL NUD16_HUMAN 3H95 GLSE-PE-E DDLEDTAVREVFEETGL NUD18_HUMAN 3G66 GRIE-PG-E TLVEAAARELWEETGI NUD18_HUMAN 3G66 GRIE-PG-E TIVEALOREVKEEAGL Q93S37_LACBA 3EXQ GHVE-VG-E PCATAAIREVFEETGL Q97QH6_STRPN 2B06 GHVE-ND-E AFAESVIREIYEETGL Q9K704_BACHD 3FK9 GKVE-AG-E SILETVKREWEETGI 80DP_HUMAN 3ZR0 GKVQ-EG-E TIPEAAVREVWEETGL 80DP_HUMAN 3SMD GAIE-LG-E TPEAVEVWEETGL Q83652_ENTFA 2FML GFVI-RN-RN-E SIEDSVLREIKEETGV 067435_AQUAE 2YYH				
Q836H1_ENTFA 2AZW GEIE-GT-E				
Q97T37_STRPN 2PQV			GETE-GT-	ETKFEATHREVI FELGT
Q9RVK2_DEIRA 1SU2 GAVE-DG-E				
NUDE_ECOLI 1VHZ				
ADPP_ECOLI 1KHZ				
Q5SJY9_THET8 2YVP GKVD-EG-E-E			GLID-PG	SVYEARNELKEEVGF
Q84CU3_THETH 1V8M				
NUDT6_HUMAN 3H95		/		
NUD1_ECOLI 3SHD				
NUD18_HUMAN 3GG6 GRME-PG-E- TIVEALQ REVKEEAGL Q03537_LACBA 3EXQ GHVE-VG-E- PCATAAIREVFEETGL Q97QH6_STRPN 2B06 GHVE-ND-E- STEETWEETGL Q9K704_BACHD 3FK9 GKME-AG-E- SILETVK REYWEETGI 80DP_HUMAN 3ZR0 GKVQ-EG-E- TIEDGARREL QEESGL A0REX4_BACAH 3SMD GALE-LG- TPEAAVIEVWEETGI Q83032_ENTFA 2FML GFVN-RN-E- SIEDSVLRETKEETGV 067435_AQUAE 2YYH GFVE-VG-E- RVEEAAAREMREETGL				
Q03S37_LACBA 3EXQ GHVE-VG-E		NUDJ_ECOLI 3SHD	GHLE-AD	ETLVEAAA <mark>R</mark> ELW <mark>EETG</mark> I
Q03S37_LACBA 3EXQ GHVE-VG-E		NUD18_HUMAN 3GG6	GRME-PG	E <mark>tivealorevkeeag</mark> l
Q97QH6_STRPN 2B06 GHVE-ND-E		Q03S37_LACBA 3EXO	GHVE-VG	E <mark>P</mark> CATAAI <mark>R</mark> EVF <mark>EETG</mark> L
Q9K704_BACHD 3FK9 GKME-AGE			GHVE-ND	EAFAESVI <mark>R</mark> EIYEETGL
80DP_HUMAN 3ZRØ GKVQ-EGETIEDGARRELQEESGL AØREX4_BACAH 3SMD GAIE-LGETPEEAVVREVWEETGL Q83Ø52_ENTFA 2FML GFVN-RNESTEDSVLRETKEETGV 067435_AQUAE 2YYH GFVE-VGERVEEAAAREMREETGL			GKME-AG	ESTI ETVKREYMEETGT
AØREX4_BACAH 3SMD			GKVO-FG	TTENGARREI GESCI
Q830S2_ENTFA 2FML				
067435_AQUAE 2YYH GFVE-VGERVEEAAAREMREETGL				
			GEVE VC	DVESAAARMOSETC
— ŲSJWUZ_BUKPI 4DYW <mark>□KVŪ</mark> -WLĒ				
		— V27MOZ_BUKPI 4DYW	UNVD-WL	PVEKAVUREILEEELGI

G

E

REU EE GU

Figure S3. Nudix box sequences in the structure-induced sequence alignment. UniProt Entry Names and PDB IDs of 78 Nudix enzymes and their Nudix box sequences are shown in the alignment. The most conserved positions are indicated in bold on top of the alignment. The top 38 enzymes have at least one substitution in one of the most conserved positions, or one insertion or deletion between the conserved positions. The Clustal X color scheme was applied to all residues.

Figure S4

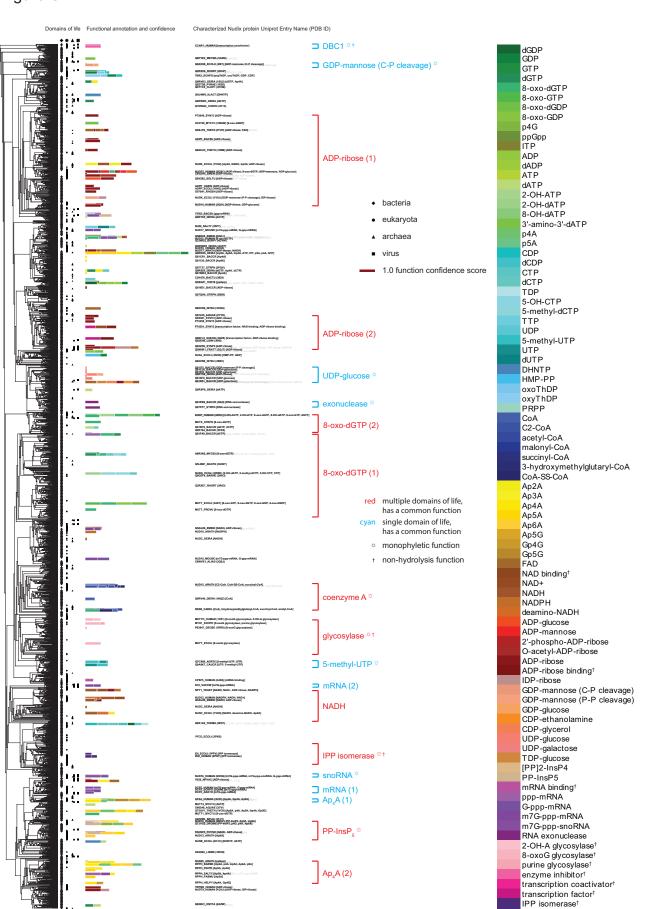


Figure S4. Phylogeny of the complete Nudix clan, with 38,950 unique Nudix domains that match the HMM of one of the Pfam families in Pfam v27.0 (Mar 2013). The phylogeny was constructed with FastTree ⁶⁴ and mid-point rooted. The branch lengths of the tree are omitted for simplicity. The protein domains with experimental annotations are the same as those in Fig. 5. The domains of life of leaves are indicated on the side of the phylogeny as follows: solid diamonds (bacteria), circles (eukaryota), triangles (archaea), or squares (virus). The clades, functions, and IDs of leaves are indicated as in Fig. 5 for protein domains with experimental annotations. When the text labels of multiple leaves overlapped, one leaf is shown in black, and others are shown in light grey. A high-resolution version is available in Table S1, Resource 2.