

# Analysis of NUDIX enzymes across fungi reveals previously unrecognized diversity

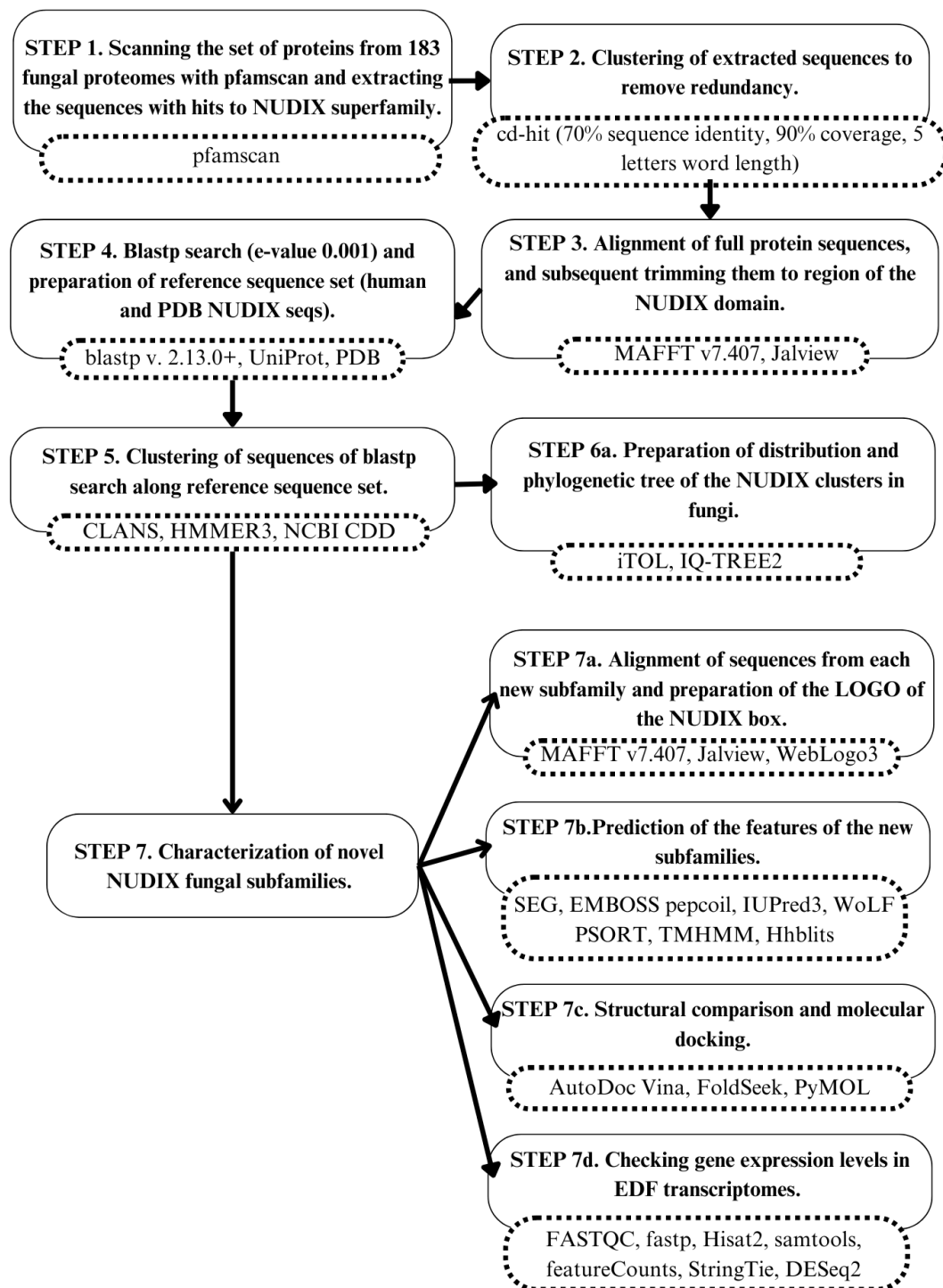
Zofia Pasterny<sup>1,2</sup>, Drishtee Barua<sup>1</sup>, Eugenio Mancera<sup>3</sup> and Anna Muszewska<sup>\*1</sup>

<sup>1</sup> Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawinskiego 5A, 02-106 Warsaw, Poland

<sup>2</sup>Department of Chemistry, Wrocław University of Technology, Wyb. Wyspińskiego 27, 50-370, Wrocław, Poland

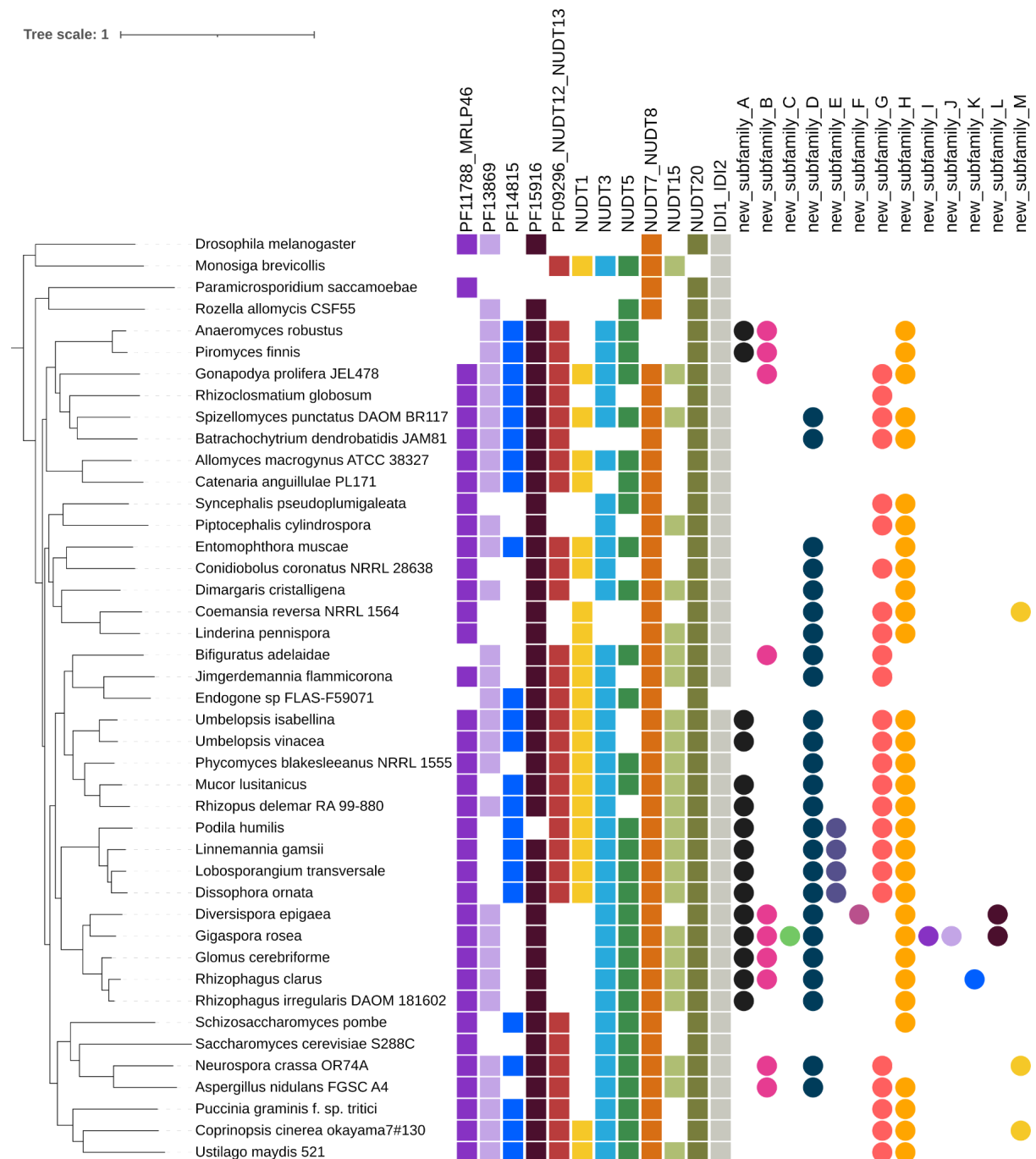
<sup>3</sup>Departamento de Ingeniería Genética, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Unidad Irapuato, Irapuato, Mexico

<b>Supplementary Figure S1. Workflow of the computational strategy to identify NUDIX proteins and assign them possible biological functions.</b>	<b>2</b>
<b>Supplementary Figure S2. Distribution of NUDIX subfamilies across selected organisms.</b>	<b>3</b>
<b>Supplementary Figure S3. The distribution of the NUDIX hydrolases from newly identified subfamilies over fungal phyla.</b>	<b>4</b>
<b>Supplementary Figure S5. Alignment of sequences from new_subfamily_D.</b>	<b>6</b>
<b>Supplementary Figure S6. Alignment of sequences from new_subfamily_F.</b>	<b>7</b>
<b>Supplementary Figure S7. Alignment of sequences from new_subfamily_A and human NUDT2 (PDB: 3U53).</b>	<b>8</b>
<b>Supplementary Figure S8. Molecular docking of the NUDIX domain with Ap4A.</b>	<b>9</b>
<b>Supplementary Figure S9. Heatmap clustering of binding affinities</b>	<b>10</b>
<b>Supplementary Figure S10. Conservation of DHNTPase residues</b>	<b>11</b>
<b>Supplementary characterization of subfamilies</b>	<b>13</b>



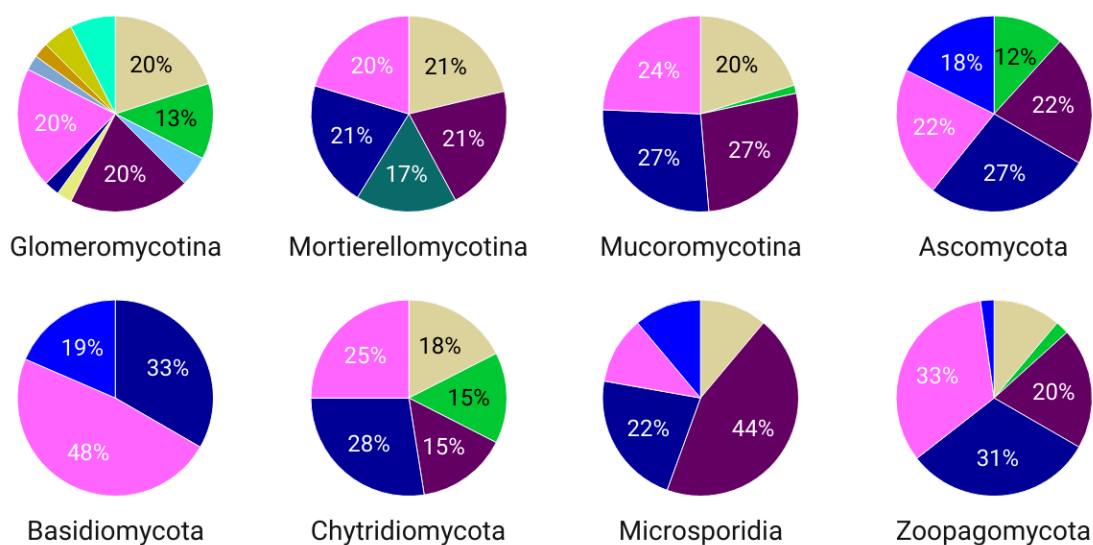
**Supplementary Figure S1. Workflow of the computational strategy to identify NUDIX proteins and assign them possible biological functions.**

Each box explains the analysis performed while the dotted boxes specify the computational tools employed for each step.



### Supplementary Figure S2. Distribution of NUDIX subfamilies across selected organisms.

Phylogenomics tree of representative taxa for each of the main fungal lineages computed with OrthoFinder using *Monosiga brevicollis* and *Drosophila melanogaster* as an outgroup annotated with the distribution of NUDIX families and subfamilies. NUDIX families and subfamilies with human homologs represented as squares, and newly identified subfamilies marked as circles.



Created with Datawrapper

**Supplementary Figure S3. The distribution of the NUDIX hydrolases from newly identified subfamilies over fungal phyla.**

Phyla with a marginal number of copies were left out.

Cluster		NUDIX box motif																									
NUDIX motif				G	x	x	x	x	x	E	x	x	x	x	x	x	x	R	E	I/L/V	x	E	E	x	G	I/L/V	
Newly identified NUDIX subfamilies	new_subfamily_A			G	R	V	I	G	D	E	D	E	L	K	C	A	L	R	E	T	L	E	I	T	G	L	
	new_subfamily_B		G	S	I	E	P	N	D	K	S	P	L	E	R	A	L	R	E	I	K	E	E	T	G	L	
	new_subfamily_C		G	S	I	E	E	T	D	K	G	I	F	K	A	A	K	R	E	C	K	E	E	T	S	Y	
	new_subfamily_D			G	M	L	D	G	S	G	D	F	A	G	K	A	A	E	E	L	K	E	E	C	G	I	
	new_subfamily_E		G	K	V	E	R	G	L	E	S	I	A	E	S	A	A	R	E	L	Q	E	E	A	F	I	
	new_subfamily_F		G	G	H	V	E	E	T	D	I	S	T	K	H	A	A	Q	R	E	V	L	E	E	T	G	I
	new_subfamily_G			G	P	N	D	P	P	L	H	N	K	I	C	A	I	R	E	T	F	E	E	S	G	V	
	new_subfamily_H		G	K	R	E	E	G	D	E	D	D	Q	E	T	A	A	R	E	T	L	E	E	I	G	L	
	new_subfamily_I		G	K	A	I	H	E	K	N	E	N	I	D	V	C	V	I	R	E	C	F	E	E	A	E	I
	new_subfamily_J			N	G	H	E	I	L	E	T	E	E	D	A	A	M	R	E	T	L	E	E	S	G	I	
	new_subfamily_K			G	K	V	E	L	G	E	S	S	I	Q	A	V	L	R	E	T	Q	E	E	T	G	I	
	new_subfamily_L			G	G	K	K	N	N	E	S	Y	E	Q	C	A	I	R	E	T	K	E	E	T	N	V	
new_subfamily_M			G	R	K	D	I	G	E	S	L	E	Q	A	A	I	R	E	T	Y	E	E	T	G	Y		
subfamilies associated with reference sequences	NUDT1			G	K	V	E	P	G	E	T	V	E	E	G	A	L	R	E	L	E	E	E	A	G	I	
	NUDT3			G	G	W	E	N	D	E	T	Q	E	E	A	A	L	R	E	T	W	E	E	A	G	V	
	NUDT5		G	L	I	D	A	G	E	T	S	P	E	Q	A	A	L	R	E	L	K	E	E	T	G	Y	
	NUDT7_NUDT8		G	G	K	R	D	P	T	D	K	S	L	L	A	T	A	L	R	E	A	E	E	E	I	G	L
	NUDT15			G	H	L	E	F	G	E	S	F	E	E	C	A	I	R	E	V	L	E	E	T	G	L	
	NUDT20			G	K	I	N	K	D	E	P	D	T	D	C	A	V	R	E	V	L	E	E	T	G	F	
	PF009296			G	F	V	E	P	G	E	S	L	E	E	A	V	R	R	E	V	K	E	E	S	G	I	
	PF15916			G	G	I	G	F	G	H	S	P	F	E	T	V	I	K	E	C	M	E	E	A	S	I	
	PF11788			G	G	V	R	V	C	E	N	L	H	E	A	A	G	R	E	L	Q	E	E	C	G	S	
	PF13869		G	D	Y	L	K	P	G	E	D	E	I	E	G	L	K	A	R	L	N	E	R	L	A	P	
IDI1_IDI2	NUDIX box is highly divergent across different species, and it does not follow the typical NUDIX motif																										
PF14815	does not contain conserved NUDIX box motif																										

**Supplementary Figure S4. NUDIX box motif conservation across all identified NUDIX subfamilies in fungi.**

	60	70	80	90	100
CEP17950.1/1-104	ILLTQQLPIPHFNFPELPA			SGDFA	GGAAKETE.EET..VQR
KAF0507845.1/1-195	VLLTQPRRLVPSLNFPEIPA		GL...	SGDFA	..
KAF0938159.1/1-297	VLLTKQARLSPSFDFPEIPA		GMLDG	SGDFA	..
KAF8946220.1/1-1336	IVLTKQPRLSPSFDFPEIPA		GMLDG	SGDFA	..
KAF8974856.1/1-275	VVLTQKARLSPSFNFPEIPA		GMLDG	SGDFA	..
KAF8986817.1/1-262	VLLTKQARLSPSFDFPEIPA		GMLDG	SGDFA	..
KAF9108265.1/1-267	VLLTKQPRLSPSFNFPEIPA		GMLDG	SGDFA	..
KAF9112066.1/1-281	VLLTKQARLSPSFDFPEIPA		GMLDG	SGDFA	..
KAF9135357.1/1-281	VVLTQKPRLSPSFNFPEIPA		GMLDG	SGDFA	..
KAF9156063.1/1-266	IVLTKQPRLSPSFDFPEIPA		GMLDG	SGDFA	..
KAF9172532.1/1-269	VVLTQKARLSPSFNFPEIPA		GMLDG	SGDFA	..
KAF9180642.1/1-279	VLLTKQARLSPSFNFPEIPA		GMLDG	SGDFA	..
KAF9180845.1/1-279	VLLTKQARLSPSFNFPEIPA		GMLDG	SGDFA	..
KAF9200127.1/1-275	VVLTQKARLSPSFNFPEIPA		GMLDG	SGDFA	..
KAF9311268.1/1-166	VLLTKQARLSPSFDFPEIPA		GMLDG	SGDFA	..
KAF9358093.1/1-276	VVLTQKARLSPSFNFPEIPA		GMLDG	SGDFA	..
KAF9359124.1/1-229	..			SGDFA	..
KAF9402252.1/1-269	VVLTQKARLSPSFNFPEIPA		GMLDG	SGDFA	..
KAF9434301.1/1-279	VLLTKQARLSPSFDFPEIPA		GMLDG	SGDFA	..
KAF9550266.1/1-266	VVLTQKPRLSPSFDFPEIPA		GMLDG	SGDFA	..
KAF9585877.1/1-287	VVLTQKPRLSPSFNFPEIPA		GMLDG	SGDFA	..
KAF9912166.1/1-280	VVLTQKPRLSPSFNFPEIPA		GMLDG	SGDFA	..
KAF9944486.1/1-214	..QARLA	PSFAFPEIPA	GMLDG	SGDFA	..
KAF9993632.1/1-276	VVLTQKARLSPSFNFPEIPA		GMLDG	SGDFA	..
KAG0032219.1/1-267	VLLTKQARLSPSFDFPEIPA		GMLDG	SGDFA	..
KAG0056299.1/1-265	VVLTQKPRLSPSFDFPEIPA		GMLDG	SGDFA	..
KAG0097359.1/1-267	VVLTQKPRLSPSFDFPEIPA		GMLDG	SGDFA	..
KAG0197881.1/1-285	VVLTQKARLSPSFDFPEIPA		GMLDG	SGDFA	..
KAG0223646.1/1-270	VVMVNAQRLA	PSFDFPEIPA	GMLDG	SGDFA	..
KAG0230969.1/1-281	VVLTQKPRLSPSFDFPEIPA		GMLDG	SGDFA	..
KAG0256308.1/1-231	..	..VPA	GMLDG	SGDFA	..
KAG0264017.1/1-272	VLLTKQARLSPSFNFPEIPA		GMLDG	SGDFA	..
KAG0275988.1/1-279	TVLTKQPRLSPSFNFPEIPA		GMLDG	SGDFA	..
KAG0298622.1/1-272	IVLTKQPRLSPSFNFPEIPA		GMLDG	SGDFA	..
KAG0328178.1/1-278	VVLTQKPRLA	PCFDFPEIPA	GMLDG	SGDFA	..
KAG0334083.1/1-280	VLLTKQARLA	PSFNFPEIPA	GMLDG	SGDFA	..
KAG0354906.1/1-267	VVLTQKPRIP	PSLDFPEIPA	GMLDG	SGDFA	..
KAG0378502.1/1-279	AVLTKQPRLSPSFNFPEIPA		GMLDG	SGDFA	..
KAG9065156.1/1-266	VVLTQKPRLSPSFNFPEIPA		GMLDG	SGDFA	..
KAG9258564.1/1-294	VVLTQKPRLSPSLAFTEIPA	GFSTSVIYDFNNLPWVL	GMLDG	SGDFA	..
KFH62508.1/1-266	VVLTQKPRLSPSFNFPEIPA		GMLDG	SGDFA	..
OAQ30095.1/1-266	VVLTQKPRLSPSFNFPEIPA		GMLDG	SGDFA	..
ORZ02061.1/1-282	VVLTQKARLA	PSFAFPEIPA	GMLDG	SGDFA	..
PVU84602.1/1-264	VVLTQKPRIP	VPDFHMLEIPA	GMLDG	SGDFA	..
PVU03001.1/1-352	VVLTQKPRLA	VPDFHMLEIPA	GMLDG	SGDFA	..
PKR03543.1/1-282	VVLTQKPRIP	VPDFHMLEIPA	GMLDG	SGDFA	..
RHZ64137.1/1-306	VVLTQKPRLA	VPDLSPPEIPA	GMLDG	SGDFA	..
RIB18197.1/1-261	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
RUP47766.1/1-1925	VVLTQKPRLA	VPDLSPPEIPA	GMLDG	SGDFA	..
KAF0507844.1/1-301	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
TPX72271.1/1-275	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
TPX62866.1/1-274	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
RKP38283.1/1-280	VVLTQKPRLA	VPDLSPPEIPA	GMLDG	SGDFA	..
RKO99357.1/1-297	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
RKO85042.1/1-265	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
RIB6387.1/1-267	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
RGB24081.1/1-229	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
RCH98193.1/1-258	ILLTKQPRIP	IPHFAPLEIPA	GMLDG	SGDFA	..
POG68840.1/1-311	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
PIA14195.1/1-289	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
PHZ12513.1/1-258	ILLTKQPRIP	IPHFAPLEIPA	GMLDG	SGDFA	..
OZJ021225.1/1-253	LLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
ORZ23424.1/1-258	ILLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
ORZ01410.1/1-251	LLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
ORX88830.1/1-251	LLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
ORX70990.1/1-230	VLLTKQARLP	VPDLSPPEIPA	GMLDG	SGDFA	..
ORX58676.1/1-252	ILLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
OBZ82911.1/1-253	ILLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
QAD78918.1/1-251	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KXN7045.1/1-253	ILLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KND01036.1/1-299	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAG8525856.1/1-238	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAG2231002.1/1-253	ILLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAG2217193.1/1-301	ILLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAG2205658.1/1-255	ILLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAG2191789.1/1-270	ILLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAG2177336.1/1-276	VLLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAG2175933.1/1-277	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAG0355481.1/1-281	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAG0200119.1/1-261	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAG0190422.1/1-259	ILLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAG0179942.1/1-259	ILLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAG0170618.1/1-465	ILLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAF9987863.1/1-280	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAF9416604.1/1-263	VVLTQKARLSPSFDFPEIPA		GMLDG	SGDFA	..
KAF9137952.1/1-266	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAF9083249.1/1-267	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAF7754877.1/1-300	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAF7731625.1/1-259	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAF3904235.1/1-331	ALLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAF3902742.1/1-329	ALLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
KAF1807386.1/1-271	ILLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
EIE81380.1/1-590	ILLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
EGY21149.1/1-286	ALLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
EGX49510.1/1-296	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
EGR46282.1/1-227	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
EGF80123.1/1-277	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
EDO65364.2/1-332	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
ED000943.1/1-206	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
EAS30480.3/1-301	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
SAL96329.1/1-256	ILLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
CEP19456.1/1-259	ILLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
CAZ80945.1/1-273	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
GES77410.1/1-311	VVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
GAN06740.1/1-271	ILLTKQPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
EAA61227.1/1-332	AVLTQKPRIP	VPDLSPPEIPA	GMLDG	SGDFA	..
consensus>70	vllt qpr . ips . fpe . pa .		gml dg	sgdf	g . a . e . . . . .
			*****	*****	*****

## Supplementary Figure S5. Alignment of sequences from new\_subfamily\_D.

The region of the NUDIX box is marked with orange stars. most of the sequences have conserved motif before the catalytic center: FPEXPA.



	1	10	20	30	40
RHZ45831.1/69-187	..	..	..	..	..
RHZ48206.1/250-371	YS	TI	..	..	..
RHZ52810.1/12-137	YT	TT	VL	YDRK	TK
RHZ53956.1/12-127	YV	SI	IL	RRPN	TP
RHZ56541.1/12-137	YV	SI	IL	RRPN	TP
RHZ57731.1/12-137	YV	SI	IL	RRPN	TP
RHZ57732.1/12-137	YV	SI	IL	RRPN	TP
RHZ58156.1/12-137	YV	SI	IL	RRPN	TP
RHZ58196.1/12-137	YT	TT	VL	HDKK	TK
RHZ62880.1/12-137	YV	SI	IL	RRPN	TP
RHZ63037.1/12-137	YT	TT	VL	YDKK	TK
RHZ66594.1/610-735	YT	TT	VL	YDKK	TK
RHZ69292.1/12-96	YT	TT	VL	YDKK	TK
RHZ72504.1/12-137	YV	SI	IL	RRPN	TP
RHZ75106.1/12-137	YT	TT	VL	YDKK	TK
RHZ76086.1/737-862	YT	TI	VL	YDRK	TK
RHZ76144.1/12-137	YV	SI	IL	RRPN	TP
RHZ77245.1/12-137	YT	TT	VL	YDRK	TK
RHZ77815.1/12-136	YV	SI	IL	RRPN	TP
RHZ78222.1/12-137	YV	SI	IL	RRPN	TP
RHZ78629.1/12-137	YV	SI	IL	RRPN	TP
RHZ78630.1/12-129	YV	SI	IL	RRPN	TP
RHZ82225.1/84-160	YT	TT	VL	YDKK	TK
RHZ82644.1/120-213	..	..	..	..	..
RHZ83734.1/12-101	YT	TT	VL	YDKK	TK
RHZ84765.1/12-117	YV	SI	IL	RRPN	TP
RHZ84767.1/12-137	YV	SI	IL	RRPN	TP
RHZ85202.1/12-137	YV	SI	IL	RRPN	TP
RHZ86637.1/82-155	YV	SI	IL	RRPN	TP
RHZ86964.1/12-137	YT	TT	VL	YDKK	TK
RHZ87345.1/12-105	YV	SI	IL	RRPN	TP
RHZ87367.1/12-137	YV	SI	IL	RRPN	TP
RHZ87385.1/12-137	YT	TT	VL	YDKK	TK
RHZ88322.1/12-120	YT	TT	VL	YDKK	TK
RHZ88644.1/120-213	YT	TT	VL	YDKK	TK
RHZ89132.1/12-137	YT	TT	VL	YDKK	TK
consensus> 70	y	..	il	..	..

	1	10	20	30	40
RHZ45831.1/69-187	..	..	..	..	..
RHZ48206.1/250-371	YS	TI	..	..	..
RHZ52810.1/12-137	YT	TT	VL	YDRK	TK
RHZ53956.1/12-127	YV	SI	IL	RRPN	TP
RHZ56541.1/12-137	YV	SI	IL	RRPN	TP
RHZ57731.1/12-137	YV	SI	IL	RRPN	TP
RHZ57732.1/12-137	YV	SI	IL	RRPN	TP
RHZ58156.1/12-137	YV	SI	IL	RRPN	TP
RHZ58196.1/12-137	YT	TT	VL	HDKK	TK
RHZ62880.1/12-137	YV	SI	IL	RRPN	TP
RHZ63037.1/12-137	YT	TT	VL	YDKK	TK
RHZ66594.1/610-735	YT	TT	VL	YDKK	TK
RHZ69292.1/12-96	YT	TT	VL	YDKK	TK
RHZ72504.1/12-137	YV	SI	IL	RRPN	TP
RHZ75106.1/12-137	YT	TT	VL	YDKK	TK
RHZ76086.1/737-862	YT	TI	VL	YDRK	TK
RHZ76144.1/12-137	YV	SI	IL	RRPN	TP
RHZ77245.1/12-137	YT	TT	VL	YDRK	TK
RHZ77815.1/12-136	YV	SI	IL	RRPN	TP
RHZ78222.1/12-137	YV	SI	IL	RRPN	TP
RHZ78629.1/12-137	YV	SI	IL	RRPN	TP
RHZ78630.1/12-129	YV	SI	IL	RRPN	TP
RHZ82225.1/84-160	YT	TT	VL	YDKK	TK
RHZ82644.1/120-213	..	..	..	..	..
RHZ83734.1/12-101	YT	TT	VL	YDKK	TK
RHZ84765.1/12-117	YV	SI	IL	RRPN	TP
RHZ84767.1/12-137	YV	SI	IL	RRPN	TP
RHZ85202.1/12-137	YV	SI	IL	RRPN	TP
RHZ86637.1/82-155	YV	SI	IL	RRPN	TP
RHZ86964.1/12-137	YT	TT	VL	YDKK	TK
RHZ87345.1/12-105	YV	SI	IL	RRPN	TP
RHZ87367.1/12-137	YV	SI	IL	RRPN	TP
RHZ87385.1/12-137	YT	TT	VL	YDKK	TK
RHZ88322.1/12-120	YT	TT	VL	YDKK	TK
RHZ88644.1/120-213	YT	TT	VL	YDKK	TK
RHZ89132.1/12-137	YT	TT	VL	YDKK	TK
consensus> 70	y	..	il	..	..

consensus> 70 y...il...t.Kiww.Kr.n.....p#keFy.HWQ.pGgH!ee.#IS.k.aA.R

★★★★★★★★★★★★★★★★

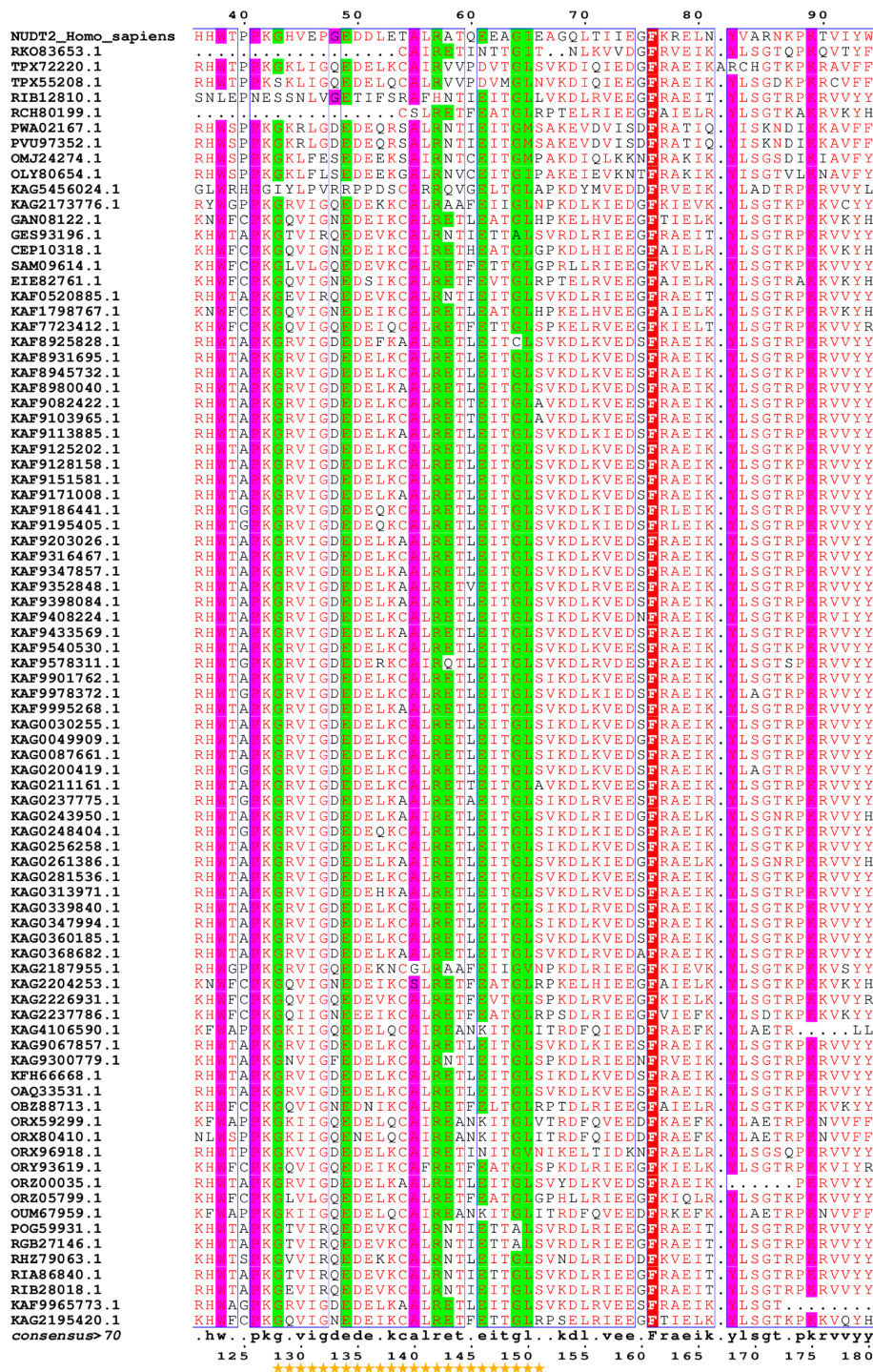
	50	60	70	80	90	100
RHZ45831.1/69-187	EV	LE	ET	GI	QPE	LE
RHZ48206.1/250-371	EV	FE	ET	GI	YKL	LE
RHZ52810.1/12-137	EV	FE	ET	GI	YKL	LE
RHZ53956.1/12-127	EV	LE	ET	GI	QPE	LE
RHZ56541.1/12-137	EV	LE	ET	GI	QPE	LE
RHZ57731.1/12-137	EV	LE	ET	GI	QPE	LE
RHZ57732.1/12-137	EV	LE	ET	GI	QPE	LE
RHZ58156.1/12-137	EV	LE	ET	GI	QPE	LE
RHZ58196.1/12-137	EI	FE	ET	GI	YKL	LE
RHZ62880.1/12-137	EV	LE	ET	GI	QPE	LE
RHZ63037.1/12-137	EI	FE	ET	GI	YKL	LE
RHZ66594.1/610-735	EV	FE	ET	GI	YKL	LE
RHZ69292.1/12-96	EV	FE	ET	GI	YKL	LE
RHZ72504.1/12-137	EV	LE	ET	GI	QPE	LE
RHZ75106.1/12-137	EV	FE	ET	GI	YKL	LE
RHZ76086.1/737-862	EV	LE	ET	GI	YKL	LE
RHZ76144.1/12-137	EV	LE	ET	GI	QPE	LE
RHZ77245.1/12-137	EV	FE	ET	GI	YKL	LE
RHZ77815.1/12-136	EV	LE	ET	GI	QPE	LE
RHZ78222.1/12-137	EV	LE	ET	GI	QPE	LE
RHZ78629.1/12-137	EV	LE	ET	GI	QPE	LE
RHZ78630.1/12-129	EV	LE	ET	GI	QPE	LE
RHZ82225.1/84-160	EV	FE	ET	GI	YKL	LE
RHZ82644.1/120-213	EV	LE	ET	GI	QPE	LE
RHZ83734.1/12-101	EV	FE	ET	GI	YKL	LE
RHZ84765.1/12-117	EV	LE	ET	GI	QPE	LE
RHZ84767.1/12-137	EV	LE	ET	GI	QPE	LE
RHZ85202.1/12-137	EI	LE	ET	GI	QPE	LE
RHZ86637.1/82-155	KV	LE	ET	GI	QPE	LE
RHZ86964.1/12-137	EV	FE	ET	GI	YKL	LE
RHZ87345.1/12-105	EV	LE	ET	GI	QPE	LE
RHZ87367.1/12-137	EV	LE	ET	GI	QPE	LE
RHZ87385.1/12-137	EV	FE	ET	GI	YKL	LE
RHZ88322.1/12-120	EV	FE	ET	GI	YKL	LE
RHZ88644.1/120-213	E	..	..	..	..	..
RHZ89132.1/12-137	EV	FE	ET	GI	YKL	LE
consensus> 70	ev	..	..	..	..	..

consensus> 70 ev.eetgi...leel#y.rtnhyf.e.ewrivhcy...k.vpeltep.emt.wqlkk.k

★★★★★★★★

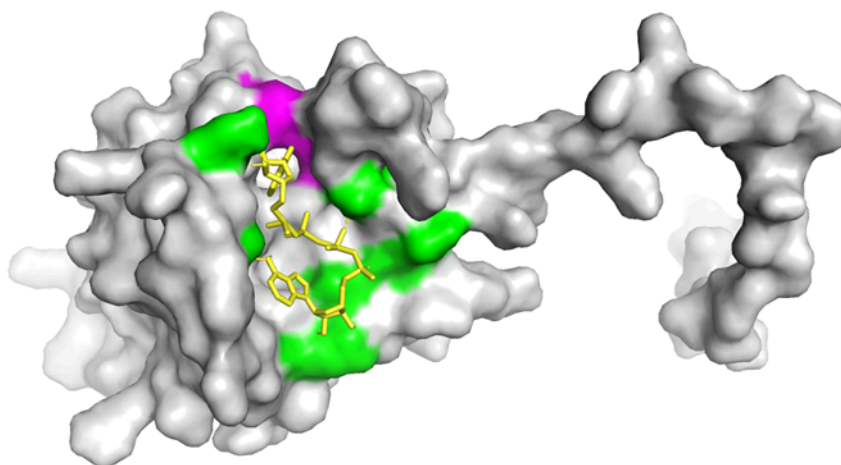
## Supplementary Figure S6. Alignment of sequences from new\_subfamily\_F.

The region of the NUDIX box is marked with orange stars. Conserved residues between all proteins are highlighted in red.



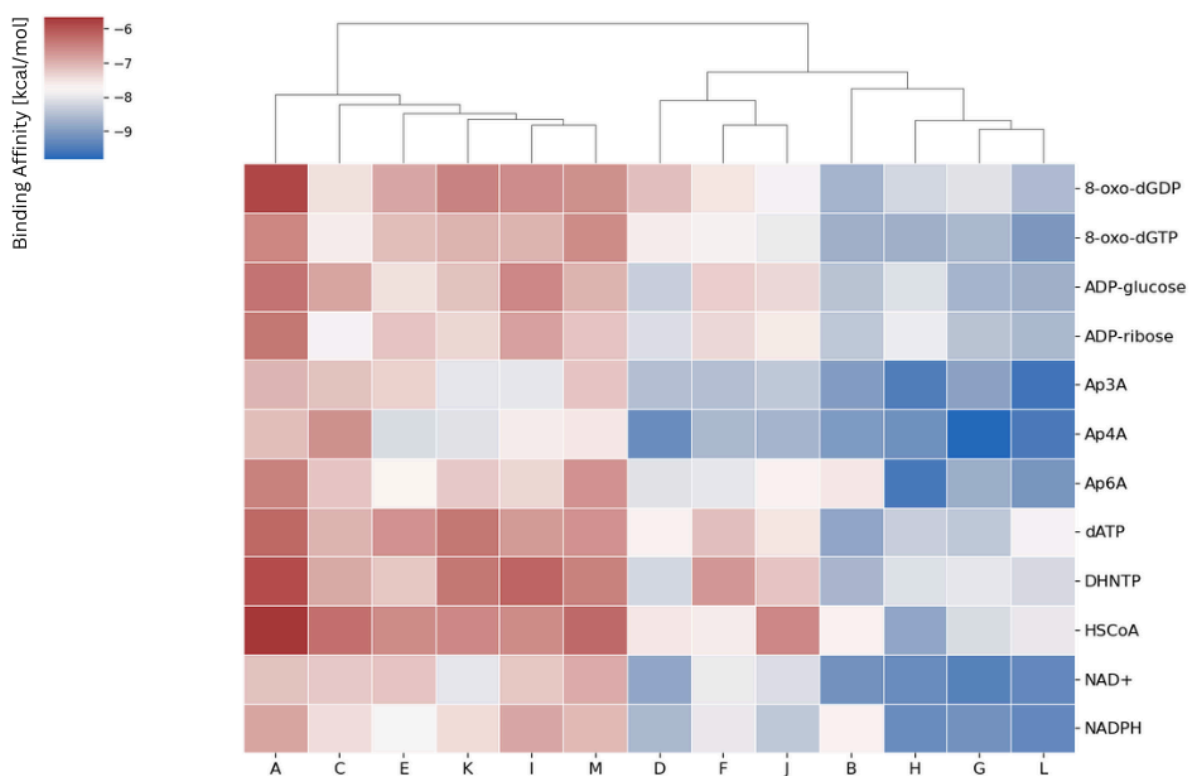


conserved residues from NUDIX box. The region of the NUDIX box is marked with orange stars.



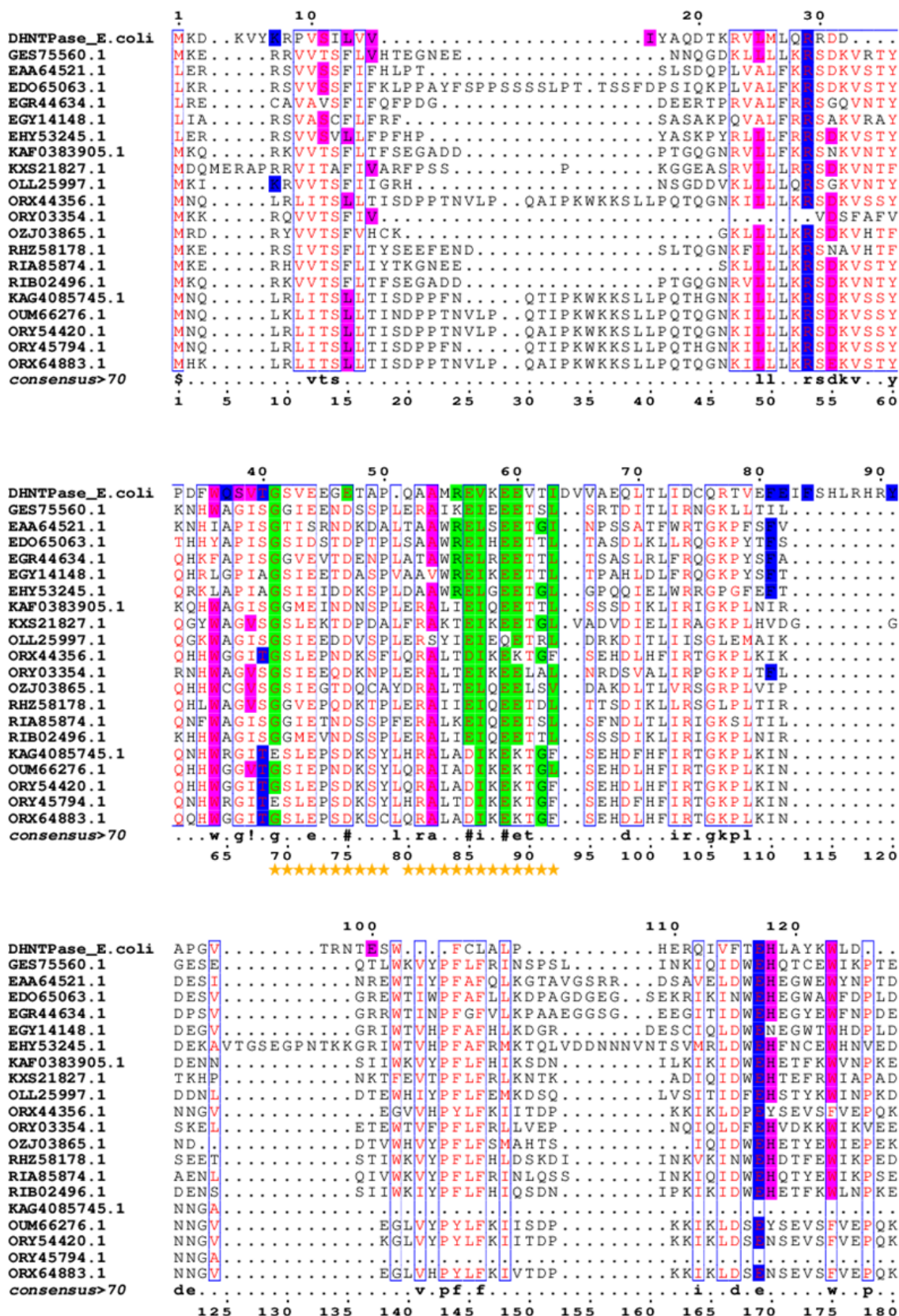
**Supplementary Figure S8. Molecular docking of the NUDIX domain with Ap4A.**

Representatives of the newly identified subfamily A have high affinity to Ap4A and conserved features of Ap4Aases. Scheme showing the molecular docking of the NUDIX hydrolase domain-containing protein (EIE82761.1) (grey) from *Rhizopus delemar* with Ap4A (yellow) at pH 9.0. Corresponding residues that lie on top of the active site cleft in human NUDT2 are marked in magenta (Tyr461 and Phe507). Other binding residues (Asp409, Lys414, Gln423, Glu437, Glu440 and Lys468) are shown in green.



### Supplementary Figure S9. Heatmap clustering of binding affinities

Molecular dockings show similarities in the binding affinities of the newly identified NUDIX fungal subfamilies. Molecular docking results for newly identified subfamilies against 12 substrates clustered by similarity between subfamilies. Results represent relative binding energy [kcal/mol].



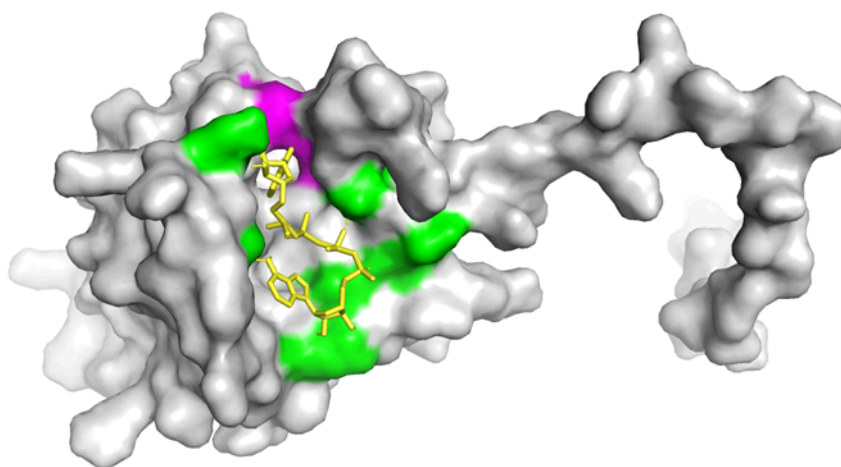
**Supplementary Figure S10. Conservation of DHNTPase residues**

Typical DHNTPase residues are conserved only within the NUDIX box in the proteins from subfamily B. MSA (Multiple Sequence Alignment) of DHNTPase from E. coli (PDB 2O1C)

and sequences from subfamily\_B. The conserved residues from the Nudix signature box are indicated in green, and the ones implicated in recognition or catalytic mechanisms are highlighted in blue. Other conserved residues are depicted in magenta. The region of the Nudix box is marked with orange stars. This Supplementary Figure S was made with the programs MAFFT v7.407 (W. Li et al., 2001; Rozewicki et al., 2019) and ESPript (Robert & Gouet, 2014). The conserved residues in DHNTPase from *E. coli* are based on (Gabelli et al., 2007).

## Supplementary characterization of subfamilies

The cluster of **subfamily\_A** includes representatives from the fungal phyla Chytridiomycota, Microsporidia, Mucoromycota, Olpidiomyota, and Zoopagomycota. Apart from the canonical NUDIX (PF00293) domain these proteins also contain two copies of the zinc finger C-x8-C-x5-C-x3-H domain (PF00642) and, in most cases, also the translation initiation factor eIF4-gamma/eIF5/eIF2-epsilon domain (PF02020). Almost all sequences from this subfamily have conserved residues typical for Ap4Ases such as NUDT2 from *Homo sapiens* (PDB: 3U53): Trp38, Pro41, Gly43, Lys89, and Tyr82 (Ge et al., 2013; Srouji et al., 2017). Moreover, ten residues downstream from the NUDIX box, these enzymes have a conserved phenylalanine which is also present in the human NUDT2 protein (see Supplementary Figures section 1). Molecular docking results of representatives of this subfamily showed high activity towards NAD<sup>+</sup>, Ap4A, and Ap3A. As proteins from this subfamily do not contain characteristics of NADH pyrophosphatases or diphosphatases (NADD) such as the SQPWPFPS motif, it is unlikely that they hydrolyze NAD<sup>+</sup>. This concludes that NUDT2 is known not to have de-NAD-ing activity (Sharma et al., 2020). On the other hand, as can be observed in Figure 6, the two residues needed for docking of Ap4A, Tyr461 and Phe507, lie on top of the active site, as they do in human NUDT2. The residues from within the NUDIX motif also interact with Ap4A (Figure 6). These findings strongly suggest that the substrate of proteins of **subfamily\_A** is Ap4A and, as human NUDT2, their role may be to maintain the intracellular level of Ap4A ensuring the viability of cells in adverse environments (Zegarra et al., 2023).





Supplementary Figure S8. Representatives of the newly identified subfamily A have high affinity to Ap4A and conserved features of Ap4Aases. Scheme showing the molecular docking of the NUDIX hydrolase domain-containing protein (EIE82761.1)(grey) from *Rhizopus delemar* with Ap4A (yellow) at pH 9.0. Corresponding residues that lie on top of the active site cleft in human NUDT2 are marked in magenta (Tyr461 and Phe507). Other binding residues (Asp409, Lys414, Gln423, Glu437, Glu440 and Lys468) are shown in green.

Sequences from the **subfamily\_B** have an elongated NUDIX box motif by one residue, and a recurrent coexisting domain, the translation initiation factor 2B subunit (PF1008) essential for eukaryotic translation initiation. In fungi, we identified members of this subfamily in the phyla Ascomycota, Chytridiomycota, Mucoromycota, and Zoopagomycota, and there are also some representatives in bacteria and archaea. The sequences from this cluster map to the DHNTPase-like enzymes (cd04664) according to the NCBI CDD database, however, they do not retain sequence motifs characteristics for such hydrolases. Furthermore, we compared the conserved residues typical for DHNTPases with the sequences of this subfamily (Supplementary Figures Section 7) and found that the residues and the characteristic alanine for DHNTPase within the NUDIX box are quite well-conserved. However, outside of the NUDIX box, the sequence is quite divergent. Moreover, typical dihydroneopterin triphosphate pyrophosphatases stack the substrate's ring into a hydrophobic pocket consisting of Phe81 and Phe84 (Gabelli et al., 2007), and these residues were not conserved in this subfamily. Finally, molecular docking of representatives of the subfamily confirmed that DHNTP is not likely a preferred substrate. Since the neighboring domains are essential for translation, we rather think that proteins from this subfamily maintain the housekeeping nature of NUDIX hydrolases.

Sequences from **subfamily\_C** are found only in species of the order *Diversisporales*. These enzymes have an elongated NUDIX box by the insertion of an aspartate (D): **G**X<sub>5</sub>Dx<sub>8</sub>**R**E<sub>x</sub><sub>2</sub>**E**E<sub>x</sub><sub>3</sub> (conserved residues from the NUDIX box are marked in the bold, see Supplementary Figures section 2). 62.5% of these hydrolases have a deoxynucleoside kinase domain (PF01712) in addition to the NUDIX motif. Their predicted subcellular localization is the nucleus suggesting that they may hydrolyze nucleotides. Proteins from **subfamily\_D** have sequence similarity to a *provisional plant NUDIX subfamily* in the CDD database (PLN03143). This may be because a small fraction of the representatives of this subfamily are from Protozoa and plants, although most of the proteins in the subfamily are fungal. Most of the proteins of **subfamily\_D** belong to early-diverging fungal phyla and they have a strongly conserved pattern in their NUDIX box: **G**M**L**Dx<sub>4</sub>Fx<sub>6</sub>E<sub>x</sub><sub>2</sub>**E**E<sub>x</sub>**G**U (conserved residues marked in bold).

Proteins of **subfamily\_E** are related to the human NUDIX enzyme NUDT1, although all members of the cluster are fungal proteins. Their position in the phylogenetic tree suggests that they may have originated as a result of a post-duplication diversification of NUDT1 paralogs. However, these enzymes differ from **do not retain all** characteristics that are distinctive of NUDT1. For example, the substrate affinity calculated through molecular docking is low for substrates that are typical of NUDT1 such as 8-oxo-dGTP, 8-oxo-dGDP, or dATP. This suggests that members of this subfamily probably operate distinctly from NUDT1.

**Subfamily\_F** groups NUDIX hydrolases only from a single species from the phylum Glomeromycota, *Diversispora epigaea*, that have the following NUDIX box signature motif: **GGH(I/V)EEDISxKxA(Q/R)REUxEEExGU** (in bold black are residues conserved from NUDIX box motif, in bold green residues conserved across most of the sequences in this family). Furthermore, this motif is preceded by two conserved lysines situated four residues apart and by a conserved histidine-tryptophan-glycine triplet of three residues from the NUDIX box (see Supplementary Figures section 4). Molecular docking showed a relatively high affinity for ApnA substrates, but the binding was not through the residues from within the NUDIX box making it unlikely that this subfamily hydrolases ApnA. All proteins are predicted to be localized in the nucleus and therefore they could be hydrolyzing nucleotides.

Both **subfamily\_G** and **subfamily\_H** were associated with human NUDT7 CoAse (cd03426) by NCBI CDD and Foldseek. Subfamily\_G has representatives from bacteria, and a few insects apart from the fungal proteins, while subfamily\_H is almost exclusively fungal with a few members from bacteria. However, molecular docking predicted that Coenzyme A is not the preferred substrate of the enzymes in these subfamilies, showing functional divergence from NUDT7.

**Subfamilies I and J** only include sequences from a single species from the phylum Glomeromycotina, *G. rosea*. According to NCBI CDD search, these enzymes have either NUDT1MTH1 (cd03427) or NUDT15MTH2 (cd04678) characteristics, which correspond to human NUDIX enzymes NUDT1 and NUDT15, respectively. In humans, these hydrolases are known to oxidize deoxyribonucleoside di- and triphosphates to preserve genomic integrity (Carter et al., 2015; Hashiguchi et al., 2018). Both subfamily genes are often found close to kinase-like domain-containing genes. However, the predicted structures of the fungal proteins from these subfamilies lack the typical NUDT15 recognition pocket and the Arg139 residue that is important for thiopurine metabolism. Furthermore, these enzymes also do not have the

conserved residues characteristic of human NUDT1 and molecular docking did not indicate high affinity towards the canonical NUDT1 or NUDT15 substrates. Even though there are structural differences between proteins from these subfamilies and NUDT1 and NUDT15, the highly conserved NUDIX box may indicate these proteins are engaged in housekeeping functions.

**subfamily\_K** groups fungal proteins of order *Glomerales*, but there are also representatives in bacteria. In addition to the canonical NUDIX domain, 57% of sequences have a retroviral aspartyl protease domain (PF08284) and 21% have a zinc knuckle domain (PF00098). The viral nature of the coexisting domains suggests that the domain architecture of this subfamily may be the consequence of the fusion of a NUDIX protein with a retrotransposon. Species of the *Rhizophagus* genus are known to have experienced massive transposon activity, causing that half of the genomic sequence to be some kind of repeat. The subcellular localization of the fungal sequences in this subfamily is predicted to be mostly the nucleus, perhaps due to the fusion with a retrotransposon. In contrast, human NUDT15, (which was annotated by NCBI CDD as a possible homolog of the proteins of this subfamily) is known to reside in the cytosol. Despite its possible association with retrotransposons, molecular docking showed that binding with 8-oxo-dGTP occurs through the residues that correspond to the binding pocket in NUDT15. In consequence, the members of this subfamily may be involved in thiopurine metabolism.

As for subfamily\_C, **subfamily\_L** only groups sequences from the phylum Glomeromycotina, particularly the *Diversisporales* order. Most NUDIX hydrolases in this cluster are predicted to be in the nucleus.

Contrary to most newly identified subfamilies, **subfamily\_M** includes proteins mainly from Ascomycota and Basidiomycota and a small subset of proteins from Bilateria and bacteria. Similar to the characterized Ap4Ases identified in the yeast species *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*, proteins of this subfamily lack the conserved tyrosine situated 16–18 amino acids after the NUDIX box (Dunn et al., 1999). The docking results ranked Ap3A and Ap4A as substrates with the lowest changes in free energy during binding, making them possible ligands. Therefore, enzymes from **subfamily\_M** might be Ap4ases involved in the diadenosine polyphosphate catabolic process, similar to the function of previously characterized yeast enzymes (Dunn et al., 1999). Based on their phylogenetic position, the members of **subfamily\_M** are related to the human

NUDT2 enzyme. According to the tree, proteins of **subfamily\_M** diverged from the same branch that gave rise to different human hydrolases from the canonical NUDIX family (NUDT2, NUDT6, and NUDT20), and also to 39S mitochondrial ribosomal proteins L46 that belong to the MRP-L46 family (PF11788), another member of the NUDIX superfamily.