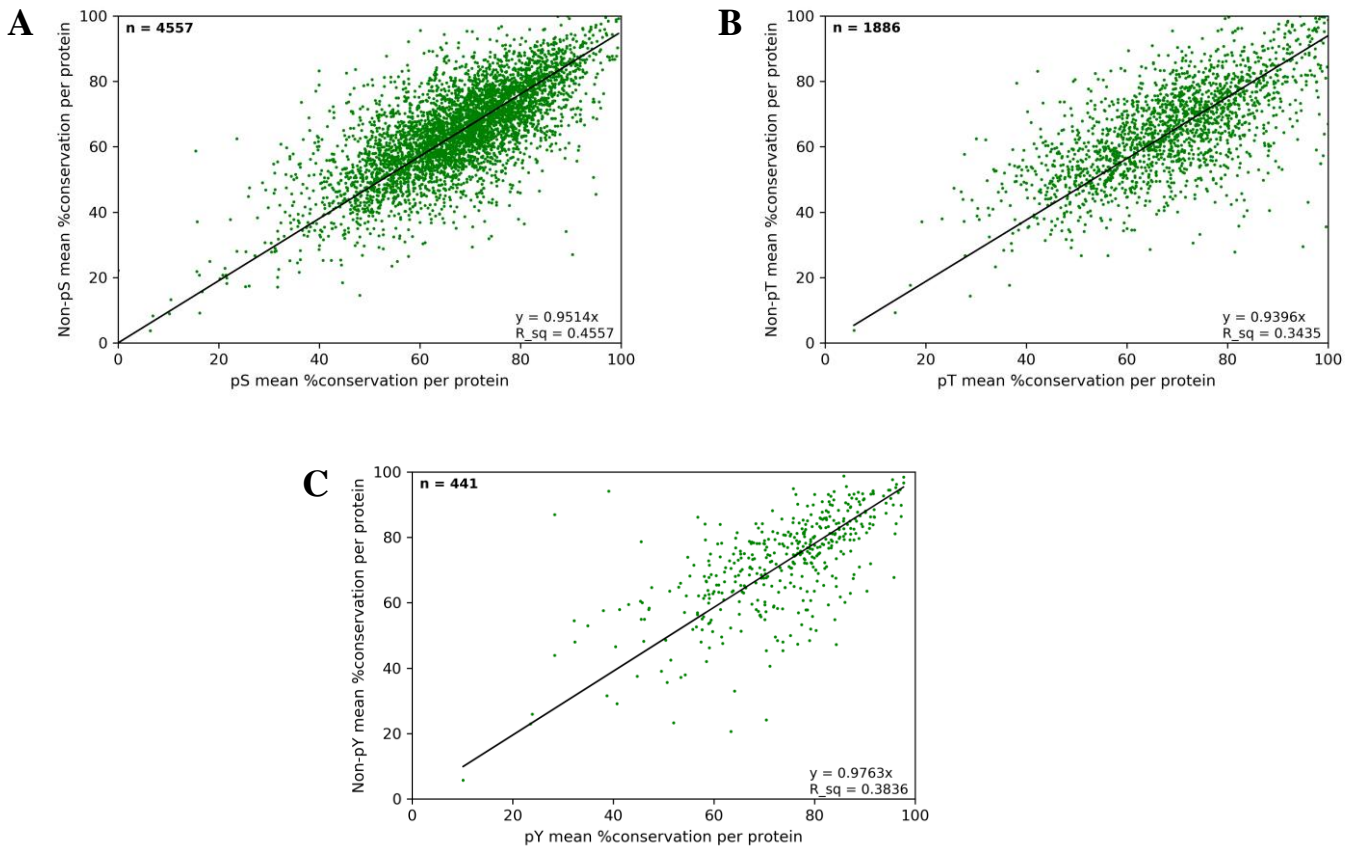


# Profiling the Human Phosphoproteome to Estimate the True Extent of Protein Phosphorylation and Conservation

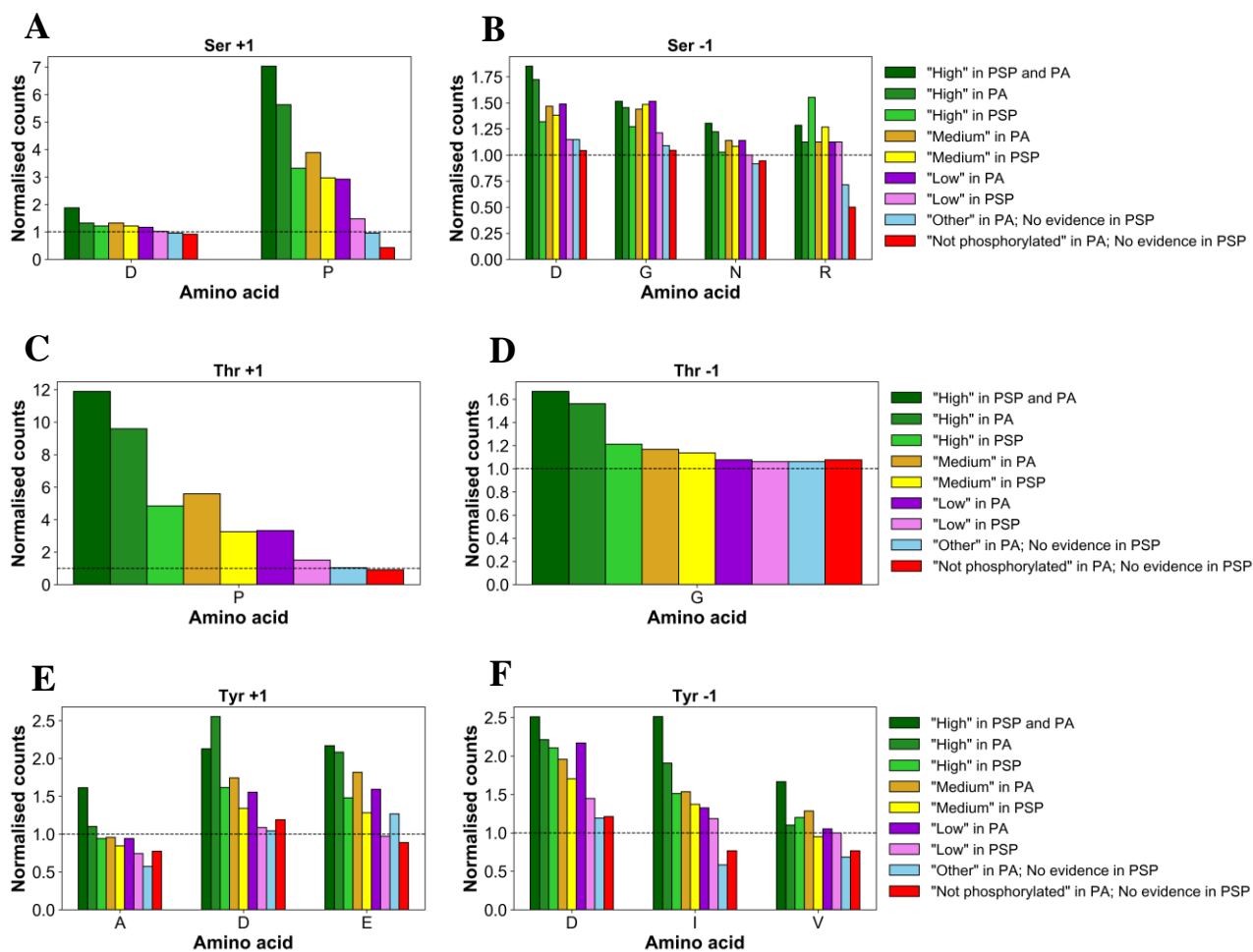
Anton Kalyuzhnyy

## Supplementary Information (SI)

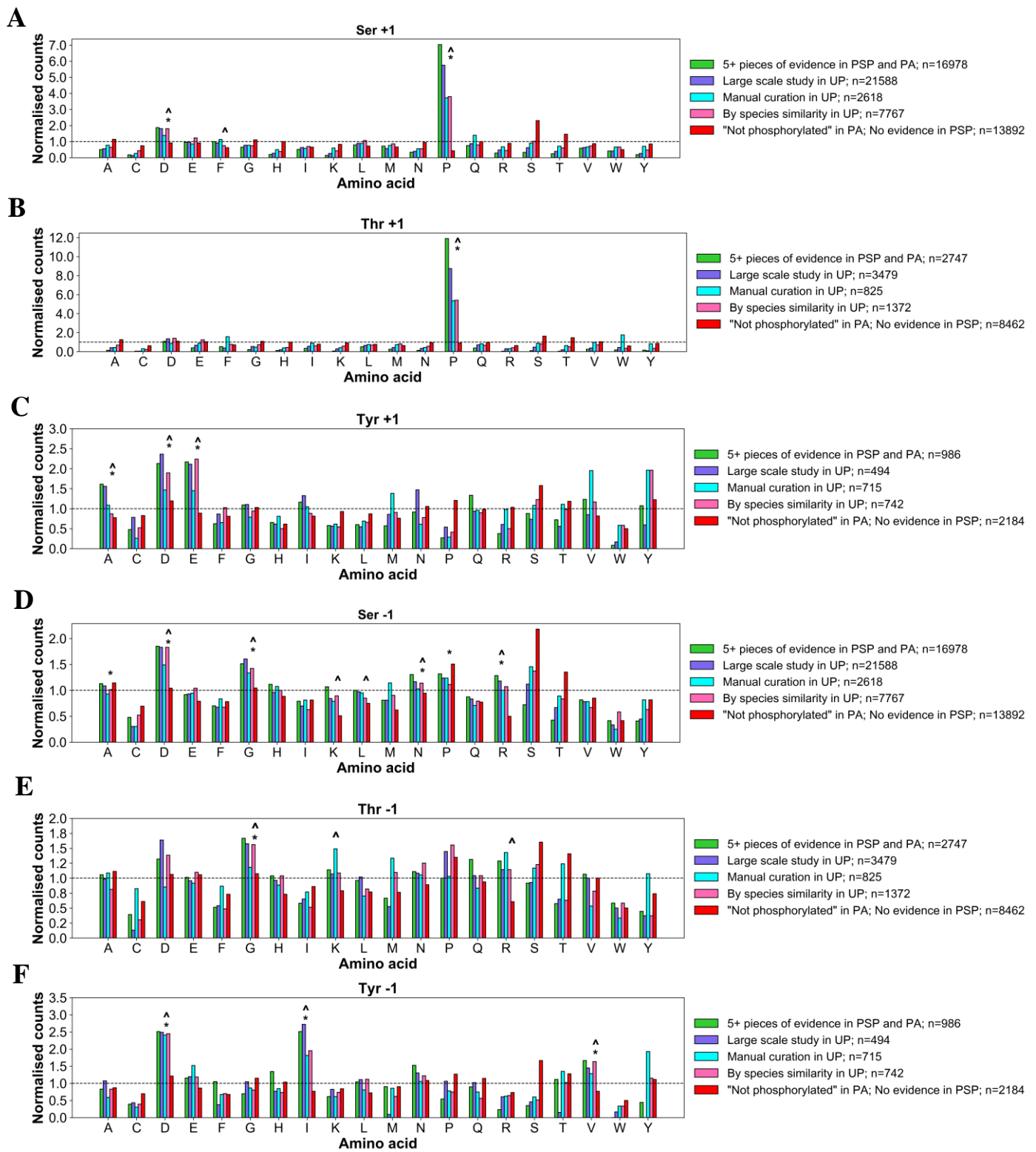
<b>Figure S1.</b> Linear regression analysis of conservation within found orthologues between phosphosites and non-phosphosites per protein.	Page 1 in SI file
<b>Figure S2.</b> Proximal site and FDR analysis performed separately for PhosphoSitePlus and PeptideAtlas sets of STY sites.	Page 2 in SI file
<b>Figure S3.</b> Proximal site and FDR analysis of STY sites with phosphorylation evidence in UniProt.	Pages 3-4 in SI file
<b>Figure S4.</b> Count of significant functional groups identified in DAVID for protein sets containing different highest ranked STY sites.	Page 5 in SI file
<b>Figure S5.</b> Top 10 functional categories for which protein sets containing different highest ranked STY sites were enriched in DAVID.	Page 6 in SI file
<b>Figure S6.</b> Conservation patterns of individual Ser, Thr and Tyr phosphosites from human proteins.	Pages 7-8 in SI file
<b>Figure S7.</b> The functional enrichment analysis by DAVID of proteins with different phosphosite conservation patterns	Pages 9-10 in SI file
<b>Table S1.</b> Filtered PeptideAtlas build with human STY sites that have at least 1 associated PSM.	Submitted separately as .zip file
<b>Table S2.</b> Filtered PhosphoSitePlus build with human STY sites from canonical protein sequences.	Submitted separately as .zip file
<b>Table S3.</b> Proteomes of eukaryotic species used in conservation analysis.	Pages 11-13 in SI file
<b>Table S4.</b> Proteins in the human proteome which were not analysed and the reasons for their exclusion.	Submitted separately as .xlsx file
<b>Table S5.</b> Summary of all STY sites in our analysis, their conservation data, proximal sites, phosphorylation likelihood and structural data.	Submitted separately as .zip file
<b>Table S6.</b> FASTA sequences of analysed proteins.	Submitted separately as .zip file
<b>Table S7.</b> Positions of secondary structures within analysed target proteins in the human proteome.	Submitted separately as .xlsx file
<b>Table S8.</b> Counts of STY sites in phosphorylation likelihood sets based on evidence in PeptideAtlas before considering evidence in PhosphoSitePlus.	Submitted separately as .xlsx file
<b>Table S9.</b> Cross-referencing sets of sites between PhosphoSitePlus and PeptideAtlas	Submitted separately as .xlsx file
<b>Table S10.</b> STY sites in human proteome with plenty of phosphorylation evidence in both PhosphoSitePlus and PeptideAtlas.	Submitted separately as .xlsx file
<b>Table S11.</b> STY conservation scores within proteins which had at least 3 phosphosites and 3 non-phosphosites.	Submitted separately as .xlsx file
<b>Table S12.</b> Conservation of STY sites in each phosphorylation likelihood set.	Submitted separately as .xlsx file
<b>Table S13.</b> Counts of amino acids adjacent to target STY sites at -1 and +1 positions within phosphorylation likelihood sets.	Submitted separately as .xlsx file
<b>Table S14.</b> Calculating phosphosite FDR within sets of STY sites ranked according to combined PhosphoSitePlus and PeptideAtlas evidence.	Submitted separately as .xlsx file
<b>Table S15.</b> Calculating phosphosite FDR within separate PhosphoSitePlus and PeptideAtlas sets of STY sites.	Submitted separately as .xlsx file
<b>Table S16.</b> Highest ranked STY site within each analysed target protein in the human proteome.	Submitted separately as .xlsx file
<b>Table S17.</b> Percentage of proteins within each ranked set linked to a certain UniProt term.	Submitted separately as .xlsx file
<b>Table S18.</b> Conservation Mastersheet summarising phosphosite conservation results.	Submitted separately as .xlsx file
<b>Table S19.</b> Predicted phosphosites across 100 eukaryotic species.	Submitted separately as .xlsx file



**Figure S1.** Mean % conservation across found orthologues of likely (A) Ser, (B) Thr, (C) Tyr phosphosites and corresponding likely non-

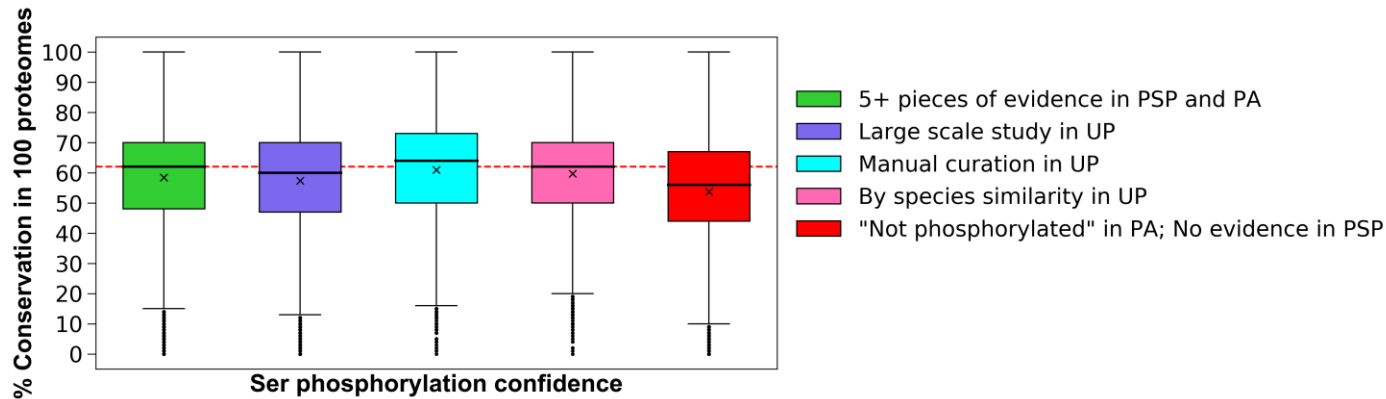
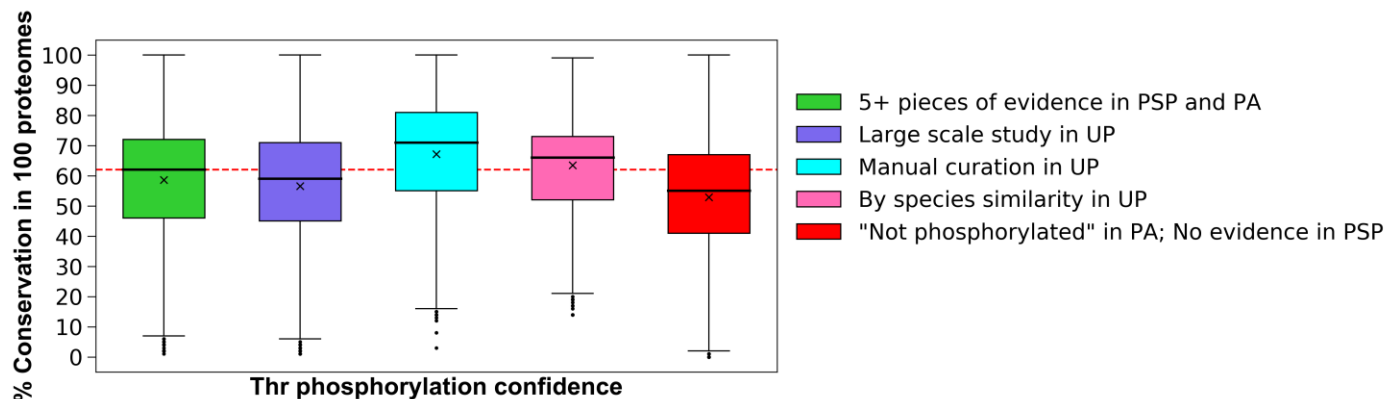


**Figure S2.** Normalised counts of proximal amino acids positioned at (A) +1 around Ser; (B) -1 around Ser; (C) +1 around Thr; (D) -1 around Thr; (E) +1 around Tyr; (F) -1 around Tyr sites of various phosphorylation likelihood based on evidence in PSP and PA, which are significantly (Bonferroni corrected p-value <0.001) enriched in the “High in PSP and PA” compared to the “Not phosphorylated” set and to the expected amino acid distribution in the human proteome (represented by dotted baseline).

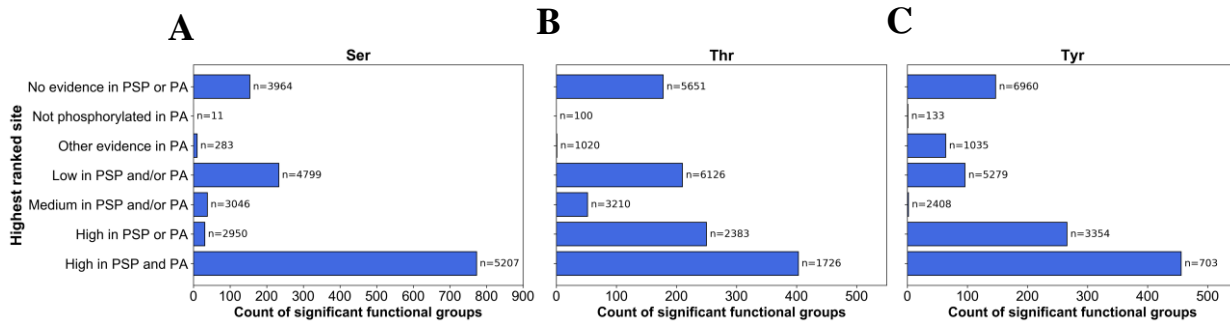


**G**

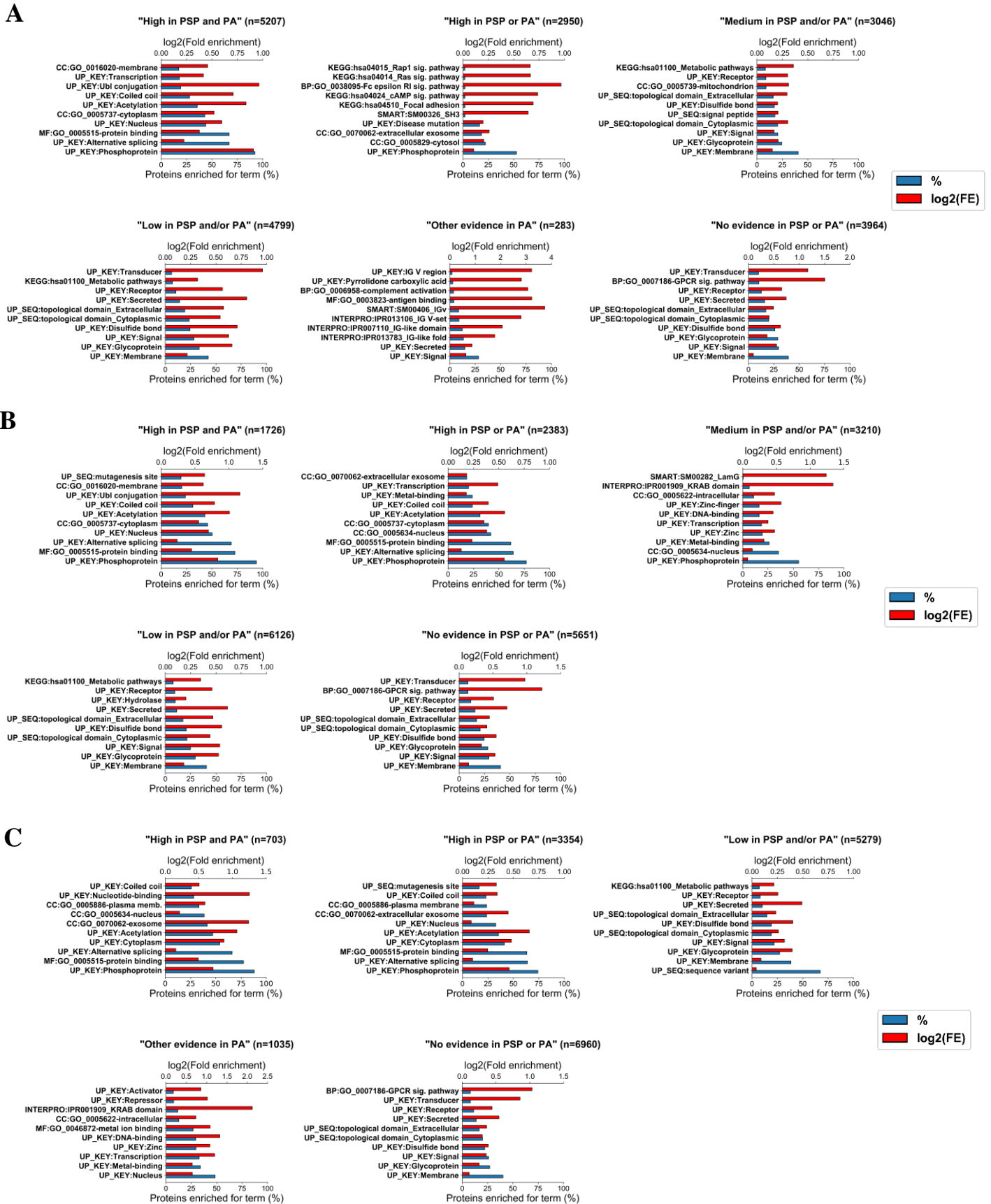
Evidence source in UniProt	Ser count	Ser % FDR (95% CI)	Ser TP count	Thr count	Thr % FDR (95% CI)	Thr TP count	Tyr count	Tyr % FDR (95% CI)	Tyr TP count
Large scale study data	21588	7 (± 8)	20077	3479	22 (± 14)	3910	494	6 (± 7)	464

**H****I**

**Figure S3.** The analysis of UniProt (UP) sites. Counts of proximal amino acids positioned at (A) +1 around Ser; (B) +1 around Thr; (C) +1 around Tyr; (D) -1 around Ser; (E) -1 around Thr; (F) -1 around Tyr sites with different phosphorylation evidence types across PhosphoSitePlus (PSP), PeptideAtlas (PA) and UP, normalised to observed distribution of those amino acids in human proteome (represented by dotted baseline fixed at 1). Significant (Bonferroni corrected p-value <0.001) enrichment of proximal amino acids in the “5+ pieces of evidence in PSP and PA” set is highlighted by the caret symbol (^) when compared against the “Not phosphorylated” set, and an asterisk symbol (\*) when compared to the expected amino acid distribution. (G) False discovery rate (FDR) and counts of estimated true positive (TP) Ser, Thr and Tyr phosphosites within a set of UniProt sites identified by large scale proteomics studies and which are found in the reference human proteome. Box plots show conservation percentages (%) across 100 eukaryotic species of human (H) Ser, (I) Thr sites categorised based on evidence type in PSP/PA and UniProt (UP). Within each box, a horizontal line represents median % conservation, an (x) symbol represents mean % conservation per group. Each box extends from the 25th to the 75th percentile of each set's distribution of conservation % values. Vertical lines extending from the boxes correspond to adjacent values. Dots (•) represent outlier values. Red line shows median % conservation in “5+ pieces of evidence in PSP and PA” set for visual comparison.



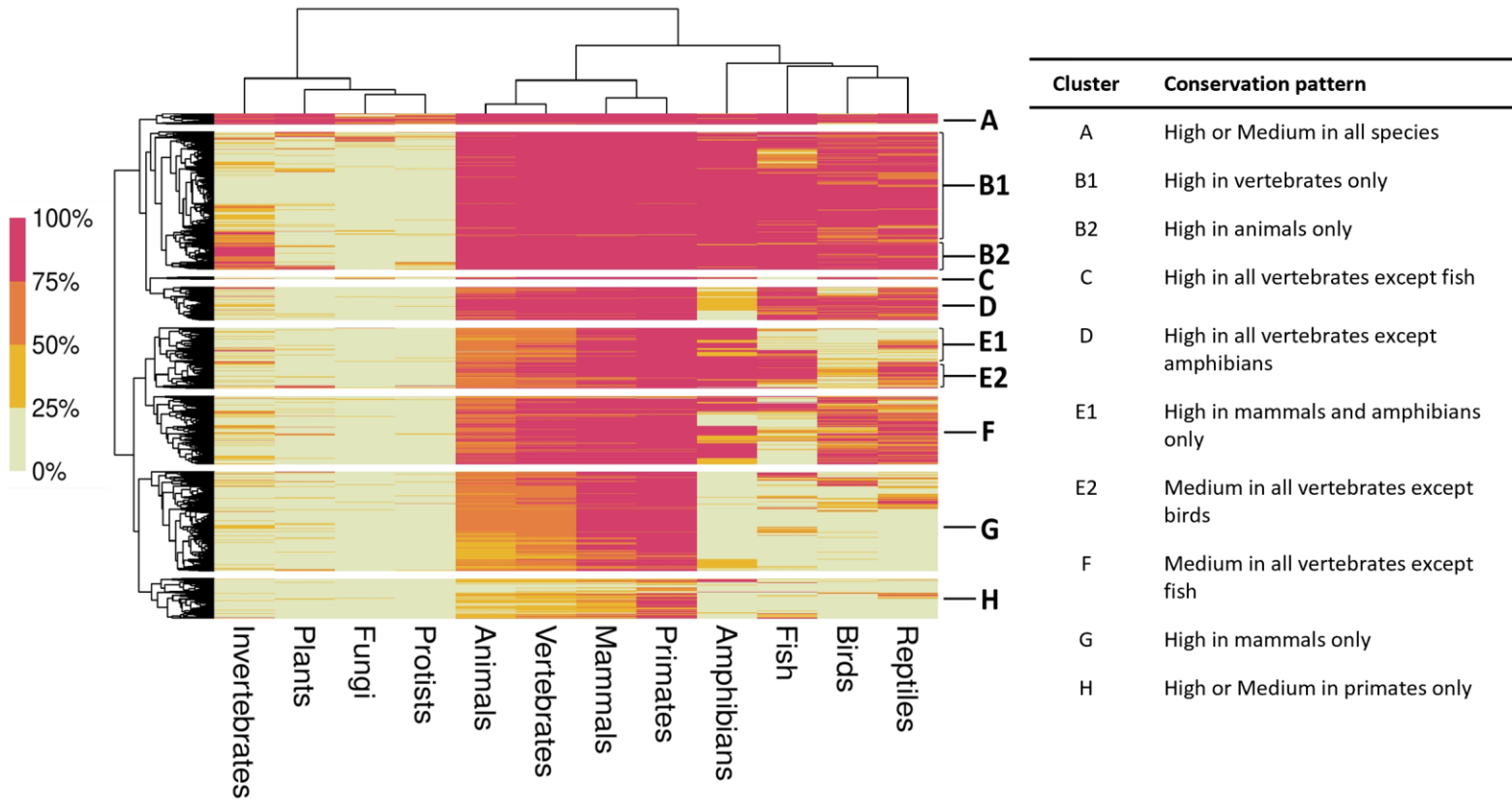
**Figure S4.** Count of significant (Benjamini–Hochberg adj. p-value <0.05) functional groups identified in DAVID for protein sets containing different highest ranked (A) Ser, (B) Thr, (C) Tyr sites based on phosphorylation likelihood sets in PSP and PA. The number of proteins in each set is presented by *n*.



**Figure S5.** Top 10 functional categories for which protein sets containing various highest ranked (A) Ser, (B) Thr, (C) Tyr sites based on the amount of available phosphorylation evidence were significantly enriched in DAVID (Benjamini–Hochberg corrected  $p$ -value  $< 0.05$ ). For each protein set, the percentage of proteins (%) enriched for a particular functional category is given as well as the  $\log_2$ (fold enrichment) for that category. The number of proteins in each set is presented by  $n$ .

A

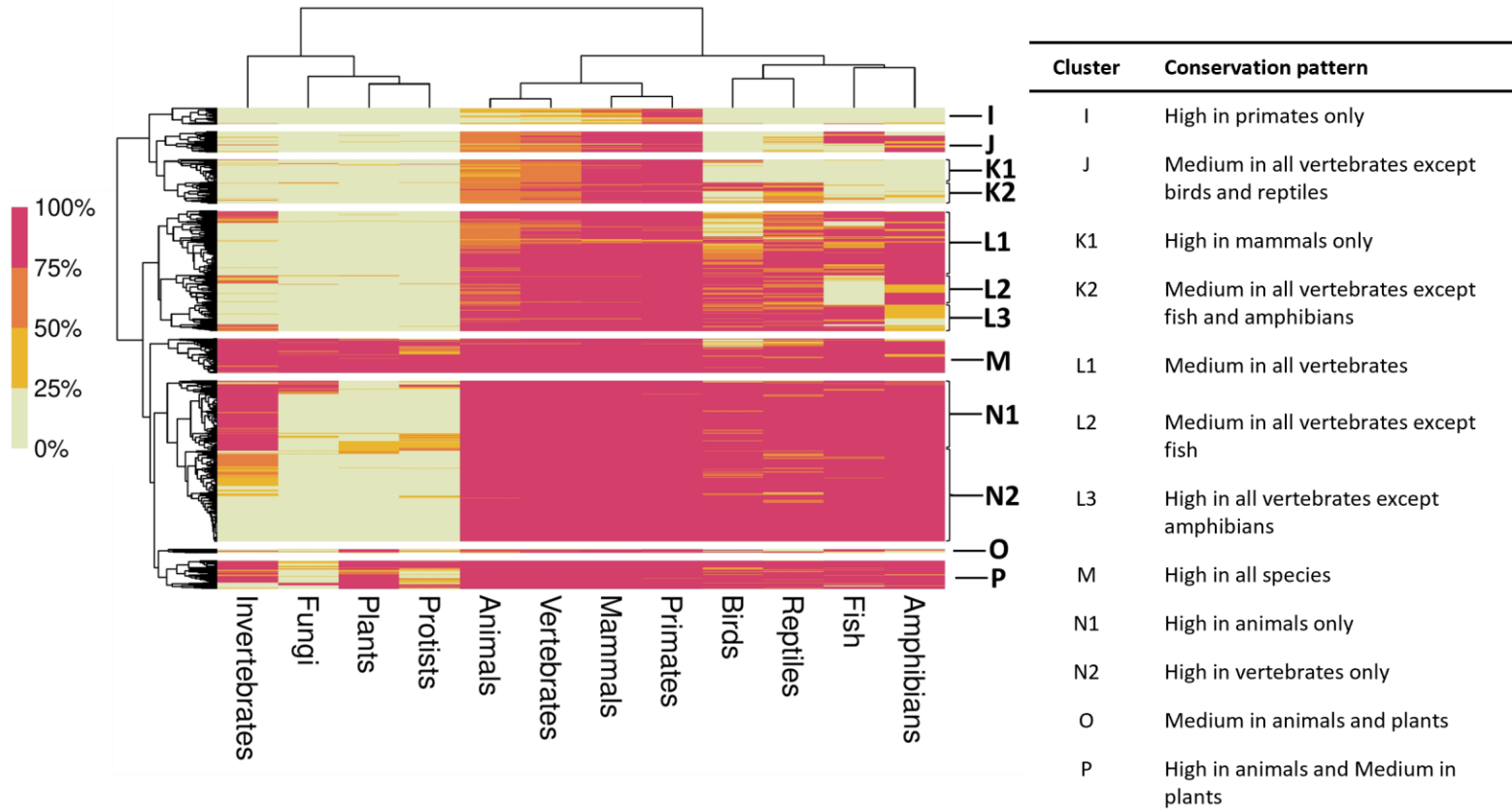
## pST conservation (%) (n=19725) within species groups





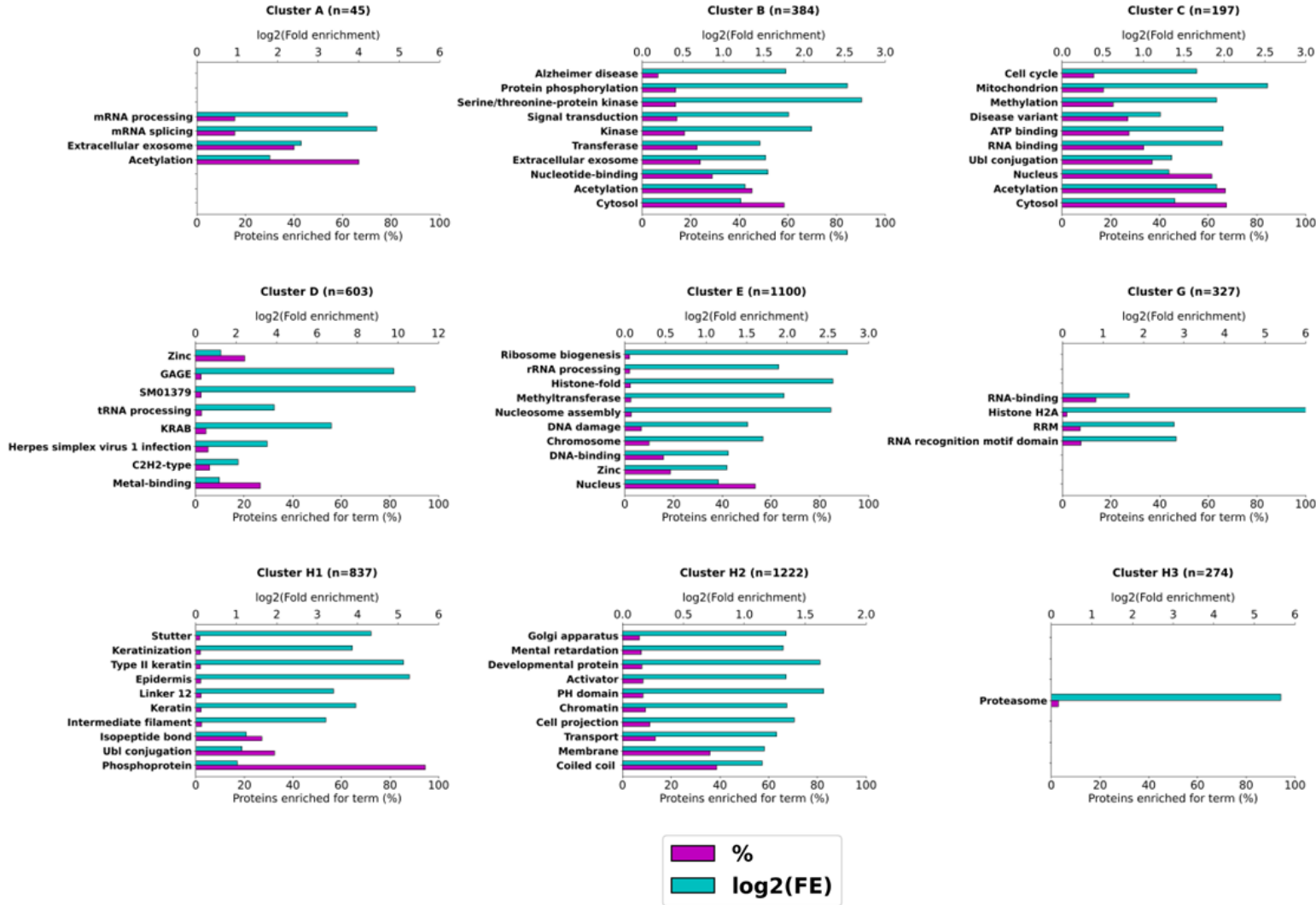
**B**

**pY conservation (%) (n=986)  
within species groups**



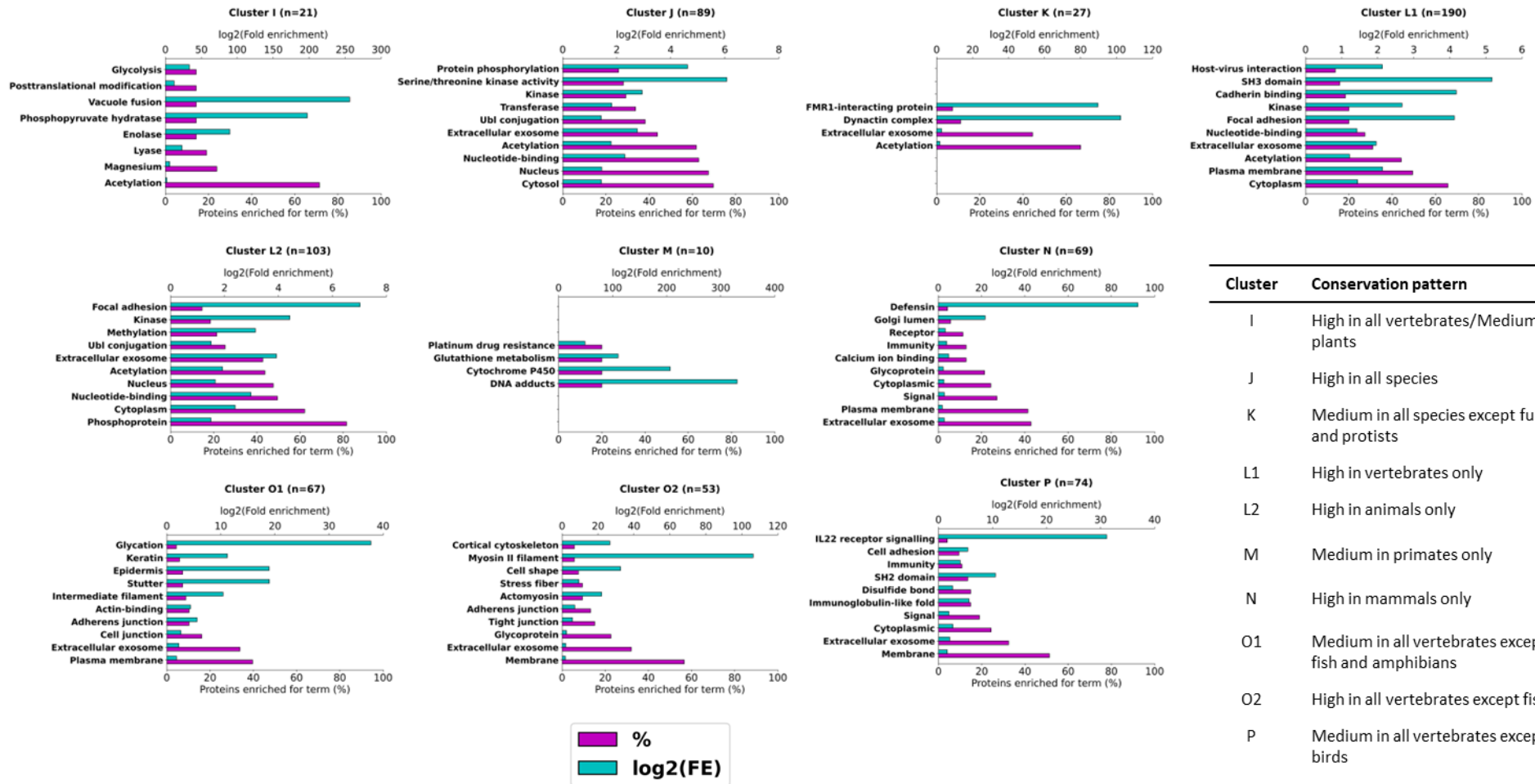
**SI Figure S6.** Conservation patterns of (A) Ser/Thr and (B) Tyr phosphosites from human proteins across the groups of eukaryotic species. Each row in the heatmap represents an individual phosphosite and its conservation across specific species groups which are separated into columns. Conservation is scored as a percentage out of all species per group and reflected by a colour gradient divided at quarterly intervals. Phosphosites were clustered based on their similarity in conservation patterns using the Euclidean distance method. For each cluster, a label is assigned which describes the most observed conservation pattern (i.e., at least 50% of sites in the cluster follow the described phosphosite conservation pattern), where high and medium conservation refers to conservation scores of  $\geq 75\%$  and  $\geq 50\%$ , respectively. The total number of analysed phosphosites is given by  $n$

A



Cluster	Conservation pattern
A	High in all vertebrates except birds
B	High in animals only
C	High or Medium in all species
D	High in primates only
E	High in mammals only
F	Medium in all vertebrates except birds (no enrichment)
G	High in mammals/Medium in reptiles
H1	High in all vertebrates except fish
H2	High in vertebrates only
H3	Medium in all vertebrates except amphibians

**B**



**SI Figure S7.** Top 10 functional categories for which protein sets with different (A) Ser/Thr and (B) Tyr phosphosite conservation patterns were significantly enriched in DAVID (Benjamini–Hochberg corrected  $p$ -value  $<0.05$ ). For each protein set, the % of proteins enriched for a particular functional category is given as well as the  $\log_2(\text{fold enrichment})$  for that set. The number of proteins in each set is presented by  $n$ .

**Table S3.** Proteomes of eukaryotic species analysed in the conservation analysis.

Organism name from Uniprot	Proteome ID	Organism ID	BUSCO score	Gene count
<i>Pan troglodytes</i> (Chimpanzee)	UP000002277	9598	C:98.3% [S:48.4%,D:49.9%],F:0.5%,M:1.1%,n:6192	23003
<i>Pan paniscus</i> (Pygmy chimpanzee)	UP000240080	9597	C:97.4% [S:54.4%,D:43%],F:1.4%,M:1.1%,n:6192	21211
<i>Gorilla gorilla gorilla</i> (Western lowland gorilla)	UP000001519	9595	C:97.3% [S:52%,D:45.3%],F:1.8%,M:1%,n:6192	21787
<i>Pongo abelii</i> (Sumatran orangutan)	UP000001595	9601	C:94.5% [S:88.5%,D:6%],F:4.2%,M:1.3%,n:6192	21992
<i>Nomascus leucogenys</i> (Northern white-cheeked gibbon)	UP000001073	61853	C:96.3% [S:57.7%,D:38.6%],F:2.6%,M:1.1%,n:6192	20753
<i>Macaca mulatta</i> (Rhesus macaque)	UP000006718	9544	C:89.6% [S:47.8%,D:41.8%],F:1.5%,M:8.9%,n:6192	21868
<i>Cercocebus atys</i> (Sooty mangabey)	UP000233060	9531	C:98% [S:48.3%,D:49.7%],F:1.1%,M:0.9%,n:6192	20874
<i>Rhinopithecus bieti</i> (Black snub-nosed monkey)	UP000233180	61621	C:95.1% [S:49.6%,D:45.5%],F:2.9%,M:2%,n:6192	20845
<i>Chlorocebus sabaues</i> (Green monkey)	UP000029965	60711	C:95.4% [S:94.4%,D:1%],F:3.4%,M:1.2%,n:6192	19136
<i>Macaca fascicularis</i> (Crab-eating macaque)	UP000233100	9541	C:98.2% [S:48.8%,D:49.3%],F:1.1%,M:0.8%,n:6192	22278
<i>Papio anubis</i> (Olive baboon)	UP000028761	9555	C:98.4% [S:51.7%,D:46.7%],F:0.8%,M:0.8%,n:6192	21559
<i>Mandrillus leucophaeus</i> (Drill)	UP000233140	9568	C:95.5% [S:55.1%,D:40.3%],F:3%,M:1.6%,n:6192	20767
<i>Saimiri boliviensis boliviensis</i> (Bolivian squirrel monkey)	UP000233220	39432	C:96.3% [S:50.1%,D:46.2%],F:2.2%,M:1.5%,n:6192	19356
<i>Callithrix jacchus</i> (White-tufted-ear marmoset)	UP000008225	9483	C:97.9% [S:52.7%,D:45.2%],F:1%,M:1.1%,n:6192	22587
<i>Aotus nancymae</i> (Ma's night monkey)	UP000233020	37293	C:97.4% [S:50.9%,D:46.5%],F:1.3%,M:1.3%,n:6192	20363
<i>Tarsius syrichta</i> (Philippine tarsier)	UP000189704	1868482	C:75.6% [S:60%,D:15.6%],F:5.7%,M:18.7%,n:6192	19956
<i>Otolemur garnettii</i> (Small-eared galago)	UP000005225	30611	C:96.8% [S:94%,D:2.8%],F:2.1%,M:1.1%,n:6192	19443
<i>Propithecus coquereli</i> (Coquerel's sifaka)	UP000233160	379532	C:92% [S:63.6%,D:28.4%],F:4.1%,M:3.9%,n:6192	17876
<i>Ictidomys tridecemlineatus</i> (Thirteen-lined ground squirrel)	UP000005215	43179	C:94.2% [S:71.1%,D:23%],F:2.6%,M:3.2%,n:6192	18440
<i>Cavia porcellus</i> (Guinea pig)	UP000005447	10141	C:94.2% [S:70.3%,D:23.9%],F:2.9%,M:2.9%,n:6192	18247
<i>Mus musculus</i> (Mouse)	UP000000589	10090	C:99.7% [S:52.1%,D:47.6%],F:0.2%,M:0.1%,n:6192	21982
<i>Oryctolagus cuniculus</i> (Rabbit)	UP000001811	9986	C:91.2% [S:84.3%,D:6.9%],F:4.6%,M:4.1%,n:6192	21178
<i>Cricetulus griseus</i> (Chinese hamster)	UP000001075	10029	C:63.7% [S:63%,D:0.7%],F:19.2%,M:17.1%,n:6192	23874
<i>Fukomys damarensis</i> (Damaraland mole rat)	UP000028990	885580	C:82.4% [S:81.9%,D:0.5%],F:7.7%,M:9.9%,n:6192	20401
<i>Mesocricetus auratus</i> (Golden hamster)	UP000189706	10036	C:77.6% [S:54.7%,D:22.9%],F:1.9%,M:20.5%,n:6192	20418
<i>Dipodomys ordii</i> (Ord's kangaroo rat)	UP000081671	10020	C:93.7% [S:66.1%,D:27.6%],F:5.4%,M:0.9%,n:6192	19730
<i>Heterocephalus glaber</i> (Naked mole rat)	UP000006813	10181	C:87% [S:85.5%,D:1.5%],F:6.7%,M:6.2%,n:6192	21445
<i>Vombatus ursinus</i> (Common wombat)	UP000314987	29139	C:96.6% [S:56.2%,D:40.4%],F:1.8%,M:1.6%,n:4104	19872
<i>Myotis lucifugus</i> (Little brown bat)	UP000001074	59463	C:95.1% [S:90%,D:5.1%],F:3.5%,M:1.4%,n:6253	19655
<i>Canis lupus familiaris</i> (Dog)	UP000002254	9615	C:97% [S:45.5%,D:51.5%],F:1.7%,M:1.3%,n:6253	20624
<i>Capra hircus</i> (Goat)	UP000291000	9925	C:98% [S:57.9%,D:40.1%],F:1.2%,M:0.8%,n:6253	21149
<i>Ovis aries</i> (Sheep)	UP000002356	9940	C:98% [S:86.8%,D:11.2%],F:1.4%,M:0.5%,n:6253	21212
<i>Sus scrofa</i> (Pig)	UP000008227	9823	C:95.3% [S:49%,D:46.3%],F:2.4%,M:2.3%,n:6253	22130
<i>Felis catus</i> (Cat)	UP000011712	9685	C:97% [S:67%,D:30%],F:1.4%,M:1.6%,n:6253	19645
<i>Ailuropoda melanoleuca</i> (Giant panda)	UP000008912	9646	C:97.4% [S:91.7%,D:5.8%],F:2%,M:0.6%,n:6253	19332
<i>Pteropus alecto</i> (Black flying fox)	UP000010552	9402	C:84.2% [S:83.7%,D:0.5%],F:9%,M:6.8%,n:6253	19520
<i>Erinaceus europaeus</i> (Western European hedgehog)	UP000079721	9365	C:95.2% [S:66.7%,D:28.4%],F:3.6%,M:1.2%,n:6253	19242
<i>Equus caballus</i> (Horse)	UP000002281	9796	C:97.8% [S:57.5%,D:40.3%],F:1.3%,M:0.9%,n:6253	20845
<i>Bos taurus</i> (Bovine)	UP000009136	9913	C:98.2% [S:34.4%,D:63.8%],F:1.2%,M:0.6%,n:6253	23844
<i>Mustela putorius furo</i> (European domestic ferret)	UP000000715	9669	C:96.2% [S:94.9%,D:1.3%],F:2.4%,M:1.4%,n:6253	19902
<i>Lipotes vexillifer</i> (Yangtze river dolphin)	UP000265300	118797	C:98.5% [S:74.2%,D:24.3%],F:1.1%,M:0.3%,n:6253	18846
<i>Leptonychotes weddellii</i> (Weddell seal)	UP000245341	9713	C:85% [S:50.5%,D:34.5%],F:14%,M:1%,n:6253	13162
<i>Ursus maritimus</i> (Polar bear)	UP000261680	29073	C:97.1% [S:71.8%,D:25.3%],F:2.6%,M:0.2%,n:6253	19368
<i>Delphinapterus leucas</i> (Beluga whale)	UP000248483	9749	C:98.6% [S:44%,D:54.6%],F:1.1%,M:0.4%,n:6253	17043

<i>Odobenus rosmarus divergens</i> (Pacific walrus)	UP000245340	9708	C:99% [S:64%,D:35%],F:0.9%,M:0.1%,n:6253	19331
<i>Physeter macrocephalus</i> (Sperm whale)	UP000248484	9755	C:86.3% [S:53.7%,D:32.6%],F:0.8%,M:12.9%,n:6253	20100
<i>Tursiops truncatus</i> (Atlantic bottle-nosed dolphin)	UP000245320	9739	C:84.1% [S:42.6%,D:41.4%],F:8.4%,M:7.6%,n:6253	17075
<i>Loxodonta africana</i> (African elephant)	UP000007646	9785	C:97.4% [S:73.3%,D:24.1%],F:1.9%,M:0.6%,n:4104	20015
<i>Trichechus manatus latirostris</i> (Florida manatee)	UP000248480	127582	C:98.1% [S:55.4%,D:42.7%],F:1.6%,M:0.3%,n:4104	19079
<i>Ornithorhynchus anatinus</i> (Duckbill platypus)	UP000002279	9258	C:75.7% [S:68%,D:7.7%],F:18.8%,M:5.5%,n:4104	21677
<i>Meleagris gallopavo</i> (Wild turkey)	UP000001645	9103	C:91.1% [S:81.4%,D:9.7%],F:5.4%,M:3.5%,n:4915	14164
<i>Taeniopygia guttata</i> (Zebra finch)	UP000007754	59729	C:95.4% [S:91.2%,D:4.3%],F:3.6%,M:0.9%,n:4915	17428
<i>Anas platyrhynchos</i> (Mallard)	UP000296049	8839	C:79.4% [S:78.5%,D:0.9%],F:7.8%,M:12.8%,n:4915	16574
<i>Dryobates pubescens</i> (Downy woodpecker)	UP000053875	118200	C:94.6% [S:93.7%,D:1%],F:2.2%,M:3.2%,n:4915	13097
<i>Tinamus guttatus</i> (White-throated tinamou)	UP000053641	94827	C:89.8% [S:88.5%,D:1.3%],F:5.5%,M:4.7%,n:4915	13377
<i>Amazona aestiva</i> (Blue-fronted Amazon parrot)	UP000051836	12930	C:86.6% [S:85.4%,D:1.2%],F:7.5%,M:5.9%,n:4915	16092
<i>Calypte anna</i> (Anna's hummingbird)	UP000054308	9244	C:95.4% [S:94.5%,D:0.9%],F:1.7%,M:2.9%,n:4915	13267
<i>Columba livia</i> (Rock dove)	UP000053872	8932	C:93% [S:78.5%,D:14.4%],F:5%,M:2.1%,n:4915	14619
<i>Callipepla squamata</i> (Scaled quail)	UP000198323	9009	C:77.6% [S:75.8%,D:1.8%],F:14.5%,M:7.9%,n:4915	16973
<i>Aptenodytes forsteri</i> (Emperor penguin)	UP000053286	9233	C:97.9% [S:97%,D:0.9%],F:0.8%,M:1.3%,n:4915	13704
<i>Opisthocomus hoazin</i> (Hoatzin)	UP000053605	30419	C:95.3% [S:94.7%,D:0.6%],F:2.1%,M:2.6%,n:4915	12773
<i>Egretta garzetta</i> (Little egret)	UP000053119	188379	C:96.7% [S:96%,D:0.8%],F:1%,M:2.3%,n:4915	13489
<i>Alligator mississippiensis</i> (American alligator)	UP000050525	8496	C:88.6% [S:67.4%,D:21.2%],F:6.8%,M:4.6%,n:3950	24656
<i>Alligator sinensis</i> (Chinese alligator)	UP000189705	38654	C:74.4% [S:55.1%,D:19.3%],F:2.4%,M:23.2%,n:3950	19111
<i>Anolis carolinensis</i> (Green anole)	UP000001646	28377	C:92.1% [S:88.4%,D:3.7%],F:5.3%,M:2.6%,n:3950	18525
<i>Pelodiscus sinensis</i> (Chinese softshell turtle)	UP000007267	13735	C:93.5% [S:79.1%,D:14.4%],F:4.9%,M:1.6%,n:3950	18109
<i>Salmo salar</i> (Atlantic salmon)	UP000087266	8030	C:97.9% [S:23.9%,D:74%],F:1.5%,M:0.6%,n:4584	47717
<i>Oncorhynchus mykiss</i> (Rainbow trout)	UP000193380	8022	C:77.1% [S:46.4%,D:30.7%],F:10.9%,M:12%,n:4584	46447
<i>Gasterosteus aculeatus</i> (Three-spined stickleback)	UP000007635	69293	C:97.5% [S:74.5%,D:23%],F:1.9%,M:0.6%,n:4584	20665
<i>Seriola dumerili</i> (Greater amberjack)	UP000261420	41447	C:97.6% [S:69.5%,D:28.1%],F:1.5%,M:0.8%,n:4584	23238
<i>Takifugu rubripes</i> (Japanese pufferfish)	UP000005226	31033	C:95.1% [S:65.4%,D:29.7%],F:2.7%,M:2.2%,n:4584	20591
<i>Xenopus laevis</i> (African clawed frog)	UP000186698	8355	C:95.6% [S:41.5%,D:54.1%],F:1.6%,M:2.8%,n:3950	43235
<i>Xenopus tropicalis</i> (Western clawed frog)	UP000008143	8364	C:62.9% [S:37.4%,D:25.4%],F:2.4%,M:34.8%,n:3950	35973
<i>Daphnia pulex</i> (Water flea)	UP000000305	6669	C:96.1% [S:93.8%,D:2.3%],F:1.4%,M:2.5%,n:1066	30118
<i>Tribolium castaneum</i> (Red flour beetle)	UP000007266	7070	C:98.6% [S:91.5%,D:7.1%],F:1.2%,M:0.2%,n:1658	16568
<i>Bombyx mori</i> (Silk moth)	UP000005204	7091	C:89.9% [S:89.4%,D:0.5%],F:6.8%,M:3.3%,n:1658	14773
<i>Anopheles darlingi</i> (Mosquito)	UP000000673	43151	C:89.4% [S:89%,D:0.4%],F:2.6%,M:7.9%,n:2799	10447
<i>Drosophila melanogaster</i> (Fruit fly)	UP000000803	7227	C:99.3% [S:38.2%,D:61.1%],F:0.4%,M:0.3%,n:2799	13790
<i>Harpegnathos saltator</i> (Jerdon's jumping ant)	UP000008237	610380	C:90.3% [S:90%,D:0.3%],F:6.4%,M:3.3%,n:4415	15029
<i>Ooceraea biroi</i> (Clonal raider ant)	UP000053097	2015173	C:95.9% [S:95.4%,D:0.4%],F:3.4%,M:0.8%,n:4415	16497
<i>Papilio xuthus</i> (Asian swallowtail butterfly)	UP000053268	66420	C:96.3% [S:94.9%,D:1.4%],F:2.4%,M:1.3%,n:1658	15265
<i>Zootermopsis nevadensis</i> (Dampwood termite)	UP000027135	136037	C:93.9% [S:92.8%,D:1.1%],F:1.6%,M:4.5%,n:1658	14539
<i>Operophtera brumata</i> (winter moth)	UP000037510	104452	C:75.9% [S:73.4%,D:2.5%],F:15%,M:9.2%,n:1658	16814
<i>Lucilia cuprina</i> (Green bottle fly)	UP000037069	7375	C:92.5% [S:91.5%,D:1%],F:1.8%,M:5.6%,n:2799	14353
<i>Saccharomyces cerevisiae</i> (strain ATCC 204508 / S288c) (Baker's yeast)	UP000002311	559292	C:98.9% [S:98.3%,D:0.6%],F:1.1%,M:0%,n:1711	6049
<i>Emericella nidulans</i> (strain FGSC A4 / ATCC 38163 / CBS 112.46 / NRRL 194 / M139) ( <i>Aspergillus nidulans</i> )	UP000000560	227321	C:93% [S:92.9%,D:0.1%],F:4.5%,M:2.5%,n:4046	10557
<i>Neurospora crassa</i> (strain ATCC 24698 / 74-OR23-1A / CBS 708.71 / DSM 1257 / FGSC 987)	UP000001805	367110	C:99% [S:88.6%,D:10.4%],F:0.8%,M:0.2%,n:3725	9759

<i>Yarrowia lipolytica</i> (strain CLIB 122 / E 150) (Yeast) ( <i>Candida lipolytica</i> )	UP000001300	284591	C:87.9% [S:87.3%,D:0.6%],F:9.8%,M:2.3%,n:1711	6449
<i>Arachis hypogaea</i> (Peanut)	UP000289738	3818	C:93.7% [S:17.2%,D:76.5%],F:1.7%,M:4.7%,n:1440	71122
<i>Musa acuminata subsp. Malaccensis</i> (Wild banana)	UP000012960	214687	C:86.7% [S:76.5%,D:10.3%],F:4.7%,M:8.5%,n:1440	36474
<i>Arabidopsis thaliana</i> (Mouse-ear cress)	UP000006548	3702	C:99.6% [S:59.1%,D:40.5%],F:0.2%,M:0.2%,n:1440	27466
<i>Oryza sativa subsp. indica</i> (Rice)	UP000007015	39946	C:94.8% [S:93.6%,D:1.2%],F:2.4%,M:2.8%,n:1440	37344
<i>Zea mays</i> (Maize)	UP000007305	4577	C:96.4% [S:49.4%,D:46.9%],F:2.1%,M:1.5%,n:1440	39400
<i>Triticum aestivum</i> (Wheat)	UP000019116	4565	C:99.4% [S:1.4%,D:98%],F:0.1%,M:0.5%,n:1440	105061
<i>Physcomitrella patens subsp. patens</i> (Moss)	UP000006727	3218	C:67.8% [S:52%,D:15.8%],F:3.3%,M:28.9%,n:1440	30857
<i>Emiliania huxleyi</i> ( <i>Pontosphaera huxleyi</i> )	UP000013827	2903	C:74.9% [S:0.7%,D:74.3%],F:14.2%,M:10.9%,n:303	35676
<i>Dictyostelium discoideum</i> (Slime mold)	UP000002195	44689	C:96% [S:92.4%,D:3.6%],F:0.7%,M:3.3%,n:303	12739
<i>Chlamydomonas reinhardtii</i> ( <i>Chlamydomonas smithii</i> )	UP000006906	3055	C:96% [S:90.1%,D:5.9%],F:2.3%,M:1.7%,n:303	17614
<i>Thalassiosira pseudonana</i> (Marine diatom)	UP000001449	35128	C:34.2% [S:33.3%,D:0.9%],F:1.3%,M:64.5%,n:234	11717
<i>Plasmodium falciparum</i> (isolate 3D7)	UP000001450	36329	C:25.6% [S:25.2%,D:0.4%],F:0.4%,M:73.9%,n:234	5376