

Supplementary materials

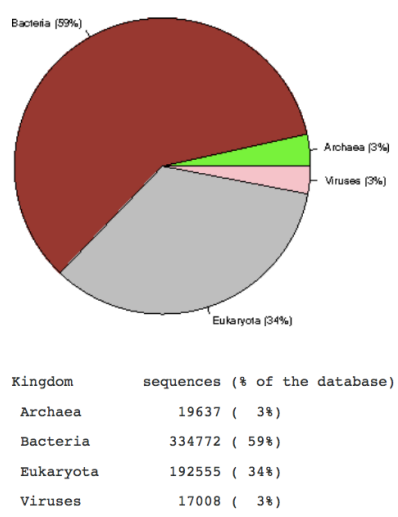


Figure 1 Taxonomic distribution of sequences within UniprotKB/Swiss-Prot release 6/2020

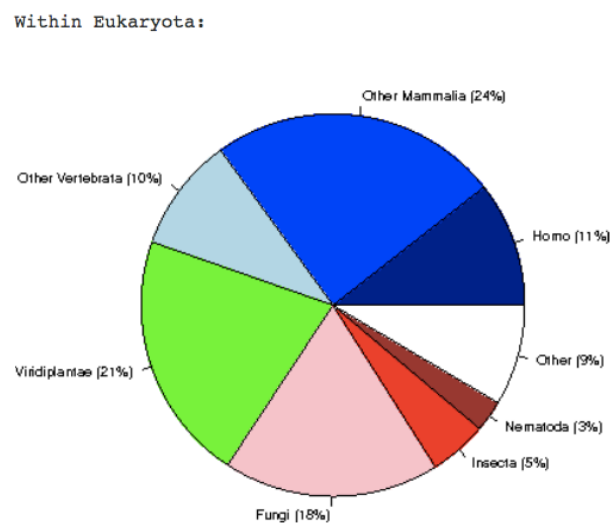


Figure 2 Taxonomic distribution within eukaryota.

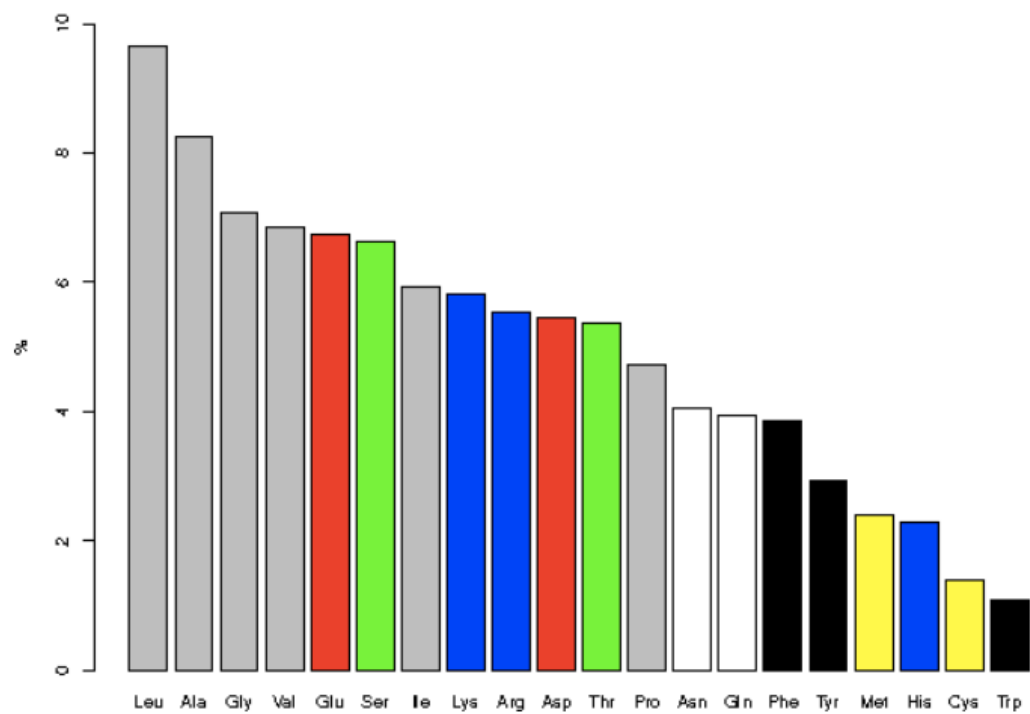


Figure 3 Amino acid composition where grey indicates aliphatic, red - acidic, green - small hydroxy, blue - basic, black - aromatic, white - amide and yellow is sulfur.

Class	Number of folds	Number of superfamilies	Number of families
a: All alpha proteins	289	516	1062
b: All beta proteins	178	370	968
c: Alpha and beta proteins (a/b)	148	246	986
d: Alpha and beta proteins (a+b)	388	565	1338
e: Multi-domain proteins (alpha and beta)	71	71	118
f: Membrane and cell surface proteins and peptides	60	119	173
g: Small proteins	98	139	274

Figure 4 Statistics of SCOPe 16.07.2020 retrieved from v2.07

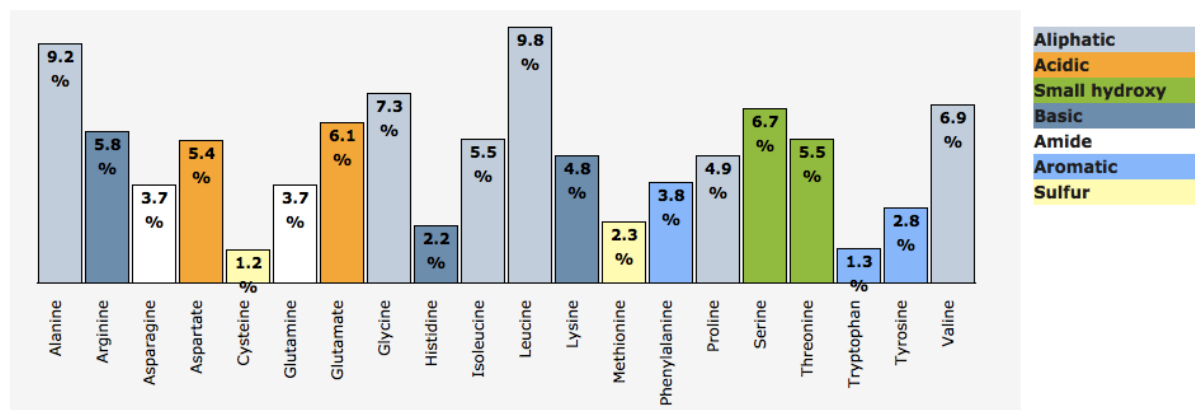


Figure 5 Amino acid composition TrEMBL 06 2020

Line type / subtype	Total number	Number of entries	Average per entry	Rank
Features (FT)	4672113		8.28	
ACT_SITE	167217	101305	0.30	10
BINDING	415853	110438	0.74	2
CA_BIND	4213	1748	0.01	36
CARBOHYD	119513	30639	0.21	15
CHAIN	572202	556666	1.01	1

COILED	22161	15336	0.04	27
COMPBIAS	59229	31869	0.11	21
CONFLICT	137066	47851	0.24	13
CROSSLNK	24088	8679	0.04	26
DISULFID	129095	34529	0.23	14
DNA_BIND	11963	10701	0.02	33
DOMAIN	206270	127088	0.37	9
HELIX	284077	25649	0.50	6
INIT_MET	17415	17367	0.03	28
INTRAMEM	2854	1308	0.01	37
LIPID	13398	8635	0.02	30
METAL	405245	97801	0.72	3
MOD_RES	255843	72851	0.45	7
MOTIF	44954	29384	0.08	23
MUTAGEN	79475	17062	0.14	18
NON_CONS	2514	818	<0.01	38
NON_STD	358	283	<0.01	39
NON_TER	12567	9638	0.02	31
NP_BIND	160248	86613	0.28	11
PEPTIDE	11992	8234	0.02	32
PROPEP	14666	12512	0.03	29
REGION	207707	96704	0.37	8
REPEAT	106965	14914	0.19	16
SIGNAL	43008	43007	0.08	24
SITE	60808	32841	0.11	20
STRAND	293820	24193	0.52	5
TOPO_DOM	145195	29543	0.26	12
TRANSIT	9257	9141	0.02	34
TRANSMEM	375583	78579	0.67	4
TURN	68593	20893	0.12	19
UNSURE	5577	841	0.01	35

VAR_SEQ	52388	22254	0.09	22
VARIANT	98566	17191	0.17	17
ZN_FING	30170	12894	0.05	25

References:

- “SCOPe 2.07: Structural Classification of Proteins — Extended.” n.d. Accessed December 11, 2020.
<https://scop.berkeley.edu/statistics/ver=2.07>.
- “UniProtKB/Swiss-Prot 2020_06.” n.d. Accessed December 11, 2020.
<https://www.uniprot.org/statistics/Swiss-Prot>.
- “UniProtKB/Swiss-Prot Release 2020_06 Statistics.” n.d. Accessed January 01, 2021.
<https://web.expasy.org/docs/relnotes/relstat.html>.