Table 1. Commonly Occurring SCOP Superfamilies in the Proteomes

	Human		Fly		Worm		Yeast		Archaea		Bacteria	
SCOP superfamily	N	R	N	R	N	R	N	R	N	R	N	R
Classic zinc finger, C2H2	5092	1	1096	1	190	10	74	9	–	269	–	–
Immunoglobulin*	1214	2	483	3	457	2	8	91	1	135	4	94
EGF/laminin	1192	3	320	4	413	4	–	–	–	–	–	–
P-loop containing nucleotide triphosphate hydrolases*	847	4	575	2	516	1	408	1	126	1	168	1
Fibronectin type III*	842	5	247	7	222	8	1	301	–	–	1	237
Cadherin	608	6	222	10	135	21	–	–	3	72	–	–
RNA-binding domain	587	7	282	5	199	9	128	3	–	–	–	420
Protein kinase-like (PK-like)*	557	8	271	6	434	3	142	2	3	72	5	82
Hemeodomain-like	334	9	144	18	145	17	32	20	1	221	17	16
Spectrin repeat	327	10	227	9	150	13	–	–	–	–	–	–
PH domain-like*	327	10	140	19	100	31	23	29	–	–	–	–
SH3 domain	304	12	105	23	70	37	29	23	–	–	–	454
EF-hand*	284	13	163	14	120	26	23	29	–	–	–	420
Ankyrin repeat	278	14	120	21	128	24	31	22	–	–	1	342
Complement control module/SCR domain	277	15	57	38	52	43	–	–	–	–	–	–
PDZ domain-like	265	16	103	24	89	32	6	120	1	169	6	64
Ligand-binding domain of low-density lipoprotein receptor	247	17	196	12	143	18	3	194	–	–	–	–
Tetratricopeptide repeat (TPR)*	215	18	171	13	115	27	98	5	4	48	16	19
RING-finger domain, C3HC4	207	19	108	22	122	25	33	19	–	–	–	–
Trp-Asp repeat (WD-repeat)	193	20	198	11	142	19	114	4	2	121	3	157
C2 domain (Calcium/lipid-binding* domain, CaLB)	186	21	68	32	89	32	32	20	–	–	–	–
NAD(P)-binding Rossmann-fold domains*	177	22	150	16	130	23	88	7	27	3	72	2
ARM repeat*	177	22	137	20	105	28	80	8	1	221	–	–
SH2 domain*	161	24	59	37	72	35	8	91	–	–	–	–
Thioredoxin-like*	152	25	148	17	148	14	50	12	8	21	18	13
C-type lectin-like*	149	26	40	53	310	5	–	–	–	–	–	454
Glucocorticoid receptor-like (DNA-binding domain)*	143	27	69	31	281	6	14	59	–	–	–	–
ConA-like lectins/glucanases*	136	28	66	34	105	28	8	91	1	169	3	157
Actin-like ATPase domain*	135	29	65	35	38	56	58	10	2	97	12	26
No. distinct proteins in proteome	28,913		13,922		16,323		6,237		2,176		2,789	
No. distinct superfamilies in proteome	546		518		482		434		328		499	

R, the rank of a superfamily within a proteome.
N, the frequency of domains within this superfamily.
*Denotes that our analysis showed that several PFAM (Bateman et al. 2000) families (and hence several INTERPRO families) are included within the single SCOP superfamily. The number of distinct proteins and the number of domains per superfamily (N) for archaea and bacteria are averages, whereas the number of distinct superfamilies are totals over the species (seven for bacteria and three for archaea).