

Table S1. Ten reference protein sets used in this study.

Genomes ^a	Proteins ^b	Domains ^c
[Prokaryotes]		
<i>Bacillus subtilis</i> strain 168 (UP000001570: 4,197; 4,185)	2,426 (1,601)	1,735
<i>Escherichia coli</i> strain K12 (UP000000625: 4,306; 4,306)	2,950 (1,949)	2,138
<i>Treponema pallidum</i> strain Nichols (UP000000811: 1,028; 620)	350 (205)	469
<i>Streptococcus pyogenes</i> serotype M1 (UP000000750: 1,690; 427)	412 (257)	533
<i>Staphylococcus epidermidis</i> strain ATCC 12228 (UP000001411: 2,461; 758)	704 (444)	844
<i>Staphylococcus aureus</i> strain COL (UP000000530: 2,680; 899)	809 (521)	917
<i>Yersinia pestis</i> Orientalis (UP000000815: 3,909; 1,021)	955 (650)	1,067
[Eukaryotes]		
<i>Drosophila melanogaster</i> (UP000000803: 22,005; 3,246)	2,591 (1,561)	2,195
<i>Mus musculus</i> (UP000000589: 50,189; 16,750)	11,920 (6,484)	5,381
<i>Saccharomyces cerevisiae</i> strain ATCC 204508 (UP000002311: 6,721; 6,721)	4,028 (2,595)	2,889

^aUniprot proteome identifier, the total number of proteins, and the number of reviewed Swiss-Prot proteins are shown in parentheses.

^bTotal number of proteins in the data set used in this study. Only proteins that have one or more identifiable domains (with E-value threshold of 1.0) and are annotated for a UniProtKB protein family are included. The number of single domain proteins is shown in parentheses.

^cTotal number of domains identified in the data set. Domains are identified without allowing overlaps.

Table S2. Clustering performance based on the ten reference protein sets.^a

Genomes	Clusters^b	MOCASSIN-prot^c	TRIBE-MCL^d	SCPS^e
<i>T. pallidum</i>	298	0.5653 (2; 188)	0.2159 (5.6; 48)	0.1667 (1.02; 24)
<i>S. pyogenes</i>	358	0.5652 (2; 231)	0.2782 (6; 72)	0.1987 (1.03; 40)
<i>S. epidermidis</i>	604	0.5928 (2; 414)	0.2374 (6; 111)	0.2727 (1.03; 99)
<i>S. aureus</i>	662	0.6017 (2; 496)	0.2371 (6; 108)	0.2996 (1.03; 110)
<i>Y. pestis</i>	829	0.6446 (2; 604)	0.1957 (6; 115)	0.2275 (1.02; 113)
<i>B. subtilis</i>	1,322	0.5602 (3; 1618)	0.4416 (5; 462)	0.5279 (1.02; 569)
<i>E. coli</i>	1,766	0.5680 (2; 1921)	0.4477 (6; 884)	0.4773 (1.02; 690)
<i>S. cerevisiae</i>	2,243	0.5248 (2; 2523)	0.4211 (5.8; 891)	0.4543 (1.02; 1002)
<i>D. melanogaster</i>	1,528	0.4638 (2; 1501)	0.3733 (5.8; 371)	0.3975 (1.03; 446)
<i>M. musculus</i>	3,702	0.3503 (2; 6936)	0.3792 (4.5; 4170)	0.3793 (1.02; 2869)

^aPerformance is measured using the overall F-measure.

^bTotal number of reference clusters.

^cOptimal number of iterations followed by number of clusters generated are shown in parentheses.

^dOptimal inflation parameter followed by number of clusters generated are shown in parentheses.

^eOptimal first eigenratio threshold followed by number of clusters are generated shown in parentheses.

Table S3. Clustering performance based on the GO annotations for the *M. musculus* protein set.^a

MOCASSIN-prot		TRIBE-MCL		SCPS	
Iteration	Prop (cluster #/size) ^b	Inflation	Prop (cluster #/size) ^b	Epsilon	Prop (cluster #/size) ^b
1	0.994, 0.998 (686/9,769)	1.1	0.993, 0.999 (414/10,767)	1.02	0.991, 0.985 (669/6,680)
2	0.988, 0.993 (1,216/9,285)	1.2	0.984, 0.993 (876/10,515)	1.03	0.991, 0.985 (669/6,680)
3	0.990, 0.995 (1,255/8,988)	1.3	0.989, 0.995 (1,045/10,398)	1.04	0.991, 0.985 (669/6,680)
4	0.989, 0.994 (1,273/9,215)	1.4	0.988, 0.994 (1,118/10,314)	1.05	0.991, 0.985 (669/6,680)
		1.5	0.986, 0.993 (1,184/10,233)		
		1.6	0.985, 0.993 (1,233/10,170)		
		1.7	0.983, 0.990 (1,267/10,107)		
		1.8	0.983, 0.989 (1,297/10,049)		
		1.9	0.985, 0.992 (1,319/10,010)		
		2.0	0.984, 0.989 (1,342/9,951)		
		2.1	0.983, 0.988 (1,360/9,907)		
		2.2	0.983, 0.987 (1,380/9,870)		
		2.3	0.982, 0.987 (1,394/9,813)		
		2.4	0.982, 0.989 (1,402/9,754)		
		2.5	0.982, 0.989 (1,412/9,734)		
		2.6	0.982, 0.982 (1,415/9,699)		
		2.7	0.982, 0.983 (1,427/9,663)		
		2.8	0.980, 0.982 (1,431/9,641)		
		2.9	0.981, 0.983 (1,431/9,613)		
		3.0	0.981, 0.982 (1,440/9,592)		
		3.1	0.981, 0.982 (1,449/9,571)		
		3.2	0.980, 0.982 (1,454/9,555)		
		3.3	0.980, 0.982 (1,458/9,545)		
		3.4	0.980, 0.981 (1,464/9,531)		
		3.5	0.980, 0.987 (1,463/9,511)		
		3.6	0.980, 0.987 (1,467/9,487)		
		3.7	0.980, 0.987 (1,475/9,482)		
		3.8	0.979, 0.986 (1,481/9,470)		
		3.9	0.978, 0.984 (1,487/9,465)		
		4.0	0.978, 0.984 (1,487/9,453)		
		4.1	0.979, 0.984 (1,491/9,440)		
		4.2	0.979, 0.984 (1,492/9,420)		
		4.3	0.979, 0.986 (1,494/9,397)		
		4.4	0.979, 0.986 (1,498/9,380)		
		4.5	0.979, 0.986 (1,498/9,367)		
		4.6	0.979, 0.986 (1,501/9,353)		
		4.7	0.979, 0.986 (1,506/9,343)		
		4.8	0.979, 0.986 (1,508/9,335)		
		4.9	0.978, 0.983 (1,510/9,322)		
		5.0	0.978, 0.983 (1,514/9,315)		
		5.1	0.978, 0.983 (1,512/9,295)		
		5.2	0.977, 0.979 (1,514/9,286)		
		5.3	0.977, 0.979 (1,516/9,280)		
		5.4	0.977, 0.980 (1,517/9,272)		
		5.5	0.977, 0.979 (1,520/9,265)		
		5.6	0.977, 0.980 (1,523/9,269)		
		5.7	0.978, 0.983 (1,524/9,262)		
		5.8	0.977, 0.980 (1,524/9,256)		
		5.9	0.977, 0.980 (1,522/9,237)		
		6.0	0.978, 0.983 (1,524/9,236)		

^aClusters including three or more proteins are included in the analysis. Significant clusters are identified as those with overrepresenting GO terms at the significant level of 0.0001.

^bProportions of significant clusters were calculated based on the cluster number and the cluster size and shown in this order. The number and the cumulative size of significant clusters are shown in parentheses. The highest proportion obtained for each method is highlighted with orange and those where the largest number of significant clusters were obtained are highlighted with blue.

Table S4. Clustering performance based on the GO annotations for the *M. musculus* protein set.^a

MOCASSIN-prot		TRIBE-MCL		SCPS	
Iteration	Prop (cluster #/size) ^b	Inflation	Prop (cluster #/size) ^b	Epsilon	Prop (cluster #/size) ^b
1	1.0, 1.0 (686, 9,769)	1.1	1.0, 1.0 (414/10,767)	1.02	1.0, 1.0 (669/6,680)
2	0.999, 1.0 (1,216, 9,285)	1.2	0.999, 1.0 (876/10,515)	1.03	1.0, 1.0 (669/6,680)
3	0.999, 1.0 (1,255, 8,988)	1.3	0.999, 1.0 (1,045/10,398)	1.04	1.0, 1.0 (669/6,680)
4	0.999, 1.0 (1,273, 9,215)	1.4	0.999, 1.0 (1,118/10,314)	1.05	1.0, 1.0 (669/6,680)
		1.5	0.998, 0.999 (1,184/10,233)		
		1.6	0.999, 1.0 (1,233/10,170)		
		1.7	0.999, 1.0 (1,267/10,107)		
		1.8	0.999, 1.0 (1,297/10,049)		
		1.9	0.999, 1.0 (1,319/10,010)		
		2.0	0.999, 1.0 (1,342/9,951)		
		2.1	0.999, 1.0 (1,360/9,907)		
		2.2	0.999, 1.0 (1,380/9,870)		
		2.3	0.999, 1.0 (1,394/9,813)		
		2.4	0.999, 1.0 (1,402/9,754)		
		2.5	0.999, 1.0 (1,412/9,734)		
		2.6	0.999, 1.0 (1,415/9,699)		
		2.7	0.999, 1.0 (1,427/9,663)		
		2.8	0.999, 1.0 (1,431/9,641)		
		2.9	0.999, 1.0 (1,431/9,613)		
		3.0	0.999, 1.0 (1,440/9,592)		
		3.1	0.999, 1.0 (1,449/9,571)		
		3.2	0.999, 1.0 (1,454/9,555)		
		3.3	0.999, 1.0 (1,458/9,545)		
		3.4	0.999, 1.0 (1,464/9,531)		
		3.5	0.999, 1.0 (1,463/9,511)		
		3.6	0.999, 1.0 (1,467/9,487)		
		3.7	0.999, 1.0 (1,475/9,482)		
		3.8	0.999, 1.0 (1,481/9,470)		
		3.9	0.999, 1.0 (1,487/9,465)		
		4.0	0.999, 1.0 (1,487/9,453)		
		4.1	0.999, 1.0 (1,491/9,440)		
		4.2	0.999, 1.0 (1,492/9,420)		
		4.3	0.999, 1.0 (1,494/9,397)		
		4.4	0.999, 1.0 (1,498/9,380)		
		4.5	0.999, 1.0 (1,498/9,367)		
		4.6	0.999, 1.0 (1,501/9,353)		
		4.7	0.999, 1.0 (1,506/9,343)		
		4.8	0.999, 1.0 (1,508/9,335)		
		4.9	0.999, 1.0 (1,510/9,322)		
		5.0	0.999, 1.0 (1,514/9,315)		
		5.1	0.999, 1.0 (1,512/9,295)		
		5.2	0.999, 1.0 (1,514/9,286)		
		5.3	0.999, 1.0 (1,516/9,280)		
		5.4	0.999, 1.0 (1,517/9,272)		
		5.5	0.999, 1.0 (1,520/9,265)		
		5.6	0.999, 1.0 (1,523/9,269)		
		5.7	0.999, 1.0 (1,524/9,262)		
		5.8	0.999, 1.0 (1,524/9,256)		
		5.9	0.999, 1.0 (1,522/9,237)		
		6.0	0.999, 1.0 (1,524/9,236)		

^aClusters including three or more proteins are included in the analysis. Significant clusters are identified as those with overrepresenting GO terms at the significant level of 0.01.

^bProportions of significant clusters were calculated based on the cluster number and the cluster size and shown in this order. The number and the cumulative size of significant clusters are shown in parentheses. The highest proportion obtained for each method is highlighted with orange and those where the largest number of significant clusters were obtained are highlighted with blue.

Table S5. Clustering performance based on the GO annotations for the *M. musculus* protein set.^a

MOCASSIN-prot		TRIBE-MCL		SCPS	
Iteration	Prop (cluster #/size) ^b	Inflation	Prop (cluster #/size) ^b	Epsilon	Prop (cluster #/size) ^b
1	0.981, 0.995 (1,249/10,895)	1.1	0.983, 0.998 (475, 10,889)	1.02	0.980, 0.983 (1,332/8,006)
2	0.980, 0.989 (2,021/10,895)	1.2	0.979, 0.991 (1,060, 10,883)	1.03	0.980, 0.983 (1,332/8,006)
3	0.981, 0.990 (2,099/10,676)	1.3	0.983, 0.994 (1,285, 10,878)	1.04	0.980, 0.983 (1,332/8,006)
4	0.980, 0.990 (2,110/10,889)	1.4	0.982, 0.992 (1,400, 10,878)	1.05	0.980, 0.983 (1,332/8,006)
		1.5	0.981, 0.992 (1,504, 10,873)		
		1.6	0.981, 0.991 (1,577, 10,858)		
		1.7	0.979, 0.988 (1,636, 10,845)		
		1.8	0.979, 0.988 (1,686, 10,827)		
		1.9	0.980, 0.990 (1,719, 10,810)		
		2.0	0.980, 0.987 (1,761, 10,789)		
		2.1	0.978, 0.986 (1,793, 10,773)		
		2.2	0.978, 0.985 (1,824, 10,758)		
		2.3	0.978, 0.985 (1,856, 10,737)		
		2.4	0.977, 0.987 (1,885, 10,720)		
		2.5	0.977, 0.986 (1,901, 10,712)		
		2.6	0.977, 0.981 (1,916, 10,701)		
		2.7	0.977, 0.981 (1,940, 10,689)		
		2.8	0.975, 0.980 (1,954, 10,687)		
		2.9	0.976, 0.981 (1,960, 10,671)		
		3.0	0.976, 0.980 (1,977, 10,666)		
		3.1	0.977, 0.981 (1,992, 10,657)		
		3.2	0.977, 0.981 (2,003, 10,653)		
		3.3	0.977, 0.981 (2,008, 10,645)		
		3.4	0.976, 0.980 (2,018, 10,639)		
		3.5	0.976, 0.985 (2,021, 10,627)		
		3.6	0.976, 0.985 (2,032, 10,617)		
		3.7	0.976, 0.985 (2,040, 10,612)		
		3.8	0.976, 0.984 (2,051, 10,610)		
		3.9	0.975, 0.982 (2,057, 10,605)		
		4.0	0.975, 0.982 (2,062, 10,603)		
		4.1	0.975, 0.982 (2,071, 10,600)		
		4.2	0.975, 0.982 (2,081, 10,598)		
		4.3	0.976, 0.984 (2,091, 10,591)		
		4.4	0.976, 0.984 (2,103, 10,590)		
		4.5	0.976, 0.984 (2,109, 10,589)		
		4.6	0.975, 0.984 (2,116, 10,583)		
		4.7	0.976, 0.984 (2,123, 10,577)		
		4.8	0.975, 0.983 (2,128, 10,575)		
		4.9	0.974, 0.981 (2,135, 10,572)		
		5.0	0.974, 0.981 (2,138, 10,563)		
		5.1	0.974, 0.981 (2,144, 10,559)		
		5.2	0.974, 0.978 (2,148, 10,554)		
		5.3	0.974, 0.978 (2,151, 10,550)		
		5.4	0.974, 0.978 (2,153, 10,544)		
		5.5	0.974, 0.978 (2,160, 10,545)		
		5.6	0.975, 0.978 (2,159, 10,541)		
		5.7	0.975, 0.981 (2,163, 10,540)		
		5.8	0.975, 0.978 (2,164, 10,536)		
		5.9	0.975, 0.979 (2,170, 10,533)		
		6.0	0.975, 0.981 (2,171, 10,530)		

^aClusters including two or more proteins are included in the analysis. Significant clusters are identified as those with overrepresenting GO terms at the significant level of 0.0001.

^bProportions of significant clusters were calculated based on the cluster number and the cluster size and shown in this order. The number and the cumulative size of significant clusters are shown in parentheses. The highest proportion obtained for each method is highlighted with orange and those where the largest number of significant clusters were obtained are highlighted with blue.

Table S6. Clustering performance based on the GO annotations for the *M. musculus* protein set.^a

MOCASSIN-prot		TRIBE-MCL		SCPS	
Iteration	Prop (cluster #/size) ^b	Inflation	Prop (cluster #/size) ^b	Epsilon	Prop (cluster #/size) ^b
1	0.994, 0.999 (1,249/10,895)	1.1	0.994, 0.999 (475/10,889)	1.02	0.992, 0.998 (1,332/8,006)
2	0.994, 0.998 (2,021/10,895)	1.2	0.994, 0.999 (1,060/10,883)	1.03	0.992, 0.998 (1,332/8,006)
3	0.994, 0.997 (2,099/10,676)	1.3	0.995, 0.999 (1,285/10,878)	1.04	0.992, 0.998 (1,332/8,006)
4	0.993, 0.997 (2,110/10,889)	1.4	0.996, 0.999 (1,400/10,878)	1.05	0.992, 0.998 (1,332/8,006)
		1.5	0.995, 0.998 (1,504/10,873)		
		1.6	0.996, 0.999 (1,577/10,858)		
		1.7	0.996, 0.999 (1,636/10,845)		
		1.8	0.996, 0.999 (1,686/10,827)		
		1.9	0.997, 0.999 (1,719/10,810)		
		2.0	0.997, 0.999 (1,761/10,789)		
		2.1	0.997, 0.999 (1,793/10,773)		
		2.2	0.997, 0.999 (1,824/10,758)		
		2.3	0.997, 0.999 (1,856/10,737)		
		2.4	0.996, 0.999 (1,885/10,720)		
		2.5	0.996, 0.999 (1,901/10,712)		
		2.6	0.996, 0.999 (1,916/10,701)		
		2.7	0.996, 0.999 (1,940/10,689)		
		2.8	0.996, 0.998 (1,954/10,687)		
		2.9	0.996, 0.998 (1,960/10,671)		
		3.0	0.996, 0.998 (1,977/10,666)		
		3.1	0.996, 0.999 (1,992/10,657)		
		3.2	0.997, 0.999 (2,003/10,653)		
		3.3	0.997, 0.999 (2,008/10,645)		
		3.4	0.997, 0.999 (2,018/10,639)		
		3.5	0.997, 0.999 (2,021/10,627)		
		3.6	0.997, 0.999 (2,032/10,617)		
		3.7	0.997, 0.999 (2,040/10,612)		
		3.8	0.997, 0.999 (2,051/10,610)		
		3.9	0.997, 0.999 (2,057/10,605)		
		4.0	0.997, 0.999 (2,062/10,603)		
		4.1	0.997, 0.999 (2,071/10,600)		
		4.2	0.997, 0.999 (2,081/10,598)		
		4.3	0.997, 0.999 (2,091/10,591)		
		4.4	0.997, 0.999 (2,103/10,590)		
		4.5	0.997, 0.999 (2,109/10,589)		
		4.6	0.997, 0.999 (2,116/10,583)		
		4.7	0.996, 0.998 (2,123/10,577)		
		4.8	0.996, 0.998 (2,128/10,575)		
		4.9	0.996, 0.998 (2,135/10,572)		
		5.0	0.996, 0.998 (2,138/10,563)		
		5.1	0.996, 0.998 (2,144/10,559)		
		5.2	0.996, 0.998 (2,148/10,554)		
		5.3	0.996, 0.998 (2,151/10,550)		
		5.4	0.996, 0.998 (2,153/10,544)		
		5.5	0.996, 0.998 (2,160/10,545)		
		5.6	0.996, 0.998 (2,159/10,541)		
		5.7	0.996, 0.998 (2,163/10,540)		
		5.8	0.996, 0.998 (2,164/10,536)		
		5.9	0.996, 0.998 (2,170/10,533)		
		6.0	0.996, 0.998 (2,171/10,530)		

^aClusters including two or more proteins are included in the analysis. Significant clusters are identified as those with overrepresenting GO terms at the significant level of 0.01.

^bProportions of significant clusters were calculated based on the cluster number and the cluster size and shown in this order. The number and the cumulative size of significant clusters are shown in parentheses. The highest proportion obtained for each method is highlighted with orange and those where the largest number of significant clusters were obtained are highlighted with blue.