| # | Keywords | Number of proteins | Number of families | Average sequence length | Z-score | P-value |
|----|---------------------------|--------------------|--------------------|-------------------------|---------|---------|
| 0 | Oxidoreductase | 14995 | 992 | 376.63 | -29.54 | 0 |
| 1 | Transferase | 26525 | 1606 | 445.17 | -24.25 | 0 |
| 2 | Lyase | 7262 | 347 | 377.92 | -22.64 | 0 |
| 3 | Hydrolase | 20464 | 1995 | 430.68 | -21.75 | 0 |
| 4 | Isomerase | 4487 | 220 | 383.98 | -14.18 | 0 |
| 5 | Glycosidase | 1826 | 244 | 444.73 | -13.98 | 0 |
| 6 | Glycosyltransferase | 2950 | 261 | 437.53 | -12.51 | 0 |
| 7 | Acyltransferase | 2239 | 179 | 402.83 | -10.85 | 0 |
| 8 | Methyltransferase | 3524 | 224 | 349.60 | -10.53 | 0 |
| 9 | Kinase | 7017 | 322 | 448.29 | -10.22 | 0 |
| 10 | Ligase | 8010 | 230 | 529.41 | -10.06 | 0 |
| 11 | Decarboxylase | 1293 | 63 | 345.26 | -9.66 | 0 |
| 12 | Monooxygenase | 1668 | 73 | 444.87 | -9.26 | 0 |
| 13 | Metalloprotease | 1100 | 109 | 553.73 | -7.89 | 0 |
| 14 | Aminopeptidase | 452 | 39 | 509.17 | -7.55 | 0 |
| 15 | Dioxygenase | 360 | 66 | 433.20 | -7.32 | 0 |
| 16 | Aminoacyl-tRNA synthetase | 3402 | 37 | 571.83 | -7.15 | 0 |
| 17 | Protease | 4423 | 380 | 549.70 | -7.1 | 0 |
| 18 | Aminotransferase | 955 | 28 | 420.27 | -6.02 | 0 |

| # | name | n | avg_len | avg_dis | z | p |
|----|----------------------------|------|---------|---------|--------|---|
| 0 | Oxidoreductase | 4126 | 472.25 | 0.28 | -41.35 | 0 |
| 1 | Hydrolase | 7564 | 614.81 | 0.51 | -26.99 | 0 |
| 2 | Lyase | 1431 | 481.37 | 0.30 | -23.62 | 0 |
| 3 | Monooxygenase | 555 | 503.36 | 0.20 | -20.29 | 0 |
| 4 | Transferase | 8846 | 631.95 | 0.55 | -19.72 | 0 |
| 5 | Ligase | 995 | 693.30 | 0.46 | -18.05 | 0 |
| 6 | Glycosyltransferase | 1134 | 551.26 | 0.40 | -17.04 | 0 |
| 7 | Glycosidase | 697 | 570.50 | 0.37 | -16.81 | 0 |
| 8 | Isomerase | 931 | 422.72 | 0.35 | -13.60 | 0 |
| 9 | Protease | 1863 | 674.42 | 0.54 | -13.20 | 0 |
| 10 | Transducer | 1703 | 482.28 | 0.41 | -12.56 | 0 |
| 11 | G-protein coupled receptor | 1385 | 465.62 | 0.39 | -12.45 | 0 |
| 12 | Acyltransferase | 867 | 531.58 | 0.42 | -11.28 | 0 |
| 13 | Decarboxylase | 195 | 488.21 | 0.25 | -10.70 | 0 |
| 14 | Aminotransferase | 202 | 451.05 | 0.24 | -10.23 | 0 |
| 15 | Aminopeptidase | 130 | 668.72 | 0.37 | -9.10 | 0 |
| 16 | Serine protease | 460 | 700.07 | 0.50 | -8.87 | 0 |
| 17 | Metalloprotease | 507 | 688.25 | 0.56 | -8.43 | 0 |
| 18 | Methyltransferase | 874 | 611.22 | 0.47 | -8.33 | 0 |
| 19 | Carboxypeptidase | 116 | 631.16 | 0.37 | -8.25 | 0 |
| 20 | Threonine protease | 138 | 246.88 | 0.18 | -7.50 | 0 |
| 21 | Dioxygenase | 366 | 622.32 | 0.48 | -7.39 | 0 |
| 22 | Serine esterase | 141 | 423.09 | 0.28 | -7.21 | 0 |
| 23 | Receptor | 3424 | 647.92 | 0.59 | -7.09 | 0 |
| 24 | Nucleotidyltransferase | 600 | 969.68 | 0.63 | -6.18 | 0 |
| 25 | Peroxidase | 221 | 457.61 | 0.40 | -5.50 | 0 |
| 26 | Serine protease inhibitor | 182 | 497.66 | 0.38 | -4.92 | 0 |
| 27 | Porin | 63 | 318.84 | 0.21 | -4.55 | 0 |
| 28 | Dipeptidase | 22 | 522.27 | 0.23 | -4.43 | 0 |
| 29 | Integrin | 71 | 1038.25 | 0.70 | -4.29 | 0 |

| | | ı | | | | |
|----------------|---|--------------------|----------------------------|----------------------|-------------------------|----------------|
| # | name | n | avg_len | avg_dis | \mathcal{Z} | p |
| 0 | catalytic activity | 28179 | 569.86 | 0.48 | -52.25 | 0 |
| 1 | oxidoreductase activity | 4986 | 450.49 | 0.30 | -39.62 | 0 |
| 2 | hydrolase activity | 11097 | 623.94 | 0.53 | -27.76 | 0 |
| | • | | | | | |
| 3 | lyase activity | 1624 | 480.74 | 0.31 | -24.47 | 0 |
| 4 | monooxygenase activity | 633 | 511.13 | 0.22 | -21.40 | 0 |
| 5 | nucleotide binding | 10750 | 705.30 | 0.61 | -19.31 | 0 |
| 6 | transporter activity | 5388 | 621.16 | 0.53 | -19.15 | 0 |
| | | | | | | |
| 7 | transferase activity | 10428 | 619.96 | 0.56 | -18.80 | 0 |
| 8 | transferase activity, trans | 1343 | 545.86 | 0.41 | -18.49 | 0 |
| 9 | ligase activity | 1097 | 675.42 | 0.46 | -16.68 | 0 |
| 10 | hydrolase activity, acting | 934 | 548.13 | 0.39 | -16.32 | 0 |
| 11 | isomerase activity | 1046 | 425.93 | 0.35 | -14.19 | 0 |
| | | | | | | |
| 12 | transferase activity, trans | 1235 | 540.97 | 0.43 | -13.31 | 0 |
| 13 | carboxy-lyase activity | 292 | 494.47 | 0.26 | -13.09 | 0 |
| 14 | peptidase activity | 2211 | 644.11 | 0.53 | -13.05 | 0 |
| 15 | G-protein coupled receptor | 1473 | 465.64 | 0.40 | -12.30 | 0 |
| - | | | | | | |
| 16 | carboxylic ester hydrolase | 628 | 482.21 | 0.37 | -12.01 | 0 |
| 17 | serine-type peptidase activity | 708 | 614.43 | 0.46 | -10.79 | 0 |
| 18 | iron-sulfur cluster binding | 706 | 516.93 | 0.42 | -10.77 | 0 |
| 19 | aminopeptidase activity | 177 | 633.80 | 0.35 | -10.74 | 0 |
| 20 | 4 iron, 4 sulfur cluster bi | 444 | 571.30 | 0.42 | -10.51 | 0 |
| | · | | | | | |
| 21 | transaminase activity | 230 | 459.43 | 0.26 | -10.46 | 0 |
| 22 | methyltransferase activity | 947 | 607.08 | 0.48 | -8.77 | 0 |
| 23 | GTP binding | 1467 | 458.88 | 0.48 | -8.55 | 0 |
| 24 | carboxypeptidase activity | 143 | 596.81 | 0.34 | -8.54 | 0 |
| | 7 | | | | | |
| 25 | metallopeptidase activity | 632 | 663.90 | 0.56 | -8.44 | 0 |
| 26 | quinone binding | 141 | 375.78 | 0.21 | -8.40 | 0 |
| 27 | ATP binding | 7585 | 801.90 | 0.69 | -8.22 | 0 |
| 28 | dioxygenase activity | 419 | 604.98 | 0.48 | -7.83 | 0 |
| 29 | carbohydrate binding | 978 | 571.03 | 0.52 | -7.43 | 0 |
| - | | | | | | |
| 30 | symporter activity | 562 | 602.99 | 0.56 | -7.14 | 0 |
| 31 | peroxidase activity | 282 | 414.57 | 0.34 | -7.11 | 0 |
| 32 | antiporter activity | 374 | 599.75 | 0.52 | -6.92 | 0 |
| 33 | antioxidant activity | 499 | 344.42 | 0.36 | -6.89 | 0 |
| 34 | • | | | | | |
| | metal ion binding | 14382 | 631.50 | 0.62 | -6.80 | 0 |
| 35 | threonine-type endopeptidas | 139 | 247.87 | 0.19 | -6.79 | 0 |
| 36 | nucleotidyltransferase acti | 892 | 805.20 | 0.59 | -6.07 | 0 |
| 37 | oxygen carrier activity | 55 | 220.55 | 0.09 | -5.47 | 0 |
| 38 | porin activity | 76 | 369.24 | 0.22 | -5.22 | 0 |
| _ | 1 , | | | | | |
| 39 | serine-type endopeptidase i | 218 | 485.65 | 0.40 | -4.99 | 0 |
| 40 | phospholipase A2 activity | 93 | 377.39 | 0.30 | -4.93 | 0 |
| 41 | phospholipase D activity | 37 | 778.27 | 0.43 | -4.87 | 0 |
| 42 | prenyltransferase activity | 128 | 369.37 | 0.32 | -4.70 | 0 |
| 43 | | | | | | 0 |
| $\overline{}$ | toxin activity | 285 | 349.98 | 0.32 | -4.59 | |
| 44 | 2 iron, 2 sulfur cluster bi | 229 | 383.88 | 0.38 | -4.38 | 0 |
| 45 | dipeptidase activity | 41 | 557.73 | 0.34 | -4.23 | 0 |
| 46 | metalloendopeptidase activity | 369 | 711.73 | 0.63 | -4.22 | 0 |
| 47 | phospholipase A2 activity (| 70 | 365.39 | 0.30 | -3.90 | 0 |
| - | | | | | | |
| 48 | chlorophyll binding | 52 | 408.21 | 0.31 | -3.88 | 0 |
| 49 | endonuclease activity | 773 | 658.33 | 0.59 | -3.85 | 0 |
| 50 | folic acid binding | 32 | 423.94 | 0.25 | -3.78 | 0 |
| 51 | antigen binding | 277 | 345.87 | 0.38 | -3.68 | 0 |
| 52 | | 116 | 464.12 | 0.46 | -3.35 | 0 |
| | translation elongation fact | | | | | |
| 53 | 3 iron, 4 sulfur cluster bi | 37 | 740.78 | 0.43 | -3.32 | 0 |
| 54 | nuclease activity | 1124 | 659.17 | 0.61 | -3.20 | 0 |
| 55 | G-protein coupled photorece | 44 | 390.23 | 0.30 | -2.99 | 0 |
| 56 | chloride channel activity | 195 | 622.49 | 0.62 | -2.67 | 0.01 |
| | · · | | | | | |
| 57 | RNA-directed 5'-3' RNA poly | 75 | 2341.31 | 0.87 | -2.57 | 0.02 |
| 58 | cobalamin binding | 45 | 877.44 | 0.53 | -2.56 | 0.01 |
| 59 | L-ascorbic acid binding | 81 | 479.84 | 0.48 | -2.40 | 0.02 |
| 60 | tRNA binding | 247 | 566.35 | 0.57 | -2.39 | 0.01 |
| 61 | peptidase inhibitor activity | 446 | 493.46 | 0.50 | -2.38 | 0.01 |
| - | • | | | | | |
| 62 | photoreceptor activity | 91 | 574.22 | 0.54 | -2.30 | 0.01 |
| | glucan endo-1,3-beta-D-gluc | 28 | 522.18 | 0.43 | -2.29 | 0.02 |
| 63 | | Γ | | 0.62 | 0.17 | 0.00 |
| _ | signal transducer activity | 4149 | 619.06 | 0.62 | -2.17 | 0.02 |
| 63 64 65 | | | | | | |
| | signal transducer activity aminoacyl-tRNA ligase activity cysteine-type endopeptidase | 4149 255 262 | 619.06 659.40 953.41 | 0.62 0.67 0.64 | -2.17 -1.93 -1.65 | 0.02 0.04 0.05 |