

| <b>Domain</b>  | <b>Pfam<br/>AC #</b> | <b>Present in<br/>number of<br/>domain<br/>architectures<br/>that have<br/>multiple<br/>origins</b> | <b>Present in<br/>number of<br/>domain<br/>architectures<br/>overall</b> |
|--|----------------------|---|--|
| Histidine kinase-, DNA gyraseB-, and HSP90-like ATPase   | PF02518              | 30  | 273  |
| SH3 domain   | PF00018              | 22  | 160  |
| PHD-finger   | PF00628              | 21  | 131  |
| His Kinase A (phosphoacceptor) domain                    | PF00512              | 21  | 168  |
| PH domain  | PF00169              | 21  | 186  |
| PDZ domain (Also known as DHR or GLGF)                   | PF00595              | 19  | 147  |
| Zinc finger, C3HC4 type (RING finger)                    | PF00097              | 17  | 142  |
| Response regulator receiver domain                       | PF00072              | 17  | 162  |
| C1 domain  | PF03107              | 16  | 42   |
| C1-like domain   | PF07649              | 16  | 47   |
| Helicase conserved C-terminal domain                     | PF00271              | 16  | 176  |
| WW domain  | PF00397              | 15  | 74   |
| EF hand  | PF00036              | 15  | 155  |
| Variant SH3 domain                                       | PF07653              | 13  | 82   |
| GAF domain   | PF01590              | 13  | 118  |
| C2 domain  | PF00168              | 13  | 123  |
| PAS fold   | PF08448              | 13  | 141  |
| Phorbol esters/diacylglycerol binding domain (C1 domain) | PF00130              | 12  | 92   |
| PAS fold   | PF08447              | 12  | 137  |
| Bromodomain  | PF00439              | 11  | 50   |
| Protein tyrosine kinase                                  | PF07714              | 11  | 155  |
| Protein kinase domain                                    | PF00069              | 11  | 233  |
| HAMP domain  | PF00672              | 10  | 59   |
| GGDEF domain   | PF00990              | 10  | 64   |
| 4Fe-4S binding domain                                    | PF00037              | 10  | 93   |
| Thrombospondin type 1 domain                             | PF00090              | 10  | 120  |
| RhoGEF domain  | PF00621              | 9   | 72   |
| EGF-like domain  | PF07974              | 8   | 123  |
| Immunoglobulin I-set domain                              | PF07679              | 8   | 167  |

|   |         |   |     |
|---|---------|---|-----|
| Fibronectin type III domain                                     | PF00041 | 8 | 203 |
| EGF-like domain   | PF00008 | 8 | 255 |
| G-patch domain  | PF01585 | 7 | 21  |
| JmjC domain   | PF02373 | 7 | 30  |
| ADAM-TS Spacer 1  | PF05986 | 7 | 43  |
| DEAD/DEAH box helicase  | PF00270 | 7 | 77  |
| Zinc knuckle  | PF00098 | 7 | 79  |
| PAS fold  | PF00989 | 7 | 165 |
| Immunoglobulin domain   | PF00047 | 7 | 194 |
| Fibronectin type II domain                                      | PF00040 | 6 | 23  |
| PWWP domain   | PF00855 | 6 | 26  |
| EAL domain  | PF00563 | 6 | 40  |
| Ion transport protein   | PF00520 | 6 | 44  |
| Myb-like DNA-binding domain                                     | PF00249 | 6 | 49  |
| ATPase family associated with various cellular activities (AAA) | PF00004 | 6 | 51  |
| Kunitz/Bovine pancreatic trypsin inhibitor domain               | PF00014 | 6 | 59  |
| Lectin C-type domain  | PF00059 | 6 | 65  |
| SH2 domain  | PF00017 | 6 | 67  |
| Phosphopantetheine attachment site                              | PF00550 | 6 | 73  |
| Zinc finger, C2H2 type  | PF00096 | 6 | 162 |
| dDENN domain  | PF03455 | 5 | 14  |
| jmjN domain   | PF02375 | 5 | 18  |
| Bacterial regulatory helix-turn-helix proteins, AraC family     | PF00165 | 5 | 26  |
| PLAC (protease and lacunin) domain                              | PF08686 | 5 | 27  |
| Elongation factor Tu GTP binding domain                         | PF00009 | 5 | 28  |
| Phosphatidylinositol-specific phospholipase C, X domain         | PF00388 | 5 | 29  |
| LIM domain  | PF00412 | 5 | 30  |
| B-box zinc finger   | PF00643 | 5 | 31  |
| SCAN domain   | PF02023 | 5 | 32  |
| PX domain   | PF00787 | 5 | 38  |
| Histidine kinase  | PF07568 | 5 | 43  |
| BTB/POZ domain  | PF00651 | 5 | 50  |
| Ras association (RalGDS/AF-6) domain                            | PF00788 | 5 | 55  |
| RhoGAP domain   | PF00620 | 5 | 56  |

|  |         |   |     |
|--|---------|---|-----|
| Integrase core domain                                    | PF00665 | 5 | 62  |
| Calponin homology (CH) domain                            | PF00307 | 5 | 64  |
| SNF2 family N-terminal domain                            | PF00176 | 5 | 75  |
| F-box domain   | PF00646 | 5 | 78  |
| von Willebrand factor type A domain                      | PF00092 | 5 | 79  |
| Immunoglobulin V-set domain                              | PF07686 | 5 | 93  |
| DIX domain   | PF00778 | 4 | 7   |
| C5HC2 zinc finger  | PF02928 | 4 | 14  |
| CNH domain   | PF00780 | 4 | 16  |
| Phosphoinositide-specific phospholipase C, ehand-like    | PF09279 | 4 | 18  |
| Guanylate kinase   | PF00625 | 4 | 21  |
| Receptor family ligand binding region                    | PF01094 | 4 | 22  |
| Sigma-54 interaction domain                              | PF00158 | 4 | 23  |
| Elongation factor Tu domain 2                            | PF03144 | 4 | 24  |
| HD domain  | PF01966 | 4 | 25  |
| DnaJ domain  | PF00226 | 4 | 26  |
| Phosphatidylinositol-specific phospholipase C, Y domain  | PF00387 | 4 | 27  |
| Carbamoyl-phosphate synthase L chain, ATP binding domain | PF02786 | 4 | 28  |
| SAM domain (Sterile alpha motif)                         | PF07647 | 4 | 32  |
| Domain of unknown function (DUF1787)                     | PF08742 | 4 | 35  |
| Protein kinase C terminal domain                         | PF00433 | 4 | 36  |
| Leucine rich repeat N-terminal domain                    | PF01462 | 4 | 39  |
| Zn-finger in Ran binding protein and others              | PF00641 | 4 | 40  |
| LysM domain  | PF01476 | 4 | 44  |
| Latrophilin/CL-1-like GPS domain                         | PF01825 | 4 | 46  |
| CBS domain pair  | PF00571 | 4 | 48  |
| Cyclic nucleotide-binding domain                         | PF00027 | 4 | 50  |
| IQ calmodulin-binding motif                              | PF00612 | 4 | 51  |
| AMP-binding enzyme                                       | PF00501 | 4 | 53  |
| KRAB box   | PF01352 | 4 | 57  |
| Laminin G domain   | PF02210 | 4 | 84  |
| Laminin EGF-like (Domains III and V)                     | PF00053 | 4 | 88  |
| Leucine Rich Repeat                                      | PF00560 | 4 | 101 |
| Calcium binding EGF domain                               | PF07645 | 4 | 123 |

|   |         |   |    |
|---|---------|---|----|
| CobQ/CobB/MinD/ParA nucleotide binding domain                   | PF01656 | 3 | 6  |
| Putative Fe-S cluster   | PF04060 | 3 | 7  |
| Inhibitor of Apoptosis domain                                   | PF00653 | 3 | 8  |
| 2OG-Fe(II) oxygenase superfamily                                | PF03171 | 3 | 9  |
| Phospholipase D Active site motif                               | PF00614 | 3 | 10 |
| Integrin beta tail domain                                       | PF07965 | 3 | 11 |
| MutS domain III   | PF05192 | 3 | 12 |
| MerR family regulatory protein                                  | PF00376 | 3 | 13 |
| Ricin-type beta-trefoil lectin domain                           | PF00652 | 3 | 14 |
| Tudor domain  | PF00567 | 3 | 15 |
| NACHT domain  | PF05729 | 3 | 16 |
| Stage II sporulation protein E (SpoIIE)                         | PF07228 | 3 | 17 |
| SWIM zinc finger  | PF04434 | 3 | 18 |
| Diacylglycerol kinase catalytic domain (presumed)               | PF00781 | 3 | 19 |
| ShK domain-like   | PF01549 | 3 | 20 |
| MyTH4 domain  | PF00784 | 3 | 21 |
| ATPase family associated with various cellular activities (AAA) | PF07728 | 3 | 25 |
| SPRY domain   | PF00622 | 3 | 26 |
| Acetyltransferase (GNAT) family                                 | PF00583 | 3 | 27 |
| HECT-domain (ubiquitin-transferase)                             | PF00632 | 3 | 29 |
| BAH domain  | PF01426 | 3 | 30 |
| Leucine rich repeat C-terminal domain                           | PF01463 | 3 | 31 |
| Double-stranded RNA binding motif                               | PF00035 | 3 | 33 |
| Reprolysin (M12B) family zinc metalloprotease                   | PF01421 | 3 | 34 |
| 7 transmembrane receptor (Secretin family)                      | PF00002 | 3 | 35 |
| Cadherin domain   | PF00028 | 3 | 37 |
| Zinc finger C-x8-C-x5-C-x3-H type (and similar)                 | PF00642 | 3 | 43 |
| KH domain   | PF00013 | 3 | 45 |
| SAM domain (Sterile alpha motif)                                | PF00536 | 3 | 49 |
| Ankyrin repeat  | PF00023 | 3 | 68 |
| Bacterial transcription activator, effector binding domain      | PF06445 | 2 | 3  |
| SGS domain  | PF05002 | 2 | 4  |

|   |         |   |    |
|---|---------|---|----|
| Pyruvate ferredoxin/flavodoxin oxidoreductase     | PF01558 | 2 | 5  |
| RasGAP C-terminus                                 | PF03836 | 2 | 6  |
| Thrombospondin C-terminal region                  | PF05735 | 2 | 7  |
| S4 domain   | PF01479 | 2 | 8  |
| Domain of Unknown Function (DUF906)               | PF06010 | 2 | 9  |
| KOW motif   | PF00467 | 2 | 10 |
| FKBP-type peptidyl-prolyl cis-trans isomerase     | PF00254 | 2 | 11 |
| GRF zinc finger                                   | PF06839 | 2 | 12 |
| Fes/CIP4 homology domain                          | PF00611 | 2 | 13 |
| R3H domain  | PF01424 | 2 | 14 |
| TrkA-N domain                                     | PF02254 | 2 | 15 |
| RNase3 domain                                     | PF00636 | 2 | 16 |
| Helix-loop-helix DNA-binding domain               | PF00010 | 2 | 17 |
| Endonuclease/Exonuclease/phosphatase family       | PF03372 | 2 | 18 |
| Tetratricopeptide repeat                          | PF07719 | 2 | 19 |
| FERM domain (Band 4.1 family)                     | PF00373 | 2 | 20 |
| HTH domain  | PF08279 | 2 | 21 |
| S1 RNA binding domain                             | PF00575 | 2 | 23 |
| Zona pellucida-like domain                        | PF00100 | 2 | 24 |
| Subtilase family                                  | PF00082 | 2 | 25 |
| Kringle domain                                    | PF00051 | 2 | 26 |
| Radical SAM superfamily                           | PF04055 | 2 | 27 |
| OB-fold nucleic acid binding domain               | PF01336 | 2 | 28 |
| RasGEF domain                                     | PF00617 | 2 | 29 |
| GCC2 and GCC3                                     | PF07699 | 2 | 30 |
| Ubiquitin family                                  | PF00240 | 2 | 31 |
| chromo' (CHRromatin Organisation MOdifier) domain | PF00385 | 2 | 33 |
| short chain dehydrogenase                         | PF00106 | 2 | 35 |
| TIR domain  | PF01582 | 2 | 36 |
| PPR repeat  | PF01535 | 2 | 37 |
| Condensation domain                               | PF00668 | 2 | 38 |
| BRCA1 C Terminus (BRCT) domain                    | PF00533 | 2 | 40 |
| haloacid dehalogenase-like hydrolase              | PF00702 | 2 | 42 |
| PKD domain  | PF00801 | 2 | 44 |

|  |         |   |    |
|--|---------|---|----|
| Low-density lipoprotein receptor domain class A                        | PF00057 | 2 | 45 |
| Homeobox domain  | PF00046 | 2 | 46 |
| Reverse transcriptase (RNA-dependent DNA polymerase)                   | PF00078 | 2 | 50 |
| UBA/TS-N domain  | PF00627 | 2 | 59 |
| RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)                | PF00076 | 2 | 87 |
| Chitinase class I  | PF00182 | 1 | 1  |
| Bacterial DNA polymerase III alpha subunit                             | PF07733 | 1 | 2  |
| Precorrin-8X methylmutase  | PF02570 | 1 | 3  |
| Voltage gated calcium channel IQ domain                                | PF08763 | 1 | 4  |
| BPS (Between PH and SH2)   | PF08947 | 1 | 5  |
| Elongation factor G, domain IV   | PF03764 | 1 | 6  |
| Carboxyl transferase domain  | PF01039 | 1 | 7  |
| Glycosyl hydrolases family 2, immunoglobulin-like beta-sandwich domain | PF00703 | 1 | 8  |
| SAND domain  | PF01342 | 1 | 9  |
| L27 domain   | PF02828 | 1 | 10 |
| Formin Homology 2 Domain   | PF02181 | 1 | 11 |
| Biotin carboxylase C-terminal domain                                   | PF02785 | 1 | 12 |
| HNH endonuclease   | PF01844 | 1 | 13 |
| FerB (NUC096) domain   | PF08150 | 1 | 14 |
| Ligand-binding domain of nuclear hormone receptor                      | PF00104 | 1 | 15 |
| Miro-like protein  | PF08477 | 1 | 16 |
| ABC transporter transmembrane region                                   | PF00664 | 1 | 17 |
| Autotransporter beta-domain  | PF03797 | 1 | 18 |
| Putative GTPase activating protein for Arf                             | PF01412 | 1 | 19 |
| Ubiquitin carboxyl-terminal hydrolase                                  | PF00443 | 1 | 20 |
| IPT/TIG domain   | PF01833 | 1 | 21 |
| Laminin Domain II  | PF06009 | 1 | 22 |
| Immunoglobulin C1-set domain   | PF07654 | 1 | 23 |
| Retrotransposon gag protein  | PF03732 | 1 | 25 |
| EMI domain   | PF07546 | 1 | 26 |
| Acyl transferase domain  | PF00698 | 1 | 27 |
| Kazal-type serine protease inhibitor domain                            | PF07648 | 1 | 30 |

|  |         |   |    |
|--|---------|---|----|
| Methyl-accepting chemotaxis protein (MCP) signaling domain | PF00015 | 1 | 31 |
| Spectrin repeat  | PF00435 | 1 | 32 |
| Zinc finger, ZZ type                                       | PF00569 | 1 | 33 |
| Beta-ketoacyl synthase, N-terminal domain                  | PF00109 | 1 | 40 |
| Beta-ketoacyl synthase, C-terminal domain                  | PF02801 | 1 | 42 |
| Myosin head (motor domain)                                 | PF00063 | 1 | 43 |
| F5/8 type C domain   | PF00754 | 1 | 50 |
| ABC transporter  | PF00005 | 1 | 54 |
| Trypsin  | PF00089 | 1 | 57 |

**Table S5.** Domains occurring in multiply originated architectures, according to the *nolimit* dataset. The first column shows the domain name and the second its Pfam accession number. The third column contains the number of architectures where multiple origins were inferred and where the domain occurs at least once, and the fourth the total number of architectures in our dataset where the domain is present.