S1 Supporting Information - Group composition

The following tables show how the originally defined groups in the reference datasets OrthoBench and COG are separated by Proteinortho (Po) and its extension PoFF. Members determines the number of genes in each reference group. Groups indicates into how many groups, the original reference group was divided by the respective approaches. Missing Genes gives rise to the number of original members that are not present in any group. Additional Genes tells how many members were added to the algorithm derived groups that were not present in the reference group before.

**Table S1.1 OrthoBench**

| Group                        | Members | Groups | Missing Genes | Additional Genes |
|------------------------------|---------|--------|---------------|------------------|
| **Low Speed of Evolution**   |         |        |               |                  |
| Split hand/Split foot 1      | 8       | 2      | 0             | 6                |
| PhD Finger family            | 12      | 1      | 0             | 0                |
| **High Speed of Evolution**  |         |        |               |                  |
| Vomeronasalulin              | 5       | 1      | 0             | 0                |
| PLUNC proteins               | 60      | 11     | 1             | 21               |
| Transient receptor channel   | 41      | 8      | 1             | 2                |
| Ionotropic glutamate receptors| 7      | 0      | 7             | 7                |
| Phosphodiesterase 4D inter. pro.| 11 | 3   | 1             | 16               |
| Vitellogenin                  | 9       | 3      | 2             | 11               |
| LIM domain kinase            | 23      | 4      | 2             | 21               |
| Otoferlin                    | 26      | 4      | 0             | 19               |
| **Low complexity regions/repeats** |   |        |               |                  |
| Sec13                        | 11      | 1      | 1             | 8                |
| Retinitis pigmentosa         | 18      | 4      | 4             | 5                |
| Ankyrin repeat               | 99      | 15     | 3             | 41               |
| Osteoclast protein           | 13      | 1      | 0             | 0                |
| Dilute myosin heavy chain    | 32      | 5      | 3             | 34               |
| Myosin heavy chain           | 41      | 5      | 5             | 50               |
| GPS domain-containing GPCRs  | 50      | 10     | 5             | 57               |
| Leucine-rich repeat          | 12      | 2      | 0             | 11               |
| Laminin alpha                | 18      | 3      | 1             | 20               |
| **Domain shuffling/evolution**|      |        |               |                  |
| Thrombospondin protein       | 54      | 9      | 4             | 15               |
| Low-density lipoprotein receptors| 11   | 1      | 2             | 2                |
| FOG- Immunoglobin related proteins | 16  | 5     | 0             | 19               |
| RNA helicase BRR2            | 10      | 1      | 0             | 4                |
| **Multigene families/Paralogy** |     |        |               |                  |
| GPCR hormone receptors       | 27      | 4      | 5             | 18               |
| Tyrosine-protein kinase      | 67      | 14     | 5             | 39               |
| Cytochrome P450               | 27      | 4      | 5             | 4                |
| Carbonyl reductases          | 24      | 3      | 1             | 1                |
| ATP-binding cassette         | 31      | 4      | 3             | 18               |
| Pyruvate dehydrogenase kinases | 41    | 3     | 1             | 6                |
| **Low Alignment Quality**    |         |        |               |                  |
| Collagen type IV alpha chain | 49      | 11     | 2             | 55               |
| Filagrin                     | 6       | 2      | 2             | 3                |
| GATA 1/2/3                   | 32      | 5      | 4             | 7                |
| Mucins                       | 60      | 14     | 15            | 35               |
| Chitinase                    | 45      | 6      | 12            | 31               |
| TRAF4                        | 13      | 1      | 2             | 3                |
| FGFR1/2/3/4                  | 41      | 11     | 19            | 97               |
| **High Alignment Quality**   |         |        |               |                  |
| Ski oncogene protein         | 23      | 5      | 6             | 23               |
| Ataxin-7-like protein        | 11      | 2      | 3             | 6                |
| Ribosomal protein S12        | 13      | 1      | 4             | 2                |
| **Random Families**          |         |        |               |                  |
| Ribosomal protein L11        | 21      | 1      | 2             | 9                |
| Methionine-R-sulfoxide reductase | 18  | 1     | 3             | 12               |
| Erythropoietin 4              | 15      | 1      | 1             | 3                |
| Proline oxidase              | 18      | 3      | 5             | 18               |
| Peroxisomal multifunctional enzyme | 13  | 3     | 4             | 9                |
| Group | Members | Groups | OrthoBench Po | PoFF Po | Po | PoFF Po | Po | PoFF Po |
|-------|---------|--------|---------------|---------|----|---------|----|---------|
| Summary | 1642 (23.5%) | 265 (3.8%) | 377 (5.4%) | 150 (9.1%) | 174 (10.6%) | 976 (39.5%) | 802 (35.3%) |

Table S1.2 COG

| Summary | Members | Groups | OrthoBench Po | PoFF Po | Po | PoFF Po | Po | PoFF Po |
|---------|---------|--------|---------------|---------|----|---------|----|---------|
| Sum | 31346 | 5143 | 5344 | 3546 | 3848 | 3201 | 2901 |
| Average/Fraction | (≠18.4%) | (≠3.0%) | (≠5.1%) | (11.3%) | (12.3%) | (10.3%) | (9.5%) |