Supplemental Material for

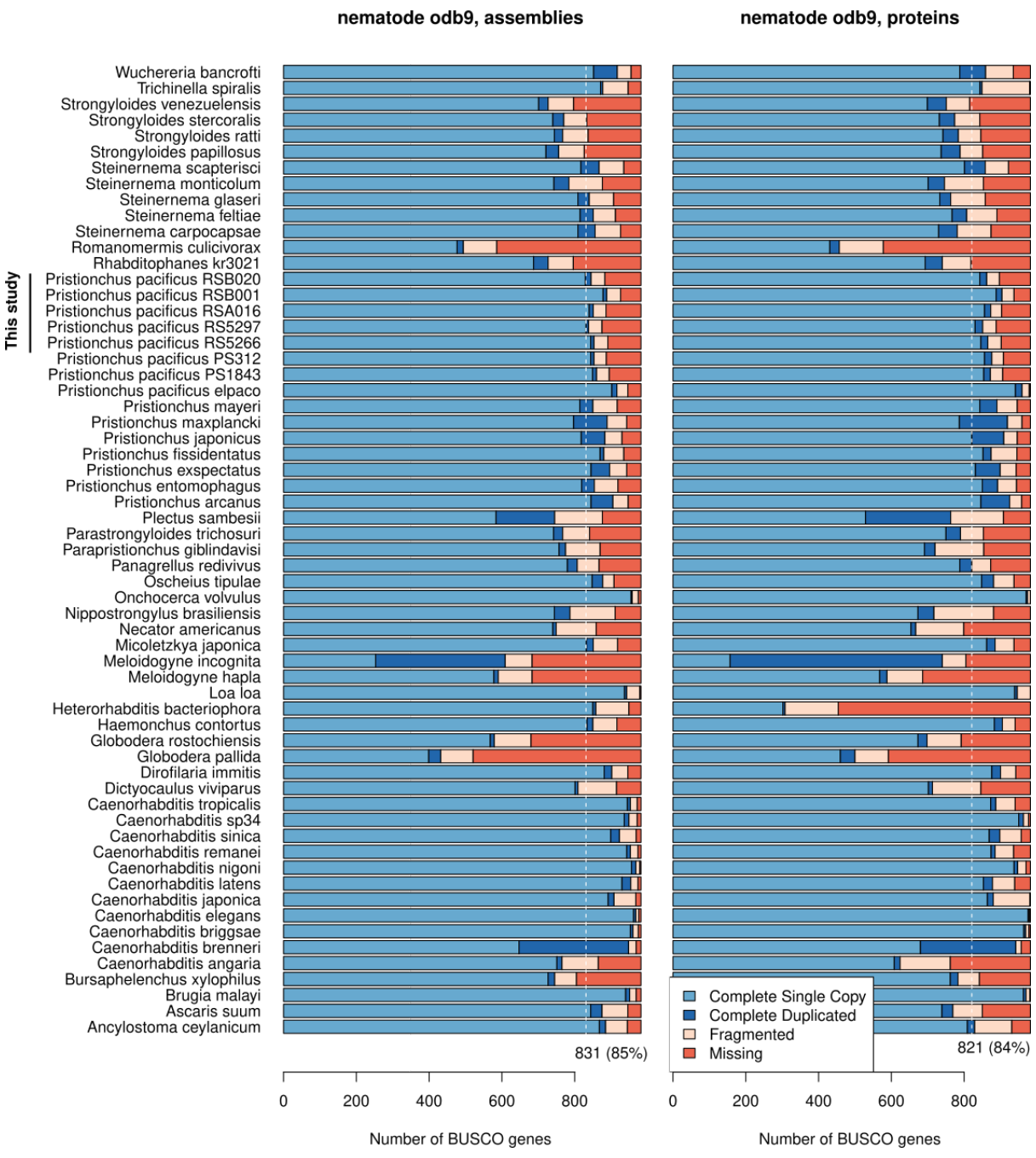
Multiple *Pristionchus pacificus* genomes reveal distinct evolutionary dynamics between *de novo* candidates and duplicated genes

Neel Prabh[1,2], Christian Rödelsperger[1,*]

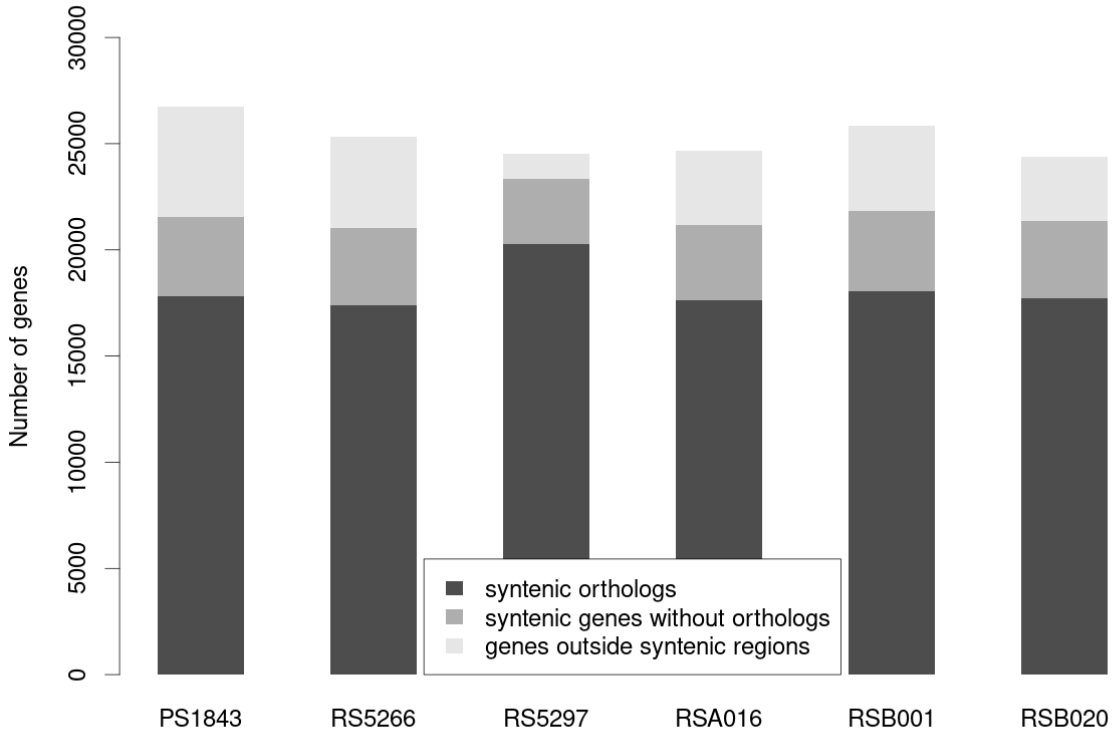
[1]Department for Integrative Evolutionary Biology, Max Planck Institute for Biology, Max-Planck-Ring 9, 72076 Tübingen, Germany

[2] Current address: Department for Evolutionary Genetics, Max Planck Institute for Evolutionary Biology, August-Thienmann-Str. 2, 24306 Plön, Germany

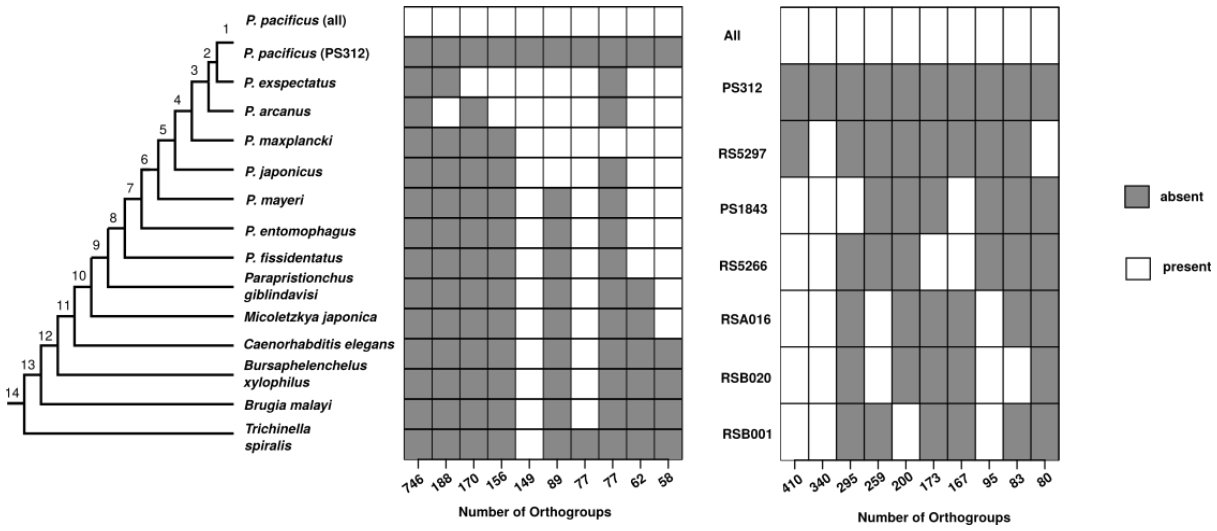
[*] Corresponding author's email address: christian.roedelsperger@tuebingen.mpg.de



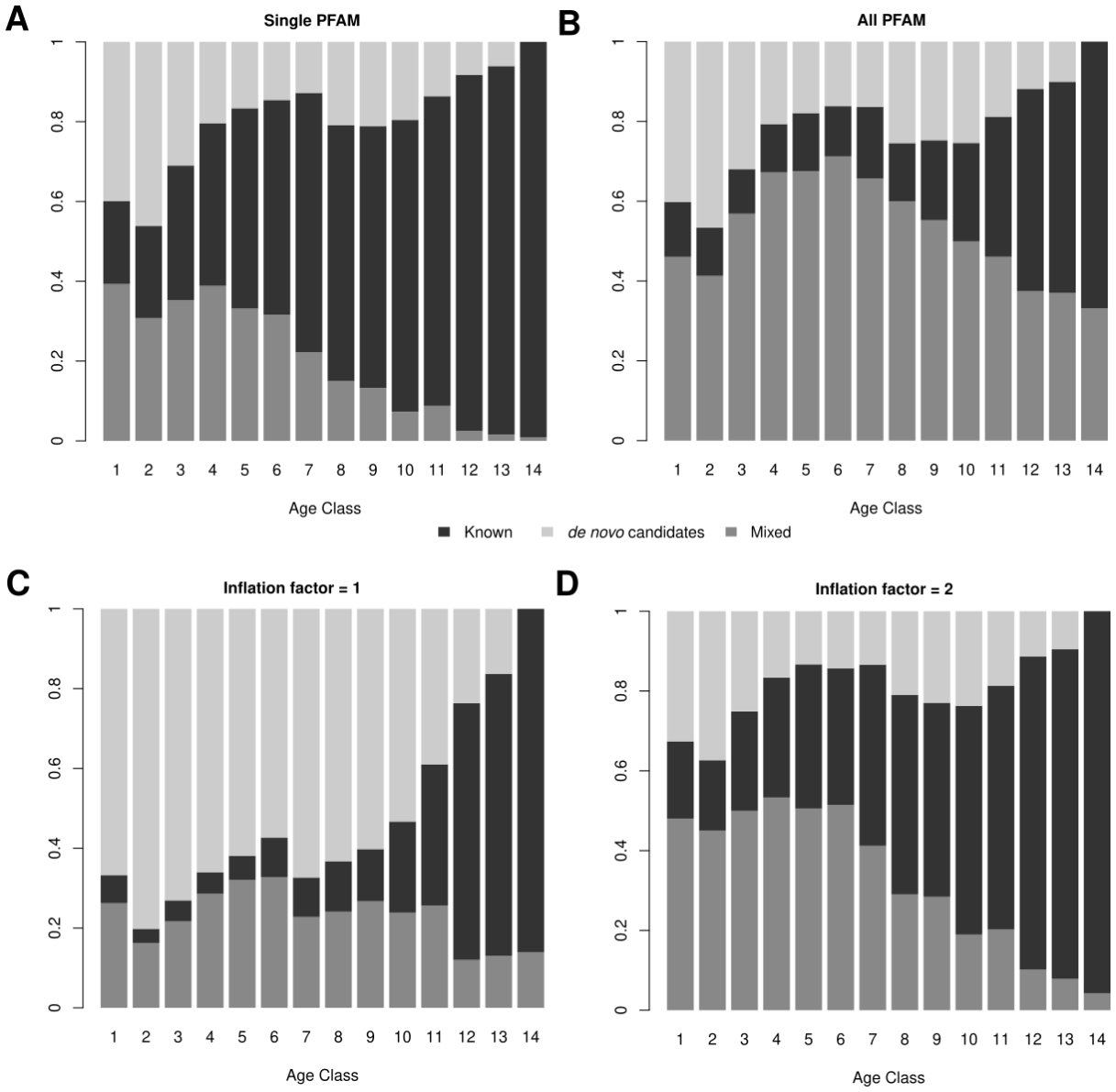
Supplemental Figure S1. Comparative evaluation of nematode genomes. Genome assemblies and protein-coding genes from this study were compared to a phylogenomic data set from 54 nematodes. The dashed line marks the median BUSCO Complete Single Copy value for the whole data set.



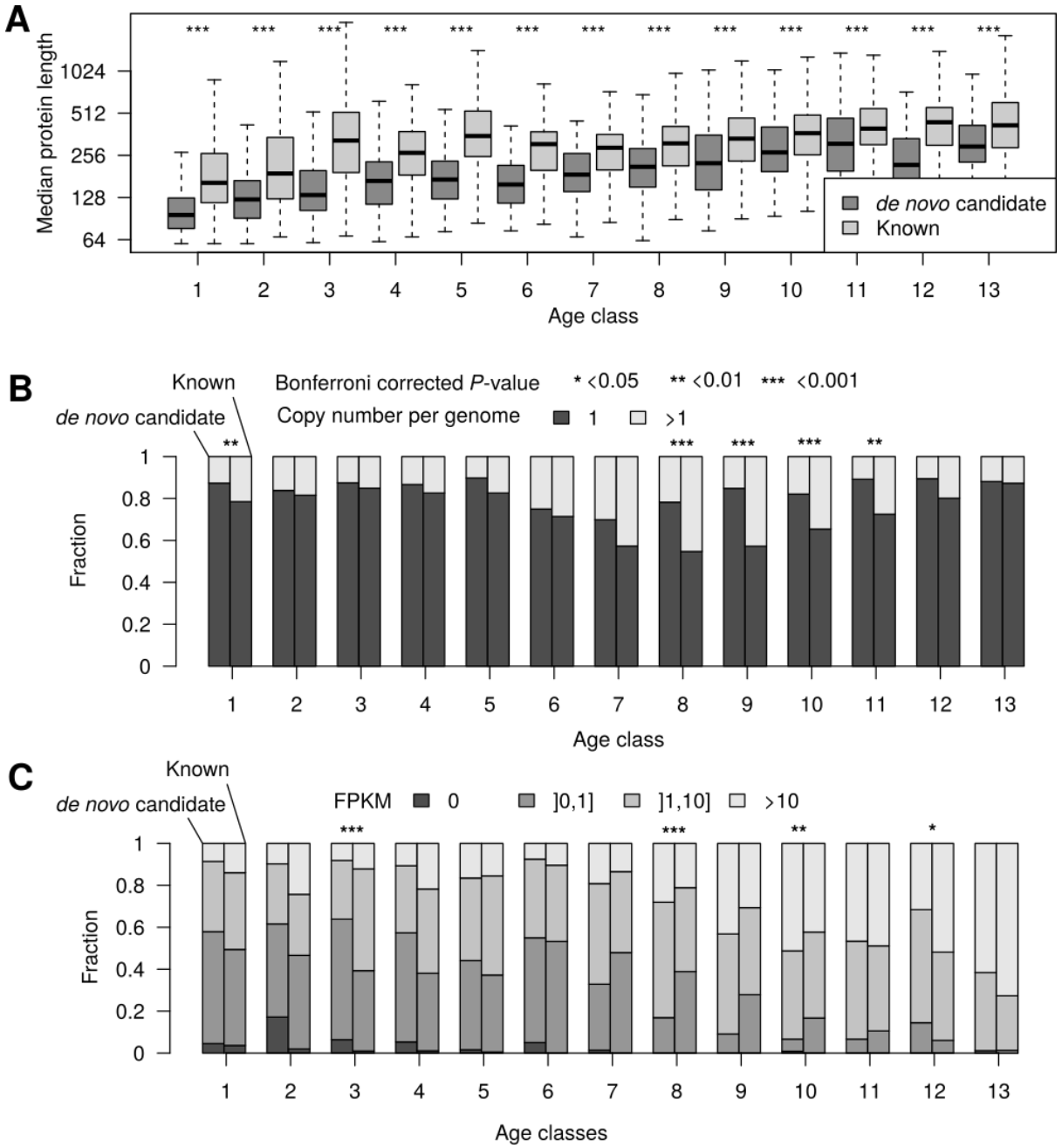
Supplemental Figure S2. Syntenic orthologs. Conserved syntenic regions were identified based on gene order alignments with the reference strain PS312 by the software CYNTENATOR. From these alignments, syntenic orthologs were defined as gene pairs in collinear blocks with protein homology. Genes that were aligned to a gap or with a mismatch to a non-homologous gene were counted as syntenic genes without ortholog.



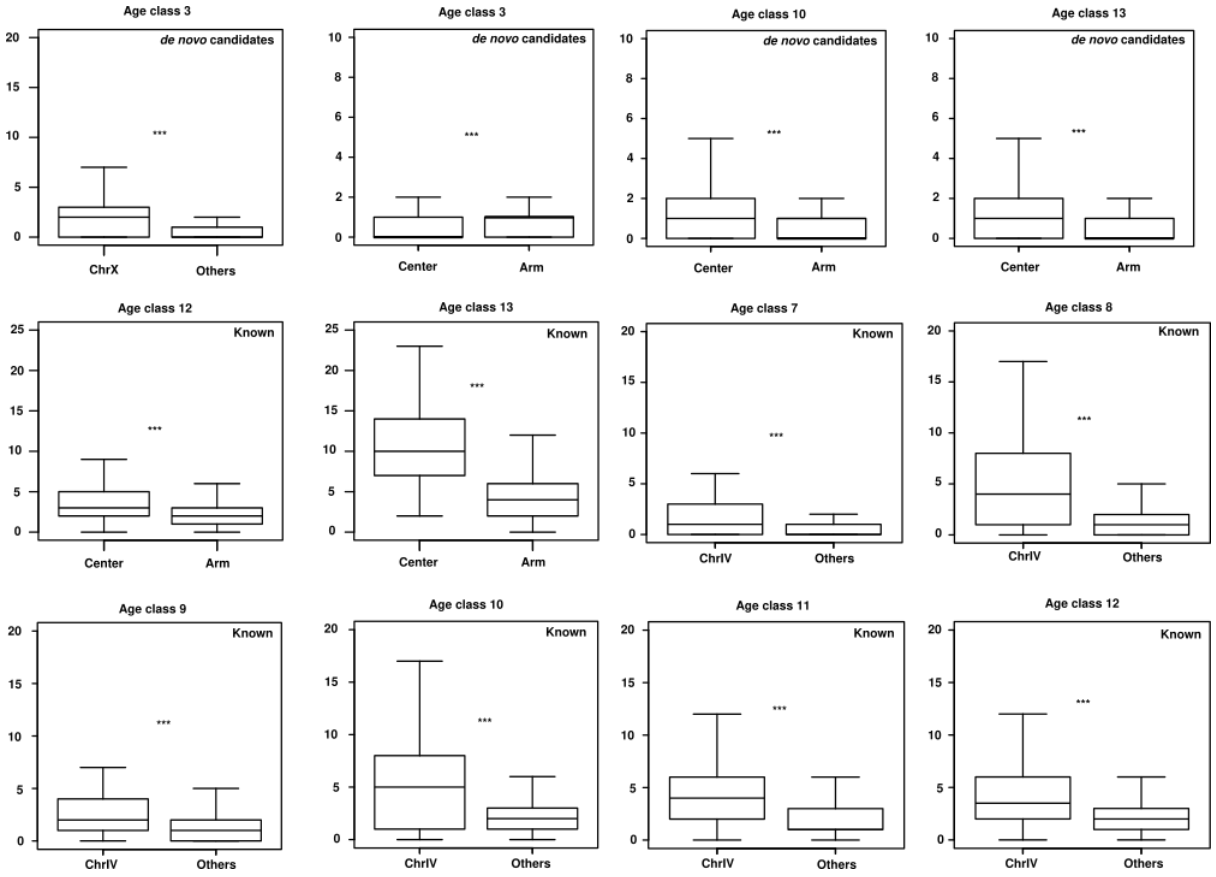
Supplemental Figure S3. Orthogroups missing in the reference strain PS312. 3,396 orthogroups that lack an ortholog in the reference strain PS312 but have been preserved in at least one of the six additional strains The heatmaps show the ten most abundant presence/absence patterns of these orthogroups across nematode species and *P. pacificus* strains.



Supplemental Figure S4. Robustness of rapid turnover hypothesis. (A) The bars show the distribution of categories of origin for different age classes. Hereby, orthogroups were defined as known, if a single sequence had a predicted protein domain in the PFAM database. (B) Orthogroups were defined as known, if all sequences had a specific PFAM domain. (C) When changing the inflation factor of the MCL algorithm from the default value of 1.5 to 1, orthogroups become less granular. This also reduced the total number of orthogroups from 37,228 to 18,085. This changes the relative contribution between known and *de novo* candidates. (D) Changing the inflation parameter to 2 resulted in 44,390 orthogroups and increased the contribution of known gene families. For panel C and D orthogroups were defined as known, if more than half of the sequences had a specific PFAM domain or had a BLASTP hit in the *D. melanogaster* or mouse data sets.



Supplemental Figure S5. **Comparison of protein length, copy number, and expression.** **(**A) Protein lengths were compared between *de* novo candidates and known gene families for different age classes. All comparisons showed that *de novo* candidates are typically shorter than members of known gene families (Wilcoxon-test, Bonferroni corrected $P < 0.001$). (B) The barplots show the fraction of orthogroups with one or more gene copies per genome for different age classes. Known gene families of age classes 1,8-11 have significantly larger fraction of multicopy orthogroups (χ^2-test, Bonferroni corrected $P < 0.05$). (C) Mean expression levels for *P. pacificus* genes were computed for mixed-stage cultures of all seven strains. The barplots show the distribution of expression levels across age classes. Age classes 3,8,10,12 show significant differences between *de novo* candidates and known genes (χ^2-test, Bonferroni corrected $P < 0.05$).



Supplemental Figure S6. **Chromosomal bias of different age classes and categories of origin.** Apart from the tendency of *de novo* candidates from age class 1 to be enriched on the Chromosome X and to be depleted in the chromosome centers, additional significant chromosomal biases were detected (Wilcoxon-test, Bonferroni corrected $P < 0.001$). The y-axis denotes the number of orthogroups of the specific gene set per 500-kb window.