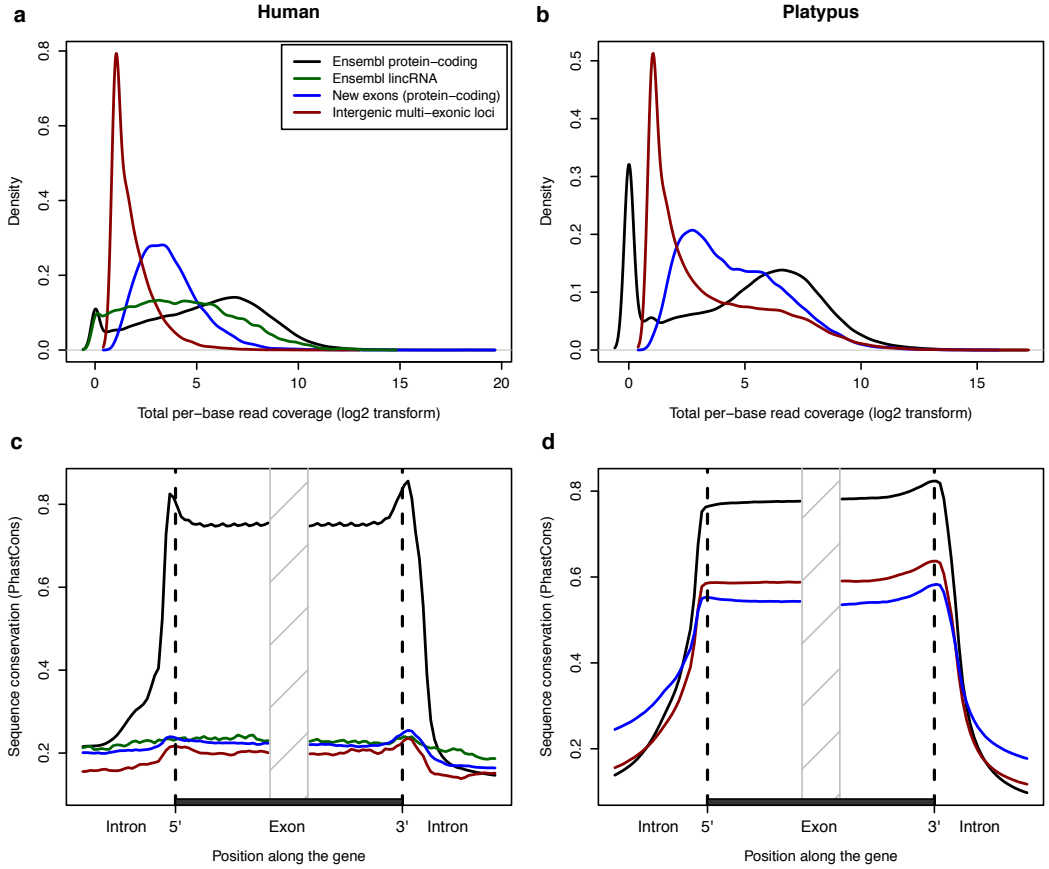
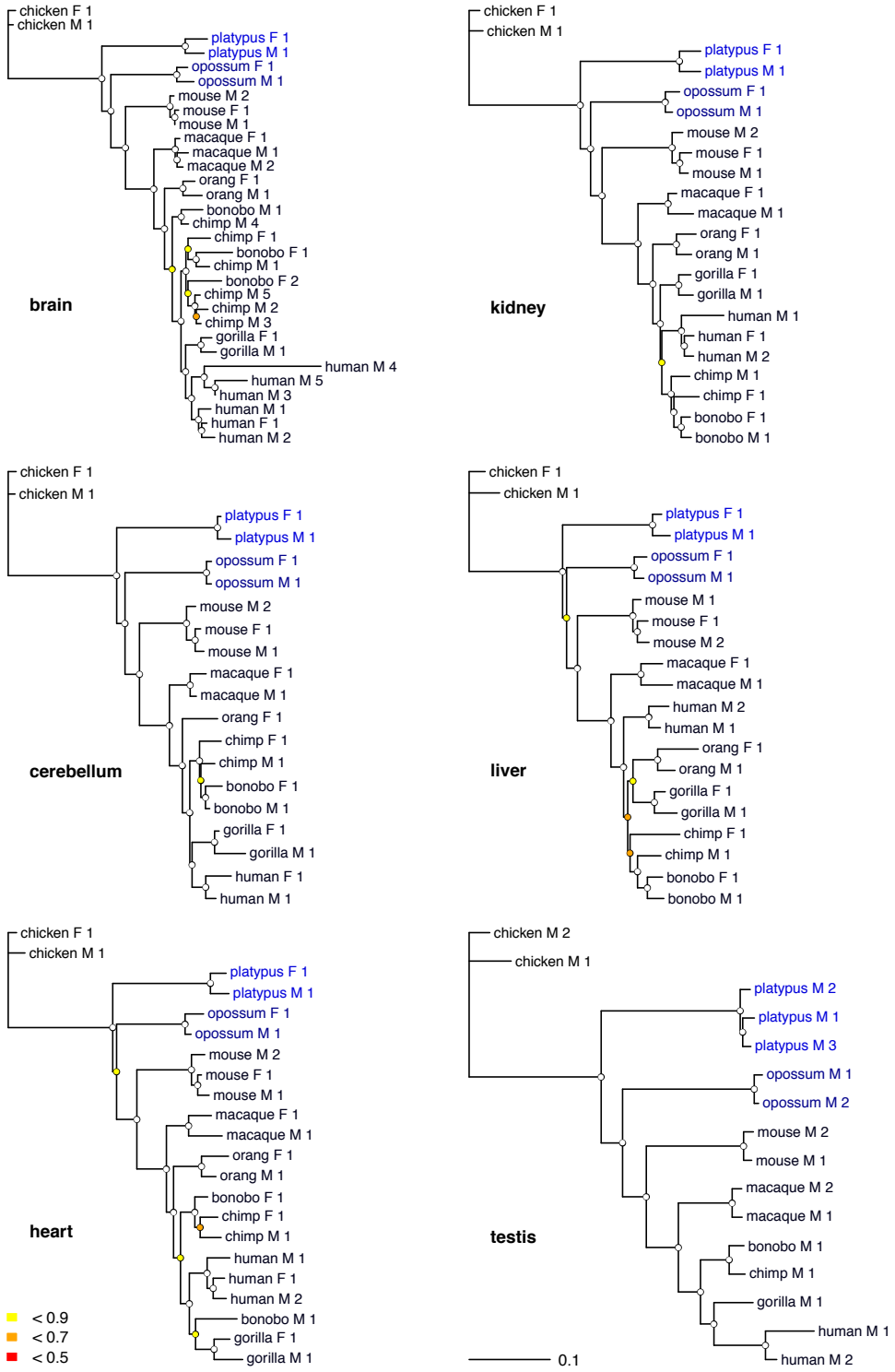
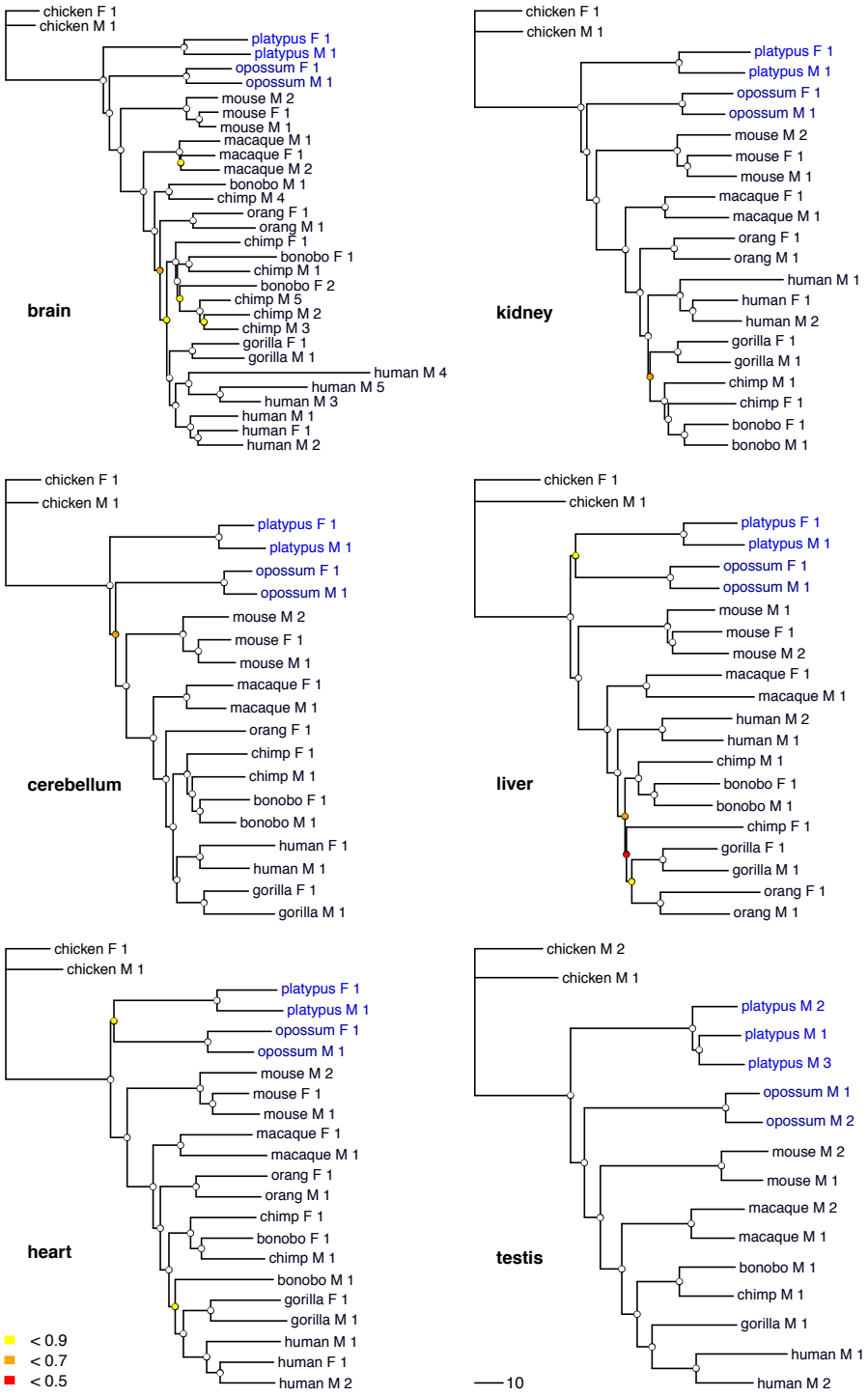
doi:10.1038/nature10532



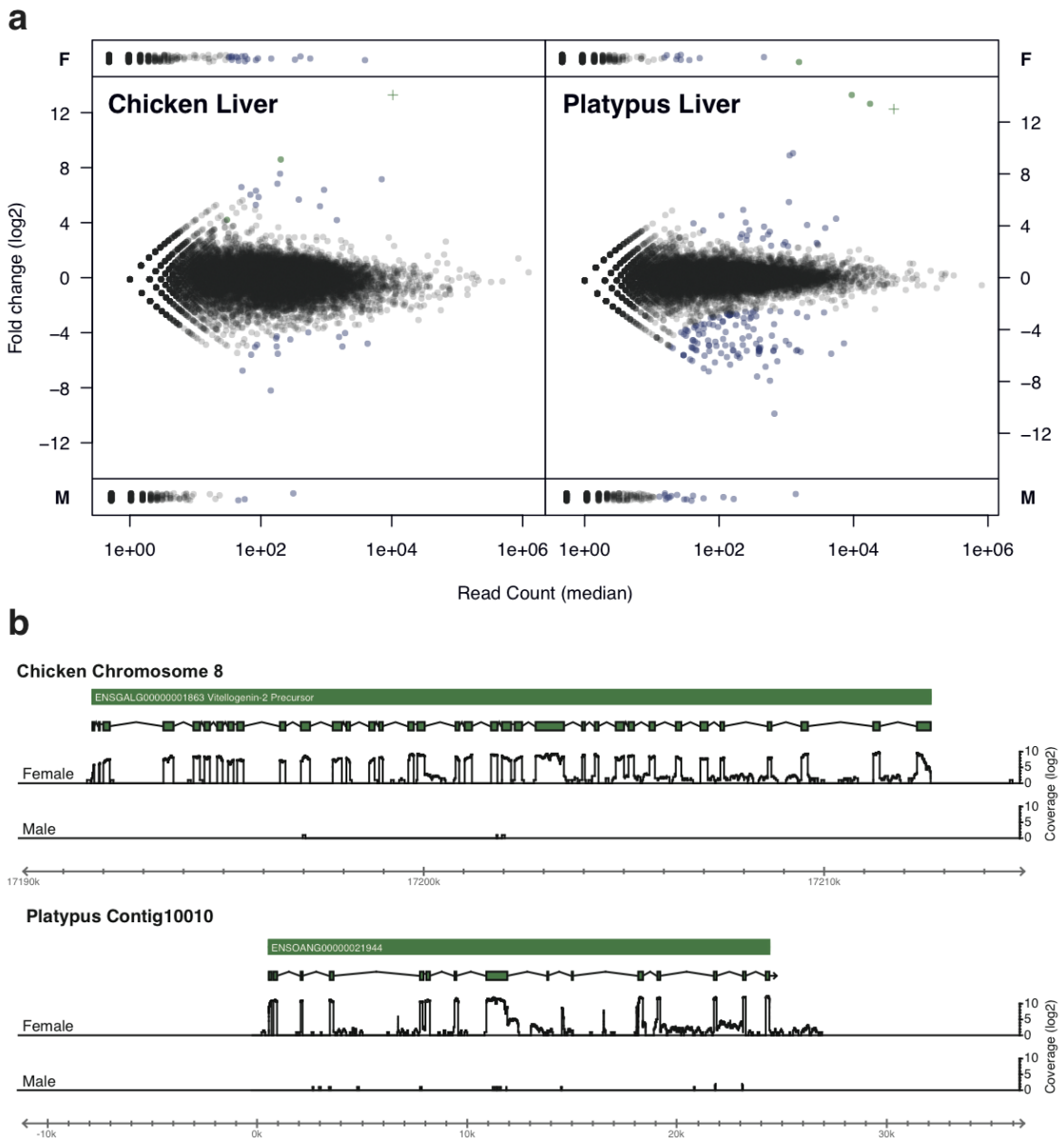
Supplementary Figure 1. Expression level and sequence conservation of newly annotated regions in the human and platypus genomes. a, **b** Density plot of the total per-base read coverage (all samples combined) for four categories: exons of Ensembl-annotated protein-coding genes (black) and large intervening noncoding RNA (lincRNA) genes (green), new exons added to previously known (Ensembl) protein-coding genes (blue), and exons of intergenic multi-exonic transcribed loci (red). The read coverage was log2-transformed (an offset of 1 was added to preserve null values). **c**, **d** Variation of the mean PhastCons score around exon boundaries, for the same categories of exons. The PhastCons score was computed for 100 positions, divided into two 50 bp segments centered on the exon boundaries, and was averaged over all exons in a class. Only internal exons with a length of at least 50 bp and with flanking introns larger than 25 bp were considered for this analysis.



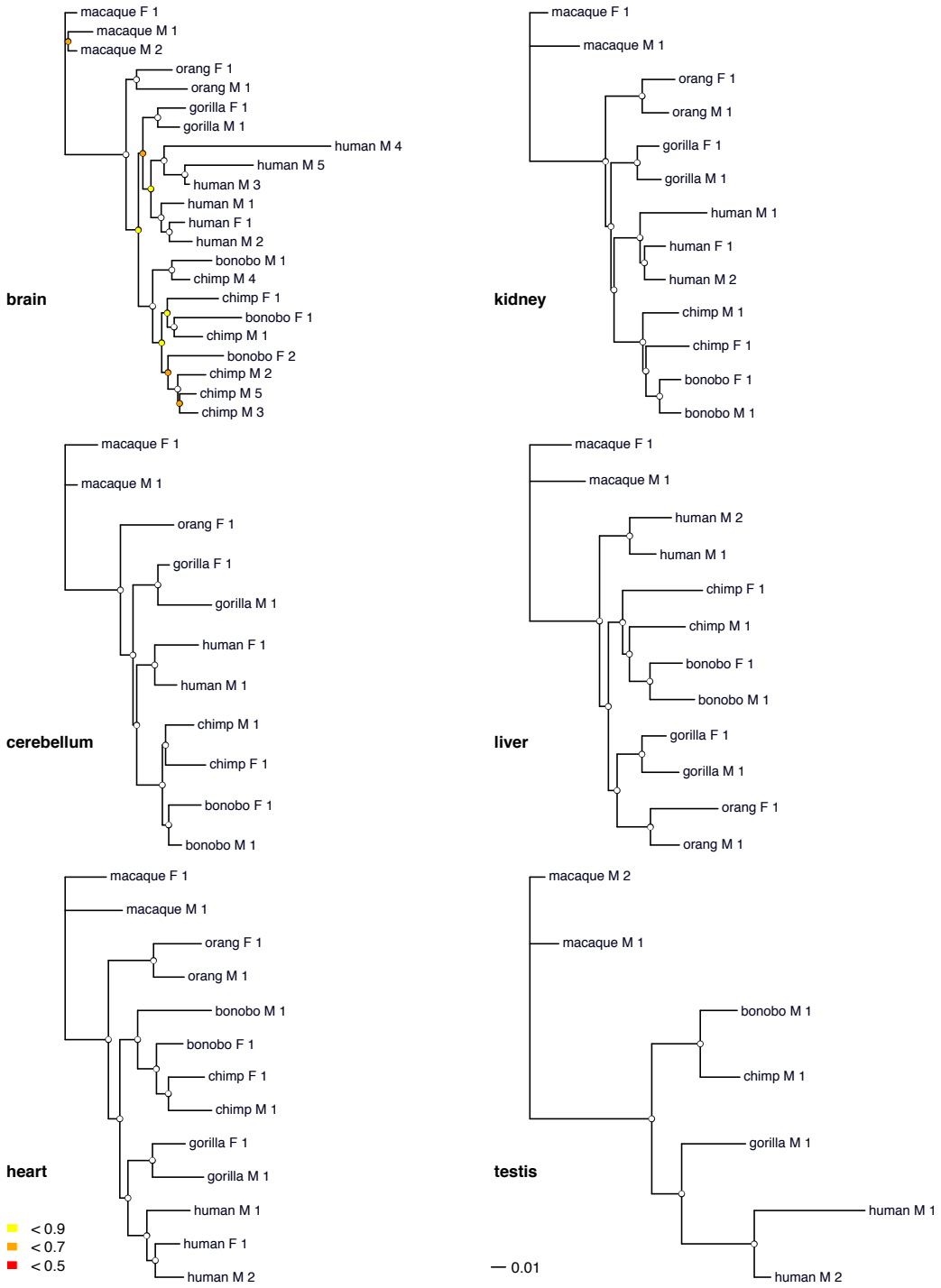
Supplementary Figure 2. Mammalian gene expression phylogenies. Neighbor-joining trees based on distance matrices comprising all pairwise expression level distances (1–*rho*, Spearman's correlation coefficient) for the six different organs. Trees are drawn to the same scale (indicated by the scale bar). Bootstrap values (i.e., proportions of replicate trees that have the branching pattern as in the majority-rule consensus tree shown) are indicated by circles at the corresponding nodes: > 0.9 (white fill), < 0.9 (yellow), < 0.7 (red). Mammalian species names are color-coded according to the three major mammalian lineages: monotremes (i.e., platypus; light blue), marsupials (opossum; dark blue), and eutherians (mice and primates; black).



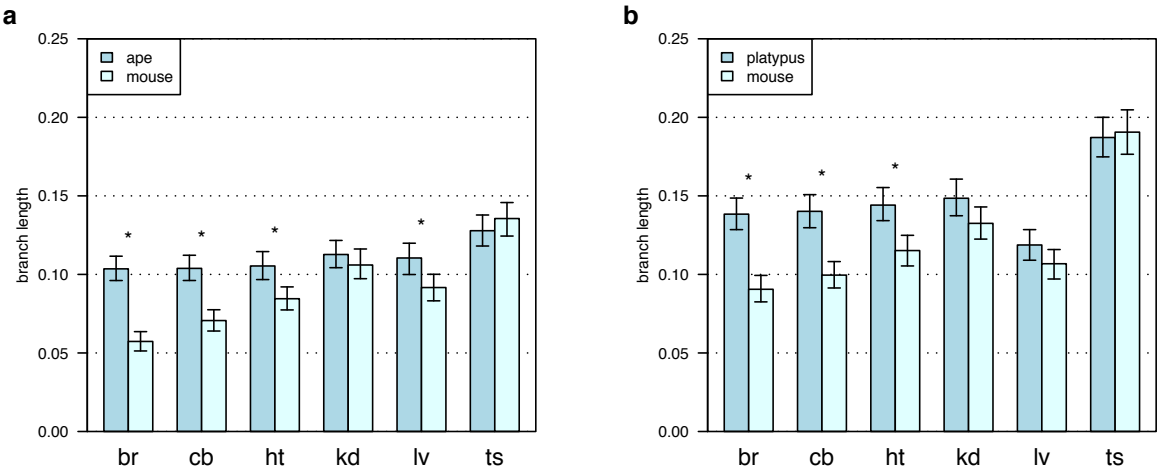
Supplementary Figure 3. Mammalian gene expression phylogenies. Neighbor-joining trees based on distance matrices comprising all pairwise expression level Euclidean distances for the six different organs. Trees are drawn to the same scale (indicated by the scale bar). Bootstrap values (i.e., proportions of replicate trees that have the branching pattern as in the majority-rule consensus tree shown) are indicated by circles at the corresponding nodes: > 0.9 (white fill), < 0.9 (yellow), < 0.7 (red). Mammalian species names are color-coded according to the three major mammalian lineages: monotremes (i.e., platypus; light blue), marsupials (opossum; dark blue), and eutherians (mice and primates; black).



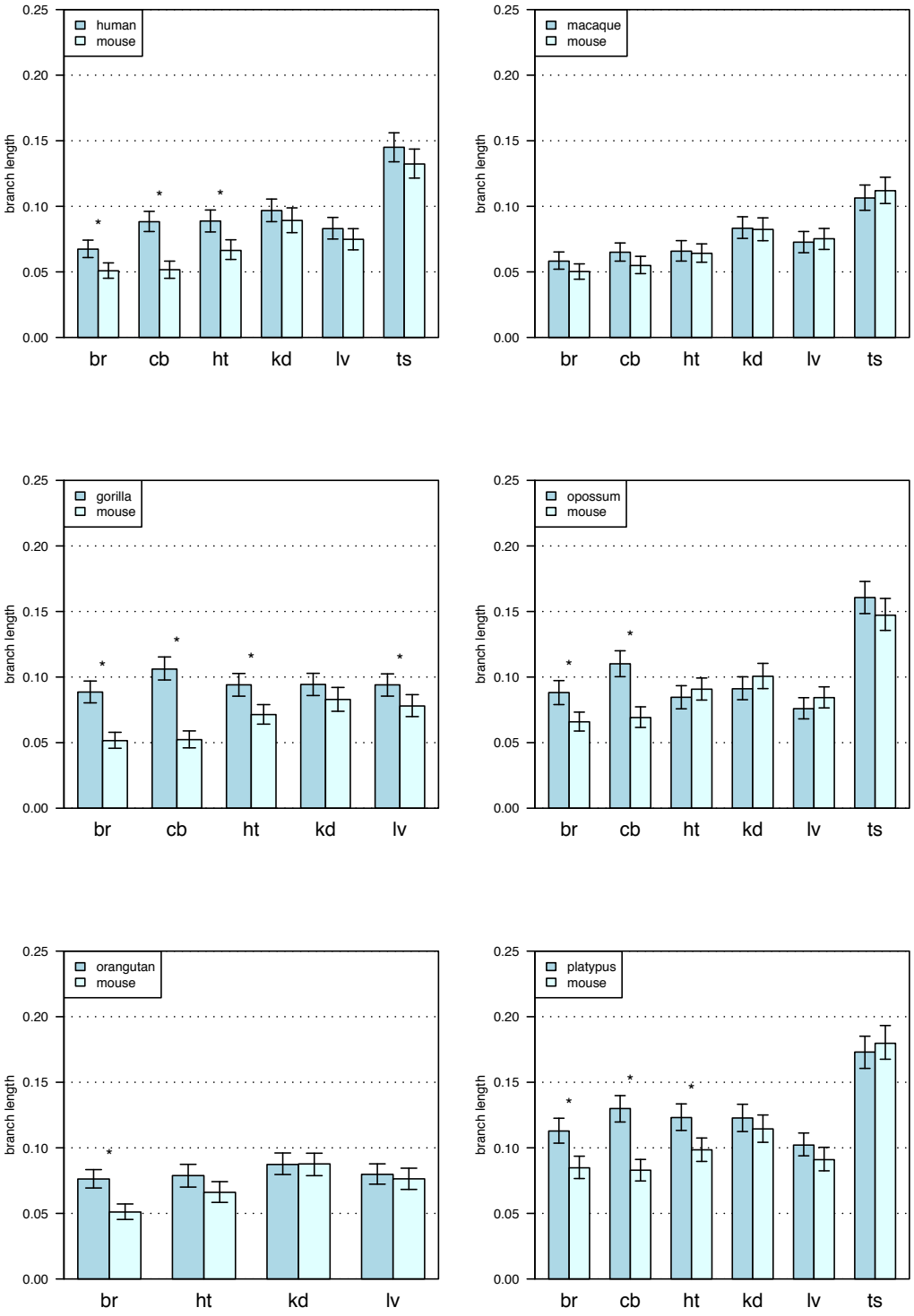
Supplementary Figure 4. Sex-biased expression of the vitellogenin egg yolk genes in platypus and chicken livers. a, Median gene expression levels (based on read counts) of genes in livers of male (M) and female (F) chicken and platypus animals (X-axis) are plotted against expression differences (shown as log2 fold-changes) between males and females (Y-axis). Top and bottom boxes contain genes with zero expression in females or males, respectively. Vitellogenin genes or gene fragments are marked in green: in the chicken plot, the green plus sign indicates the *VIT2* gene (further illustrated in subpanel b), the more central green circle indicates the *VIT1* gene, and the less central green circle represents *VIT3* (*VIT2* and *VIT3* have significantly higher expression levels in females, see also Supplementary Table 41); in the platypus plot, the plus sign indicates the contig containing the major part of the predicted platypus *VIT* gene, whereas the three green circles represent contigs containing the remaining platypus *VIT* exons. **b**, Ensembl annotation of chicken *VIT2* and a major part of platypus *VIT* and the corresponding read coverage in females and males.



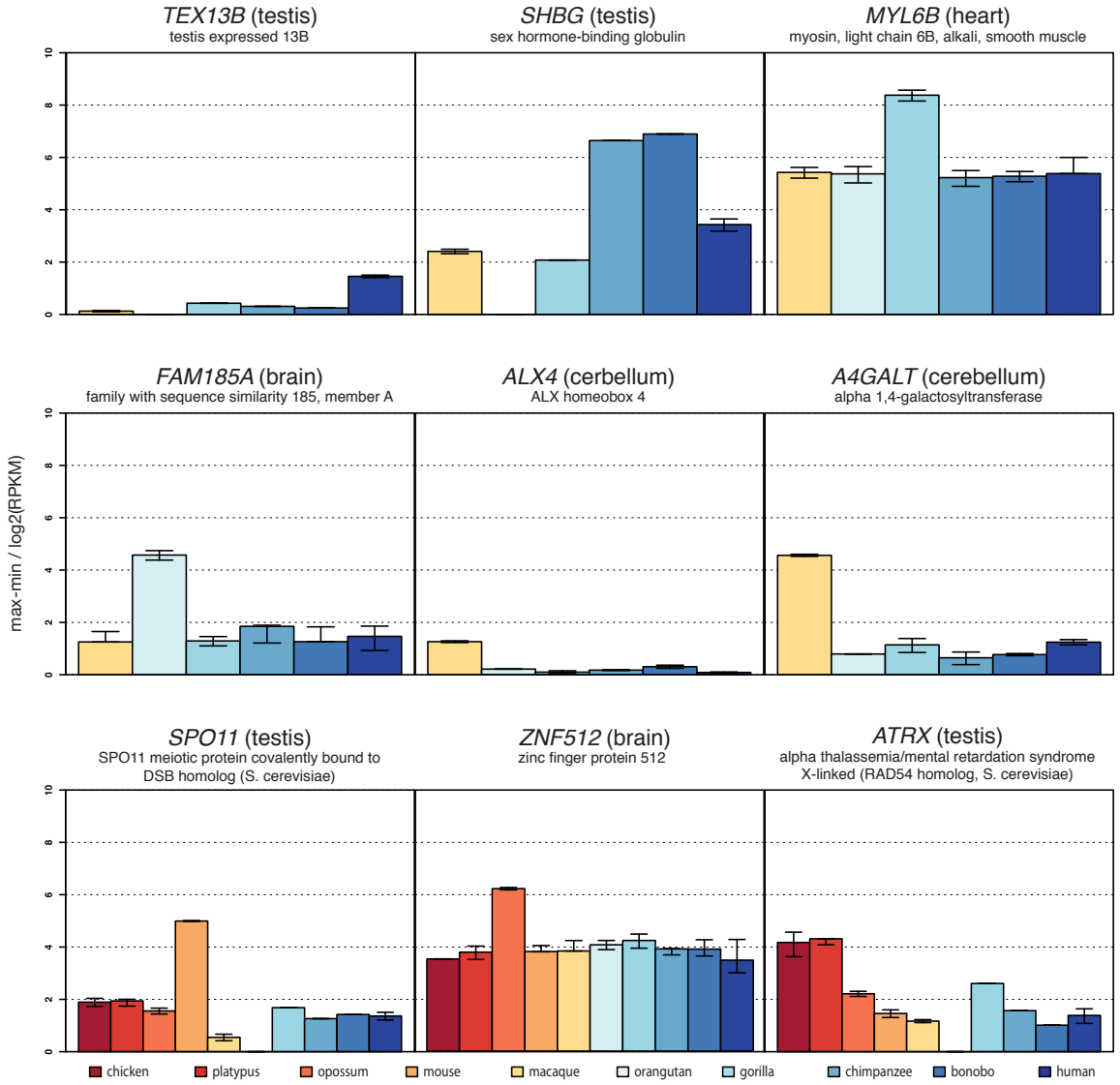
Supplementary Figure 5. Primate gene expression phylogenies. Neighbor-joining trees based on distance matrices comprising all pairwise expression level distances (1–rho, Spearman's correlation coefficient) for the six different organs. Note that the underlying gene expression values were calculated based on constitutive orthologous exons (from the 13,277 1–1 orthologous genes) that are perfectly aligned (no gaps permitted) between the six primate species (see Supplementary Note section 1.6 for details). Thus, we rule out any gene expression variation that might arise when gene expression levels are computed for potentially different sequences (e.g., as a result of annotation or biological/genomic differences) among 1–1 orthologs from different species. Bootstrap values (i.e., proportions of replicate trees that have the branching pattern as in the majority-rule consensus tree shown) are indicated by circles at the corresponding nodes: > 0.9 (white fill), < 0.9 (yellow), < 0.7 (red).



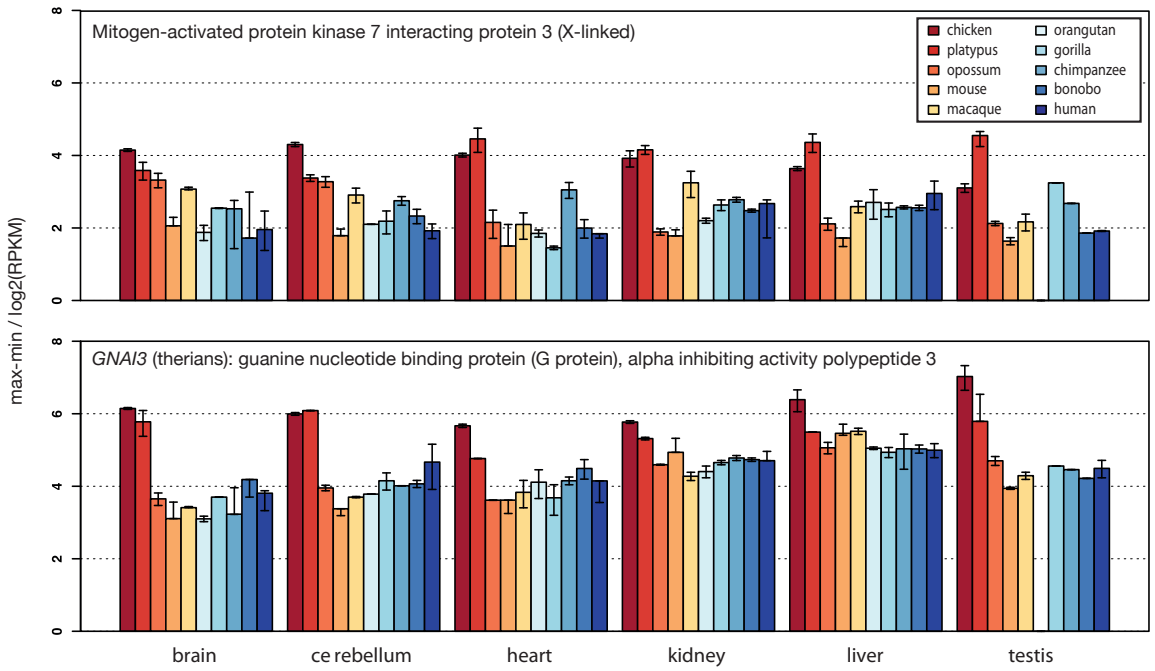
Supplementary Figure 6. Comparisons of branch lengths among mammalian lineages. a, **b**, Bootstrap pseudosamples (1,000) were generated that each comprise 5,636 sets of orthologues from one randomly sampled individual per species or group of species (in the case of apes). For each of these pseudosamples, Neighbor-joining trees of two species and the outgroup (chicken) were constructed; the total length of the branch (or branches) from the last common ancestral node to the respective terminal nodes was assessed for each of these trees. 95% confidence intervals are shown. Statistical significance was assessed by counting the number of trees in which one or the other species/lineage had longer branch lengths: [*] indicates two-tailed $P < 0.05$ (i.e., one lineage with longer branches in more than 97.5% of replicates; after Bonferroni correction for the six tests performed for the six tissues in each comparison). To factor out within-species branch length differences, similar analyses were performed that measure branch lengths from the last common ancestral node of a given species-pair to the last common ancestral node of individuals for each of the species in the pair (i.e., terminal branches were not considered in this analysis; Supplementary Fig. 6).



Supplementary Figure 7. Comparisons of branch lengths among mammalian lineages. Bootstrap pseudosamples (1,000) were generated that each comprise 5,636 sets of orthologues from one randomly sampled individual per species. For each of these pseudosamples, Neighbor-joining trees of two species and the outgroup (chicken) were constructed. To factor out within-species branch length differences, the total length of the branch (or branches) from the last common ancestral node of the given species pair to the most recent common ancestral nodes of individuals for each species in the pair, respectively, was assessed for each of these trees. 95% confidence intervals are shown. Statistical significance was assessed by counting the number of trees in which one or the other species/lineage had longer branch lengths: [*] indicates two-tailed $P < 0.05$ (i.e., one lineage with longer branches in more than 97.5% of replicates; after Bonferroni correction for the six tests performed for the six tissues in each comparison).



Supplementary Figure 8. Genes with shifts in optimal expression values (individual tissues). Examples of genes that evolved new optimal expression levels in a given mammalian lineage or species (as detected using a phylogenetic maximum likelihood procedure – see main text and Supplementary Note) are shown (details regarding these genes and all other statistically significant expression changes in mammals are provided in Supplementary Tables 9-24): *TEX13B* (ENSG00000170925), specifically expressed in spermatogonia (i.e., pre-meiotic cells), was significantly upregulated in human testis; *SHBG* (ENSG00000129214), a key gene in testosterone metabolism evolved high expression in chimpanzee and bonobo testes; *MYL6B* (ENSG00000196465), encoding a myosin light chain subunit, became upregulated in gorilla heart; *FAM185A* (ENSG00000222011), a gene with unknown function evolved significantly higher expression in orangutan brain; *A4GALT* (ENSG00000128274; involved in glycolipid metabolism) and *ALX4* (ENSG00000052850; previously implicated in human cranial phenotypes) shifted to lower expression in great ape cerebellum; *SPO11* (ENSG00000054796; key for double strand formation during meiotic recombination) evolved high expression in mouse testes; *ZNF512* (ENSG00000243943; zinc finger protein 512, a potential transcription factor) shifted to high expression in opossum brain; and *ATRX* (ENSG00000085224; various molecular functions associated with e.g. chromosome segregation; involved in mental disorders) became downregulated in therian testis. Information about all genes was derived from the OMIM database (http://www.ncbi.nlm.nih.gov/omim).



Supplementary Figure 9. Genes with expression level shifts (multiple tissues). Examples of genes that evolved new optimal expression levels in multiple therian tissues are shown (details regarding these genes and all other statistically significant expression changes are provided in Supplementary Tables 9-24). *MAP3K7IP3* (an X-linked gene; ENSG00000157625) evolved significantly lower expression in therian heart and kidney and tends to have lower expression in eutherian/therian liver, testis, brain and cerebellum. *GNAI3* (a Gi alpha-like protein; ENSG00000065135) became significantly downregulated in the brain and cerebellum and tends to show reduced expression in the other therian tissues as well.