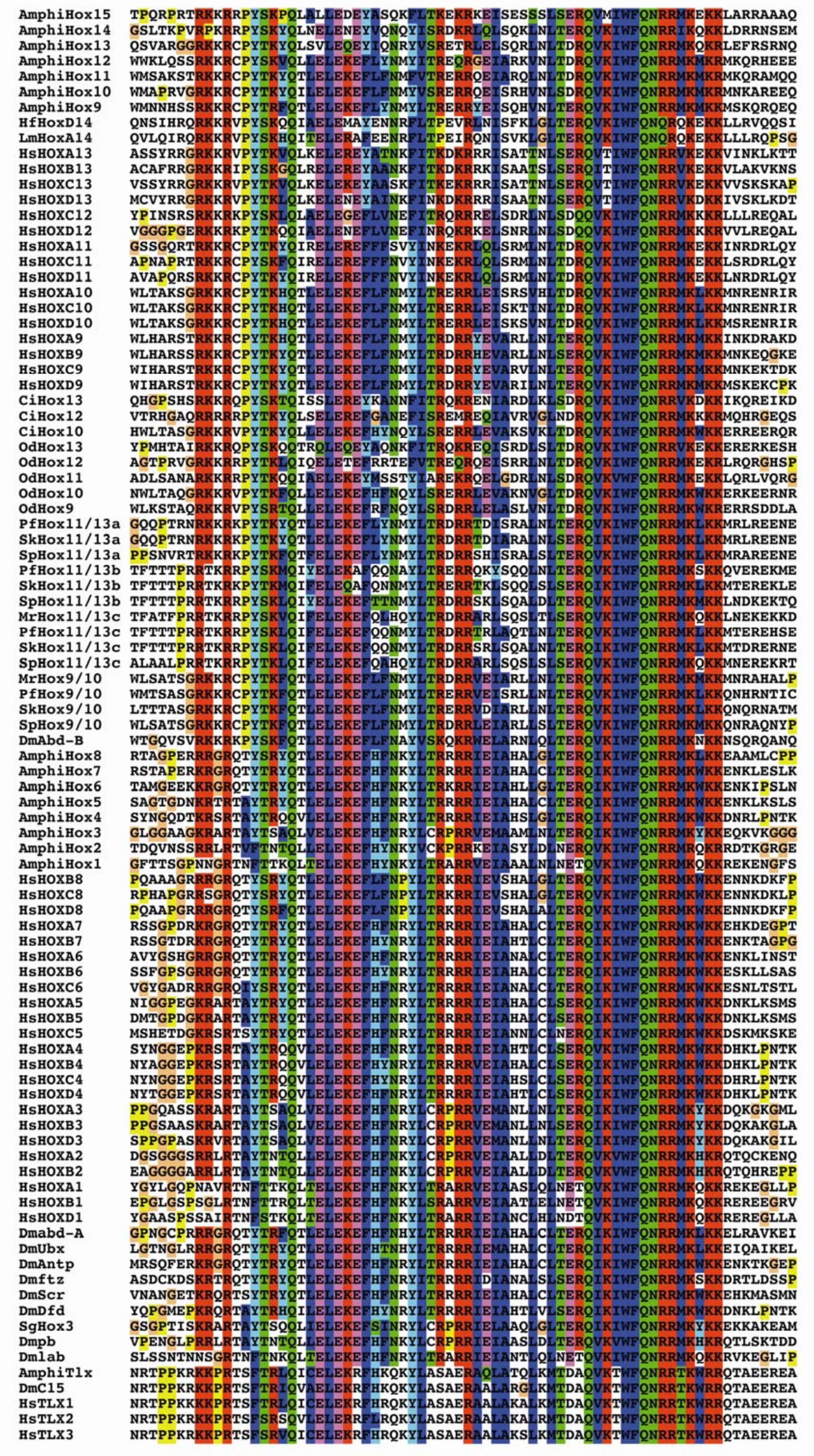
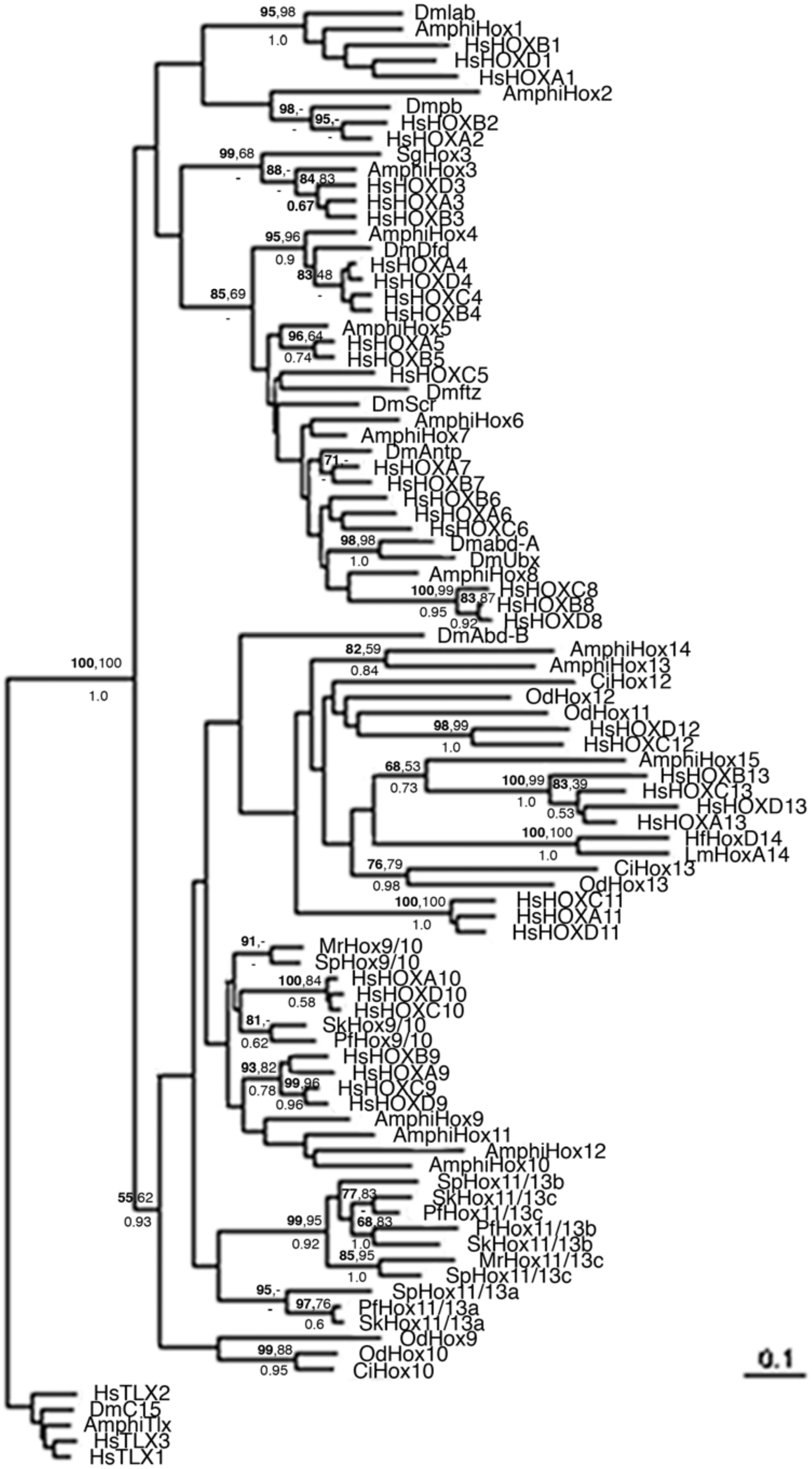
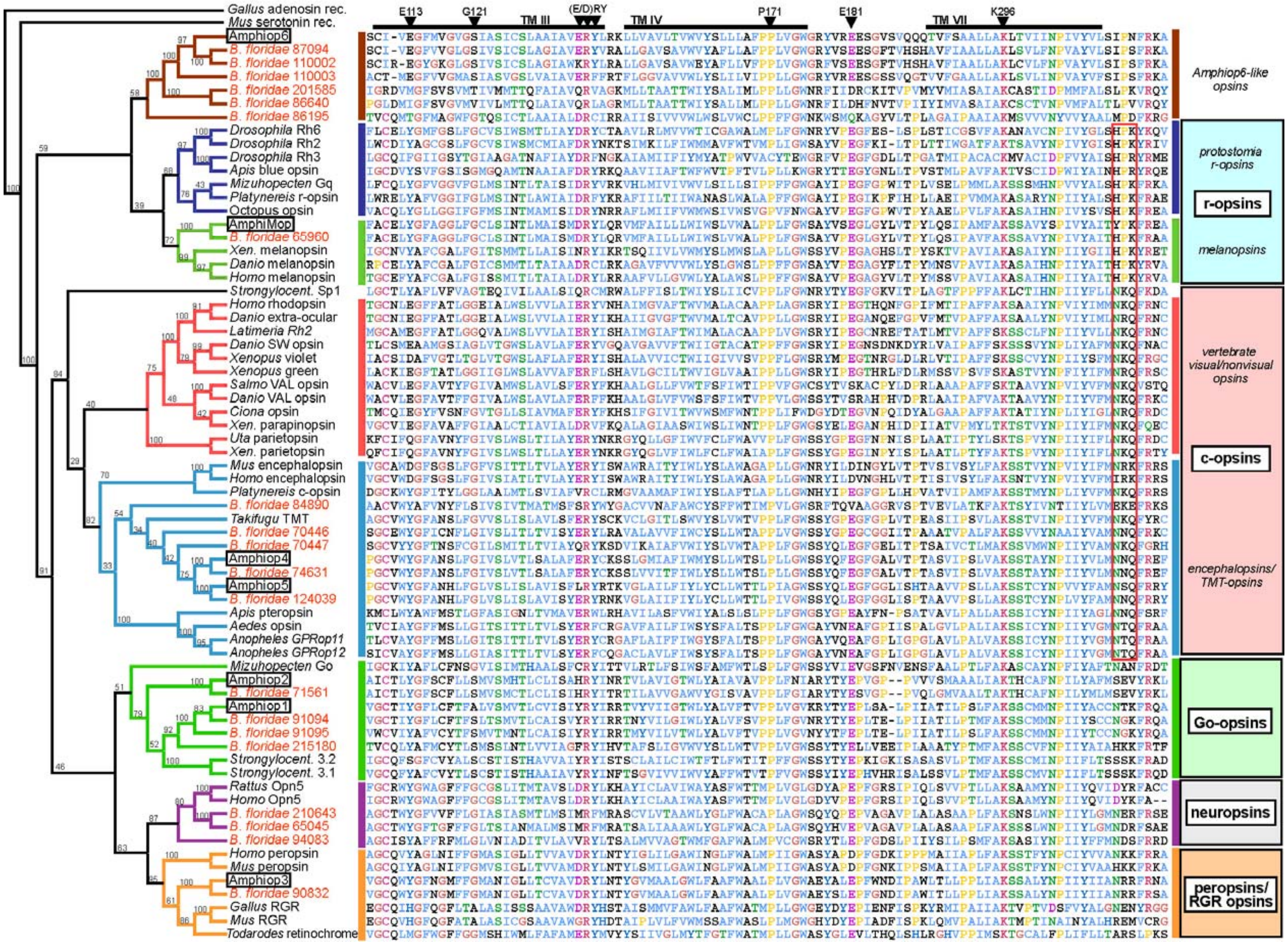


```
AmphiHox15   TPQRPRTRKKRRPYSKPQLALLEDEYASQKFLTKEKRKEISESSSLSERQVMIWFQNRRMKEKKLARRAAAQ
AmphiHox14   GSLTKPVRPKRRPYSKYQLNELENEYVQNQYISRDKRLQLSQKLNLTERQVKIWFQNRRIKQKKLDRRNSEM
AmphiHox13   QSVARGGRKKRCPYTKYQLSVLEQEYIQNRYVSRETRLELSQRLNLTDRQVKIWFQNRRMKQKRLEFRSRNQ
AmphiHox12   WWKLQSSRKKRCPYSKVQLLELEKEFLYNMYITREQRGEIARKVNLTDRQVKIWFQNRRMKMKRMKQRHEEE
AmphiHox11   WMSAKSTRKKRCPYTKYQTLELEKEFLFNMFVTRERRQEIARQLNLTDRQVKIWFQNRRMKMKRMKQRAMQQ
AmphiHox10   WMAPRVGRKKRCPYTKYQILELEKEFLFNMYVSRERRQEISRHVNLSDRQVKIWFQNRRMKMKRMNKAREEQ
AmphiHox9    WMNNHSSRKKRCPYTRFQTLELEKEFLYNMYLTRERRYEISQHVNLTERQVKIWFQNRRMKMKKMSKQRQEQ
HfHoxD14     QNSIHRQRKKRVPYSKQQIAELEMAYENNRFLTPEVRLNISFKLGLTERQVKIWFQNQRQKEKKLLRVQQSI
LmHoxA14     QVLQIRQRKKRVPYSKHQITELERAFEENRFLTPEIRQNISVKLGLTERQVKIWFQNQRQKEKKLLLRQPSG
HsHOXA13     ASSYRRGRKKRVPYTKVQLKELEREYATNKFITKDKRRRISATTNLSERQVTIWFQNRRVKEKKVINKLKTT
HsHOXB13     ACAFRRGRKKRIPYSKGQLRELEREYAANKFITKDKRRKISAATSLSERQITIWFQNRRVKEKKVLAKVKNS
HsHOXC13     VSSYRRGRKKRVPYTKVQLKELEKEYAASKFITKEKRRRISATTNLSERQVTIWFQNRRVKEKKVVSKSKAP
HsHOXD13     MCVYRRGRKKRVPYTKLQLKELENEYAINKFINKDKRRRISAATNLSERQVTIWFQNRRVKDKKIVSKLKDT
HsHOXC12     YPINSRSRKKRKPYSKLQLAELEGEFLVNEFITRQRRRELSDRLNLSDQQVKIWFQNRRMKKKRLLLREQAL
HsHOXD12     VGGGPGERKKRKPYTKQQIAELENEFLVNEFINRQKRKELSNRLNLSDQQVKIWFQNRRMKKKRVVLREQAL
HsHOXA11     GSSGQRTRKKRCPYTKYQIRELEREFFFSVYINKEKRLQLSRMLNLTDRQVKIWFQNRRMKEKKINRDRLQY
HsHOXC11     APNAPRTRKKRCPYSKFQIRELEREFFFNVYINKEKRLQLSRMLNLTDRQVKIWFQNRRMKEKKLSRDRLQY
HsHOXD11     AVAPQRSRKKRCPYTKYQIRELEREFFFNVYINKEKRLQLSRMLNLTDRQVKIWFQNRRMKEKKLNRDRLQY
HsHOXA10     WLTAKSGRKKRCPYTKHQTLELEKEFLFNMYLTRERRLEISRSVHLTDRQVKIWFQNRRMKLKKMNRENRIR
HsHOXC10     WLTAKSGRKKRCPYTKHQTLELEKEFLFNMYLTRERRLEISKTINLTDRQVKIWFQNRRMKLKKMNRENRIR
HsHOXD10     WLTAKSGRKKRCPYTKHQTLELEKEFLFNMYLTRERRLEISKSVNLTDRQVKIWFQNRRMKLKKMSRENRIR
HsHOXA9      WLHARSTRKKRCPYTKHQTLELEKEFLFNMYLTRDRRYEVARLLNLTERQVKIWFQNRRMKMKKINKDRAKD
HsHOXB9      WLHARSSRKKRCPYTKYQTLELEKEFLFNMYLTRDRRHEVARLLNLSERQVKIWFQNRRMKMKKMNKEQGKE
HsHOXC9      WIHARSTRKKRCPYTKYQTLELEKEFLFNMYLTRDRRYEVARVLNLTERQVKIWFQNRRMKMKKMNKEKTDK
HsHOXD9      WIHARSTRKKRCPYTKYQTLELEKEFLFNMYLTRDRRYEVARILNLTERQVKIWFQNRRMKMKKMSKEKCPK
CiHox13      QHGPSHSRKKRQPYSKTQISSLEREYKANNFITRQKRENIARDLKLSDRQVKIWFQNRRVKDKKIKQREIKD
CiHox12      VTRHGAQRRRRRPYTKYQLSELEREFGANEFISREMREQIAVRVGLNDRQVKIWFQNRRMKKKRMQHRGEQS
CiHox10      HWLTASGRKKRVPYTKYQLLELEKEFHYNQYLSRERRLEVAKSVKLTDRQVKIWFQNRRMKWKKERREERQR
OdHox13      YPMHTAIRKKRQPYSKQQTRQLEQEYAQNKFITRQKREQISRDLSLTDRQVKIWFQNRRVKEKKERERKESH
OdHox12      AGTPRVGRKKRRPYTKLQIQELETEFRRTEFVTREQRQEISRRLNLTDRQVKIWFQNRRMKEKRLRQRGHSP
OdHox11      ADLSANARKKRRPYTKQQIAELEKEYMSSTYIAREKRQELGDRLNLSDRQVKVWFQNRRMKEKKLQRLVQRG
OdHox10      NWLTAQGRKKRVPYTKFQLLELEKEFHFNQYLSRERRLEVAKNVGLTDRQVKIWFQNRRMKWKKERKEERNR
OdHox9       WLKSTAQRKKRVPYSRTQLLELEKEFRFNQYLSRDRRLELASLVNLTDRQVKIWFQNRRMKWKRERRSDDLA
PfHox11/13a  GQQPTRNRKKRKPYTKYQTLELEKEFLYNMYLTRDRRTDISRALNLTERQVKIWFQNRRMKLKKMRLREENE
SkHox11/13a  GQQPTRNRKKRKPYTKYQTLELEKEFLYNMYLTRDRRTDIARALNLSERQIKIWFQNRRMKLKKMRLREENE
SpHox11/13a  PPSNVRTRKKRKPYTKFQTFELEKEFLYNMYLTRDRRSHISRALSLTERQVKIWFQNRRMKLKKMRAREENE
PfHox11/13b  TFTTTPRRTKRRPYSKMQIYELEKAFQQNAYLTRERRQKYSQQLNLTERQVKIWFQNRRMKSKKQVEREKME
SkHox11/13b  TFTTTPRRTKRKPYTKMQIFELEQAFQNNMYLTRERRTKLSQQLSLSERQIKIWFQNRRMKLKKMTEREKLE
SpHox11/13b  TFTTTPRRTKRRPYSKLQIYELEKEFTTNMYLTRDRRSKLSQALDLTERQVKIWFQNRRMKMKKLNDKEKTQ
MrHox11/13c  TFATFPRRTKRRPYSKVQIFELEKEFQLHQYLTRDRRARLSQSLTLSERQVKIWFQNRRMKQKKLNEKEKKD
PfHox11/13c  TFTTTPRRTKRRPYSKLQIFELEKEFQQNMYLTRDRRTRLAQTLNLTERQVKIWFQNRRMKLKKMTEREHSE
SkHox11/13c  TFTTTPRRTKRRPYSKLQIFELEKEFQQNMYLTRDRRSRLSQALNLTERQIKIWFQNRRMKLKKMTDRERNE
SpHox11/13c  ALAALPRRTKRRPYTKLQIFELEKEFQAHQYLTRDRRARLSQSLSLSERQVKIWFQNRRMKQKKMNEREKRT
MrHox9/10    WLSATSGRKKRCPYTKFQTLELEKEFLFNMYLTRDRRVEIARLLNLTERQVKIWFQNRRMKMKKMNRAHALP
PfHox9/10    WMTSASGRKKRCPYTKFQTLELEKEFLFNMYLTRERRVEISRLLNLTERQVKIWFQNRRMKLKKQNHRNTIC
SkHox9/10    LTTTASGRKKRCPYTKFQTLELEKEFLFNMYLTRERRVDIARLLNLTERQVKIWFQNRRMKLKKQNQRNATM
SpHox9/10    WLSATSGRKKRCPYTKFQTLELEKEFLFNMYLTRDRRLEIARLLSLTERQVKIWFQNRRMKMKKQNRAQNYP
DmAbd-B      WTGQVSVRKKRKPYSKFQTLELEKEFLFNAYVSKQKRWELARNLQLTERQVKIWFQNRRMKNKKNSQRQANQ
AmphiHox8    RTAGPERRRGRQTYSRYQTLELEKEFHFNKYLTRRRRIEIAHALGLTERQIKIWFQNRRMKLKKEAAMLCPP
AmphiHox7    RSTAPERKRGRQTYTRYQTLELEKEFHFNKYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKENKLESLK
AmphiHox6    TAMGEEKKRGRQTYTRYQTLELEKEFHFNKYLTRKRRIEIAHLLGLTERQIKIWFQNRRMKWKKENKIPSLN
AmphiHox5    SAGTGDNKRTRTAYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKENKLKSLS
AmphiHox4    SYNGQDTKRSRTAYTRQQVLELEKEFHFNRYLTRRRRIEIAHSLGLTERQIKIWFQNRRMKWKKDNRLPNTK
AmphiHox3    GLGGAAGKRARTAYTSAQLVELEKEFHFNRYLCRPRRVEMAAMLNLTERQIKIWFQNRRMKYKKEQKVKGGG
AmphiHox2    TDQVNSSRRLRTVFTNTQLLELEKEFHYNKYVCKPRRKEIASYLDLNERQVKIWFQNRRMRQKRRDTKGRGE
AmphiHox1    GFTTSGPNNGRTNFTTKQLTELEKEFHYNKYLTRARRVEIAAALNLNETQVKIWFQNRRMKQKKREKENGFS
HsHOXB8      PQAAAGRRRGRQTYSRYQTLELEKEFLFNPYLTRKRRIEVSHALGLTERQVKIWFQNRRMKWKKENNKDKFP
HsHOXC8      RPHAPGRRSGRQTYSRYQTLELEKEFLFNPYLTRKRRIEVSHALGLTERQVKIWFQNRRMKWKKENNKDKLP
HsHOXD8      PQAAPGRRRGRQTYSRFQTLELEKEFLFNPYLTRKRRIEVSHALALTERQVKIWFQNRRMKWKKENNKDKFP
HsHOXA7      RSSGPDRKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEHKDEGPT
HsHOXB7      RSSGTDRKRGRQTYTRYQTLELEKEFHYNRYLTRRRRIEIAHTLCLTERQIKIWFQNRRMKWKKENKTAGPG
HsHOXA6      AVYGSHGRRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIANALCLTERQIKIWFQNRRMKWKKENKLINST
HsHOXB6      SSFGPSGRRGRQTYTRYQTLELEKEFHYNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKESKLLSAS
HsHOXC6      VGYGADRRRGRQIYSRYQTLELEKEFHFNRYLTRRRRIEIANALCLTERQIKIWFQNRRMKWKKESNLTSTL
HsHOXA5      NIGGPEGKRARTAYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLSERQIKIWFQNRRMKWKKDNKLKSMS
HsHOXB5      DMTGPDGKRARTAYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLSERQIKIWFQNRRMKWKKDNKLKSMS
HsHOXC5      MSHETDGKRSRTSYTRYQTLELEKEFHFNRYLTRRRRIEIANNLCLNERQIKIWFQNRRMKWKKDSKMKSKE
HsHOXA4      SYNGGEPKRSRTAYTRQQVLELEKEFHFNRYLTRRRRIEIAHTLCLSERQVKIWFQNRRMKWKKDHKLPNTK
HsHOXB4      NYAGGEPKRSRTAYTRQQVLELEKEFHYNRYLTRRRRVEIAHALCLSERQIKIWFQNRRMKWKKDHKLPNTK
HsHOXC4      NYNGGEPKRSRTAYTRQQVLELEKEFHYNRYLTRRRRIEIAHSLCLSERQIKIWFQNRRMKWKKDHRLPNTK
HsHOXD4      NYTGGEPKRSRTAYTRQQVLELEKEFHFNRYLTRRRRIEIAHTLCLSERQIKIWFQNRRMKWKKDHKLPNTK
HsHOXA3      PPGQASSKRARTAYTSAQLVELEKEFHFNRYLCRPRRVEMANLLNLTERQIKIWFQNRRMKYKKDQKGKGML
HsHOXB3      PPGSAASKRARTAYTSAQLVELEKEFHFNRYLCRPRRVEMANLLNLSERQIKIWFQNRRMKYKKDQKAKGLA
HsHOXD3      SPPGPASKRVRTAYTSAQLVELEKEFHFNRYLCRPRRVEMANLLNLTERQIKIWFQNRRMKYKKDQKAKGIL
HsHOXA2      DGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALLDLTERQVKVWFQNRRMKHKRQTQCKENQ
HsHOXB2      EAGGGGARRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALLDLTERQVKVWFQNRRMKHKRQTQHREPP
HsHOXA1      YGYLGQPNAVRTNFTTKQLTELEKEFHFNKYLTRARRVEIAASLQLNETQVKIWFQNRRMKQKKREKEGLLP
HsHOXB1      EPGLGSPSGLRTNFTTRQLTELEKEFHFNKYLSRARRVEIAATLELNETQVKIWFQNRRMKQKKREREEGRV
HsHOXD1      YGAASPSSAIRTNFSTKQLTELEKEFHFNKYLTRARRIEIANCLHLNDTQVKIWFQNRRMKQKKREREGLLA
Dmabd-A      GPNGCPRRRGRQTYTRFQTLELEKEFHFNHYLTRRRRIEIAHALCLTERQIKIWFQNRRMKLKKELRAVKEI
DmUbx        LGTNGLRRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLKKEIQAIKEL
DmAntp       MRSQFERKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKENKTKGEP
Dmftz        ASDCKDSKRTRQTYTRYQTLELEKEFHFNRYITRRRRIDIANALSLSERQIKIWFQNRRMKSKKDRTLDSSP
DmScr        VNANGETKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEHKMASMN
DmDfd        YQPGMEPKRQRTAYTRHQILELEKEFHYNRYLTRRRRIEIAHTLVLSERQIKIWFQNRRMKWKKDNKLPNTK
SgHox3       GSGPTISKRARTAYTSQQLIELEKEFSINRYLCRPRRIELAAQLGLTERQIKIWFQNRRMKYKKEKKAKEAM
Dmpb         VPENGLPRRLRTAYTNTQLLELEKEFHFNKYLCRPRRIEIAASLDLTERQVKVWFQNRRMKHKRQTLSKTDD
Dmlab        SLSSNTNNSGRTNFTNKQLTELEKEFHFNRYLTRARRIEIANTLQLNETQVKIWFQNRRMKQKKRVKEGLIP
AmphiTlx     NRTPPKRKKPRTSFTRLQICELEKRFHKQKYLASAERAQLATQLKMTDAQVKTWFQNRRTKWRRQTAEEREA
DmC15        NRTPPKRKKPRTSFTRIQVAELEKRFHKQKYLASAERAALARGLKMTDAQVKTWFQNRRTKWRRQTAEEREA
HsTLX1       NRTPPKKKKPRTSFTRLQICELEKRFHRQKYLASAERAALAKALKMTDAQVKTWFQNRRTKWRRQTAEEREA
HsTLX2       NRTPPKRKKPRTSFSRSQVLELERRFLRQKYLASAERAALAKALRMTDAQVKTWFQNRRTKWRRQTAEEREA
HsTLX3       NRTPPKRKKPRTSFSRVQICELEKRFHRQKYLASAERAALAKSLKMTDAQVKTWFQNRRTKWRRQTAEEREA
```

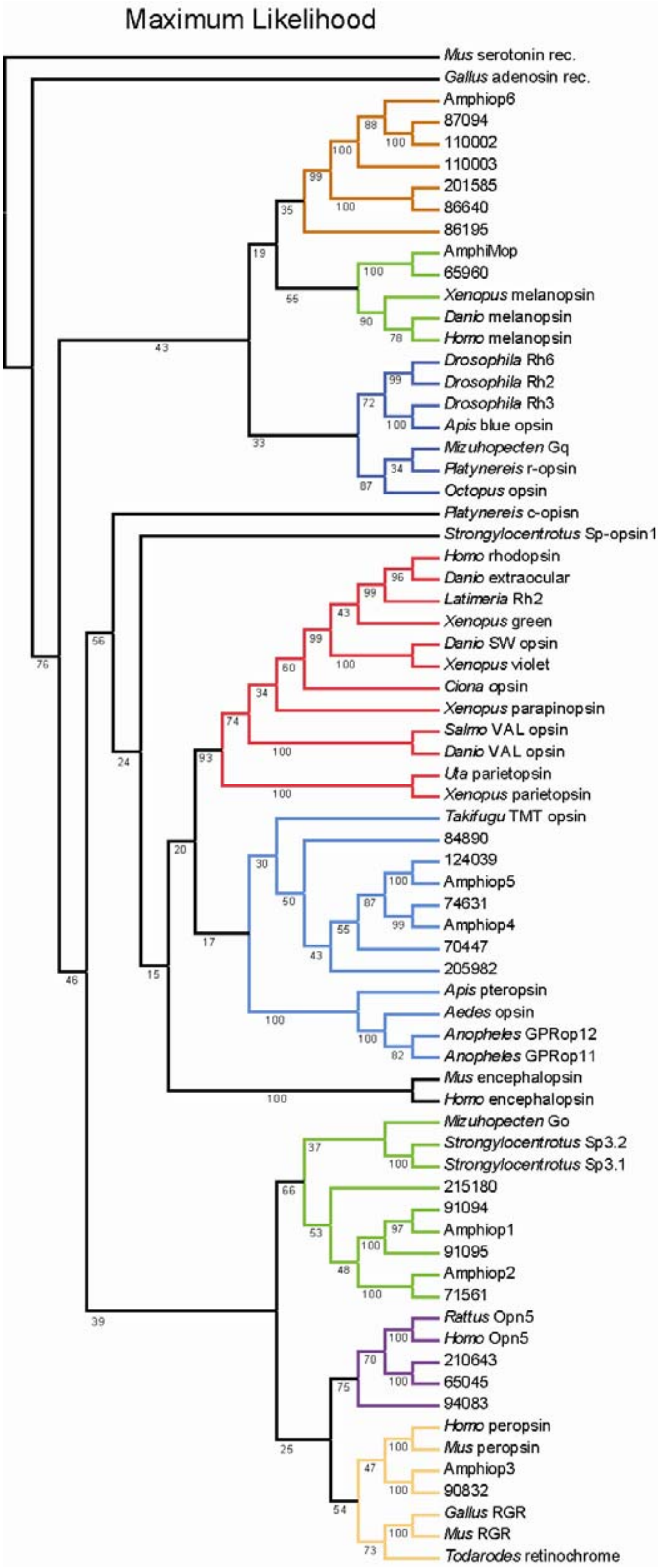
**Supplemental Figure S1**. Alignment of AmphiHox amino acid sequences to Hox proteins of other selected metazoans, plus Tlx sequences as phylogenetic outgroups. The alignment includes the entire homeodomain plus 6 further amino acids at either end. AmphiHox15 has a Posterior Hox type of sequence, but it is not strikingly more similar to any particular Posterior Hox protein than any other.



**Supplemental Figure S2**. Phylogenetic tree of AmphiHox and selected deuterostome Hox proteins. There is no significant statistical support for assignment of particular Posterior AmphiHox proteins (AmphiHox9-15) to particular vertebrate paralogy groups. Also, there is no significant statistical support for either wholescale independent duplications of ambulacrarian (Echinodermata + Hemichordata) versus chordate Posterior genes (Peterson, 2004; Cameron et al., 2006), nor for clear orthology relationships of each Posterior gene across deuterostome phyla and sub-phyla (Deuterostome Posterior Flexibility; Ferrier et al., 2000). We suggest these results reflect a mixture of independent duplications on certain lineages alongside rapid rates of evolution that obscure orthology relationships amongst deuterostome Posterior Hox genes. The tree backbone shown is derived by Neighbour Joining (NJ), inferred using the alignment in Fig. S1 (homeodomain plus 6 amino acid flanks); the topology is largely consistent with those inferred by maximum likelihood (ML) and Bayesian methods. Values at the nodes are given where bootstrap support is 70% or above, or posterior probability is 0.95 or higher; all values are shown at the basal Posterior Hox node and around the AmphiHox15 node. NJ support values are shown in bold above the node, and to their right ML support values are shown in plain text; Bayesian support values are shown below the node. The NJ tree was inferred from a JTT distance matrix; ML and Bayesian trees assumed the RtRev model of protein evolution. ML analysis was run with 4 gamma categories of rate variation among sites. The Bayesian MCMC was run for 1,250,000 generations and sampled every 100 generations after discarding trees from the burn-in phase.



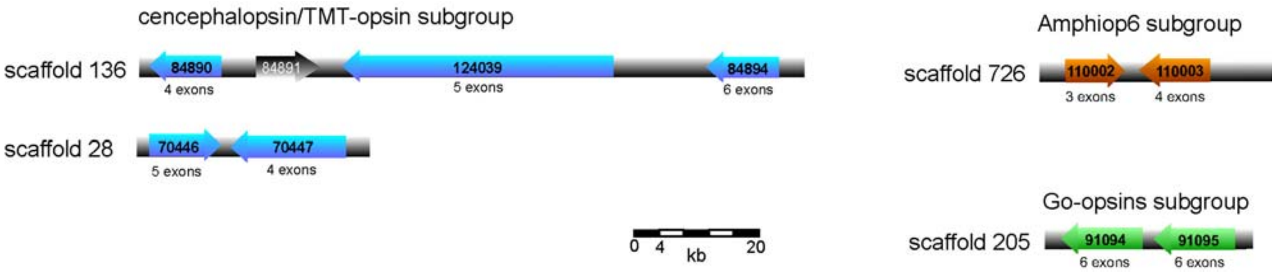
Supplemental Figure S3A.



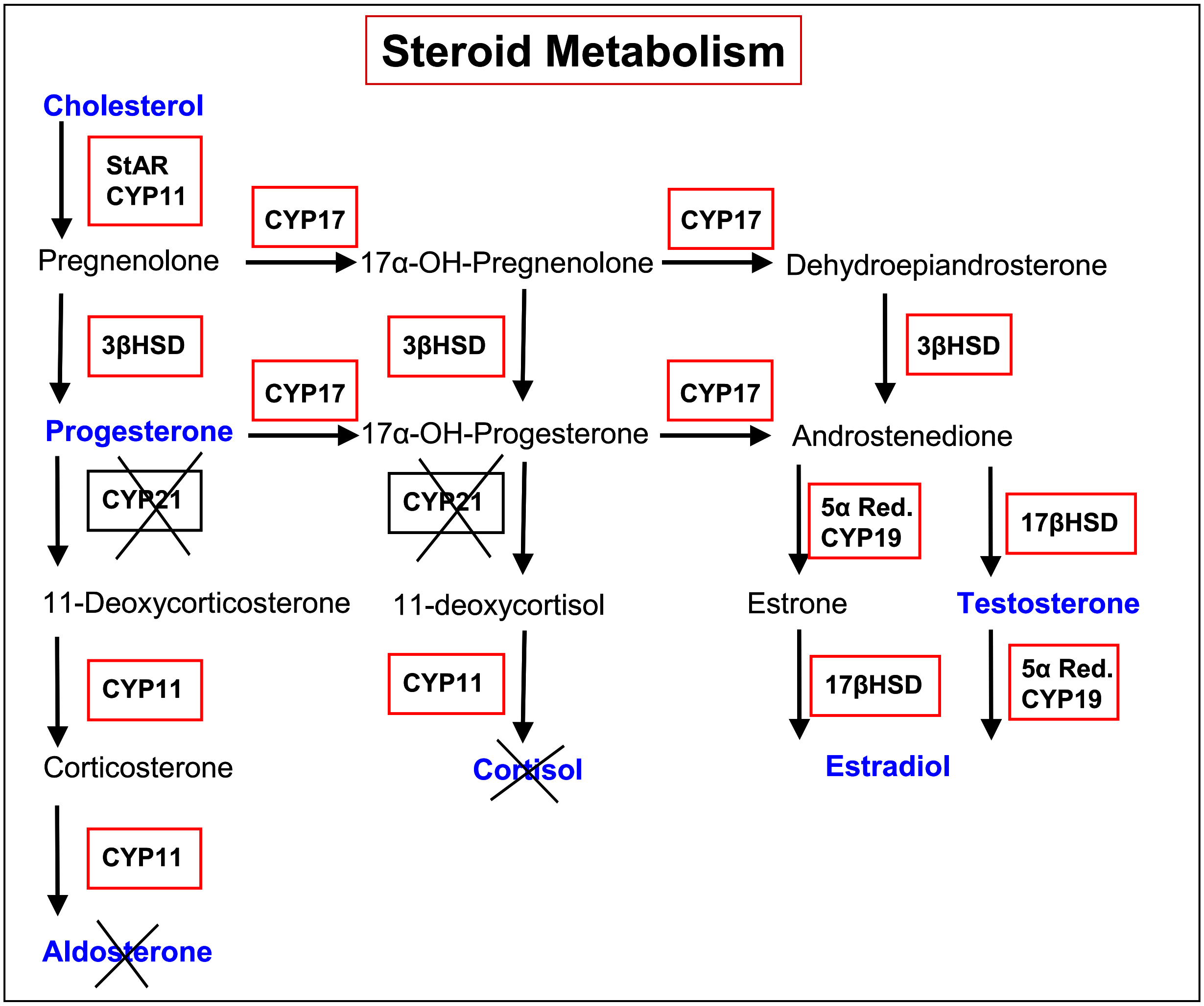
**Supplemental Figure S3B.**

**Supplementary Figure S3.** Phylogenetic tree and alignments of the opsin family.The phylogenetic tree was inferred by the neighbor-joining method using murine adenosine and chicken serotonine receptors as outgroup sequences. The numbers above each node represent the percentage of bootstrap probability based on 500 replicates. Putative *B. floridae* opsins clustered consistently in tree inferred by the Maximum likelihood method. Previously published *B. belcheri* opsins are black-framed and opsins identified in this work are typed in red and numbered according to the *B. foridae* v1.0 gene model ID from which the protein sequence was derived. Multiple sequence alignment of the opsins represented in the NJ tree: The black lines above the alignment represent the extent of transmembrane helices III, IV and VII. The color lines on both sides of the alignment mark out opsin subfamilies. Amino acid fingerprint allowing opsin classification is indicated by black arrowhead and numbered according to bovine rhodopsin protein sequence. The lysine residue K296 is critical for covalent binding of retinal via Schiff base linkage, which is stabilized by counterion E113 or E181 (Terakita 2004). The (E/D)RY triade highly conserved amongst GPCRs is important for G-protein interacition (Franke 2002). The residues G121, P171, W175 were found to be evolutionary trace residues typical for opsin family but not for GPCRs in general (Madabushi et al. 2004). The position of “HPK” and “NKQ” motif shared among rhabdomeric and ciliary opsins, respectively (Arendt, 2004) are red-framed. The classification of opsin families is represented in the color boxes on the right side.

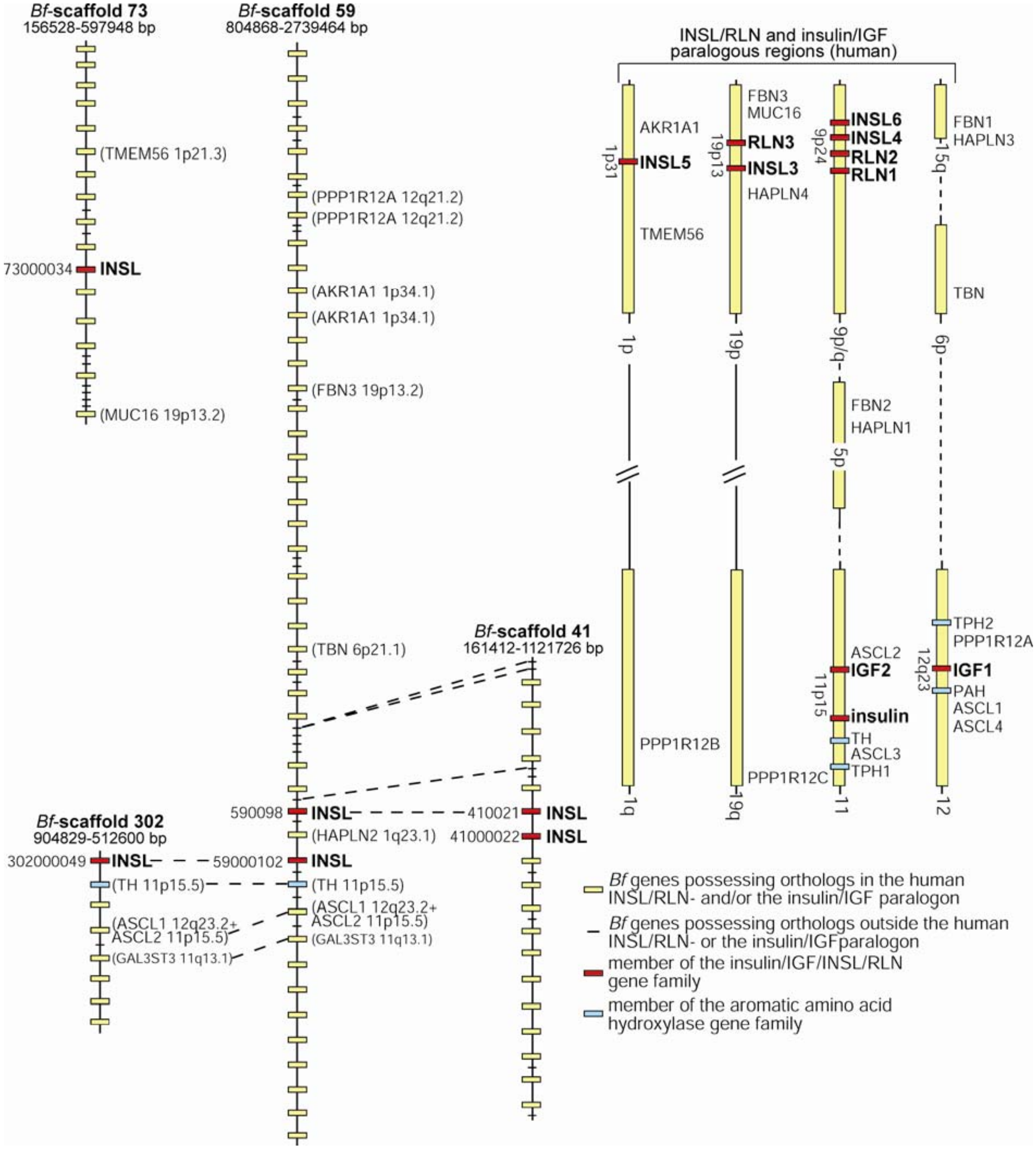
B. Putative *B. floridae* opsins clustered consistently in tree inferred by the Maximum likelihood method. Numbered sequences are from *B. floridae* as in A.



**Supplementary Figure S4**: Gene organization of putative opsins from c-opsin, the Amphiop6 group and the Go-opsin groups, suggesting their origin by tandem gene duplication.



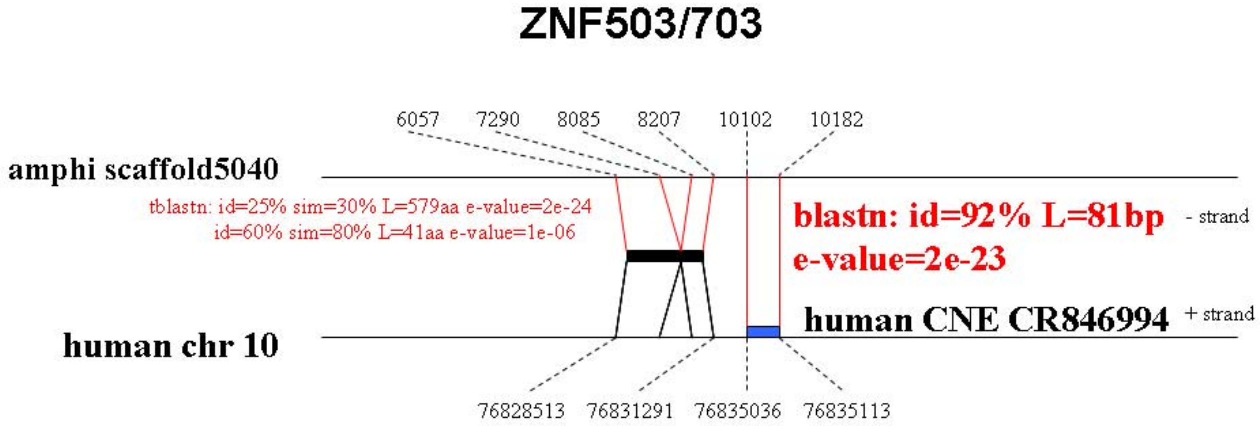
**Supplemental Figure S5**. Summary of the amphioxus gene complement involved in steroid metabolism. The existence of amphioxus homologs of a given enzyme is boxed with a red line. Genes that could not be identified in amphioxus genome are boxed in black and have a cross through them. The main steroids synthesized by this biosynthetic pathway are highlighted in blue. The lack of a CYP21 homolog suggests that amphioxus lacks the ability to synthesize aldosterone and cortisol.



## Supplemental Figure S6.

## Conserved synteny between the chromosomal regions hosting amphioxus insulin/IGF/relaxin-type homologues and chromosomal regions hosting insulin-relaxin gene family in the human genome.

Schematic organization of four *Branchiostoma floridae* (*Bf*) scaffolds hosting insulin-like (INSL) genes; bars in red show the last digits of *Branchiostoma floridae* gene model ID from assembly v1.0). Each flanking gene in *B. floridae* that possesses orthologues within the insulin-like/relaxin (INSL/RLN) and/or insulin/IGF paralogons of the human genome is depicted in yellow, whereas genes with orthologue(s) elsewhere in the human genome are indicated by a dash. Sequences that do not possess clear human orthologs (i.e. 37200005, 50000083, 674000009) are not displayed. The last two of these (50000083 and 674000009) are likely alleles, making, together with the four genes depicted in this figure, a total of six INSL/RLN genes in amphioxus, Members of aromatic amino acid hydroxylases, linked with insulin/IGF genes in *Ciona* and vertebrate genomes, are shown in blue. Selected *B. floridae* genes with orthologues in the human paralogous regions are indicated. Where a gene possesses two human co-orthologs, only those located within the human paralogons are shown. Based on shared gene content, scaffolds 302, 59 and 41 are probably linked and derive from two *B. floridae* haplotypes. Dashed lines link potential alleles. Further details of the human INSL/RLN and insulin/IGF paralogons are in Olinski *et al*., 2006. The complete list of all human orthologue(s) of each *B. floridae* gene model can be found in Supplementary Table S5.



**Supplemental Figure S7.** Conserved non-coding elements adjacent the ZNF503/703 genes in amphioxus and human.

Blast searches of the amphioxus genome with the 1373 non-coding sequences conserved among vertebrate genomes (Woolfe et al. 2005) revealed 293 with matches in the amphioxus genome (e-value <0.1). Of these, CR846994 has an e-value of $2e^{-23}$ and CR84671 with an e-value of $e^{-16}$. The best hit, CR846994, is close to the human paralogues ZNF503 and ZNF703, and the single amphioxus ZNF503/703 gene. In contrast, a similar search of the Ciona intestinalis genome gave only 70 hits with an e-value <0.1, with the best one, CR846545, having a score of $e^{-4}$. Black bar = coding region.