東海大學生命科學系
碩士論文

指導教授:陳仁祥 博士

Ren-Shiang Chen, Ph.D.

動物鈣離子通道γ次單元之演化分析
Evolutionary analysis of calcium channel γ subunits in animals

研究生:賴彥明

Yen-Ming Lai

中華民國 104 年 7 月 28 日

東海大學生命科學系碩士論文

動物鈣離子通道γ次單元之演化分析

Evolutionary analysis of calcium channel γ subunits in animals

研究生:賴彥明

Yen-Ming Lai

指導教授:陳仁祥 博士

Ren-Shiang Chen, Ph.D.

中華民國104 年 7月28日

東海大學生命科學系
碩士論文學位考試審定書

生命科學系碩士班研究生 賴彥明 君所撰寫之論文

（中文）

動物鈣離子通道γ次單元之演化分析

（英文）

Evolutionary analysis of calcium channel γ subunits in animals

經本委員會審定通過，特此證明。

學位考試委員會　　召集人 劉少倫（簽名）

委　員 廖培鈞

陳仁祥

中華民國 104 年 7 月 14 日

致謝

嗯...哈哈!終於完成這篇論文啦，Thank god!對於我原本是一個生物資訊的門外漢，直到完成這篇題目。首先最要感謝的是 陳仁祥 老師，因為他的引領帶我進入生物資訊，讓我有機會可以對生資這個領域有更多的認識。陳老師是一位非常特別的老師，就我從學生的角度，不能僅用科學家來描述老師，而他不僅是一位神經學家，對離子通道有深入的探究，同時也是教育學家，對教學的態度與熱忱是讓學生感動的。另要感謝的老師是 劉少倫 老師，這篇論文許多生資方法技術都是倚賴劉老師的幫助完成，說劉老師是我的第二指導老師根本也不為過，非常感謝劉老師花不少的時間與耐心指導。再者是 廖培鈞 老師也是這篇論文要感謝的人，感謝老師對論文的指導，對於任何建議評價，我認為都是非常可貴的，也將是研究將來努力的方向。當然，兩年一路上在實驗室的夥伴也是我要感謝的人，欣璇學姊、聖威、亦騏、乙鳴、松翰、芝禮、宜君及映柔等學弟妹的幫助與勉勵。最後，我要感謝我的家人，你們的叮嚀與鼓勵，是我人生快樂的泉源，對了!還有我的跑鞋，也是要感謝的對象，謝謝它們帶領我每周末的city run，讓我有能量可以大步邁向下一個旅程。

Table of Contents

中文摘要

電壓敏感型鈣離子通道由 α_1、$\alpha_2\delta$、β 與 γ 次單元組成，α_1 是構成孔道的次單元，$\alpha_2\delta$、β 與 γ 則是輔助用的次單元。$\alpha_2\delta$ 與 β 次單元功能上為正向調控因子，會幫助 α_1運送到細胞膜上並加強通道的活化。相反的，實驗證據顯示 γ 次單元功能上分歧不一。在十個 γ 次單元中，γ_1與 γ_6屬於通道的負向調控因子，抑制肌肉細胞的鈣離子通道電流。另一方面，γ_2、γ_3、γ_4、γ_5、γ_7 及 γ_8 被認為是 transmembrane AMPA receptor regulatory proteins (TARPs)。至於 TMEM114 與 TMEM235(近來被發現的 γ 家族成員)，其生理功能仍了解甚少。為何這十個 γ 基因是如此相似但他們生理功能卻有相當大的分歧?由於現有的實驗證據無法提供充足的線索，因此我們轉向利用生物資訊方法進行分析，試圖推敲出 γ 基因家族的演化歷程。藉由在 26 個物種進行 protein-protein BLAST，發現這群 calcium channel γ (CACNG)基因在八目鰻與硬骨魚之間，數量倍增而演化出現存的十個 CACNG 分子，也各自獨立分成 4 個單系群。在脊椎動物演化過程，γ 基因在不同的物種中獨立地消失與複製。雖然我們親緣關係樹分析基本上與先前結果一致，但是無脊椎動物序列之存在顯示脊椎與無脊椎動物的 γ 基因擁有共同祖先，最早可以追溯到兩側對稱動物。有趣的是，在染色體

地圖(chromosome map)中，PKC 跟 γ 基因緊鄰在一起，意味著這群蛋白可能跟鈣離子的濃度恆定或蛋白質磷酸化有關。TMEM114 及 TMEM235 也緊鄰 GRIN 基因及 CACNA 基因，意指 TMEM114 及 TMEM235 可能與 GRIN 及 CACNA 有功能上的交互作用。藉由搜尋在 CACNG 基因附近的同源染色體片段，我們修正了先前提出的 CACNG 基因演化途徑。在演化速率分析中，(γ_4, γ_8)及(TMEM114,TMEM235)在第二次染色體複製後，展現出顯著的氨基酸非同義置換，暗示他們可能在動物中演化出其分歧功能。此外，在硬骨魚中多出一套的 γ(γ_1, γ_2, γ_3, γ_5及 γ_7)，顯示他們也可能獲得新的功能。在此一研究中，我們希望可以洞悉 γ 基因的演化歷史，解釋現今動物 γ 功能的差異性，進而提供我們未來以實驗方法驗證 γ 基因功能時實驗設計上的洞見。

English Abstract

Voltage-dependent calcium channels (VDCCs) are comprised of pore-forming α_1 subunits as well as three other auxiliary subunits: $\alpha_2\delta$, β and γ subunits. The $\alpha_2\delta$ and β subunits are positive regulators of VDCCs that enhance membrane insertion of the α_1 subunits and channel activation. In contrast, the functions of the γ subunits are not completely established, because experimental data have suggested functional diversity. Out of the ten members of the γ subunits, γ_1 and γ_6 are negative regulators of VDCCs that inhibit calcium current in muscle cells. In contrast, the γ_2, γ_3, γ_4, γ_5, γ_7 and γ_8 are known as the transmembrane AMPA receptor regulatory proteins (TARPs). As for TMEM114 and TMEM235 (two new members of the family), their physiological functions remain largely unknown. While these ten γ genes are the closest homologs within mammalian genomes, why are their functions so diverse? Because experimental paradigms have not provided enough clues, we turn to bioinformatical analysis for evolutionary insight. By conducting protein-protein BLAST

between twenty-six animal species, we found that several γ genes emerged by gene and chromosome duplications between hyperoartia and osteichthyes, evolving into the ten currently known γ genes that are clustered into four monophyletic groups in vertebrate. In vertebrate lineages, γ genes were independently lost and duplicated. Although our phylogenetic analysis is consistent with previous results, the invertebrate sequences demonstrate that vertebrate and invertebrate γ's share common ancestors in as far back as the bilaterians. Interestingly, γ genes are almost always associated with PKC genes on chromosome, suggesting that the functions of γ proteins are related to the homeostasis of calcium or protein phosphorylation. TMEM114 and TMEM235 genes are closely located with GRIN and CACNA1 genes on chromosome, implying that TMEM114 and TMEM235 may functionally interact with GRIN and CACNA1. By searching the paralogous chromosome segments around CACNG genes, we revised the evolution pathways that was previously proposed. In evolutionary rate analysis, (γ_4, γ_8) and (TMEM114, TMEM235) exhibited significantly nonsynonymous substitution after the 2nd round of chromosome duplication, implying that their functions have diverged in the animal lineage. In addition, the additional copies of γ (γ_1, γ_2, γ_3, γ_5 and γ_7) in osteichthyes may have acquired novel functions. By elucidating the historical events that produced these ten γ genes, we hope to contribute to the explanation of the functional diversity of calcium channel γ subunits and to provide insight for the experimental design of functional verification of the ten γ proteins in the future.

Introduction

Voltage-dependent calcium channels

The voltage-dependent calcium channels (Ca^{2+} channels) were first discovered by Paul Fatt and Bernard Katz in crustacean when they found that muscle was still excitable when extracellular sodium is present calcium presentence (Fatt and Katz, 1953). During the 1960s, Albrecht Fleckenstein identified nifedipine as a Ca^{2+} channel antagonist, which is based on dihydropyridine (DHP) molecules (Dolphin, 2006). After DHPs were prevalently applied in Ca^{2+}channels researches, the DHPs created a new era of cloning and purifying Ca^{2+} channels. Meanwhile, pharmacological categories and current types of Ca^{2+} channels were clearly defined based on the response to various toxins (Dolphin, 2006).

First voltage-clamp recording of two types Ca^{2+} channels were demonstrated in starfish (Hagiwara et al., 1975). About a decade later, two different currents components were designated high-voltage activated (HVA) and low-voltage activated (LVA) Ca^{2+} currents in mammalian sensory neurons (Carbone and Lux, 1984; Fedulova et al., 1985). Furthermore, pharmacological test showed that certain HVA channels in skeleton muscle, smooth muscle, heart and neurons are sensitive to DHPs and they are called L-type Ca^{2+}channels (Hess et al., 1984). L-type Ca^{2+} channel has slow voltage-dependent inactivation and long-lasting activation (Tsien et al., 1988). In contrast to the biophysical property of HVA Ca^{2+} channels, LVA channels, also called T-type Ca^{2+} channels, are activated at much more negative potentials, inactivated rapidly, deactivated slowly and having smaller conductance (Nowycky et al., 1985).

The other HVA channels insensitive to DHPs were isolated in dorsal root ganglion neurons by single channel and whole-cell recording (Nowycky et al., 1985). These novel Ca^{2+} channels termed N-type Ca^{2+} channels were blocked by peptide ω-conotoxin GVIA and related peptide toxins (Olivera et al., 1994; Tsien et al., 1988) and have intermediate activation and rate of inactivation (Nowycky et al., 1985). Specifically, their voltage of activation are more negative than L-type but more positive than T-type, and rate of in activation is faster than L-type but slower than T-type. After pharmacological characterization, three other types Ca^{2+}channels were unveiled in neuronal cells. P-type Ca^{2+} current was distinguished by its high affinity to spider toxin ω-agatoxin IVA in Purkinje neurons (Llinas and Yarom, 1981; Llinas et al., 1989; Mintz et al., 1992). Q-type Ca^{2+} current was recorded in cerebellar granule neurons (Randall and Tsien, 1995) and blocked by toxin ω-agatoxin IVA with low sensitivity. However, P-type and Q-type are combined as P/Q, probably due to splicing variants or association with different auxiliary subunits (Bourinet et al., 1999). The other Ca^{2+} current is R-types or Residual that is insensitive to most Ca^{2+} channel antagonist but the peptide SNX-482 derived from tarantula (Newcomb et al., 1998). The expression of L-type and T-type Ca^{2+} channels are distributed in various tissues but, P/Q-type and R-type Ca^{2+} channels are confined to nervous tissue.

Subunits of voltage-dependent calcium channels

The era of purification of Ca^{2+} channels was created with DHPs. It was thanks to DHPs that some Ca^{2+} channels are termed DHP receptors. William Catterall, whose laboratory was one of the key groups to contribute to these studies at that time, identified

Ca^{2+} channels as heteromeric proteins. The initial stoichiometry of purified Ca^{2+} channels from skeleton muscle was composed of five components: α_1 subunit (170kDa), α_2 subunit (150kDa), β subunit (52kDa), γ subunit (32kDa) and δ subunit (17-25kDa) (Takahashi et al., 1987). The α_1 subunit, the pore-forming structure, is bound to DHP and have four homologous transmembrane domains, in which each domain contains six segments. Overall, scientists have identified ten members of α_1 so far. The α_1 subunits can be classified into three groups: Ca_V 1.X, Ca_V 2.X and Ca_V 3.X. "Ca" refers to the ion permeant, Ca^{2+} ions and "$_V$" corresponds to its biophysical activator, voltage (Ertel et al., 2000). Four members of Ca_V 1.X and three members of Ca_V 2.X belong to LVA. Conversely, the other three Ca_V 3.X subunits are HVA.

Apart from α_1 subunits, the other subunits are auxiliary proteins. Initially, the position of α_2 on SDS-PAGE was close to α_1 in non-reducing condition because disulfide bonds bridge δ and α_2. Nowadays, the α_2 nomenclature is still used. Interestingly, α_2 and δ are encoded in the same gene. After the polypeptide is split into two mature forms by post-translational proteolysis, they are covalently linked by disulfide bond (De Jongh et al., 1990). The functions of $\alpha_2\delta$ subunits co-expressed with α_1 subunit have been analyzed by electrophysiological method in Xenopus oocytes. Although the $\alpha_2\delta$ subunit effect on α_1 subunit varied in heterologous expression system, the $\alpha_2\delta$ subunit is consider as a positive regulator on the α_1 subunit. It strengthens current densities, accelerates activation and inactivation kinetics and causes hyperpolarizing shift in voltage dependent of inactivation (Singer et al., 1991; Williams et al., 1992).

Not only is the $\alpha_2\delta$ subunit a positive regulator but also the β subunit enhances biophysical properties of the α_1 subunit. The β subunit increases current density of the α_1 subunit by raising open probability and enhancing the α_1 subunit trafficking to plasma membrane (Dolphin, 2003). In addition, it shifts the activation threshold to more negative voltage (Birnbaumer et al., 1998).

Functional diversity of the γ subunits

During the studies of Ca^{2+} channel auxiliary subunits, γ subunits, the smallest molecule among components, perplexed the scientists due to their functional diversity. γ_1 was the first biochemically identified γ member from skeleton muscle (Takahashi et al., 1987). The role of γ_1 was demonstrated as a negative regulator to the α_1 subunit. It inhibits the calcium current in native mouse skeleton myotubes, demonstrated by using a genetic knockout mouse (Arikkath et al., 2003; Freise et al., 2000; Held et al., 2002; Jay et al., 1990). The subsequently discovered γ members, γ_2, γ_3, γ_4, γ_5, γ_6, γ_7 and γ_8, thanks to the human genome project, were cloned from neuronal and muscular tissues (Arikkath and Campbell, 2003; Black, 2003; Flucher et al., 2005; Kang and Campbell, 2003). Because of their sequences homology to γ_1, these eight genes were assumed to have a common ancestral gene. Not only is γ_1 a negative regulator, γ_6, the subunit with highest similarity to γ_1, was also shown to suppress LVA calcium current in native cardiomyocytes and a heterologous expression system (Hansen et al., 2004; Lin et al., 2008). Although the other γ (γ_2, γ_3, γ_4, γ_5, γ_7, γ_8) subunits were associated with Ca^{2+} channels and have subtle influences on the biophysical properties calcium current (Kang et al., 2001; Klugbauer et al., 2000; Letts et al., 1998; Moss et al., 2002; Rousset et al., 2001), these γ subunits did not alter the calcium currents in native cells (Moss et al., 2002; Schnell et al., 2002). Unexpectedly, those γ was recognized as transmembrane AMPA receptor regulatory proteins (TARPs) (Chen et al., 1999; Hashimoto et al., 1999; Kato et al., 2008; Kato et al., 2007; Tomita et al., 2003). In both biochemical and electrophysiological systems, the results unambiguously illustrated that TARP γ (γ_2, γ_3, γ_4, γ_5, γ_7, γ_8) subunits have strong interaction with AMPA receptors, enhance AMPA expression to membrane though through trafficking and lead to current increase in neurons.

With the new terminology TARPs, a controversy was created. The research groups that work on TARPs almost abandoned the γ subunit nomenclature. In contrast, scientists who devoted themselves to γ_1 or γ_6 studies stick to their original notion. This controversy leads to the loss of interest to those subunits.

According to structural prediction, the γ subunits share a common architecture with four-transmembrane-domains proteins which are included in the pfam00822 superfamily. Especially, they possess a GLW motif in the first extracellular loop that is exactly the same to the claudin protein family (Chen et al., 2007). Moreover, phylogenetic analysis suggested that the evolution of γ subunits came from an ancestor gene on account of tandem duplication and chromosomal duplication (Burgess et al., 2001; Chu et al., 2001). Putting functional (biochemical and electrophysiological) and computational (bioinformatics and phylogenetic) analyses together, γ subunits are a "highly divergent family" which is conflict to our straightforward concept: Proteins in same family share similar biological functions.

New members of the γ subunit family

Interestingly, before reaching a settlement of this debate, transmembrane protein (TMEM) 114 and TMEM235, two novel genes, were classified into the γ subunit family. TMEM114 was first identified from a human congenital cataract case (Jamieson et al., 2007). Somewhat surprisingly, deletion of TMEM114 gene is not involved in cataract formation in a boy and his father (Gai et al., 2014). Although the role of TMEM114 in eye development remain unclear, blocking the function TMEM114 cause microphthalmia in

Xenopus tropicalis (Maher et al., 2011). Both TMEM114 and TMEM235 RNA are expressed in developing eye and neural tissues (Maher et al., 2011), but functions of these two genes need to be further examined.

Evolutionary analysis of the γ subunit family

To rationalize divergent functions in the calcium channel γ subunits (CACNG), we dedicated ourselves to re-examine the evolution of calcium channel γ subunits family in animals. We hoped to find clues from the evolutionary history of calcium channel γ subunits. It might help scientist to understand the source of their functional divergence. In particular, the new member in CACNG family, TMEM114 and TMEM235, are worthy to investigate further, because the functions of TMEM114 and TMEM235 are still unknown.

With BLAST (basic local alignment search tool), we examined available model organisms and well-sequenced organism genome as possible as we could. We started from human and look as far as into the *Caenorhabditis elegans* genome. These hundreds of sequences were reassessed and further processed. The work flow of the evolutionary analysis was showed in Fig. 1. We displayed the relationship of the sequences through phylogenetic trees. Burgess et al., 1999 (Burgess et al., 1999) and Chu et al., 2001 (Chu et al., 2001) proposed a model of evolution of γ subunits, we refined and updated this model based on an expanded list of animals whose chromosome map became accessible recently. Evolutionary rate analysis was also carried out in γ subunits. Hopefully, the study will help shed light on these highly divergent γ subunits. It might establish a new perspective and lead scientists to redefine the calcium channel γ subunits family. Furthermore, we would acquire new insight into the biological roles of the γ_1, γ_6, TMEM114 and TMEM235 subunits.

Materials and Methods

Sequence retrieval and trimming

γ-related, γ, TMEM114 and TMEM235 coding sequences were retrieved through protein-protein BLAST (basic local alignment search tool) and references. The source of sequences are Ensembl (www.ensembl.org), NCBI (https://www.ncbi.nlm.nih.gov/) and JGI (http://genome.jgi.doe.gov/). The mouse γ, TMEM114 and TMEM235 peptides were used as queries in local BLAST. The subjects were various animals which are *Hydra magnipapillata, Aplysia californica, Caenorhabditis elegans, Daphnia pulex, Drosophila melanogaster, Capitella teleta, Helobdella robusta, Strongylocentrotus purpuratus, Saccoglossus kowalevskii, Ciona intestinalis, Branchiostoma floridae, Petromyzon marinus, Callorhinchus milii, Danio rerio, Oreochromis niloticus, Takifugu rubripes, Latimeria chalumnae, Xenopus tropicalis, Anolis carolinensis, Chrysemys picta bellii, Gallus gallus, Monodelphis domestica, Canis familiaris, Bos taurus* and *Oryctolagus cuniculus* (Appendix II). The program performed local BLAST and E-values were set between 1e-1 and 1e-21. After obtaining the various results, the program performed reverse alignment, constructed a phylogenetic tree with query sequences (mouse sequences) and selected the E-value result in case of expected phylogenetic tree. Because most these retrieved sequences were predicted, multiple sequences alignment with well annotated (mammal sequences or zebrafish sequences) sequences is necessary to spot questionable alignment or annotations. The alignment tools include MUSCLE (V3.6) (Edgar, 2004), CLUSTALW version 1.83 (http://www.genome.jp/tools/clustalw/) and PRNAK (V.100311). For each multiple sequence alignment, every uncertain predicted sequence was examined seriously. Meanwhile, amino acids alignment was also considered. If these predicted sequences seemed incorrect, they would be manually edited and a note was taken. Future cDNA sequences would be necessary to justify these manual editing. Out of the 72 sequences inspected, 25 sequences were edited in 6 species.

Phylogenetic analysis

The phylogenetic tree was based on nucleotides sequences from BLAST. Not all species were analyzed in the phylogenetic analysis. Only sequence of representative animals in evolutionary stage were implemented. However, because *Oreochromis niloticus* and *Takifugu rubripes* possess two complete sets of γ genes, these two species were included. There were three different software used for phylogenetic analysis, MEGA (molecular evolutionary genetics analysis) (Tamura et al., 2013), GARLI (genetic algorithm for rapid likelihood inference) and MrBayse (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The phylogenetic trees were performed with maximum likelihood, GTR+G model and 50 bootstrap in MEGA. After the results were viewed carefully from MEGA, GARLI and MrBayse constructed the phylogenetic trees further. The bootstrap was raised to 1000 in GARLI. The number of generation was 10000000 and the chain was sampled at 1000 in MrBayse.

To rule out the possibility that the long branch attraction in our phylogenetic tree morphology is a result of poor alignment using nucleotide sequences, we also performed phylogenetic analysis using protein sequences by MEGA.

Syntenic blocks CACNG, TMEM114 and TMEM235 on chromosomes

The syntenic blocks around γ, TMEM114, and TMEM235 genes were identified from Ensembl and NCBI. Taking each γ, TMEM114, and TMEM235 genes as a center, we searched paralogous regions around γ, TMEM114, and TMEM235 genes on chromosomes in each available species. The chromosome maps we constructed were from *Caenorhabditis elegans, Drosophila melanogaster, Callorhinchus milii, Takifugu rubripes, Danio rerio, Xenopus tropicalis, Gallus gallus, Canis familiaris, Bos taurus*, *Oryctolagus cuniculus, Mus musculus* and *Homo sapiens.*

Gene structure and transmembrane domains analysis

The information of intron and exon structure was obtained from Ensembl and NCBI. Transmembrane domains were predicted by TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) (Krogh et al., 2001; Sonnhammer et al., 1998). The visualization of gene structure and transmembrane domains were constructed using

PowerPoint. The analyzed γ genes in terms of gene structure and transmembrane domains were the same as those in phylogenetic analysis.

Functional site prediction of the γ subunits

All γ subunits are transmembrane proteins that consist of four transmembrane domains (hydrophobic region). They also share two extracellular loops, one intracellular loop, C-terminal and N-terminal (hydrophilic region). Prosite (http://prosite.expasy.org/) (Sigrist et al., 2013) is a database of domains, family and functional sites. Previous analysis using Prosite have successfully discovered distinct functional sites carried in each of the clusters with in the γ subunit family (Chen et al., 2007). In the present study, we included the newly addition into the family, TMEM114 and TMEM235, and updated the results.

Likelihood ratio test (LTR) for positive selection

PAML (Phylogenetic Analysis by Maximum Likelihood) 4.7 (Xu and Yang, 2013; Yang, 2007) was used in positive selection analysis. Animal species included in this analysis are: *Callorhinchus milii, Danio rerio, Oreochromis niloticus, Takifugu rubripes, Latimeria chalumnae, Xenopus tropicalis, Anolis carolinensis, Chrysemys picta bellii, Gallus gallus, Canis familiaris, Bos Taurus, Oryctolagus cuniculus*, *Mus musculus, Rattus Norvegicus* and *Homo sapiens*. Total number of sequences used was 134. All analysis in PAML was codeML. Each γ subgroup was analyzed with codons respectively in pairwise comparison and site model. Pairwise comparison estimates dN/dS (ω) between two γ sequences. Simply, comparison between two γ sequences showed how related they are. We selected site models, which comprises M0, M1 (neutral), M2 (selection), M7 (beta), and M8 (beta & ω), to performed the LTR test. M1-M2 and M7-M8 comparison was tested to validate whether each γ subunit encounter positive selection. M1 and M7 were null models without positive selection, whereas M2 and M8 were alternative model with positive selection. We also analyzed two different treefiles. One was based on gene tree established by MrBayse and the other was species tree.

Because the difference of tree length between paralogous γ subunits, we further conducted branch model test. The branch model was used to show the asymmetric evolution between paralogous γ subunits: (γ_2, γ_3), (γ_4, γ_8) and (TMEM114, TMEM235), respectively. In addition, because *Danio rerio, Oreochromis niloticus* and *Takifugu rubripes* have additional γ sets of subunits, we would like to evaluate if the additional copies accumulate mutation and acquire new function. Therefore, each of group γ subunit was tested for branch model. There were three time period that we would like to test: after duplication, after speciation and between duplication and speciation (Fig. 2). Those time period were assigned as foreground lineage. The test 1 were set as null hypothesis without positive selection, whereas the test 2 were alternative selection with positive selection. Lnl1 and Lnl2 were maximum likelihood value from test 1and test 2. All the likelihood ratio were calculate by

$$\mathbf{2} \times (\mathbf{Lnl2} - \mathbf{Lnl1}),$$

and were compared against χ^2 distribution. If test1 and test2 have statistically significant difference, the ω on interested foreground lineage in test 2 will be checked whether the two ω on interested foreground lineage were different from each other.

Results

<u>γ subunits were independently lost and duplicated in vertebrate lineages</u>

Earlier evolutionary studies of calcium channel γ subunits were based on mouse, rat, and human sequences, not including the TMEM114 and TMEM235 subunits (Burgess et al., 2001; Chu et al., 2001). Subsequent studies have suggested that γ subunits were derived from tandem duplication and at least two rounds of chromosome duplications in animal evolution (Kasahara, 2007). In order to reveal the evolution of γ subunits, we must explore γ genes in other animals. Fortunately, various species genome dataset are getting more complete and accessible. They provide us an opportunity to include a wide range of species in our evolutionary analysis. By performing BLAST, we acquired γ-related, γ, TMEM114 and TMEM235 nucleotides/amino acids sequences from various species genome datasets. The results showed that γ genes, TMEM114 and TMEM235 exist in many species.

With protein-protein BLAST, we retrieved hundreds of γ coding sequences from the one of earliest representative animals, *C.elegans*, to *H.sapiens,* (Appendix I). Before urochordata, most of the proteins were not functionally identified and were annotated with serial number or as γ-like proteins. It seems that we did not find any γ proteins in invertebrate animals. The first γ gene that appeared in the vertebrate lineage was the lamprey (*P. marinus*) γ_1. In cartilaginous fish and later vertebrate species, the currently known γ's, includingTMEM114 and TMEM235, genes appeared. It suggests that γ, TMEM114 and TMEM235 genes rapidly evolved from one into ten γ genes. Additionally, γ genes have double copies in several osteichthyes. Specifically, tilapia (*O. niloticus*) and fugu (*T. rubripes*) had two copies of TMEM235 genes but no TMEM114 gene. It indicates that TMEM235 may compensate for TMEM114's function. Furthermore, γ, TMEM114 and TMEM235 genes were independently lost and duplicated in many animals. *G. gallus*, for example, had two copies of γ_1, γ_4, and γ_5 genes besides the fact that there was no TMEM235 gene in chickens. Because number of CACNG genes are variable in animal lineage, CACNG genes may not be critical for their survival.

γ subunits have their most recent common ancestor in bilaterians

Earlier phylogenetic studies of the calcium channel γ subunits employed only the mouse, rat, and human γ sequences (Burgess et al., 2001; Chu et al., 2001; Maher et al., 2011). The γ subunits appear to form a monophyletic group, with claudins proteins as their outgroup anchor. To explore deeper into their evolutionary history, we collected sequences from various vertebrate and invertebrate species, carefully inspected and manually edited the coding sequences (see Methods), and analyzed their phylogenetic relationship. Although the morphology of the phylogenetic tree appears consistent with the literature, invertebrate sequences (that are similar to vertebrate γ subunits) can be found between the (γ_2, γ_3, γ_4, γ_8) and (γ_5, γ_7) clusters (Fig. 3-4) and between the (γ_5, γ_7) and the branch that lead to (γ_1, γ_6, TMEM114, TMEM235). By comparing the results with known evolutionary tree of animals, the γ subunits can be traced back to their most recent common ancestor (MRCA) in bilaterians (~555mya). It can be seen that phylogenetic trees constructed with MEGA (Fig. 3) or MrBayes (Fig. 4) showed similar results. The morphology of both trees showed little difference, except that position of (γ_1, γ_6) and (TMEM114, TMEM235) were ambiguous. Apart from vertebrate sequences, the distribution of invertebrate sequences did not cluster with any other γ or TMEM subunits in the both trees (Fig. 4-5). Most of invertebrate sequences were clustered together and indicated long branch attraction. At the same time, the bootstrap values and posterior probability were less than fifty percent in the clade of invertebrate sequences.

When we performed phylogenetic analysis using protein sequences by MEGA, we obtained almost identical result (data not showed) as using nucleotide sequences. This result ruled out the possibility that the long branch attraction in our phylogenetic tree morphology was due to poor alignment with nucleotide sequences.

Because the long branch attraction and low credibility in the invertebrate sequences, most of the invertebrate sequences were removed and the phylogenetic analysis performed again. MEGA (Fig. 5), MrBayse (Fig. 6), and GARLI (Fig. 7) showed nearly the same results. The vertebrate CACNG family was grouped into four clades: (γ_2, γ_3, γ_4, γ_8), (γ_5, γ_7), (TMEM114, TMEM235) and (γ_1, γ_6).

Since previous literature referred to three γ-related (H.robusta HELRODRAFT 190537, S.kowalevskii CACNG5-like and C.teleta CAPTEDRAFT 155151) invertebrate sequences, the three invertebrate sequences were kept for the following analysis. As for the previous results, none of these invertebrate sequences belonged to any of the γ or TMEM subunit groups. As a result of this view, the vertebrate γ subunits shall not be viewed as members of single family. Instead, the γ subunits and TMEM114 and TMEM235 represent four (4) independently evolved monophyletic groups within the vertebrate lineage. Taken together, the γ subunits and related sequences in bilaterians form a polyphyletic group (instead of a monophyletic group) starting from ~555mya in bilaterians.

In addition, the most similar regions among CACNG family, transmembrane regions, also were analyzed for their phylogenetic relationship (Appendix IIIa-IIIc). Basically, the results were consistent with full length analysis. It suggested each γ and TMEM sequences had the same substitution probability.

Evolutionary pathway of the CACNG genes

Chu et al. and Burgess et al. proposed respectively a duplication model of γ genes in 2001 (Burgess et al., 2001; Chu et al., 2001). They suggested that γ genes have experienced several tandem duplication and two chromosome duplication in animals. To refine the evolutionary pathway model, we took advantage of chromosome maps of several species that recently became available. The previous model illustrated the duplication pathway of 8 γ genes only, whereas in our model we seek to include the two new members, TMEM114 and TMEM235.

Chromosome maps of Elephant shark, fugu, zebrafish, frog, chicken, dog, cattle, rabbit, mouse and human (Appendix VIa-VIj) were retrieved from online database (see Methods) and searched for the ten γ genes. Each of the tables showed paralogous regions around γ, TMEM114 and TMEM235 genes on chromosomes. There were many genes in paralogous regions, such as SSTR (somatostain), CACNA1 (calcium channel α_1 subunits), SYNGR (synaptogyrin), GRIN (ionotropic glutamate receptor), PRKC (protein kinase C) and TBX (T-box transcription factor). Basically, the paralogous regions around γ genes robustly clustered to one another, suggesting the two round whole genome duplication hypothesis during vertebrate evolution (Kasahara, 2007). Specifically, γ genes were almost always associated with PKC genes on chromosomes, indicating strongly conserved syntenic blocks.

Furthermore, syntenic blocks demonstrated distribution of γ, TMEM114 and TMEM235 genes among different species (Fig. 8). It illustrated translocation and tandem duplication events in different species. For example, γ_3 gene, which was localized on one chromosome with TMEM114 gene, underwent translocation event in mouse. Tandem duplication event occurred to chicken's γ_1, γ_4 and γ_5 genes. Interestingly, chromosome 3 in zebrafish concentrated multiple γ genes and paralogous genes into a large block.

With the chromosome maps from different kinds of species and the information of syntenic blocks (Fig. 8) in hand, together with the knowledge of homology from phylogenetic analysis (Fig 3-7), we were able to refine the evolutionary pathway model (Fig. 9). In this revised model, emergence of γ genes experienced several tandem duplications. Following three steps of tandem duplications γ_1, γ_4, γ_5 and TMEM235 genes were generated on a single chromosome. This proto-chromosome went through the 1st round of whole genome duplication to yield four more γ genes. However, gene deletions led to the loss of two γ genes around γ_2 or γ_3. After the 2nd round of whole genome duplication, γ genes were now distributed on four chromosomes. However, the newly created copies of TMEM114 and TMEM235 were lost due to yet understood reasons. Finally, the patterns of γ gene distribution on chromosomes were similar in most animals that we examined.

CACNG gene structure and transmembrane domains

γ subunits belonged to pam00822 family that include also claudins, proteins that form tight junctions in epithelia (Van Itallie and Anderson, 2006). Structurally, γ subunits were predicted with four transmembrane domains and contain a GLW motif in the first extracellular loop. Results from TMHMM predication of transmembrane domains were combined with exon-intron predications, and displayed as in Appendix IV.

The grey bars represented length of amino acids sequences (Appendix IV). In intron-exon structure and structure prediction, the result indicated that most γ genes have three to four introns (closed triangle) and four transmembrane domains (closed black bar). Additionally, a few problems of exon sequences were suspected because their exons length were different from other orthologous genes (indicated with “?”). Red crosses represented regions where their amino acids were not used in our phylogenetic and evolutionary rate analyses.

Functional site prediction of the γ subunits

Previous studies have predicted membrane topology and putative functional site with γ proteins (Chen et al., 2007). They indicated that each subgroup of γ, (γ_1, γ_6), (γ_2, γ_3, γ_4, γ_8) and (γ_5, γ_7), contains a specific subset of putative functional sites. However, the analysis did not include the two novel members, TMEM114 and TMEM235.

After multiple alignment with the ten mouse CACNG subunits, we labeled the predicted motifs with different colors (Appendix V).

In contrast to transmembrane domain, extracellular loops and inner loops were predicted with a few motifs. For example, there were two N-glycosylation sites, NRSQ and NVTV, in the first extracellular loop of TMEM114. The C-terminal sequence of TMEM235 contain a predicted N-myristoylation site. Additionally, the consensus motifs of TMEM was protein kinase C phosphorylation site on C-terminal sequence.

CACNG evolutionary rate analysis

γ subunits were a functionally divergent family. (γ_1, γ_6) are real calcium channel regulators, which inhibited HVA and LVA calcium channels current, respectively (Arikkath et al., 2003; Hansen et al., 2004; Lin et al., 2008). However, (γ_2, γ_3, γ_4, γ_8,) are transmembrane AMPA receptor regulatory proteins (TARPs), which are involved in AMPA receptor trafficking and AMPA receptor current regulation, whereas (γ_5, γ_7) are type II TARPs that modulated GluR2-containing AMPA receptor. Additionally, copy number of γ, TMEM114 and TMEM235 are variable in animal lineage. Due to γ's functional divergence and variable copy number, we wondered if some of the γ subunits may accumulate mutations and acquire new functions.

In the beginning, pairwise comparison was implemented for each of the γ subunits. Pairwise comparison could detect positive selection in each pair of sequences if the dN/dS value is >1. The result indicated that none of the γ subunits experienced positive selection, as the highest dN/dS, between fugu $γ_6$ b and dog $γ_6$, is simply 0.4925 (Appendix VII).

Next, site model was used to evaluate whether positive selection acted on each group of γ subunits. The site model is composed of multiple models, M0, M1, M2, M7 and M8. Each model calculated different parameters. M0 provided tree length, dN, dS and dN/dS. M1, M2, M7 and M8 estimated a likelihood respectively. Each group of γ was analyzed with two different trees, gene tree (Table I) and species tree (Table II). The results showed that tree input did not influenced the parameters. For example, tree length were close to each other no matter which tree input was. Furthermore, M1-M2 comparison revealed no positive selection of γ subunits ($p \geqq 0.05$). $γ_7$, however, had significant difference between M7 and M8 ($p < 0.05$), suggesting positive selection. Nevertheless, the highlighted amino acids for positive selection in M8 were located in gap-rich regions. The gap-rich regions were due to differences sequence in length because γ_7 sequences in zebrafish, fugu and tilapia are longer than the other γ_7. Moreover, TMEM235 had significant difference between M7 and M8 in gene tree, but the indicated sites were situated in gap region in some TMEM235 sequences.

We also inspected the tree length between γ subunits because it can be another indicator of asymmetric evolution. We picked the (γ_2, γ_3), (γ_4, γ_8) and (TMEM114, TMEM235) to do further analysis with branch model. γ_4 and γ_8 are the most similar paralogous subunits in CACNG family, so are (TMEM114, TMEM235), and (γ_2, γ_3). For example, we wanted to understand whether one of the γ_4 and γ_8 obtained more nonsynonymous substitution and the pairs (γ_2, γ_3) and (TMEM114, TMEM235) received a new function. We set three time point as foreground linages, which were after duplication, after speciation, and the time period between duplication and speciation (Fig. 2). The results showed that (γ_2, γ_3), (γ_4, γ_8) and (TMEM114, TMEM235) (Table III-V) encountered asymmetric evolution after duplication and after speciation. Briefly, (γ_2, γ_3), (γ_4, γ_8) and (TMEM114, TMEM235) encountered significantly nonsynonymous substitution after the 2nd round of chromosome duplication, implying that their functions have diverged in the animal lineage.

In addition, the three osteichthyes, tilapia, fugu and zebrafish have additional set of γ and TMEM235 genes (except for zebrafish, which do not have two copies of TMEM114 and TMEM235) that are possibly derived from the 3rd round of whole-genome duplication (Kasahara, 2007). Thus, we have been wondering whether one of the copy γ accumulated nonsynonymous mutation. The analysis also tested for three time point as foreground linages, which were after duplication, after speciation, and the period time of between duplication and speciation (Fig. 2). The results indicated that γ_1, γ_2, γ_3, γ_5 and γ_7 (Appendix VIIIa, VIIIb, VIIIc, VIIIe, and VIIIg) encountered asymmetric evolution after duplication and after speciation. In summary, some of the additional copies of γ (γ_1, γ_2, γ_3, γ_5 and γ_7) may have acquired novel functions in osteichthyes.

Discussion

Bioinformatics as a tool to reveal functional insights

The animals we chose in the study were model organisms and iconic species in animal evolutionary starting with mouse γ, TMEM114 and TMEM235 amino acids sequences as query in the protein-protein BLAST, our phylogenetic analysis showed that γ, TMEM114 and TMEM235 genes independently evolved in animals as several clusters (Appendix I). The copy number of γ, TMEM114 and TMEM235 genes were not the same in each species. For example, we did not find γ_6, γ_7 and γ_8 in elephant shark.

Maher et al. (Maher et al., 2011) suggested that two novel proteins, TMEM114 and TMEM235, belong to the CACNG family. Functional knockdown of TMEM114 gene expression led to microphthalmia in *X.tropicalis tropicalis*. TMEM114 gene was first identified by chromosomal translocation on 16p13.3 in a congenital cataract family (Jamieson et al., 2007). This chromosomal translocation lies at the promoter region of TMEM114 and it may cause dysregulation of TMEM114 expression (Jamieson et al., 2007). However, heterozygous deletion of TMEM114 gene did not cause cataract (Gai et al., 2014). To date, the function of TMEM114 and TMEM235 remain unknown. Fortunately, the rapid development of bioinformatical dataset and tools in recent years provided us an opportunity to reexamine the evolution of CACNG subunits in animals, which may shed light on the functional differentiation of the CACNG family. Chen et al. (Chen et al., 2007) demonstrated that conserved motifs within each cluster of γ subunits supported their functional divergence. Thus, further bioinformatical analysis may lead us to a new perspective for studying TMEM114 and TMEM235 in the future.

Inaccurate terminology and sequences annotations in databases

When manually curing the sequences retrieved with BLAST, we discovered numerous potential problems in sequences annotations. By inspecting sequences one by one as mentioned in the methods, we found that the annotation of some genome datasets was not compete and that each database has their own way of organizing datasets. For instance, we obtained two "novel proteins" in chicken, but these two sequences are actual γ_5

orthologs. Similarly, two γ_7 orthologs with ambiguous names were found in the dog genome. The elephant shark TMEM235 sequence in Ensembl is incorrect labelled as TMEM114.

As for problems in sequences annotation, several species genome datasets did not contain hundred percent complete sequences. γ_1 in lamprey, for example, was suspected lacking exon 1 and 2 when comparing with the other orthologs (Appendix IX.). Because the lamprey γ_1 represents the earliest calcium channel γ genes in animal history, it will be sequenced in the future by experimental approaches. Among all the sequences that we considered as having potential errors, we picked six sequences worthy to be verified (Appendix IX).

Phylogenetic relationship of the CACNG family

Three studies have reported the CACNG family phylogenetic relationship. In 2001, Burgess et al. (Burgess et al., 2001) analyzed human CACNG subunits. In the same year, 2001, human, mouse, and rat CACNG subunits were jointly included in a phylogenetic analysis (Chu et al., 2001). Both results presented the γ subunits as a monophyletic protein family anchored by claudins. Interestingly, after a decade, two novel members, TMEM114 and TMEM235, were classified into CACNG family (Maher et al., 2011). The position of TMEM114 and TMEM235 were situated at between (γ_1, γ_6) and (γ_5, γ_7) on the phylogenetic tree. The clustering of each subgroup of the CACNG family, (γ_2, γ_3, γ_4, γ_8), (γ_5, γ_7), (TMEM114, TMTM235) and (γ_1, γ_6), on phylogenetic tree, corresponds well with to their physiological functions. (γ_2, γ_3, γ_4, γ_8) were type I TARPs. (γ_5, γ_7) were type II TARPs. (γ_1, γ_6) were real calcium channel current regulators. In other words, it is hard to speculate the physiological functions of TMEM114 and TMEM235.

When it comes to methods used to produce the phylogenetic tree of the CACNG family, previous studies applied neighbor-joining method, parsimony method, distance method, maximum likelihood and maximum parsimony (Burgess et al., 1999; Burgess et al., 2001; Chu et al., 2001; Maher et al., 2011). In the present study, we used maximum likelihood and bayesian methods. Additionally, CACNG family orthologs from many species were performed. We chose CACNG orthologs only form representative species in the animal evolution since it is not necessary to include all possible species. Previous analyses were based on protein sequences, whereas we used coding cDNA to establish phylogenetic trees.

We also performed phylogenetic analysis using protein sequences by MEGA. The result barely showed any difference from that obtained with nucleotide sequences. So we ruled out that poor alignment within coding regions as the main factor to affect CACNG distribution in our phylogenetic tree (Fig. 3). Instead, the long branch attraction in the phylogenetic trees possibly reflected ancient invertebrate γ-related sequences that had diverged away from their vertebrate cousins

Our results were in very good agreement with the previous literature (Fig. 5-7). Although maximum likelihood (MEGA) (Fig. 3) and bayesian method (MrBayes) (Fig. 4) have a slight difference in the position of ($\gamma_{1,}$ γ_6) and (TMEM114, TMEM235),this may simply reflect the difference in their methodology of likelihood calculation. Because of long branch attraction and low credibility, we removed most of the invertebrate sequences. However, three invertebrate γ-related sequences analyzed in the literature was kept (Moran and Zakon, 2014). These invertebrate γ-related proteins showed 41%~46% similarity to vertebrate γ subunits (Moran and Zakon, 2014). When used for reciprocal BLAST query against the human genome, the subjects were all γ subunits, implying that the three invertebrate sequences were indeed early γ-related proteins in invertebrates (Moran and Zakon, 2014). These three γ-related sequences did not clustered into any clades in vertebrate γ our trees, again suggesting that MRCA of the γ's was bilaterians, because bilaterians were the common ancestor of vertebrate and invertebrate. Our results also supported previous inference (Moran and Zakon, 2014) that γ subunits independently evolved in the bilaterians lineage. More importantly, γ have evolved into four monophyletic groups in vertebrate lineages. Therefore, we strongly recommend that this functionally divergent protein groups not to be viewed as a single protein family. Instead, the nomenclature of these proteins need to be officially revised. Specifically, we propose to rename TARP γ_2, γ_3, γ_4, γ_5, γ_7 and γ_8 as TARP2, TARP3, TARP4, TARP5, TARP7 and TARP8, respectively to avoid functional connections with calcium channels. As for TMEM114 and TMEM235, more appropriate names should be adopted when their functions are better elucidated. We believed that reclassifying the CACNG proteins would help to reignite interests in their researches, especially for TMEM114 and TMEM235.

On a side note, we also performed phylogenetic analysis with the most consensus regions, the four transmembrane domains regions, and obtained nearly identical phylogenetic tree to full length sequences (Appendix IIIa-IIIc).

Evolutionary pathways of the CACNG genes

Susumu Ohno proposed that one or two rounds of whole genome duplication (2R hypothesis) occurred before the emergence of vertebrates (Ohno, 1970). When a gene experienced duplication, one copy may be allowed to accumulate more mutations. As a result, the gene may acquire a new function. The 2R hypothesis became an important postulation for explaining the development of immune system. For example, multiple copies of the major histocompatibility complex were considered evidence of the 2R hypothesis (Kasahara, 2007; Ohno, 1970). Chu et al and Burgess et al (Burgess et al., 1999; Burgess et al., 2001; Chu et al., 2001) both incorporate two rounds of chromosome duplication in their models for the evolutionary pathways of the CACNG genes. Interestingly, Chu and colleagues (Chu et al., 2001) suggested an alternative pathway (A_2) for generating three consecutive γ genes on the same chromosome segment. Whereas previous models relied on mouse, rat, and human chromosome maps, our refined model was based on chromosome maps of many vertebrate species (Fig. 9). If the alternative pathways A_2 were correct, then gene γ_a and γ_a' should be the closet homologue as a result of the unequal crossing-over (Chu et al., 2001). If it were true, γ_1 and γ_6 would be most similar to γ_5 and γ_7. However, because γ_5 and γ_7, the two descendent genes in real life, are closer to γ_4 and γ_8, we ruled out the possibility. Thus, the model we purposed was also compatible with the model of Burgess et al. (Burgess et al., 2001). Furthermore, TMEM114 and TMEM235 were included in our model. We believed that refined model was more accurate after examining γ, TMEM114 and TMEM235 genes distribution on chromosomes from various species. Because the lamprey $γ_1$ is the earliest γ gene in animals that we could find, $γ_1$ was taken as the earliest γ gene in our model. Because TMEM235 genes are strongly associated with $γ_1$ genes on the same chromosome in many species. We put $γ_1$ and TMEM235 as the direct descendants of the proto-$γ_a$ gene after the first tandem duplication (Fig. 9). Because some of our analyses suggested ($γ_4$, $γ_5$) are closer to (TMEM114, TMEM235) (Fig. 5-7), while others suggested that ($γ_4$, $γ_5$) are closer to ($γ_1$, $γ_6$) (Fig. 4), we considered $γ_1$ and TMEM235 as two alternative source genes from which $γ_4$ or $γ_5$ was derived (Fig. 9). In syntentic blocks, γ genes were almost always associated with PKC genes on chromosomes, suggesting that the functions of γ proteins are related to the homeostasis of calcium or protein phosphorylation. Also, TMEM114 and TMEM235 genes were closely located with GRIN2 and CACNA genes on chromosomes, implying that TMEM114 and TMEM235 may functionally interact with GRIN2 (NMDA receptor subunits) and

CACNA. These findings therefore provide useful insights for our research designs in the future.

Burgess et al (Burgess et al., 1999) made comprehensive comparison of paralogous genes around CACNG genes on human chromosomes and came up with their evolutionary pathway model. Although we had access to a wider ranges of animal chromosome maps, they were still not sufficient to reconstruct the “original chromosome” containing all the ancestral CACNG genes. The difficulty lies in the fact that in some species the CACNG genes are shown on “scaffold” rather than actually numbered chromosomes. In other words, better annotated complete datasets of chromosomes are required to reconstruct the ancestral chromosome of the CACNG genes. At present, our best knowledge is that zebrafish chromosome 3 and the longest scaffold containing (TMEM235, γ_1, γ_4, γ_5) may be the closest to the original chromosome of CACNG genes.

Evolutionary rate analysis

Evolutionary rate analysis have not been performed in the CACNG family before. In this study, we performed pairwise comparison, site model and branch model. The order of workflow was: pairwise comparison, site model and branch model. We used pairwise comparison and site model to broadly investigate whether in each γ subfamily there are any positive results or implications. We further used branch model to analyze specific branches. Because pairwise comparison did not show any significant results, we applied the site model to evaluate whether some group of γ's contain nonsynonymous substitutions. Although γ_7 had significant result in the M7-M8 comparison, it did not provide us with any meaningful amino acid site in M8, as those sites were located at gap-rich region, where the γ_7 sequences in zebrafish, fugu and tilapia are longer than the other γ_7's. As for TMEM235, our result indicated that TMEM235 had significant difference between M7 and M8 in gene tree as well. But the sites considered positive was again located in gap regions, indicating possible false positives.

Because the tree length indicated that there are asymmetric results in the site model, (γ_2, γ_3), (γ_4, γ_8) and (TMEM114, TMEM235) were analyzed in branch model (Table III-V). Notably, TMEM114 and TMEM235 had asymmetric evolution. It suggested that the function of TMEM114 may be different from that of TMEM235. This provides a hint for our following electrophysiological studies and expression distribution in zebrafish. Interestingly, the species we chose contain three osteichthyes, tilapia, fugu and zebrafish, which might have experienced one additional round of genome duplication, a very possible explanation of the additional set of CACNG genes. Thus, we tested whether the additional copies could accumulate mutations and obtain novel functions. The results suggested that, following chromosome duplication, the additional copies of γ (γ_1, γ_2, γ_3, γ_5 and γ_7) may have obtained new functions in osteichthyes.

Tissue distribution and subcellular localization of CACNG proteins

Except for TMEM114 and TMEM235, which have unknown functions, the variations in the physiological functions of the CACNG proteins are consistent with their tissues distribution patterns (Burgess et al., 2001; Chen et al., 2007; Chu et al., 2001;

Fukaya et al., 2005). As for TMEM114 and TMEM235, they were both located in brain, eye, and spinal cord in human (Maher et al., 2011), but their expression stage and subcellular location are slightly different. Because TMEM114 and TMEM235 are both expressed in neural tissue and are close to GRIN (NMDA receptor) gene and CACNA1 (calcium channel α_1 subunits) on chromosomes, we raised a hypothetical question: Can TMEM114 and TMEM235 subunits act on GRIN2 and CACNA1 to modulate their current properties? In our laboratory, we will seek to test that hypothesis in the future. Although subcellular localization of TMEM114 and TMEM235 have been performed in cultured cells (Maher et al., 2011), we will examine their tissue expression in zebrafish. We expect that TMEM114 and TMEM235 may have slightly different expression patterns.

Murine TMEM114 and TMEM235 were identified as glycoprotein, but TMEM235 contains an atypical glycosylation N-X-C motif (Maher et al., 2011). In our sequence analysis with PROSITE (Appendix V), the result suggested that there is a consensus protein kinase C phosphorylation site in the C-terminals of TMEM114 and TMEM235. The possible functional implication of this site can be further explored in the future. Additionally, γ_8 has a longer C-terminal. By using this additional fragment to perform protein-protein BLAST, we did not find any proteins but γ_8 and unnamed proteins. Thus, we ruled out the possibility that this additional fragment was from other genes (i.e. this addition fragment is specific to γ_8.).

Functional segregation following gene duplications

Because gene duplication is one of the mechanism to acquire novel function for duplicated genes, multiple paralogous genes were produced in the same genome. Whole genome duplication is a major source of gene divergence. The other small scale of gene duplications are tandem duplication, segmental duplication and duplicative retroposition (Liu et al., 2011; Liu et al., 2014; Qiu et al., 2014). After gene duplication, the genes had several options: subfunctionalization and neofunctionalization, to remain in the genome. Additionally, subfunctionalization and neofunctionalization are usually accompanied reciprocal expression pattern.

As for the CACNG family evolution, their evolution contain multiple tandem duplications and two rounds of whole chromosome duplications (Burgess et al., 1999; Burgess et al., 2001; Chu et al., 2001). According to their sequences similarity, CACNG were grouped into four subgroups: (γ_1, γ_6), (γ_2, γ_3, γ_4, γ_8), (γ_5, γ_7) and (TMEM114, TMEM235) whose functions are unknown. Their functional subgroups also correspond to their sequences similarity, except for (TMEM114, TMEM235). γ_1 and γ_6 are real calcium channel γ subunits that modulate calcium channel currents in muscle tissues. It is reasonable to consider that γ_1 and γ_6 experienced subfunctionalization, as γ_1 acts on high-voltage activation channel, whereas γ_6 modulates low-voltage activation channel (Arikkath et al., 2003; Held et al., 2002; Lin et al., 2008). In contrast, another set of paralogous genes, (γ_2, γ_3, γ_4, γ_8) and (γ_5, γ_7), are AMPA receptor regulators (Tomita et al., 2003). The fact that γ_5 and γ_7 are classified as type II TARPs may imply that (γ_2, γ_3, γ_4, γ_8) and (γ_5, γ_7) encountered neofunctionalization in neurons after gene duplication. All that been said, it is hard to speculate the function of the original γ, which is the γ before the first tandem gene duplication. At this stage, what we are interested in are TMEM114 and TMEM235 because their functions remain unclear and their expression patterns may imply subcellular relocalization or subfunctionalization.

References

Arikkath, J., and Campbell, K.P. (2003). Auxiliary subunits: essential components of the voltage-gated calcium channel complex. Current Opinion in Neurobiology *13*, 298-307.

Arikkath, J., Chen, C.C., Ahern, C., Allamand, V., Flanagan, J.D., Coronado, R., Gregg, R.G., and Campbell, K.P. (2003). Gamma 1 subunit interactions within the skeletal muscle L-type voltage-gated calcium channels. The Journal of biological chemistry *278*, 1212-1219.

Birnbaumer, L., Qin, N., Olcese, R., Tareilus, E., Platano, D., Costantin, J., and Stefani, E. (1998). Structures and functions of calcium channel beta subunits. Journal of bioenergetics and biomembranes *30*, 357-375.

Black, J.L., 3rd (2003). The voltage-gated calcium channel gamma subunits: a review of the literature. Journal of bioenergetics and biomembranes *35*, 649-660.

Bourinet, E., Soong, T.W., Sutton, K., Slaymaker, S., Mathews, E., Monteil, A., Zamponi, G.W., Nargeot, J., and Snutch, T.P. (1999). Splicing of alpha 1A subunit gene generates phenotypic variants of P- and Q-type calcium channels. Nature neuroscience *2*, 407-415.

Burgess, D.L., Davis, C.F., Gefrides, L.A., and Noebels, J.L. (1999). Identification of three novel Ca(2+) channel gamma subunit genes reveals molecular diversification by tandem and chromosome duplication. Genome research *9*, 1204-1213.

Burgess, D.L., Gefrides, L.A., Foreman, P.J., and Noebels, J.L. (2001). A cluster of three novel Ca2+ channel gamma subunit genes on chromosome 19q13.4: evolution and expression profile of the gamma subunit gene family. Genomics *71*, 339-350.

Carbone, E., and Lux, H.D. (1984). A low voltage-activated, fully inactivating Ca channel in vertebrate sensory neurones. Nature *310*, 501-502.

Chen, L., Bao, S., Qiao, X., and Thompson, R.F. (1999). Impaired cerebellar synapse maturation in waggler, a mutant mouse with a disrupted neuronal calcium channel gamma subunit. Proceedings of the National Academy of Sciences of the United States of America *96*, 12132-12137.

Chen, R.-S., Deng, T.-C., Garcia, T., Sellers, Z.M., and Best, P.M. (2007). Calcium channel gamma subunits: a functionally diverse protein family. Cell Biochem Biophys *47*, 178-186.

Chu, P.-J., Robertsonb, H.M., and Best, P.M. (2001). Calcium channel g subunits provide insights into the evolution

of this gene family. Gene, 37-48.

De Jongh, K.S., Warner, C., and Catterall, W.A. (1990). Subunits of purified calcium channels. Alpha 2 and delta are encoded by the same gene. The Journal of biological chemistry *265*, 14738-14741.

Dolphin, A.C. (2003). Beta subunits of voltage-gated calcium channels. Journal of bioenergetics and biomembranes *35*, 599-620.

Dolphin, A.C. (2006). A short history of voltage-gated calcium channels. British journal of pharmacology *147 Suppl 1*, S56-62.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research *32*, 1792-1797.

Ertel, E.A., Campbell, K.P., Harpold, M.M., Hofmann, F., Mori, Y., Perez-Reyes, E., Schwartz, A., Snutch, T.P., Tanabe, T., Birnbaumer, L., *et al.* (2000). Nomenclature of voltage-gated calcium channels. Neuron *25*, 533-535.

Fatt, P., and Katz, B. (1953). The electrical properties of crustacean muscle fibres. The Journal of physiology *120*, 171-204.

Fedulova, S.A., Kostyuk, P.G., and Veselovsky, N.S. (1985). Two types of calcium channels in the somatic membrane of new-born rat dorsal root ganglion neurones. The Journal of physiology *359*, 431-446.

Flucher, B.E., Obermair, G.J., Tuluc, P., Schredelseker, J., Kern, G., and Grabner, M. (2005). The role of auxiliary dihydropyridine receptor subunits in muscle. Journal of muscle research and cell motility *26*, 1-6.

Freise, D., Held, B., Wissenbach, U., Pfeifer, A., Trost, C., Himmerkus, N., Schweig, U., Freichel, M., Biel, M., Hofmann, F*., et al.* (2000). Absence of the gamma subunit of the skeletal muscle dihydropyridine receptor increases L-type Ca2+ currents and alters channel inactivation properties. The Journal of biological chemistry *275*, 14476-14481.

Fukaya, M., Yamazaki, M., Sakimura, K., and Watanabe, M. (2005). Spatial diversity in gene expression for VDCCgamma subunit family in developing and adult mouse brains. Neuroscience research *53*, 376-383.

Gai, D., Nicholl, J., Waters, W., Barnett, C.P., and Yu, S. (2014). Interstitial deletion at chromosome 16p13.2 involving TMEM114 (transmembrane protein 114) in a boy and his father without cataract. American journal of medical genetics Part A *164A*, 834-836.

Hagiwara, S., Ozawa, S., and Sand, O. (1975). Voltage clamp analysis of two inward current mechanisms in the egg cell membrane of a starfish. The Journal of general physiology *65*, 617-644.

Hansen, J.P., Chen, R.S., Larsen, J.K., Chu, P.J., Janes, D.M., Weis, K.E., and Best, P.M. (2004). Calcium channel gamma6 subunits are unique modulators of low voltage-activated (Cav3.1) calcium current. Journal of molecular and cellular cardiology *37*, 1147-1158.

Hashimoto, K., Fukaya, M., Qiao, X., Sakimura, K., Watanabe, M., and Kano, M. (1999). Impairment of AMPA receptor function in cerebellar granule cells of ataxic mutant mouse stargazer. The Journal of neuroscience : the official journal of the Society for Neuroscience *19*, 6027-6036.

Held, B., Freise, D., Freichel, M., Hoth, M., and Flockerzi, V. (2002). Skeletal muscle L-type Ca(2+) current modulation in gamma1-deficient and wildtype murine myotubes by the gamma1 subunit and cAMP. The Journal of physiology *539*, 459-468.

Hess, P., Lansman, J.B., and Tsien, R.W. (1984). Different modes of Ca channel gating behaviour favoured by dihydropyridine Ca agonists and antagonists. Nature *311*, 538-544.

Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics *17*, 754-755.

Jamieson, R.V., Farrar, N., Stewart, K., Perveen, R., Mihelec, M., Carette, M., Grigg, J.R., McAvoy, J.W., Lovicu, F.J., Tam, P.P.*, et al.* (2007). Characterization of a familial t(16;22) balanced translocation associated with congenital cataract leads to identification of a novel gene, TMEM114, expressed in the lens and disrupted by the translocation. Human mutation *28*, 968-977.

Jay, S.D., Ellis, S.B., McCue, A.F., Williams, M.E., Vedvick, T.S., Harpold, M.M., and Campbell, K.P. (1990). Primary structure of the gamma subunit of the DHP-sensitive calcium channel from skeletal muscle. Science *248*, 490-492.

Kang, M.G., and Campbell, K.P. (2003). Gamma subunit of voltage-activated calcium channels. The Journal of biological chemistry *278*, 21315-21318.

Kang, M.G., Chen, C.C., Felix, R., Letts, V.A., Frankel, W.N., Mori, Y., and Campbell, K.P. (2001). Biochemical and biophysical evidence for gamma 2 subunit association with neuronal voltage-activated Ca2+ channels. The Journal of biological chemistry *276*, 32917-32924.

Kasahara, M. (2007). The 2R hypothesis: an update. Current opinion in immunology *19*, 547-552.

Kato, A.S., Siuda, E.R., Nisenbaum, E.S., and Bredt, D.S. (2008). AMPA receptor subunit-specific regulation by a distinct family of type II TARPs. Neuron *59*, 986-996.

Kato, A.S., Zhou, W., Milstein, A.D., Knierman, M.D., Siuda, E.R., Dotzlaf, J.E., Yu, H., Hale, J.E., Nisenbaum, E.S., Nicoll, R.A.*, et al.* (2007). New transmembrane AMPA receptor regulatory protein isoform, gamma-7, differentially regulates AMPA receptors. The Journal of neuroscience : the official journal of the Society for Neuroscience *27*, 4969-4977.

Klugbauer, N., Dai, S., Specht, V., Lacinova, L., Marais, E., Bohn, G., and Hofmann, F. (2000). A family of gamma-like calcium channel subunits. FEBS letters *470*, 189-197.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol *305*, 567-580.

Letts, V.A., Felix, R., Biddlecome, G.H., Arikkath, J., Mahaffey, C.L., Valenzuela, A., Bartlett, F.S., 2nd, Mori, Y., Campbell, K.P., and Frankel, W.N. (1998). The mouse stargazer gene encodes a neuronal Ca2+-channel gamma subunit. Nature genetics *19*, 340-347.

Lin, Z., Witschas, K., Garcia, T., Chen, R.S., Hansen, J.P., Sellers, Z.M., Kuzmenkina, E., Herzig, S., and Best, P.M. (2008). A critical GxxxA motif in the gamma6 calcium channel subunit mediates its inhibitory effect on Cav3.1 calcium current. The Journal of physiology *586*, 5349-5366.

Liu, S.L., Baute, G.J., and Adams, K.L. (2011). Organ and cell type-specific complementary expression patterns and regulatory neofunctionalization between duplicated genes in Arabidopsis thaliana. Genome biology and evolution *3*, 1419-1436.

Liu, S.L., Pan, A.Q., and Adams, K.L. (2014). Protein subcellular relocalization of duplicated genes in Arabidopsis. Genome biology and evolution *6*, 2501-2515.

Llinas, R., and Yarom, Y. (1981). Electrophysiology of mammalian inferior olivary neurones in vitro. Different types of voltage-dependent ionic conductances. The Journal of physiology *315*, 549-567.

Llinas, R.R., Sugimori, M., and Cherksey, B. (1989). Voltage-dependent calcium conductances in mammalian neurons. The P channel. Annals of the New York Academy of Sciences *560*, 103-111.

Maher, G.J., Hilton, E.N., Urquhart, J.E., Davidson, A.E., Spencer, H.L., Black, G.C., and Manson, F.D. (2011). The cataract-associated protein TMEM114, and TMEM235, are glycosylated transmembrane proteins that are distinct from claudin family members. FEBS letters *585*, 2187-2192.

Mintz, I.M., Adams, M.E., and Bean, B.P. (1992). P-type calcium channels in rat central and peripheral neurons. Neuron *9*, 85-95.

Moran, Y., and Zakon, H.H. (2014). The evolution of the four subunits of voltage-gated calcium channels: ancient roots, increasing complexity, and multiple losses. Genome biology and evolution *6*, 2210-2217.

Moss, F.J., Viard, P., Davies, A., Bertaso, F., Page, K.M., Graham, A., Canti, C., Plumpton, M., Plumpton, C., Clare, J.J.*, et al.* (2002). The novel product of a five-exon stargazin-related gene abolishes Ca(V)2.2 calcium channel expression. The EMBO journal *21*, 1514-1523.

Newcomb, R., Szoke, B., Palma, A., Wang, G., Chen, X., Hopkins, W., Cong, R., Miller, J., Urge, L., Tarczy-Hornoch, K.*, et al.* (1998). Selective peptide antagonist of the class E calcium channel from the venom of the tarantula Hysterocrates gigas. Biochemistry *37*, 15353-15362.

Nowycky, M.C., Fox, A.P., and Tsien, R.W. (1985). Three types of neuronal calcium channel with different calcium agonist sensitivity. Nature *316*, 440-443.

Ohno, S. (1970). Evolution by Gene Duplication. New York: Springer-Verlag.

Olivera, B.M., Miljanich, G.P., Ramachandran, J., and Adams, M.E. (1994). Calcium channel diversity and neurotransmitter release: the omega-conotoxins and omega-agatoxins. Annual review of biochemistry *63*, 823-867.

Qiu, Y., Liu, S.L., and Adams, K.L. (2014). Frequent changes in expression profile and accelerated sequence evolution of duplicated imprinted genes in arabidopsis. Genome biology and evolution *6*, 1830-1842.

Randall, A., and Tsien, R.W. (1995). Pharmacological dissection of multiple types of Ca2+ channel currents in rat cerebellar granule neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience *15*, 2995-3012.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics *19*, 1572-1574.

Rousset, M., Cens, T., Restituito, S., Barrere, C., Black, J.L., 3rd, McEnery, M.W., and Charnet, P. (2001). Functional roles of gamma2, gamma3 and gamma4, three new Ca2+ channel subunits, in P/Q-type Ca2+ channel expressed in Xenopus oocytes. The Journal of physiology *532*, 583-593.

Schnell, E., Sizemore, M., Karimzadegan, S., Chen, L., Bredt, D.S., and Nicoll, R.A. (2002). Direct interactions between PSD-95 and stargazin control synaptic AMPA receptor number. Proceedings of the National Academy of Sciences of the United States of America *99*, 13902-13907.

Sigrist, C.J., de Castro, E., Cerutti, L., Cuche, B.A., Hulo, N., Bridge, A., Bougueleret, L., and Xenarios, I. (2013). New and continuing developments at PROSITE. Nucleic acids research *41*, D344-347.

Singer, D., Biel, M., Lotan, I., Flockerzi, V., Hofmann, F., and Dascal, N. (1991). The roles of the subunits in the function of the calcium channel. Science *253*, 1553-1557.

Sonnhammer, E.L., von Heijne, G., and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol *6*, 175-182.

Takahashi, M., Seagar, M.J., Jones, J.F., Reber, B.F., and Catterall, W.A. (1987). Subunit structure of dihydropyridine-sensitive calcium channels from skeletal muscle. Proceedings of the National Academy of Sciences of the United States of America *84*, 5478-5482.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular biology and evolution *30*, 2725-2729.

Tomita, S., Chen, L., Kawasaki, Y., Petralia, R.S., Wenthold, R.J., Nicoll, R.A., and Bredt, D.S. (2003). Functional studies and distribution define a family of transmembrane AMPA receptor regulatory proteins. The Journal of cell biology *161*, 805-816.

Tsien, R.W., Lipscombe, D., Madison, D.V., Bley, K.R., and Fox, A.P. (1988). Multiple types of neuronal calcium channels and their selective modulation. Trends in neurosciences *11*, 431-438.

Van Itallie, C.M., and Anderson, J.M. (2006). Claudins and epithelial paracellular transport. Annu Rev Physiol *68*, 403-429.

Williams, M.E., Feldman, D.H., McCue, A.F., Brenner, R., Velicelebi, G., Ellis, S.B., and Harpold, M.M. (1992). Structure and functional expression of alpha 1, alpha 2, and beta subunits of a novel human neuronal calcium channel subtype. Neuron *8*, 71-84.

Xu, B., and Yang, Z. (2013). PAMLX: a graphical user interface for PAML. Molecular biology and evolution *30*, 2723-2724.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution *24*, 1586-1591.

List of tables, figures, and appendices

Table I. CACNG site model analysis with gene free

Gene tree	Sequences number	M0				M1-M2 comparison		M7-M8 comparison	
		Tree length	dN/dS	dN	dS	2*(Lnl2-Lnl1)	p-value	2*(Lnl2-Lnl1)	p-value
CACNG1	16	8.5349	0.15973	1.1631	7.2814	0	1	0.510284	0.774806
CACNG2	14	5.25909	0.04618	0.2704	5.8548	0	1	3.436488	0.179381
CACNG3	12	12.24084	0.05789	0.7575	13.0861	2E-06	0.999999	1.381826	0.501118
CACNG4	13	10.72179	0.08051	0.8611	10.6958	2E-06	0.999999	0.00624	0.996883
CACNG5	15	6.35593	0.05223	0.355	6.7978	0	1	0.005556	0.997226
CACNG6	16	10.16713	0.1797	1.5309	8.5195	0	1	0.45282	0.797391
CACNG7	15	6.56445	0.06662	0.4641	6.9667	0.008826	0.995597	13.3591	0.001256
CACNG8	13	7.53485	0.12662	0.8995	7.1039	0	1	0.001146	0.999427
TMEM114	8	6.34149	0.16995	0.8979	5.2832	0	1	0.512458	0.773965
TMEM235	12	10.93991	0.31792	2.3235	7.3084	5.593118	1	5.593118	0.06102

Table II. CACNG site model analysis with species tree

Species tree	Sequences number	M0				M1-M2 comparison		M7-M8 comparison	
		Tree length	dN/dS	dN	dS	2*(Lnl2-Lnl1)	p-value	2*(Lnl2-Lnl1)	p-value
CACNG1	16	9.32493	0.16365	1.2943	7.909	0	1	2.317522	0.313875
CACNG2	14	5.23207	0.04434	0.2569	5.793	0	1	0.001298	0.999351
CACNG3	12	12.39625	0.07342	0.9514	12.9586	0	1	1.73395	0.420221
CACNG4	13	10.94704	0.06972	0.7725	11.08	0	1	0.00032	0.99984
CACNG5	15	6.76415	0.05702	0.4082	7.1585	0	1	0.808956	0.667325
CACNG6	16	10.40955	0.1986	1.6772	8.4453	0	1	2.451238	0.293576
CACNG7	15	7.0353	0.06859	0.5126	7.4733	1.055772	0.589851	17.90425	0.000129
CACNG8	13	7.93481	0.13274	0.9831	7.4065	0	1	0.003916	0.998044
TMEM114	8	6.54774	0.1654	0.9101	5.5021	0	1	0.709556	0.701329
TMEM235	12	10.52262	0.30361	2.1733	7.158	0	1	6.552298	0.037773

Table III. CACNG2 and CACNG3 LRT statistics of branch model

CACNG2_CACNG3	Number of sequences	Test 1	Test 2	2*(Lnl2-Lnl1)	p-value
after duplication	26	-13383.79829	-13375.83587	15.92482	6.59E-05
after speciation	26	-13353.89169	-13344.68317	9.208523	0.002409
between duplication and speciation	26	-13409.74599	-13409.26152	0.484469	0.486405

Table IV. CACNG4 and CACNG8 LRT statistics of branch model

CACNG4_CACNG8	Number of sequences	Test 1	Test 2	2*(Lnl2-Lnl1)	p-value
after duplication	26	-17889.49512	-17867.38245	44.22534	2.92668E-11
after speciation	26	-17879.77671	-17857.08229	45.38883	1.61553E-11
between duplication and speciation	26	-17870.41057	-17868.81223	3.196698	0.073787107

Table V. TMEM114 and TMEM235 LRT statistics of branch model

TMEM114_TMEM235	Number of sequences	Test 1	Test 2	2*(Lnl2-Lnl1)	p-value
after duplication	20	-11914.85982	-11901.29475	27.13014	1.9E-07
after speciation	20	-11914.53602	-11899.46044	30.15115	4E-08
between duplication and speciation	20	-11907.91188	-11907.8566	0.110568	0.739498

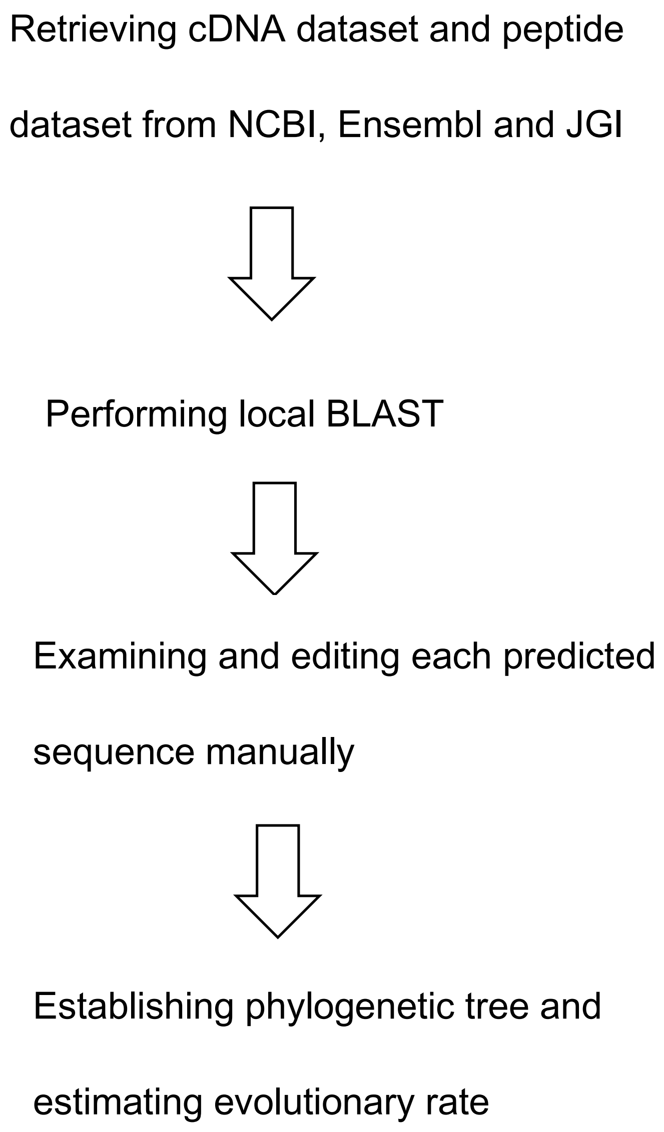


Figure 1. Work flow of the evolutionary analysis. It displays the order of tasks for this study. In the beginning, all the sequences were retrieved from databases with local BLAST. After examining and editing, we performed various evolutionary analyze.

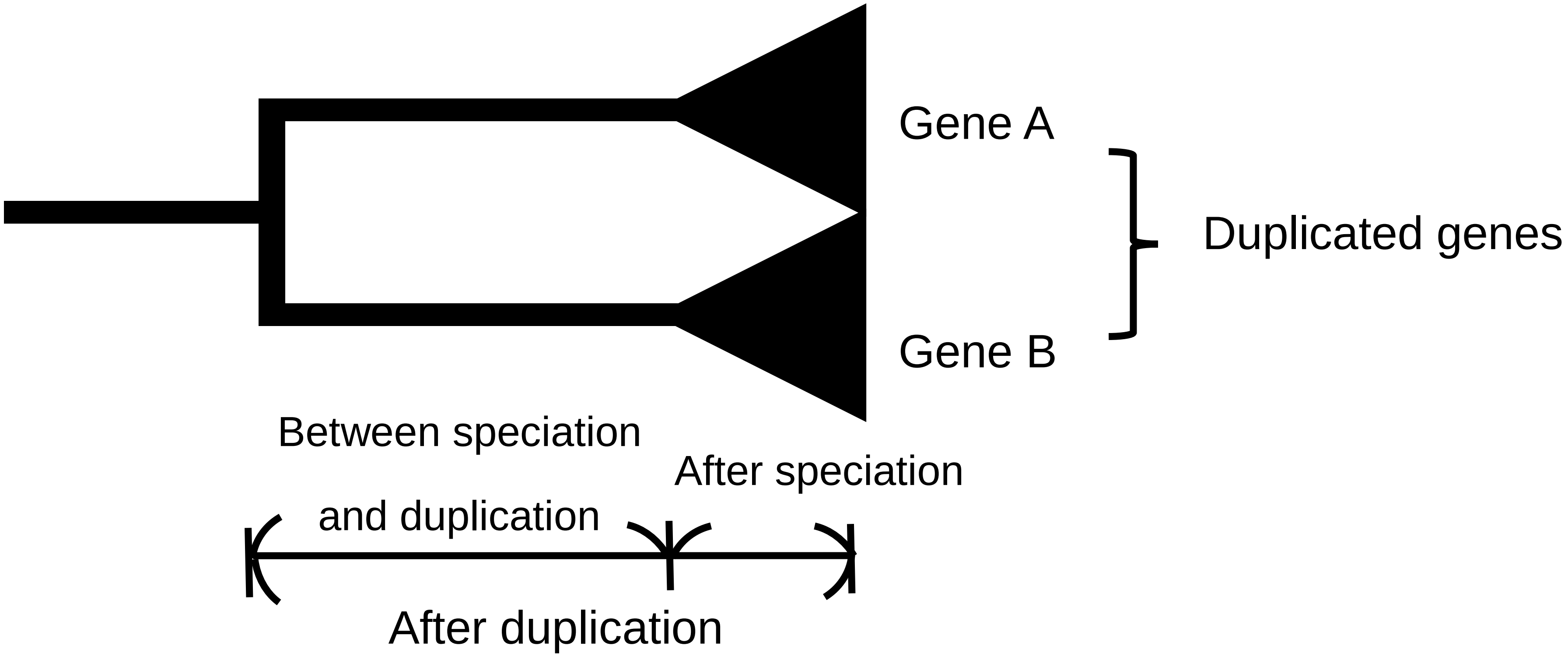


Figure 2. The different time points tested in branch model for duplicated genes. The duplicated genes were analyzed in different time points, after duplication, after speciation and between speciation and duplication included. Gene A and gene B are duplicated genes that may be paralogous or orthologous.

Figure 3.

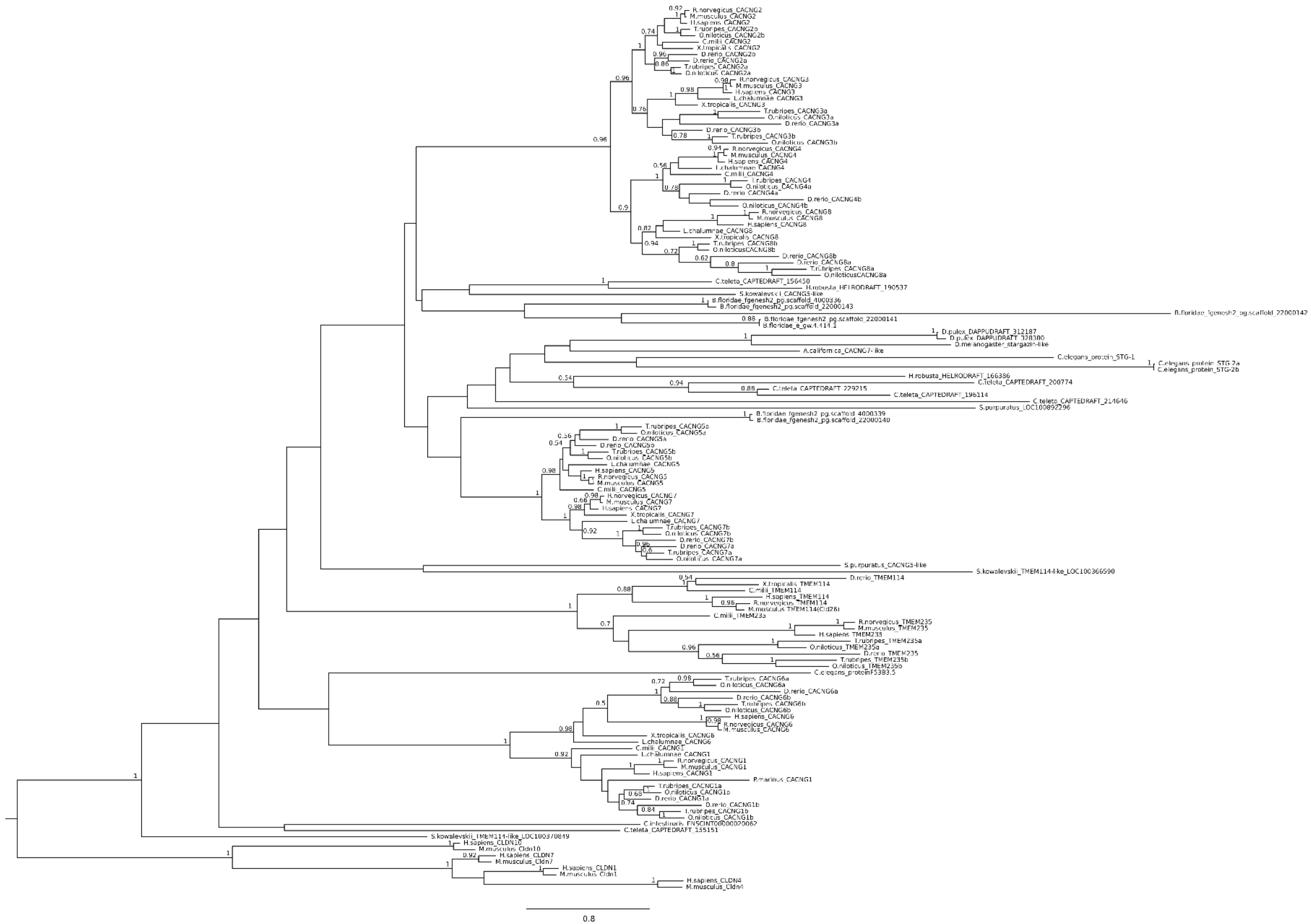


Figure 3. The phylogenetic tree of all the γ subunits by MEGA. It demonstrates phylogenetic relationship of CACNG family and γ-related subunits in invertebrate. The distribution of CACNG were classified into multiple clades (γ_2, γ_3, γ_4, γ_8), (γ_5, γ_7), (TMEM114, TMEM235) and (γ_1, γ_6). However, none of γ-related subunits were grouped into any CACNG family. The bootstrap value did not show >0.5 (50%) on most of branches. The tree was rooted with mouse's and human's claudin subunits, claudin 1, claudin 4, claudin 7 and claudin 10.

Figure 4.

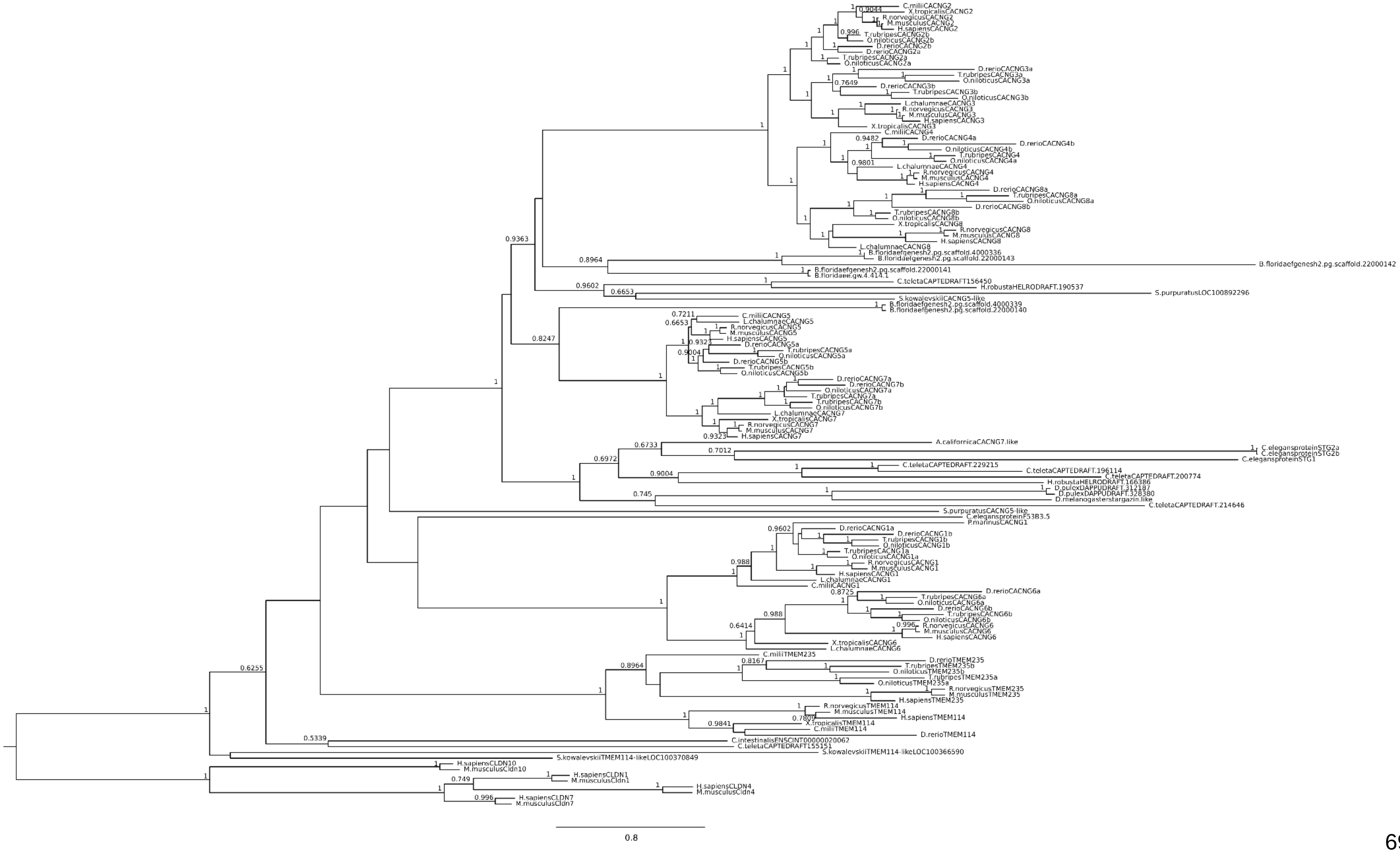

C.miliiCACNG2
X.tropicalisCACNG2
R.norvegicusCACNG2
M.musculusCACNG2
H.sapiensCACNG2
T.rubripesCACNG2b
O.niloticusCACNG2b
D.rerioCACNG2b
D.rerioCACNG2a
T.rubripesCACNG2a
O.niloticusCACNG2a
D.rerioCACNG3a
T.rubripesCACNG3a
O.niloticusCACNG3a
D.rerioCACNG3b
T.rubripesCACNG3b
O.niloticusCACNG3b
L.chalumnaeCACNG3
R.norvegicusCACNG3
M.musculusCACNG3
H.sapiensCACNG3
X.tropicalisCACNG3
C.miliiCACNG4
D.rerioCACNG4a
D.rerioCACNG4b
O.niloticusCACNG4b
T.rubripesCACNG4
O.niloticusCACNG4a
L.chalumnaeCACNG4
R.norvegicusCACNG4
M.musculusCACNG4
H.sapiensCACNG4
D.rerioCACNG8a
T.rubripesCACNG8a
O.niloticusCACNG8a
D.rerioCACNG8b
T.rubripesCACNG8b
O.niloticusCACNG8b
X.tropicalisCACNG8
R.norvegicusCACNG8
M.musculusCACNG8
H.sapiensCACNG8
L.chalumnaeCACNG8
B.floridaefgenesh2.pg.scaffold.4000336
B.floridaefgenesh2.pg.scaffold.22000143
B.floridaefgenesh2.pg.scaffold.22000142
B.floridaefgenesh2.pg.scaffold.22000141
B.floridaee.gw.4.414.1
C.teletaCAPTEDRAFT156450
H.robustaHELRODRAFT.190537
S.purpuratusLOC100892296
S.kowalevskiiCACNG5-like
B.floridaefgenesh2.pg.scaffold.4000339
B.floridaefgenesh2.pg.scaffold.22000140
C.miliiCACNG5
L.chalumnaeCACNG5
R.norvegicusCACNG5
M.musculusCACNG5
H.sapiensCACNG5
D.rerioCACNG5a
T.rubripesCACNG5a
O.niloticusCACNG5a
D.rerioCACNG5b
T.rubripesCACNG5b
O.niloticusCACNG5b
D.rerioCACNG7a
D.rerioCACNG7b
O.niloticusCACNG7a
T.rubripesCACNG7a
T.rubripesCACNG7b
O.niloticusCACNG7b
L.chalumnaeCACNG7
X.tropicalisCACNG7
R.norvegicusCACNG7
M.musculusCACNG7
H.sapiensCACNG7
A.californicaCACNG7.like
C.elegansproteinSTG2a
C.elegansproteinSTG2b
C.elegansproteinSTG1
C.teletaCAPTEDRAFT.229215
C.teletaCAPTEDRAFT.196114
C.teletaCAPTEDRAFT.200774
H.robustaHELRODRAFT.166386
D.pulexDAPPUDRAFT.312187
D.pulexDAPPUDRAFT.328380
D.melanogasterstargazin.like
C.teletaCAPTEDRAFT.214646
S.purpuratusCACNG5-like
C.elegansproteinF53B3.5
P.marinusCACNG1
D.rerioCACNG1a
D.rerioCACNG1b
T.rubripesCACNG1b
O.niloticusCACNG1b
T.rubripesCACNG1a
O.niloticusCACNG1a
R.norvegicusCACNG1
M.musculusCACNG1
H.sapiensCACNG1
L.chalumnaeCACNG1
C.miliiCACNG1
D.rerioCACNG6a
T.rubripesCACNG6a
O.niloticusCACNG6a
D.rerioCACNG6b
T.rubripesCACNG6b
O.niloticusCACNG6b
R.norvegicusCACNG6
M.musculusCACNG6
H.sapiensCACNG6
X.tropicalisCACNG6
L.chalumnaeCACNG6
C.miliiTMEM235
D.rerioTMEM235
T.rubripesTMEM235b
O.niloticusTMEM235b
T.rubripesTMEM235a
O.niloticusTMEM235a
R.norvegicusTMEM235
M.musculusTMEM235
H.sapiensTMEM235
R.norvegicusTMEM114
M.musculusTMEM114
H.sapiensTMEM114
X.tropicalisTMEM114
C.miliiTMEM114
D.rerioTMEM114
C.intestinalisENSCINT00000020062
C.teletaCAPTEDRAFT155151
S.kowalevskiiTMEM114-likeLOC100366590
S.kowalevskiiTMEM114-likeLOC100370849
H.sapiensCLDN10
M.musculusCldn10
H.sapiensCLDN1
M.musculusCldn1
H.sapiensCLDN4
M.musculusCldn4
H.sapiensCLDN7
M.musculusCldn7
0.8

Figure 4. The phylogenetic tree of all the γ subunits by MrBayse. It demonstrates phylogenetic relationship of CACNG family and γ-related subunits in invertebrate. The distribution of CACNG were classified into multiple clades (γ_2, γ_3, γ_4, γ_8), (γ_5, γ_7), (TMEM114, TMEM235) and (γ_1, γ_6). However, none of γ-related subunits were grouped into any CACNG family. The posterior probability did not show >0.5 (50%) on many branches. The tree was rooted with mouse's and human's claudin subunits, claudin 1, claudin 4, claudin 7 and claudin 10.

Figure 5.

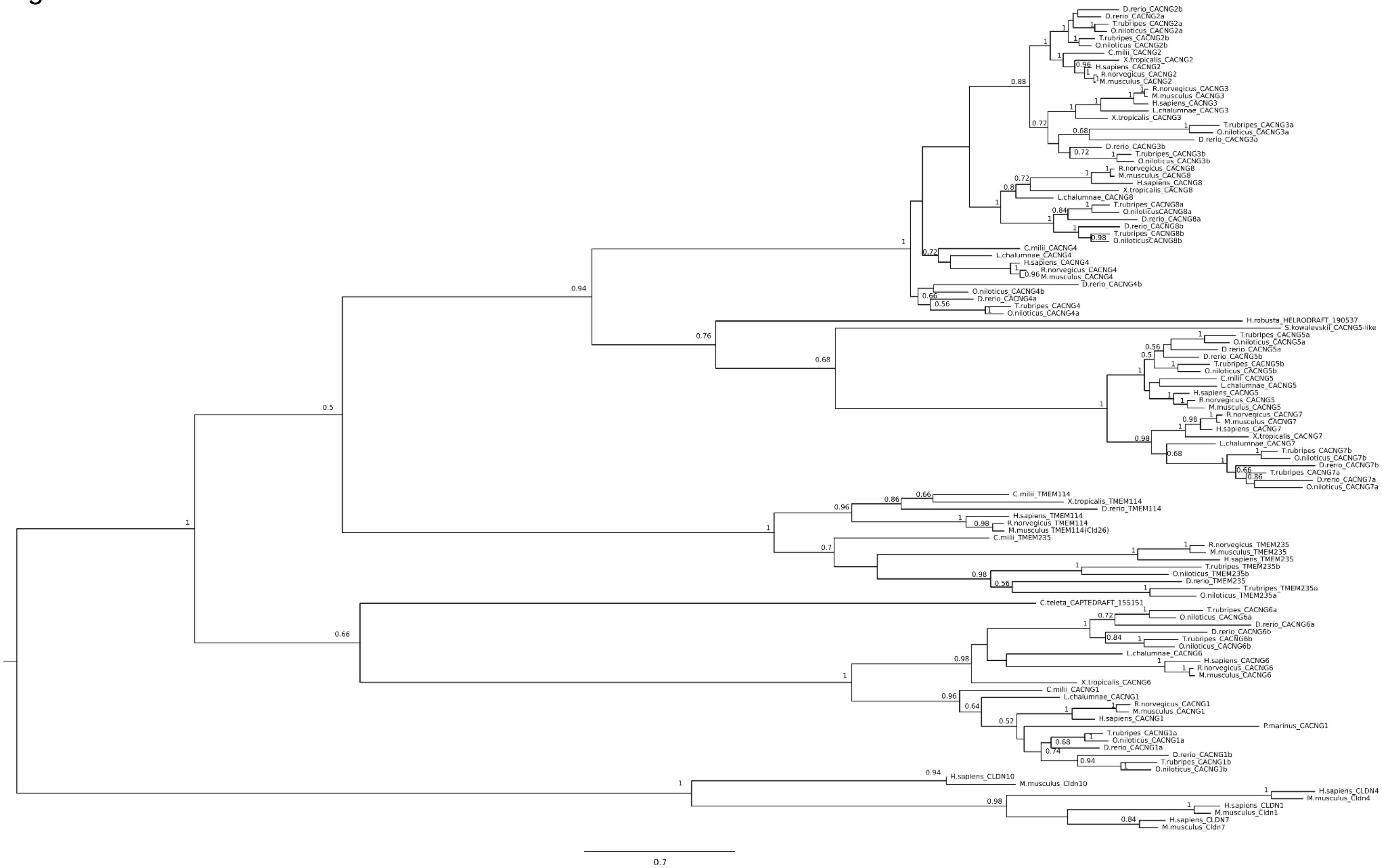

D.rerio_CACNG2b
D.rerio_CACNG2a
T.rubripes_CACNG2a
O.niloticus_CACNG2a
T.rubripes_CACNG2b
O.niloticus_CACNG2b
C.milii_CACNG2
X.tropicalis_CACNG2
H.sapiens_CACNG2
R.norvegicus_CACNG2
M.musculus_CACNG2
R.norvegicus_CACNG3
M.musculus_CACNG3
H.sapiens_CACNG3
L.chalumnae_CACNG3
X.tropicalis_CACNG3
T.rubripes_CACNG3a
O.niloticus_CACNG3a
D.rerio_CACNG3a
D.rerio_CACNG3b
T.rubripes_CACNG3b
O.niloticus_CACNG3b
R.norvegicus_CACNG8
M.musculus_CACNG8
H.sapiens_CACNG8
X.tropicalis_CACNG8
L.chalumnae_CACNG8
T.rubripes_CACNG8a
O.niloticusCACNG8a
D.rerio_CACNG8a
D.rerio_CACNG8b
T.rubripes_CACNG8b
O.niloticusCACNG8b
C.milii_CACNG4
L.chalumnae_CACNG4
H.sapiens_CACNG4
R.norvegicus_CACNG4
M.musculus_CACNG4
D.rerio_CACNG4b
O.niloticus_CACNG4b
D.rerio_CACNG4a
T.rubripes_CACNG4
O.niloticus_CACNG4a
H.robusta_HELRODRAFT_190537
S.kowalevskii_CACNG5-like
T.rubripes_CACNG5a
O.niloticus_CACNG5a
D.rerio_CACNG5a
D.rerio_CACNG5b
T.rubripes_CACNG5b
O.niloticus_CACNG5b
C.milii_CACNG5
L.chalumnae_CACNG5
H.sapiens_CACNG5
R.norvegicus_CACNG5
M.musculus_CACNG5
R.norvegicus_CACNG7
M.musculus_CACNG7
H.sapiens_CACNG7
X.tropicalis_CACNG7
L.chalumnae_CACNG7
T.rubripes_CACNG7b
O.niloticus_CACNG7b
D.rerio_CACNG7b
T.rubripes_CACNG7a
D.rerio_CACNG7a
O.niloticus_CACNG7a
C.milii_TMEM114
X.tropicalis_TMEM114
D.rerio_TMEM114
H.sapiens_TMEM114
R.norvegicus_TMEM114
M.musculus_TMEM114(Cld26)
C.milii_TMEM235
R.norvegicus_TMEM235
M.musculus_TMEM235
H.sapiens_TMEM235
T.rubripes_TMEM235b
O.niloticus_TMEM235b
D.rerio_TMEM235
T.rubripes_TMEM235a
O.niloticus_TMEM235a
C.teleta_CAPTEDRAFT_155151
T.rubripes_CACNG6a
O.niloticus_CACNG6a
D.rerio_CACNG6a
D.rerio_CACNG6b
T.rubripes_CACNG6b
O.niloticus_CACNG6b
L.chalumnae_CACNG6
H.sapiens_CACNG6
R.norvegicus_CACNG6
M.musculus_CACNG6
X.tropicalis_CACNG6
C.milii_CACNG1
L.chalumnae_CACNG1
R.norvegicus_CACNG1
M.musculus_CACNG1
H.sapiens_CACNG1
P.marinus_CACNG1
T.rubripes_CACNG1a
O.niloticus_CACNG1a
D.rerio_CACNG1a
D.rerio_CACNG1b
T.rubripes_CACNG1b
O.niloticus_CACNG1b
H.sapiens_CLDN10
M.musculus_Cldn10
H.sapiens_CLDN4
M.musculus_Cldn4
H.sapiens_CLDN1
M.musculus_Cldn1
H.sapiens_CLDN7
M.musculus_Cldn7
0.7

Figure 5. The phylogenetic tree of vertebrate γ subunits by MEGA. It demonstrates phylogenetic relationship of CACNG family and three γ-related subunits in invertebrate. The distribution of CACNG were classified into multiple clades (γ2, γ3, γ4, γ8), (γ5, γ7), (TMEM114, TMEM235) and (γ1, γ6). However, the three γ-related, H.robusta HELRODRAFT 190537, S.kowalevskii CACNG5-like and C.teleta CAPTEDRAFT 155151 subunits were not grouped into any CACNG family. The bootstrap value showed >0.5 (50%) on most of branches. The tree was rooted with mouse’s and human’s claudin subunits, claudin 1, claudin 4, claudin 7 and claudin 10.

Figure 6.

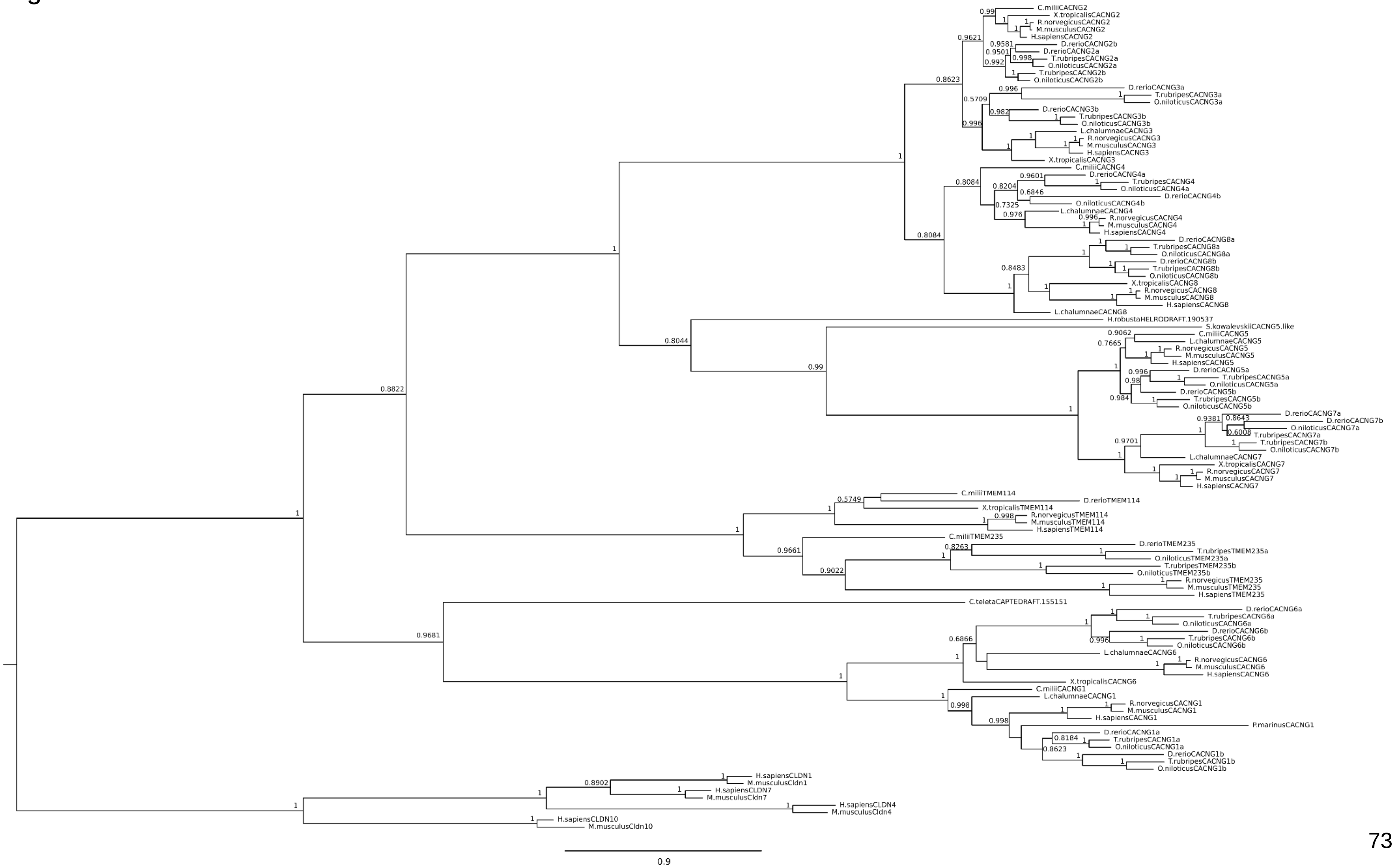

C.miliiCACNG2
X.tropicalisCACNG2
R.norvegicusCACNG2
M.musculusCACNG2
H.sapiensCACNG2
D.rerioCACNG2b
D.rerioCACNG2a
T.rubripesCACNG2a
O.niloticusCACNG2a
T.rubripesCACNG2b
O.niloticusCACNG2b
D.rerioCACNG3a
T.rubripesCACNG3a
O.niloticusCACNG3a
D.rerioCACNG3b
T.rubripesCACNG3b
O.niloticusCACNG3b
L.chalumnaeCACNG3
R.norvegicusCACNG3
M.musculusCACNG3
H.sapiensCACNG3
X.tropicalisCACNG3
C.miliiCACNG4
D.rerioCACNG4a
T.rubripesCACNG4
O.niloticusCACNG4a
D.rerioCACNG4b
O.niloticusCACNG4b
L.chalumnaeCACNG4
R.norvegicusCACNG4
M.musculusCACNG4
H.sapiensCACNG4
D.rerioCACNG8a
T.rubripesCACNG8a
O.niloticusCACNG8a
D.rerioCACNG8b
T.rubripesCACNG8b
O.niloticusCACNG8b
X.tropicalisCACNG8
R.norvegicusCACNG8
M.musculusCACNG8
H.sapiensCACNG8
L.chalumnaeCACNG8
H.robustaHELRODRAFT.190537
S.kowalevskiiCACNG5.like
C.miliiCACNG5
L.chalumnaeCACNG5
R.norvegicusCACNG5
M.musculusCACNG5
H.sapiensCACNG5
D.rerioCACNG5a
T.rubripesCACNG5a
O.niloticusCACNG5a
D.rerioCACNG5b
T.rubripesCACNG5b
O.niloticusCACNG5b
D.rerioCACNG7a
D.rerioCACNG7b
O.niloticusCACNG7a
T.rubripesCACNG7a
T.rubripesCACNG7b
O.niloticusCACNG7b
L.chalumnaeCACNG7
X.tropicalisCACNG7
R.norvegicusCACNG7
M.musculusCACNG7
H.sapiensCACNG7
C.miliiTMEM114
D.rerioTMEM114
X.tropicalisTMEM114
R.norvegicusTMEM114
M.musculusTMEM114
H.sapiensTMEM114
C.miliiTMEM235
D.rerioTMEM235
T.rubripesTMEM235a
O.niloticusTMEM235a
T.rubripesTMEM235b
O.niloticusTMEM235b
R.norvegicusTMEM235
M.musculusTMEM235
H.sapiensTMEM235
C.teletaCAPTEDRAFT.155151
D.rerioCACNG6a
T.rubripesCACNG6a
O.niloticusCACNG6a
D.rerioCACNG6b
T.rubripesCACNG6b
O.niloticusCACNG6b
L.chalumnaeCACNG6
R.norvegicusCACNG6
M.musculusCACNG6
H.sapiensCACNG6
X.tropicalisCACNG6
C.miliiCACNG1
L.chalumnaeCACNG1
R.norvegicusCACNG1
M.musculusCACNG1
H.sapiensCACNG1
P.marinusCACNG1
D.rerioCACNG1a
T.rubripesCACNG1a
O.niloticusCACNG1a
D.rerioCACNG1b
T.rubripesCACNG1b
O.niloticusCACNG1b
H.sapiensCLDN1
M.musculusCldn1
H.sapiensCLDN7
M.musculusCldn7
H.sapiensCLDN4
M.musculusCldn4
H.sapiensCLDN10
M.musculusCldn10
0.9

Figure 6. The phylogenetic tree of vertebrate γ subunits by MrBayse. It demonstrates phylogenetic relationship of CACNG family and three γ-related subunits in invertebrate. The distribution of CACNG were classified into multiple clades (γ2, γ3, γ4, γ8), (γ5, γ7), (TMEM114, TMEM235) and (γ1, γ6). However, the three γ-related, H.robusta HELRODRAFT 190537, S.kowalevskii CACNG5-like and C.teleta CAPTEDRAFT 155151 subunits were not grouped into any CACNG family. The posterior probability showed >0.5 (50%) on most of branches. The tree was rooted with mouse's and human's claudin subunits, claudin 1, claudin 4, claudin 7 and claudin 10.

Figure 7.

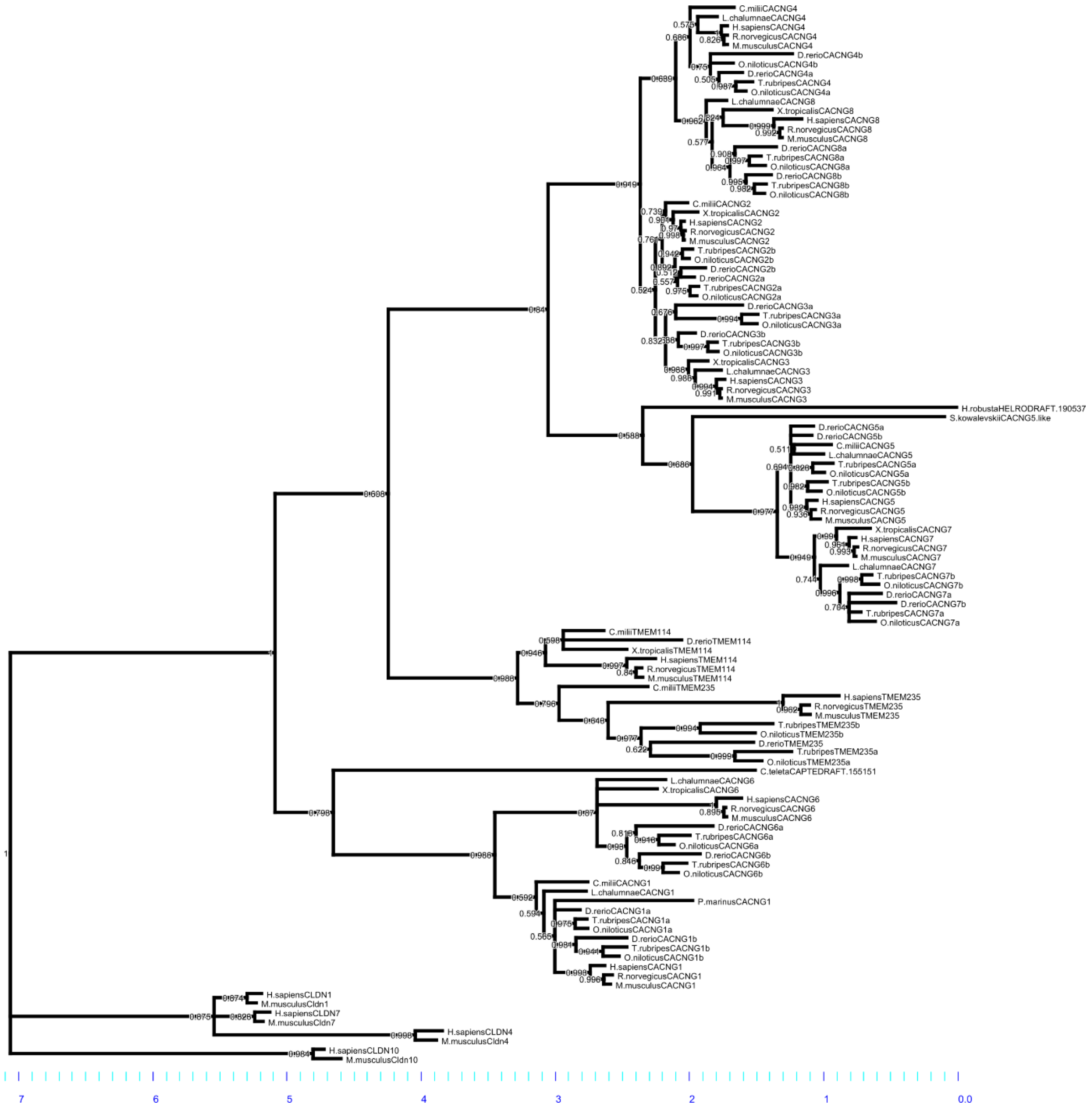
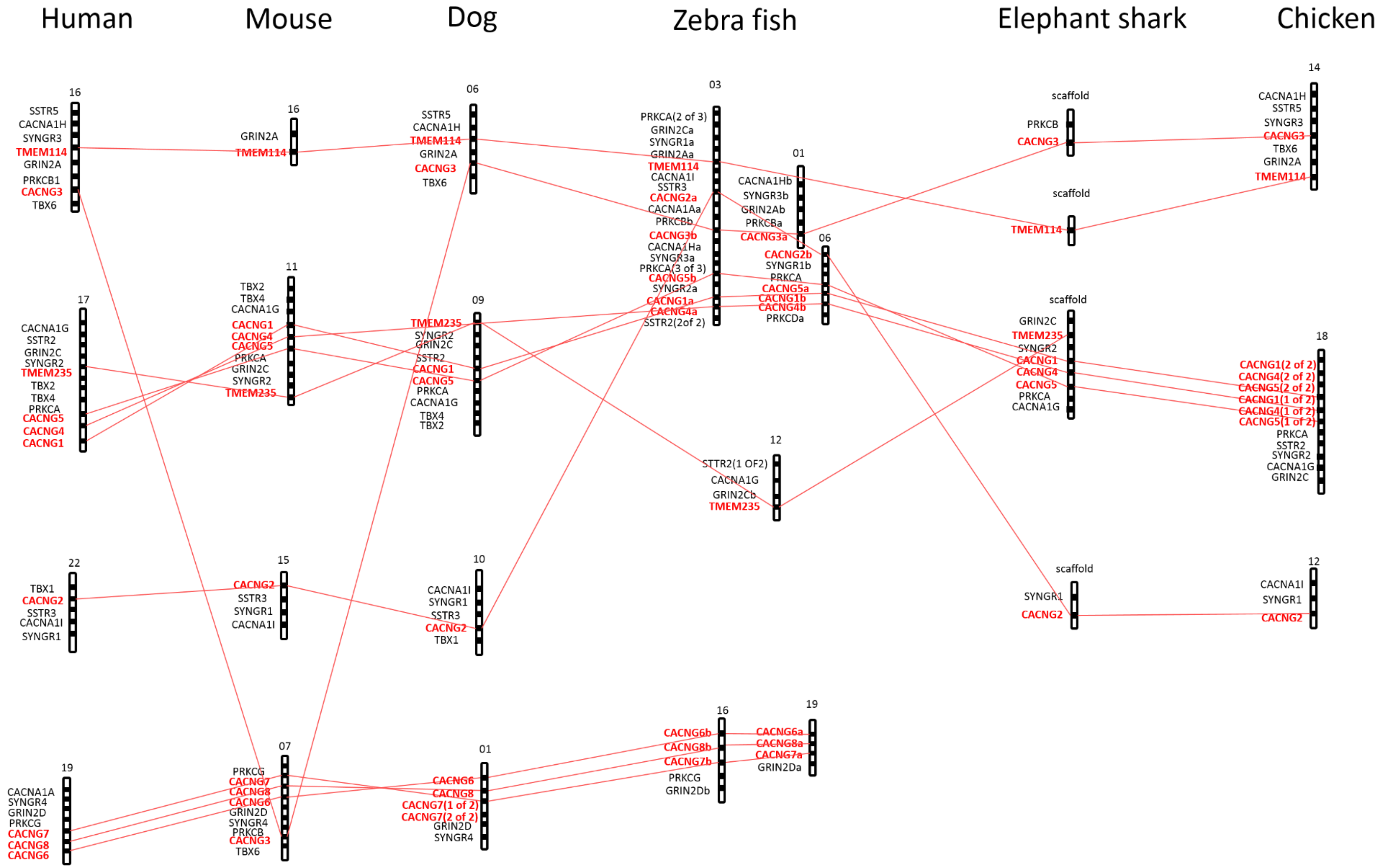

C.miliiCACNG4
L.chalumnaeCACNG4
H.sapiensCACNG4
R.norvegicusCACNG4
M.musculusCACNG4
D.rerioCACNG4b
O.niloticusCACNG4b
D.rerioCACNG4a
T.rubripesCACNG4
O.niloticusCACNG4a
L.chalumnaeCACNG8
X.tropicalisCACNG8
H.sapiensCACNG8
R.norvegicusCACNG8
M.musculusCACNG8
D.rerioCACNG8a
T.rubripesCACNG8a
O.niloticusCACNG8a
D.rerioCACNG8b
T.rubripesCACNG8b
O.niloticusCACNG8b
C.miliiCACNG2
X.tropicalisCACNG2
H.sapiensCACNG2
R.norvegicusCACNG2
M.musculusCACNG2
T.rubripesCACNG2b
O.niloticusCACNG2b
D.rerioCACNG2b
D.rerioCACNG2a
T.rubripesCACNG2a
O.niloticusCACNG2a
D.rerioCACNG3a
T.rubripesCACNG3a
O.niloticusCACNG3a
D.rerioCACNG3b
T.rubripesCACNG3b
O.niloticusCACNG3b
X.tropicalisCACNG3
L.chalumnaeCACNG3
H.sapiensCACNG3
R.norvegicusCACNG3
M.musculusCACNG3
H.robustaHELRODRAFT.190537
S.kowalevskiiCACNG5.like
D.rerioCACNG5a
D.rerioCACNG5b
C.miliiCACNG5
L.chalumnaeCACNG5
T.rubripesCACNG5a
O.niloticusCACNG5a
T.rubripesCACNG5b
O.niloticusCACNG5b
H.sapiensCACNG5
R.norvegicusCACNG5
M.musculusCACNG5
X.tropicalisCACNG7
H.sapiensCACNG7
R.norvegicusCACNG7
M.musculusCACNG7
L.chalumnaeCACNG7
T.rubripesCACNG7b
O.niloticusCACNG7b
D.rerioCACNG7a
D.rerioCACNG7b
T.rubripesCACNG7a
O.niloticusCACNG7a
C.miliiTMEM114
D.rerioTMEM114
X.tropicalisTMEM114
H.sapiensTMEM114
R.norvegicusTMEM114
M.musculusTMEM114
C.miliiTMEM235
H.sapiensTMEM235
R.norvegicusTMEM235
M.musculusTMEM235
T.rubripesTMEM235b
O.niloticusTMEM235b
D.rerioTMEM235
T.rubripesTMEM235a
O.niloticusTMEM235a
C.teletaCAPTEDRAFT.155151
L.chalumnaeCACNG6
X.tropicalisCACNG6
H.sapiensCACNG6
R.norvegicusCACNG6
M.musculusCACNG6
D.rerioCACNG6a
T.rubripesCACNG6a
O.niloticusCACNG6a
D.rerioCACNG6b
T.rubripesCACNG6b
O.niloticusCACNG6b
C.miliiCACNG1
L.chalumnaeCACNG1
P.marinusCACNG1
D.rerioCACNG1a
T.rubripesCACNG1a
O.niloticusCACNG1a
D.rerioCACNG1b
T.rubripesCACNG1b
O.niloticusCACNG1b
H.sapiensCACNG1
R.norvegicusCACNG1
M.musculusCACNG1
H.sapiensCLDN1
M.musculusCldn1
H.sapiensCLDN7
M.musculusCldn7
H.sapiensCLDN4
M.musculusCldn4
H.sapiensCLDN10
M.musculusCldn10
7
6
5
4
3
2
1
0.0

Figure 7. The phylogenetic tree of vertebrate γ subunits by GARLI. It demonstrates phylogenetic relationship of CACNG family and three γ-related subunits in invertebrate. The distribution of CACNG were classified into multiple clades (γ2, γ3, γ4, γ8), (γ5, γ7), (TMEM114, TMEM235) and (γ1, γ6). However, the three γ-related, H.robusta HELRODRAFT 190537, S.kowalevskii CACNG5-like and C.teleta CAPTEDRAFT 155151 subunits were not grouped into any CACNG family. The bootstrap value showed >0.5 (50%) on most of branches. The tree was rooted with mouse's and human's claudin subunits, claudin 1, claudin 4, claudin 7 and claudin 10.

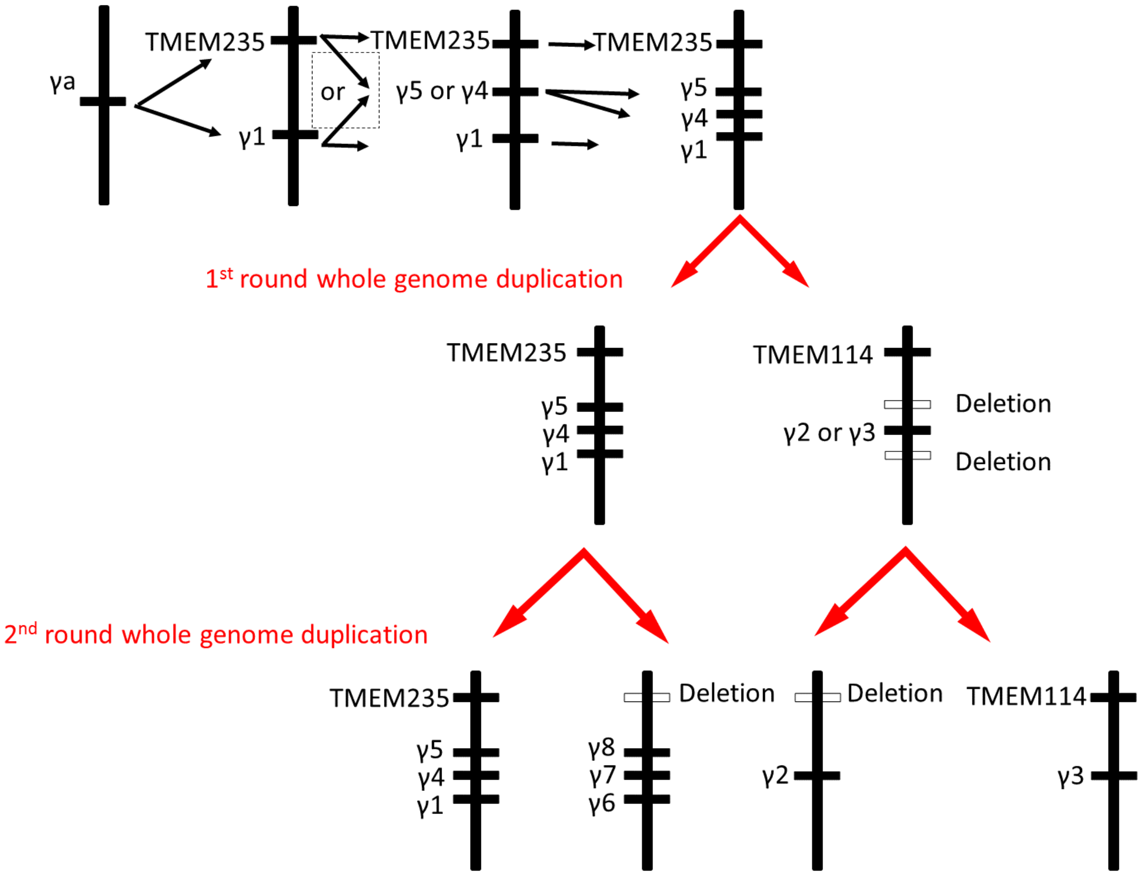
Figure 8.



Human
Mouse
Dog
Zebra fish
Elephant shark
Chicken
16
SSTR5
CACNA1H
SYNGR3
TMEM114
GRIN2A
PRKCB1
CACNG3
TBX6
17
CACNA1G
SSTR2
GRIN2C
SYNGR2
TMEM235
TBX2
TBX4
PRKCA
CACNG5
CACNG4
CACNG1
22
TBX1
CACNG2
SSTR3
CACNA1I
SYNGR1
19
CACNA1A
SYNGR4
GRIN2D
PRKCG
CACNG7
CACNG8
CACNG6
16
GRIN2A
TMEM114
11
TBX2
TBX4
CACNA1G
CACNG1
CACNG4
CACNG5
PRKCA
GRIN2C
SYNGR2
TMEM235
15
CACNG2
SSTR3
SYNGR1
CACNA1I
07
PRKCG
CACNG7
CACNG8
CACNG6
GRIN2D
SYNGR4
PRKCB
CACNG3
TBX6
06
SSTR5
CACNA1H
TMEM114
GRIN2A
CACNG3
TBX6
09
TMEM235
SYNGR2
GRIN2C
SSTR2
CACNG1
CACNG5
PRKCA
CACNA1G
TBX4
TBX2
10
CACNA1I
SYNGR1
SSTR3
CACNG2
TBX1
01
CACNG6
CACNG8
CACNG7(1 of 2)
CACNG7(2 of 2)
GRIN2D
SYNGR4
03
PRKCA(2 of 3)
GRIN2Ca
SYNGR1a
GRIN2Aa
TMEM114
CACNA1I
SSTR3
CACNG2a
CACNA1Aa
PRKCBb
CACNG3b
CACNA1Ha
SYNGR3a
PRKCA(3 of 3)
CACNG5b
SYNGR2a
CACNG1a
CACNG4a
SSTR2(2of 2)
01
CACNA1Hb
SYNGR3b
GRIN2Ab
PRKCBa
CACNG3a
06
CACNG2b
SYNGR1b
PRKCA
CACNG5a
CACNG1b
CACNG4b
PRKCDa
12
STTR2(1 OF2)
CACNA1G
GRIN2Cb
TMEM235
16
CACNG6b
CACNG8b
CACNG7b
PRKCG
GRIN2Db
19
CACNG6a
CACNG8a
CACNG7a
GRIN2Da
scaffold
PRKCB
CACNG3
scaffold
TMEM114
scaffold
GRIN2C
TMEM235
SYNGR2
CACNG1
CACNG4
CACNG5
PRKCA
CACNA1G
scaffold
SYNGR1
CACNG2
14
CACNA1H
SSTR5
SYNGR3
CACNG3
TBX6
GRIN2A
TMEM114
18
CACNG1(2 of 2)
CACNG4(2 of 2)
CACNG5(2 of 2)
CACNG1(1 of 2)
CACNG4(1 of 2)
CACNG5(1 of 2)
PRKCA
SSTR2
SYNGR2
CACNA1G
GRIN2C
12
CACNA1I
SYNGR1
CACNG2

Figure 8. The syntenic blocks of CACNG family and their surrounding genes on animal chromosomes. All CACNG family are labeled red color and connected with their orthologs. The results show that CACNG family experienced tandem duplications and translocations. For example, dog γ_7 genes have two copies. Frequently, the same of sets of paralogous genes are found on multiple chromosomes.

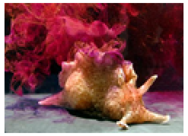
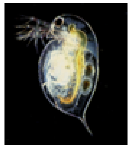
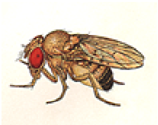
Figure 9.



γa
TMEM235
γ1
or
TMEM235
γ5 or γ4
γ1
TMEM235
γ5
γ4
γ1
1st round whole genome duplication
TMEM235
γ5
γ4
γ1
TMEM114
Deletion
γ2 or γ3
Deletion
2nd round whole genome duplication
TMEM235
γ5
γ4
γ1
Deletion
γ8
γ7
γ6
Deletion
γ2
TMEM114
γ3

Figure 9. CACNG family evolutionary pathway model. In this model, the ancestral γ gene went through multiple tandem duplications and two rounds of chromosome duplication. After the 1st round of whole genome duplication, each chromosome contained three γ's and a TMEM genes. Nevertheless, two γ genes were lost in one of chromosome. Finally, the γ genes were distributed on four chromosomes after the 2nd round of genome duplication. One copy of each TMEM114 and TMEM235 was lost.

Appendix I. γ and related proteins from various species that were retrieved from protein-protein BLAST. 201 proteins from 27 animals were listed.

	Species	Proteins
	*C. elegans* (Nematoda)	STG-1 STG-2a STG-2b F53B3.5
	*A. californica* (Mollusca)	γ7-like
	C. teleta (Annelids)	CAPTEDRAFT_229215 CAPTEDRAFT_155151 CAPTEDRAFT_196114 CAPTEDRAFT_214646 CAPTEDRAFT_200774 CAPTEDRAFT_156450
	H. Robusta (Annelids)	HELRODRAFT_166386 HELRODRAFT_190537
	D. pulex (Arthropoda)	DAPPUDRAFT_312187 DAPPUDRAFT_328380
	*D. melanogaster* (Arthropoda)	STG-like

S.kowalevskii
(Hemichordata)

CACNG5-like TMEM114-like LOC100366590

TMEM114-like LOC100370849

S. purpuratus
(Echinodermata)

LOC100892296 γ5-like

Chordata

B.Lanceolatum
(Cephalochordata)

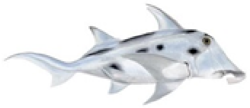
fgenesh2_pg.scaffold_4000336 fgenesh2_pg.scaffold_4000339
fgenesh2_pg.scaffold_22000141 fgenesh2_pg.scaffold_22000143
fgenesh2_pg.scaffold_22000142 fgenesh2_pg.scaffold_22000140
e_gw.4.414.1

C. intestinalis
(Urochordata)

ENSCINT00000020062

P. marinus
(Hyperoartia)

γ1

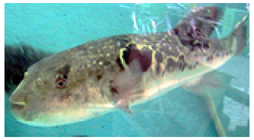


C. milli (Chorndrichthyes)

γ1 γ2 γ3 γ4 γ5 T114 T235

O. niloticus (Osteichthyes)

γ1a γ1b γ6a γ6b γ2a γ2b γ3a γ3b γ4a γ4b γ8a γ8b γ5a γ5b γ7a γ7b T235a T235b



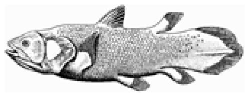
γ1a γ1b γ6a γ6b γ2a γ2b γ3a γ3b γ4a γ4b γ8a γ8b γ5a γ5b γ7a γ7b T235a T235b

T. rubripes (Osteichthyes)

γ1a γ1b γ6a γ6b γ2a γ2b γ3a γ3b γ4a γ4b γ8a γ8b γ5a γ5b γ7a γ7b T114 T235

D. rerio (Osteichthyes)

γ1 γ6 γ2 γ3 γ4 γ8 γ5 γ7



L. *chalumnae* (Sarcopterygii)

γ1 γ6 γ2 γ3 γ8 γ7 T114

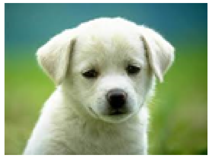
X. Tropicalis

A.Carolinensis (Reptilia) γ1 γ6 γ2 γ4 γ8 γ5 γ7 T114 T235

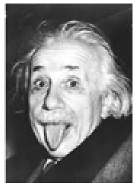
C. picta bellii (Reptilia) γ1 γ6 γ2 γ3 γ4 γ8 γ5 γ7 T114 T235

G. gallus (Aves) γ1a γ1b γ2 γ3 γ4a γ4b γ5a γ5b T114

M.domestica (Didelphimorphia) γ1 γ6-like γ2 γ3 γ4 γ7 T114



C. lupus familiaris (Carnivora) γ1 γ6 γ2 γ3 γ5 γ7a γ7b T114 T235

	*B. taurus* (Artiodactyla)	γ1	γ6	γ2	γ3	γ4	γ8	γ5	γ7	T114	T235
	*M. musculus* (Rodentia)	γ1	γ6	γ2	γ3	γ4	γ8	γ5	γ7	T114	T235
	*R. norvegicus* (Rodentia)	γ1	γ6	γ2	γ3	γ4	γ8	γ5	γ7	T114	T235
	*O. cuniculus* (Lagomorpha)	γ1	γ6	γ2	γ3	γ4		γ5	γ7	T114	T235
	*H. sapiens* (Primates)	γ1	γ6	γ2	γ3	γ4	γ8	γ5	γ7	T114	T235

Appendix II. The accession number of sequences used in this study. [a] The ID retrieved from Ensemble is showed initially "EN". [b] The ID comes to "XP", "NP", "EL" and "EF" from NCBI. [c] The ID obtained from JGI are marked with (JGI).

Organism	Gene	Ensemble ID[a]/Accession number[b]/JGI ID[C]
Caenorhabditis elegans	Protein STG-1	NP_001021976.1
	Protein STG-2b	NP_001024556.2
	Protein F53B3.5	NP_508499.3
	Protein STG-2a	NP_001024555.2
Aplysia californica	CACNG7-like	XP_005101013.1
Capitella teleta	CAPTEDRAFT_155151	ELT88457.1
	CAPTEDRAFT_229215	ELT93066.1
	CAPTEDRAFT_156450	ELU09240.1
	CAPTEDRAFT_196114	ELT88463.1
	CAPTEDRAFT_200774	ELU06940.1
Helobdella robusta	HELRODRAFT_190537	XP_009012613.1
	HELRODRAFT_166386	XP_009031568.1
Daphnia pulex	DAPPUDRAFT_312187	EFX87646.1
	DAPPUDRAFT_328380	EFX70146.1
Drosophila melanogaster	Stargazin-like protein	NP_001027082.1
Saccoglossus kowalevskii	CACNG5-like	XP_006816113.1
	TMEM114-like LOC100366590	XP_002732261.1
	TMEM114-like LOC100370849	XP_002734618.1
Strongylocentrotus purpuratus	LOC100892296	XP_003726426.1
	CACNG5-like	XP_003726972.1
Saccoglossus kowalevskii	CACNG5-like	XP_006816113.1
	TMEM114-like LOC100366590	XP_002732261.1
	TMEM114-like LOC100370849	XP_002734618.1
Branchiostoma floridae	fgenesh2_pg.scaffold_4000336	64581(JGI)
	fgenesh2_pg.scaffold_22000143	69205(JGI)
	fgenesh2_pg.scaffold_4000339	64584(JGI)

	fgenesh2_pg.scaffold_22000140	69202(JGI)
	fgenesh2_pg.scaffold_22000141	69203(JGI)
	e_gw.4.414.1	199899(JGI)
	fgenesh2_pg.scaffold_22000142	69204(JGI)

Petromyzon marinus	CACNG1	ENSPMAT00000005800
Callorhinchus milii	CACNG1	XP_007886924.1
	CACNG2	XP_007907969.1
	CACNG3	XP_007904006.1
	CACNG4	XP_007886925.1
	CACNG5	XP_007886926.1
	TMEM114	XP_007891968.1
	TMEM235	XP_007886845.1
Danio rerio	CACNG1a	ENSDART00000074199
	CACNG1b	ENSDART00000006843
	CACNG2a	ENSDART00000041388
	CACNG2b	ENSDART00000013939
	CACNG3a	ENSDART00000055023
	CACNG3b	ENSDART00000110126
	CACNG4a	ENSDART00000112043
	CACNG4b	ENSDART00000057325
	CACNG5a	ENSDART00000151099
	CACNG5b	ENSDART00000154367
	CACNG6a	ENSDART00000138310
	CACNG6b	ENSDART00000067741
	CACNG7a	ENSDART00000147128
	CACNG7b	ENSDART00000103891
	CACNG8a	ENSDART00000136842
	CACNG8b	ENSDART00000103894
	TMEM114	ENSDART00000055528
	TMEM235	ENSDART00000129501
Takifugu rubripes	CACNG1a	ENSTRUT00000026530
	CACNG1b	ENSTRUT00000001579
	CACNG2a	ENSTRUT00000000035
	CACNG2b	ENSTRUT00000000276
	CACNG3a	ENSTRUT00000023938
	CACNG3b	ENSTRUT00000025324
	CACNG4a	ENSTRUT00000026314

	CACNG4b (sequences not complete)	ENSTRUT00000001656
	CACNG5a	ENSTRUT00000001815
	CACNG5b	ENSTRUT00000041630
	CACNG6a	ENSTRUT00000015507
	CACNG6b	ENSTRUT00000012985
	CACNG7a	ENSTRUT00000016663
	CACNG7b	ENSTRUT00000011150
	CACNG8a	ENSTRUT00000015943
	CACNG8b	ENSTRUT00000011792
	TMEM235a	ENSTRUT00000029587
	TMEM235b	ENSTRUT00000043030
Oreochromis niloticus	CACNG1a	ENSONIT00000002829
	CACNG1b	ENSONIT00000008181
	CACNG2a	ENSONIT00000017424
	CACNG2b	ENSONIT00000012228
	CACNG3a	ENSONIT00000004180
	CACNG3b	ENSONIT00000019782
	CACNG4a	ENSONIT00000002830
	CACNG4b	ENSONIT00000008179
	CACNG5a	ENSONIT00000008177
	CACNG5b	ENSONIT00000021651
	CACNG6a	ENSONIT00000008877
	CACNG6b	ENSONIT00000016117
	CACNG7a	ENSONIT00000008882
	CACNG7b	ENSONIT00000016113
	CACNG8a	ENSONIT00000008880
	CACNG8b	ENSONIT00000016114
	TMEM235a	ENSONIT00000006698
	TMEM235b	ENSONIT00000025015
Latimeria *chalumnae*	CACNG1	ENSLACT00000011958
	CACNG2	ENSLACT00000003050
	CACNG3	ENSLACT00000017204
	CACNG4	ENSLACT00000011214
	CACNG5	ENSLACT00000010480
	CACNG6	ENSLACT00000005635
	CACNG7	ENSLACT00000007957
	CACNG8	ENSLACT00000006738

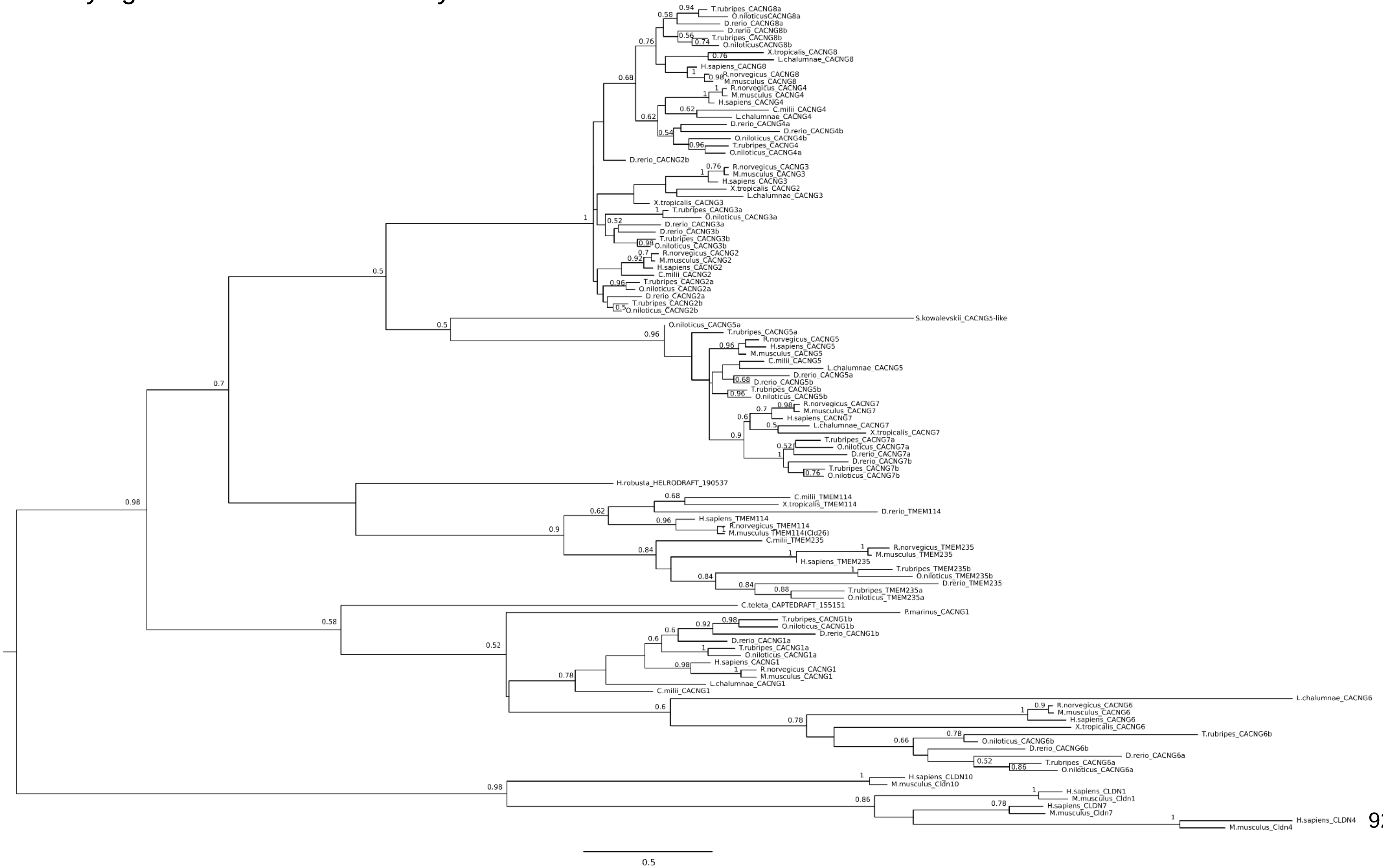
Xenopus *tropicalis*	CACNG1	ENSXETT00000033623
	CACNG2	ENSXETT00000061038
	CACNG3	ENSXETT00000035126
	CACNG4	ENSXETT00000055785
	CACNG6	NP_001096674.1
	CACNG7	ENSXETT00000022140
	CACNG8	ENSXETT00000022144
	TMEM114	ENSXETT00000019015
Anolis carolinensis	CACNG1	ENSACAT00000000195
	CACNG2	ENSACAT00000016334
	CACNG4	ENSACAT00000000192
	CACNG5	ENSACAT00000011349
	CACNG6	ENSACAT00000016560
	CACNG7	ENSACAT00000016599
	CACNG8	ENSACAT00000023146
	TMEM114	ENSACAT00000006934
	TMEM235	ENSACAT00000030370
Chrysemys picta bellii	CACNG1	XP_005282989.1
	CACNG2	XP_005302483.1
	CACNG3	XP_005288973.1
	CACNG4	XP_005282990.1
	CACNG5	XP_005314438.1
	CACNG6	XP_005312222.1
	CACNG7	XP_005312223.1
	CACNG8	XP_008174321.1
	TMEM114	XP_005295310.1
	TMEM235	XP_005283199.1
Gallus gallus	CACNG1a	ENSGALT00000006197
	CACNG1b	ENSGALT00000045934
	CACNG2	ENSGALT00000042940
	CACNG3	ENSGALT00000046134
	CACNG4a	ENSGALT00000042721
	CACNG4b	ENSGALT00000045133
	Novel(CACNG5a)	ENSGALT00000006226
	Novel(CACNG5b)	ENSGALT00000043235
	TMEM114	ENSGALT00000044917
Monodelphis domestica	CACNG1	ENSMODT00000005080
	CACNG2	ENSMODT00000037853

	CACNG3	ENSMODT00000020751
	CACNG4	ENSMODT00000005096
	CACNG6-like	XP_001378996.3
	CACNG7	ENSMODT00000017904
	TMEM114	ENSMODT00000007319
Canis lupus familiaris	CACNG1	ENSCAFT00000018075
	CACNG2	ENSCAFT00000049065
	CACNG3	ENSCAFT00000027856
	CACNG5	ENSCAFT00000047939
	CACNG6	ENSCAFT00000004269
	Novel(CACNG7) (2 of 2)	ENSCAFT00000004270
	Novel(CACNG7) (1 of 2) (analysis in evolutionary rate)	ENSCAFT00000046764
	CACNG8	ENSCAFT00000046946
	TMEM114	ENSCAFT00000043747
	TMEM235	ENSCAFT00000023435
Bos taurus	CACNG1	ENSBTAT00000009930
	CACNG2	ENSBTAT00000014423
	CACNG3	ENSBTAT00000001841
	CACNG4	ENSBTAT00000000779
	CACNG5	ENSBTAT00000016331
	CACNG6	ENSBTAT00000028860
	CACNG7	ENSBTAT00000009878
	CACNG8	ENSBTAT00000009879
	TMEM114	ENSBTAT00000037479
	TMEM235	ENSBTAT00000034726
Rattus norvegicus	CACNG1	ENSRNOT00000004349
	CACNG2	ENSRNOT00000008414
	CACNG3	ENSRNOT00000016632
	CACNG4	ENSRNOT00000004383
	CACNG5	ENSRNOT00000004578
	CACNG6	ENSRNOT00000019211
	CACNG7	ENSRNOT00000019683
	CACNG8	ENSRNOT00000019300
	TMEM114	ENSRNOT00000003749
	TMEM235	ENSRNOT00000074369
Mus musculus	CACNG1	ENSMUST00000021065

	CACNG2	ENSMUST00000019290
	CACNG3	ENSMUST00000084615
	CACNG4	ENSMUST00000021066
	CACNG5	ENSMUST00000039071
	CACNG6	ENSMUST00000183200
	CACNG7	ENSMUST00000181686
	CACNG8	ENSMUST00000092351
	TMEM114(Cldn26)	ENSMUST00000023400
	TMEM235	ENSMUST00000093905
	CLDN1	NP_057883.1
	CLDN4	NM_009903.2
	CLDN7	NP_001180548.1
	CLDN10	NP_001153568.1
Oryctolagus cuniculus	CACNG1	ENSOCUT00000008846
	CACNG2	ENSOCUT00000027502
	CACNG3	ENSOCUT00000003984
	CACNG4	ENSOCUT00000014215
	CACNG5	XP_008270034.1
	CACNG6	ENSOCUT00000000795
	CACNG7	ENSOCUT00000028081
	TMEM114	ENSOCUT00000021642
	TMEM235	ENSOCUT00000030307
Homo sapines	CACNG1	ENST00000226021
	CACNG2	ENST00000300105
	CACNG3	ENST00000005284
	CACNG4	ENST00000262138
	CACNG5	ENST00000533854
	CACNG6	ENST00000252729
	CACNG7	ENST00000391767
	CACNG8	ENST00000270458
	TMEM114	ENST00000620492
	TMEM235	ENST00000421688
	CLDN1	NP_066924.1
	CLDN4	NM_001305.4
	CLDN7	NP_001171951.1
	CLDN10	NP_001153572.1

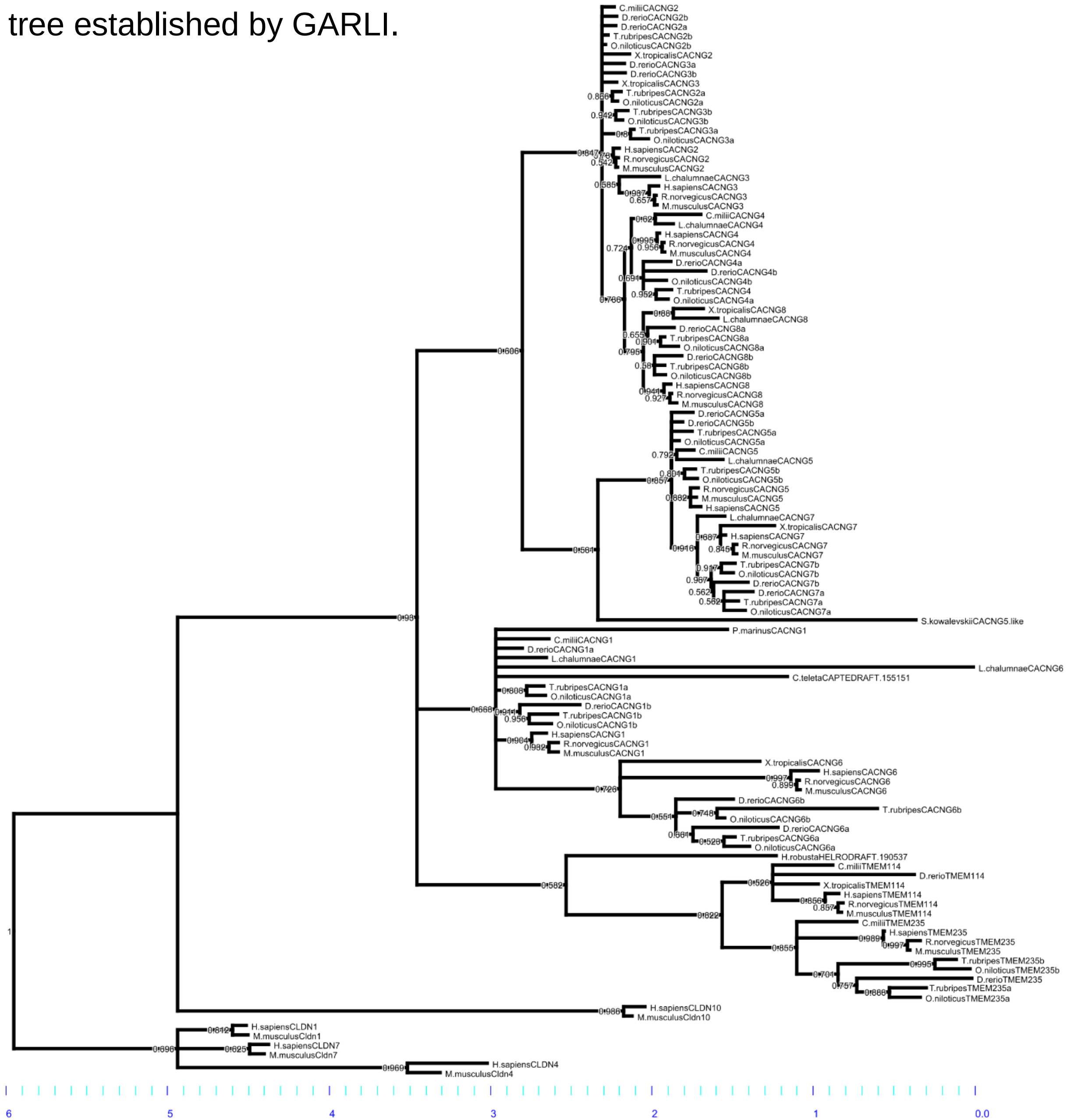
Appendix III phylogenetic trees of transmembrane residues of vertebrate γ subunits.

IIIa. Phylogenetic tree established by MEGA.

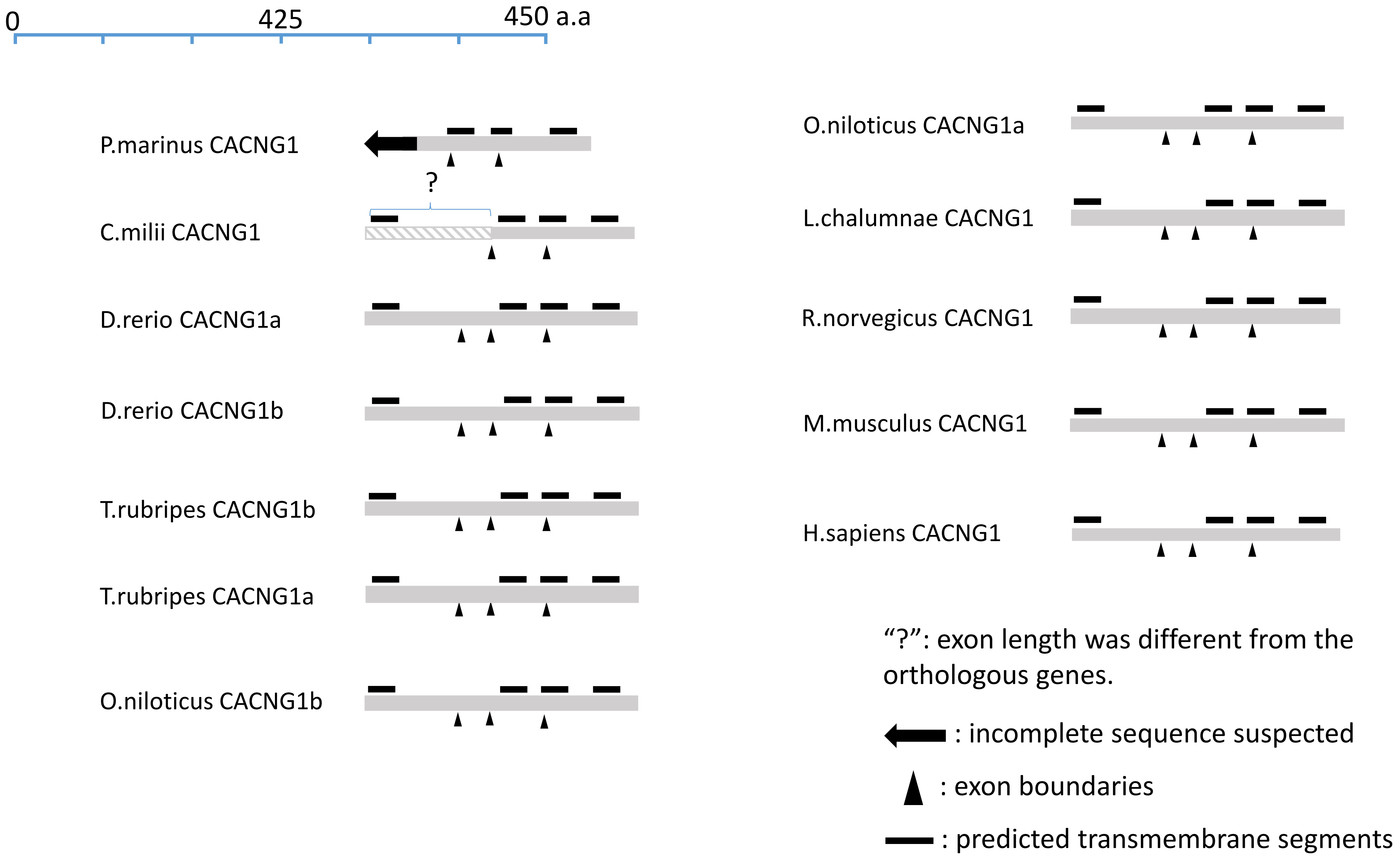


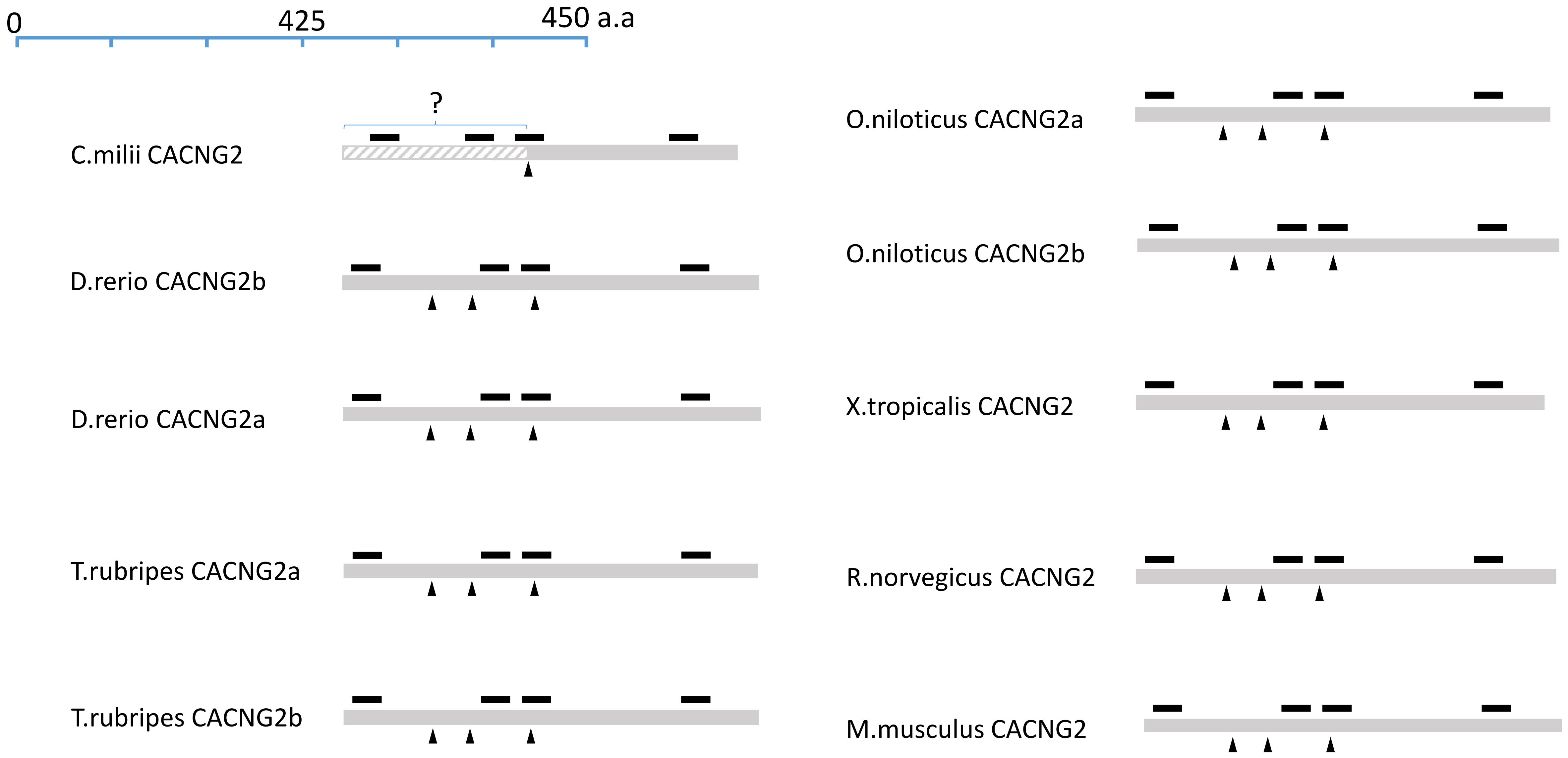
IIIb Phylogenetic tree established by MrBayse.

IIIc. Phylogenetic tree established by GARLI.



Appendix IV. Gene structures of CACNG family in animals. Most of γ subunits contain four transmembrane domains ( ) and three to four introns ( ). A few γ subunits were suspected that their exon length have problems because their exons length were different from the other orthologous genes (?). ( ) is indicated removed sequences in our study. The result indicated that the sequences we used or not used in our analysis.

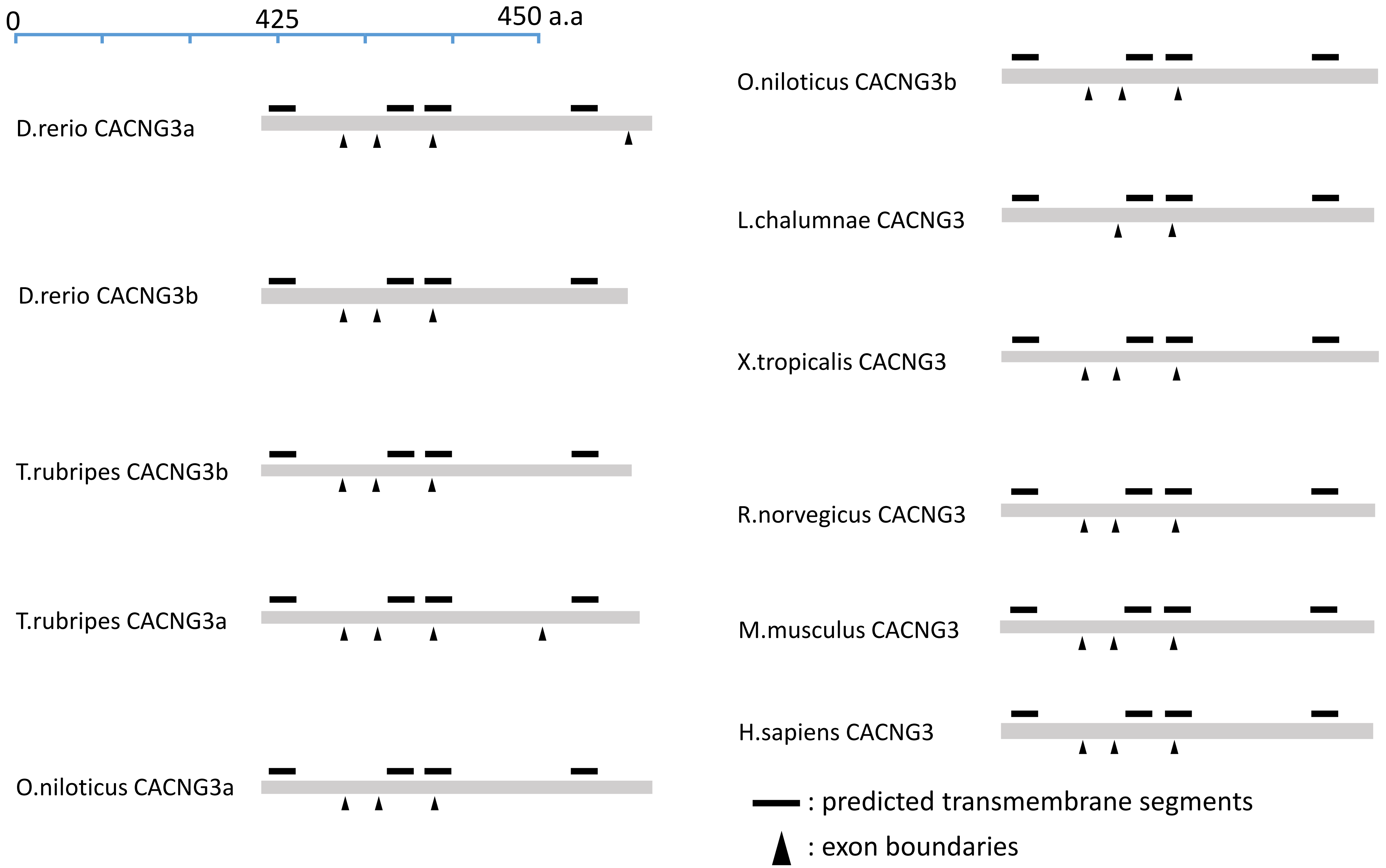

0
425
450 a.a
P.marinus CACNG1
?
C.milii CACNG1
D.rerio CACNG1a
D.rerio CACNG1b
T.rubripes CACNG1b
T.rubripes CACNG1a
O.niloticus CACNG1b
O.niloticus CACNG1a
L.chalumnae CACNG1
R.norvegicus CACNG1
M.musculus CACNG1
H.sapiens CACNG1
“?”: exon length was different from the orthologous genes.
: incomplete sequence suspected
: exon boundaries
: predicted transmembrane segments

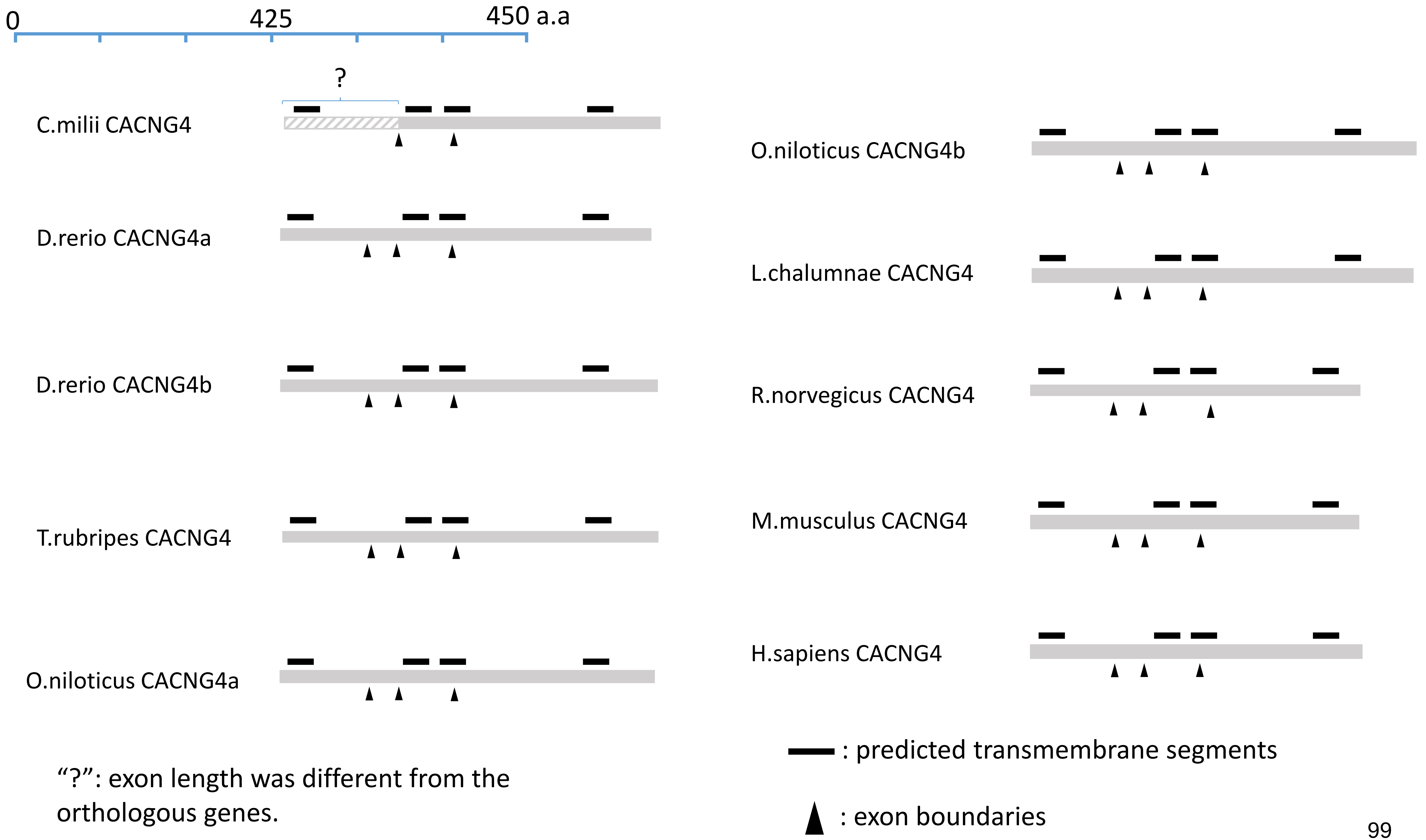


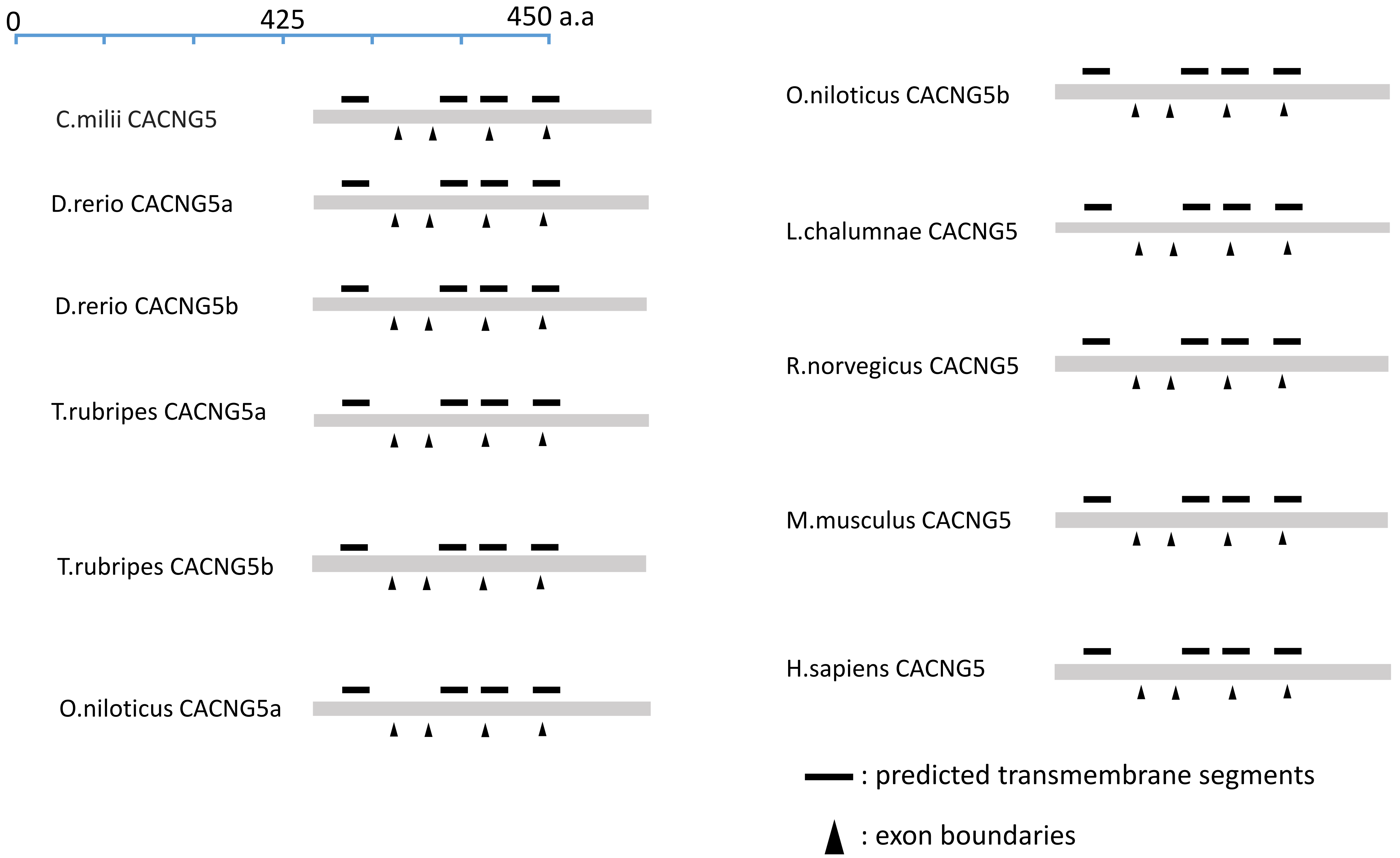
: predicted transmembrane segments

: exon boundaries

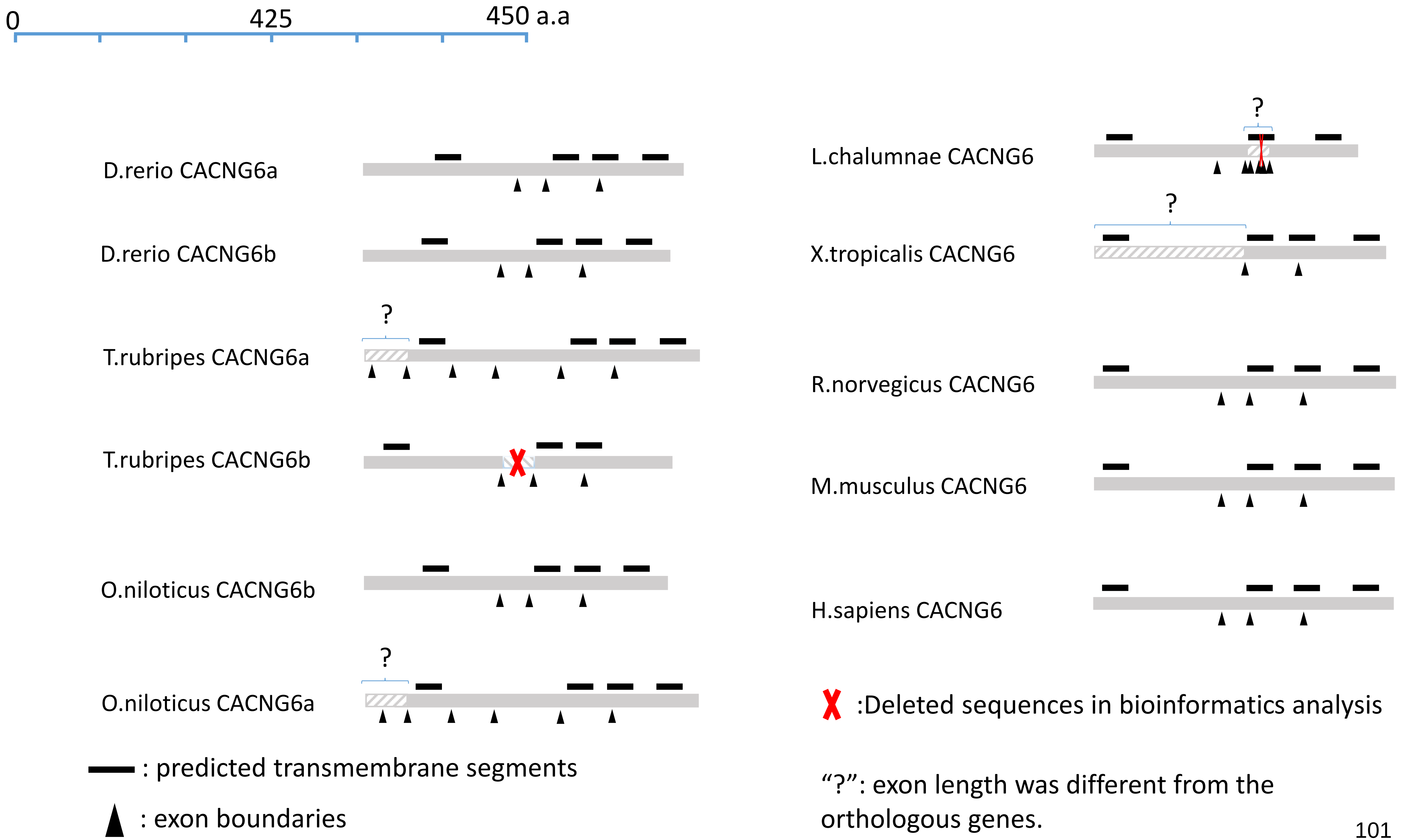
“?”: exon length was different from the orthologous genes.

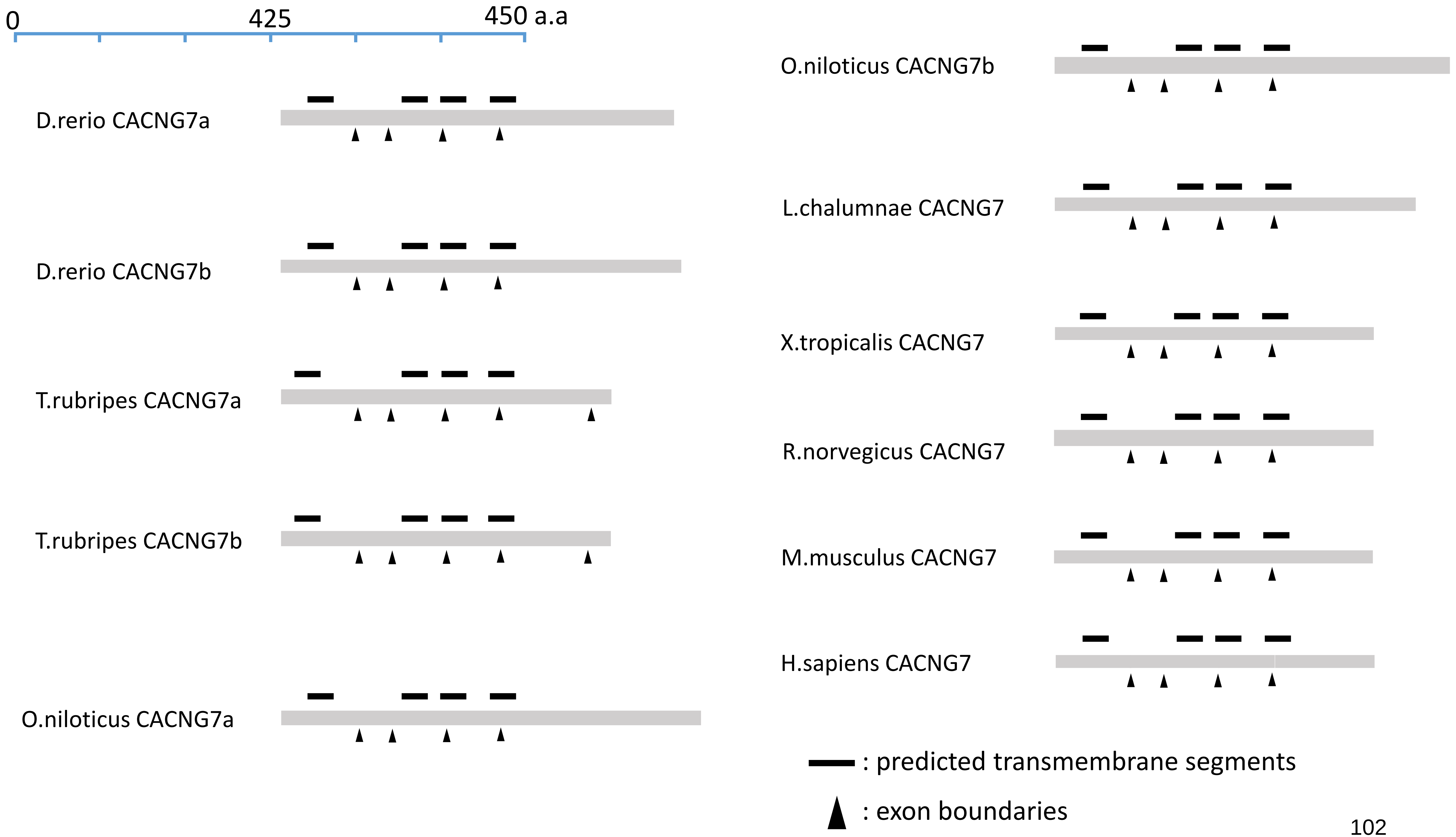

0
425
450 a.a
D.rerio CACNG3a
D.rerio CACNG3b
T.rubripes CACNG3b
T.rubripes CACNG3a
O.niloticus CACNG3a
O.niloticus CACNG3b
L.chalumnae CACNG3
X.tropicalis CACNG3
R.norvegicus CACNG3
M.musculus CACNG3
H.sapiens CACNG3
: predicted transmembrane segments
: exon boundaries

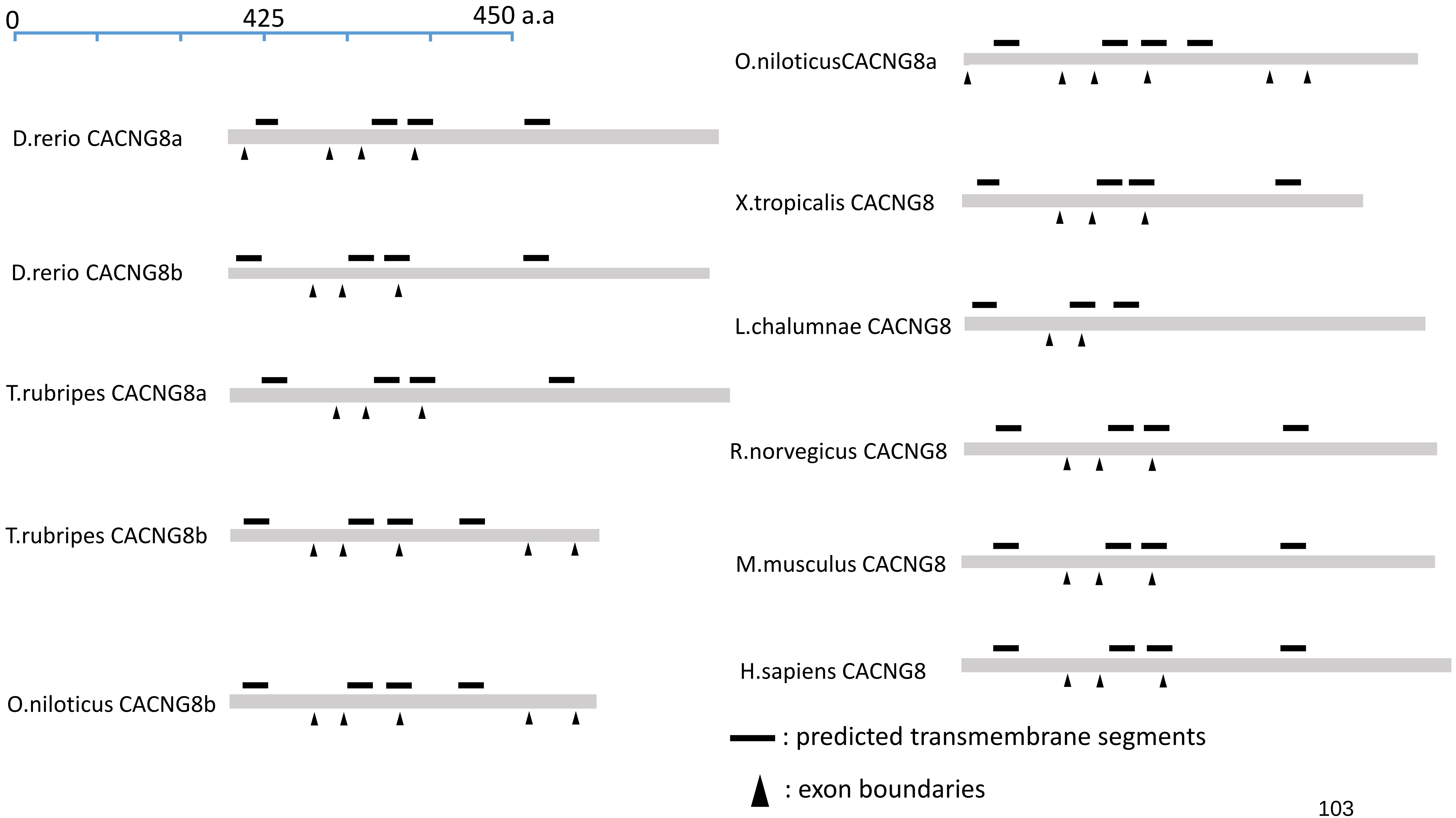

0
425
450 a.a
?
C.milii CACNG4
D.rerio CACNG4a
D.rerio CACNG4b
T.rubripes CACNG4
O.niloticus CACNG4a
O.niloticus CACNG4b
L.chalumnae CACNG4
R.norvegicus CACNG4
M.musculus CACNG4
H.sapiens CACNG4
“?”: exon length was different from the orthologous genes.
: predicted transmembrane segments
: exon boundaries



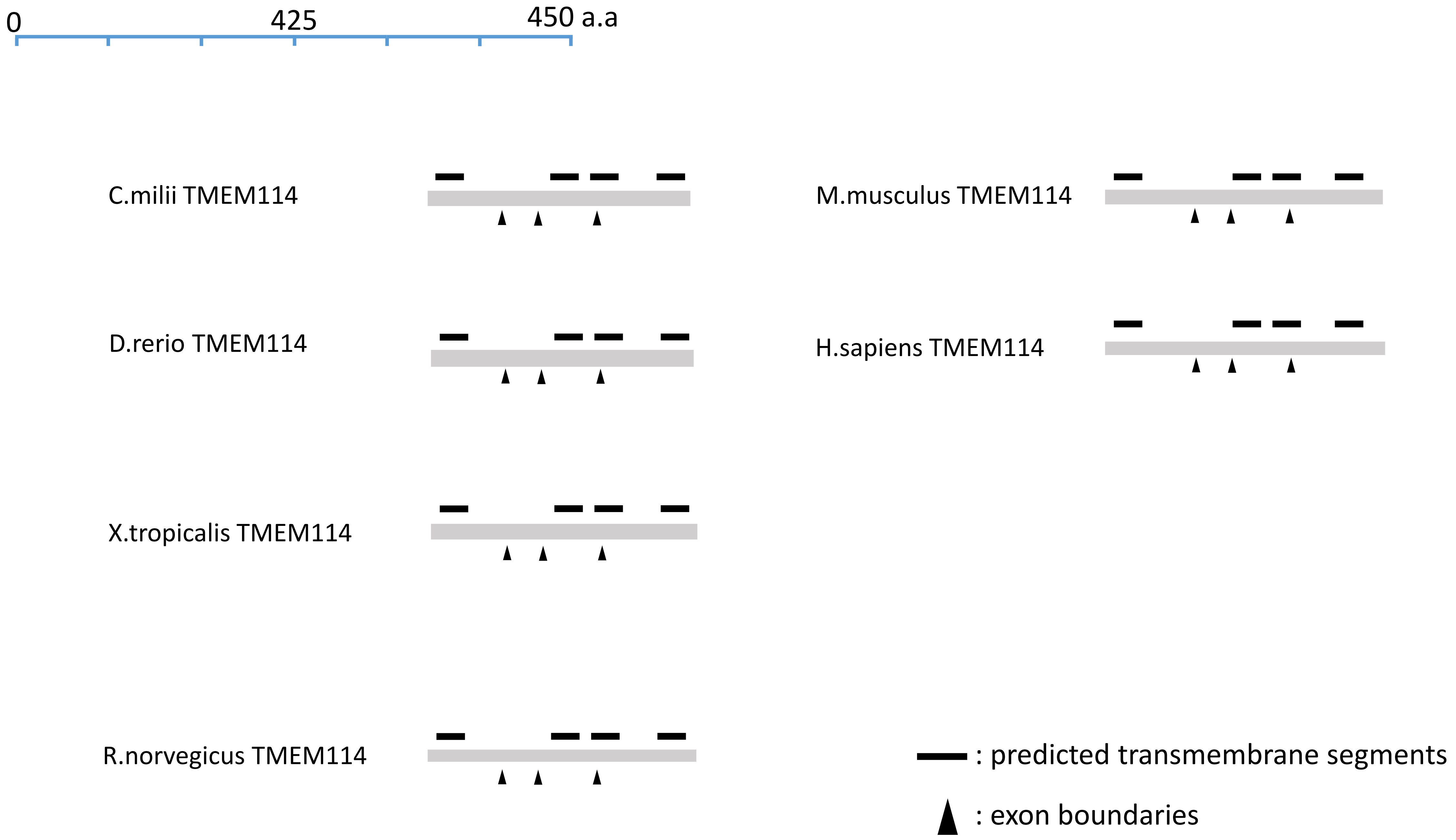
0
425
450 a.a
C.milii CACNG5
D.rerio CACNG5a
D.rerio CACNG5b
T.rubripes CACNG5a
T.rubripes CACNG5b
O.niloticus CACNG5a
O.niloticus CACNG5b
L.chalumnae CACNG5
R.norvegicus CACNG5
M.musculus CACNG5
H.sapiens CACNG5
: predicted transmembrane segments
: exon boundaries

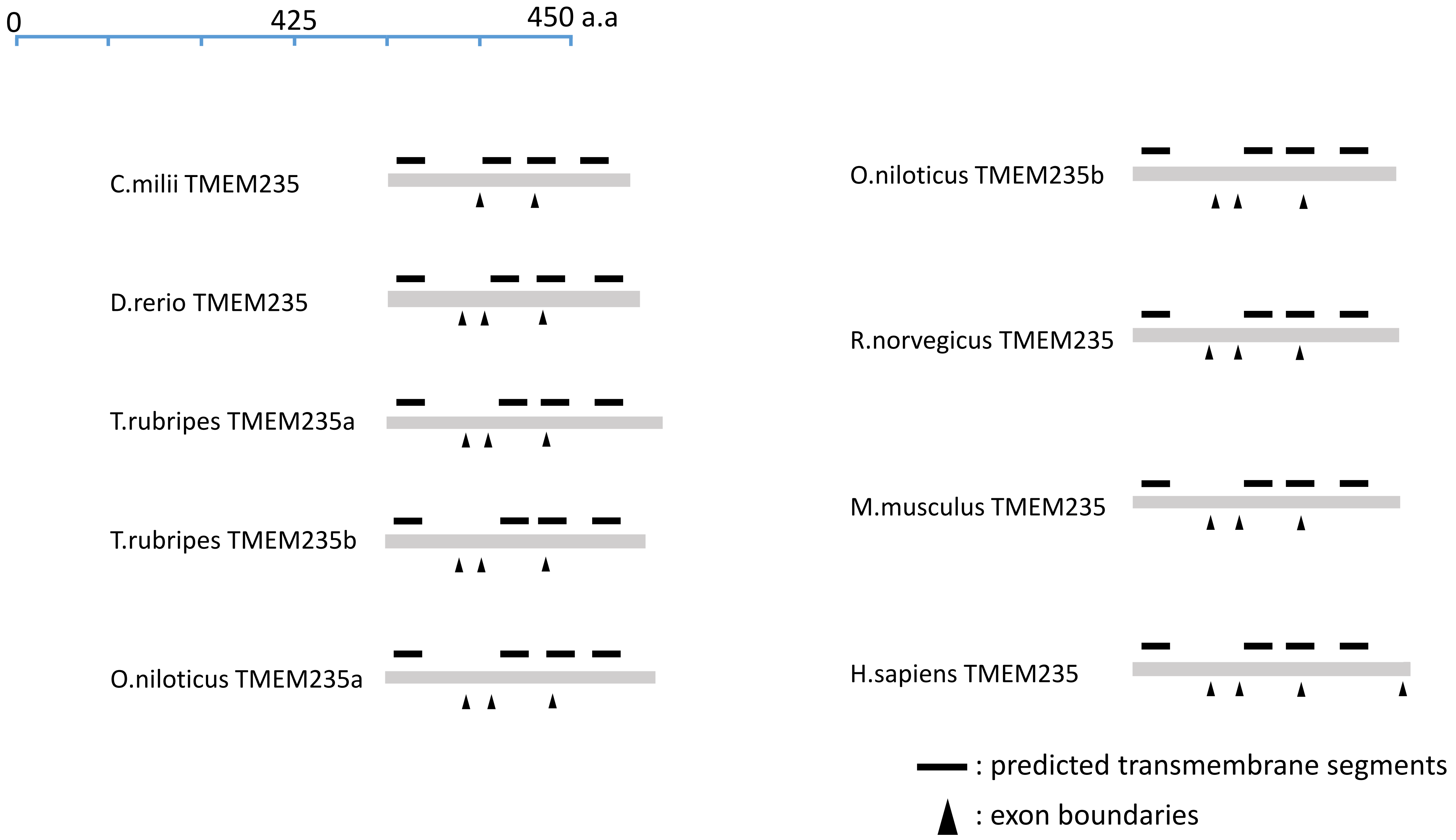

0
425
450 a.a
D.rerio CACNG6a
D.rerio CACNG6b
?
T.rubripes CACNG6a
T.rubripes CACNG6b
O.niloticus CACNG6b
?
O.niloticus CACNG6a
?
L.chalumnae CACNG6
?
X.tropicalis CACNG6
R.norvegicus CACNG6
M.musculus CACNG6
H.sapiens CACNG6
: predicted transmembrane segments
: exon boundaries
:Deleted sequences in bioinformatics analysis
"?": exon length was different from the orthologous genes.


0
425
450 a.a
D.rerio CACNG7a
D.rerio CACNG7b
T.rubripes CACNG7a
T.rubripes CACNG7b
O.niloticus CACNG7a
O.niloticus CACNG7b
L.chalumnae CACNG7
X.tropicalis CACNG7
R.norvegicus CACNG7
M.musculus CACNG7
H.sapiens CACNG7
: predicted transmembrane segments
: exon boundaries



0
425
450 a.a
D.rerio CACNG8a
D.rerio CACNG8b
T.rubripes CACNG8a
T.rubripes CACNG8b
O.niloticus CACNG8b
O.niloticusCACNG8a
X.tropicalis CACNG8
L.chalumnae CACNG8
R.norvegicus CACNG8
M.musculus CACNG8
H.sapiens CACNG8
: predicted transmembrane segments
: exon boundaries


0
425
450 a.a
C.milii TMEM114
D.rerio TMEM114
X.tropicalis TMEM114
R.norvegicus TMEM114
M.musculus TMEM114
H.sapiens TMEM114
: predicted transmembrane segments
: exon boundaries


0
425
450 a.a
C.milii TMEM235
D.rerio TMEM235
T.rubripes TMEM235a
T.rubripes TMEM235b
O.niloticus TMEM235a
O.niloticus TMEM235b
R.norvegicus TMEM235
M.musculus TMEM235
H.sapiens TMEM235
: predicted transmembrane segments
: exon boundaries

Appendix V. Putative functional sites in CACNG amino acid sequences. The multiple amino acids sequences alignments were composed of ten subunits from mouse CACNG family. Each color represented different putative functional sites. The most similar segments, transmembrane domains, were underlined and used boldfaced.

Appendix V putative functional sites in mouse CACNG proteins


```
                             10        20        30        40        50        60        70        80
                     ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
mTMEM114        1    ---------- ---------- ---------- MRVRLGALAG AAALSGALSF VLLAAAIGTD FWYIIDTERL ERSSQRMRDQ  50
mTMEM235        1    ---------- ---------- ---------- -MALLATLLL SAALGALLSF ALLAAAVASD YWYILEVADA GGL-------  42
mgamma1         1    ---------- ---------- ---------- MSQTKTAKVR VTLFFILVGG VLAMVAVVTD HWAVLSPHLE HH--------  42
mgamma6         1    MMWSNFFMQE EDRRRTAVGR RRAQEQQNLG LTPEREGKIK LGLLVAIVGA TLAVLAVGTE FWVELNTYKT NG--------  72
mgamma5         1    ---------- ---------- ---------M SACGRKALTL LSSVFAVCGL GLLGIAVSTD YWLYLEEGII LPQNQS----  47
mgamma7         1    ---------- ---------- ---------M SHCSSRALTL LSSVFGACGL LLVGIAVSTD YWLYMEEGTV LPQNQT----  47
mgamma2         1    ---------- ---------- ---------M GLFDRGVQML LTTVGAFAAF SLMTIAVGTD YWLY-SRGVC KTKSVSENET  50
mgamma3         1    ---------- ---------- ---------M RMCDRGIQML ITTVGAFAAF SLMTIAVGTD YWLY-SRGVC RTKSTSDNET  50
mgamma4         1    ---------- ---------- ---------M VRCDRGLQML LTTAGAFAAF SLMAIAIGTD YWLYSSAHIC NGTNLTM-DD  50
mgamma8         1    ---------- --------ME SLKRWNEERG LWCEKGVQVL LTTIGAFSAF GLMTIAISTD YWLYTRALIC NTTNLTAGDD  62
mclaudin1       1    ---------- ---------- ---------- --MANAGLQL LGFILASLGW IGSIVSTALP QWKIYSYAGD N---------  39
Clustal Consensus 1                                              .        :      *                  2

                             90        100       110       120       130       140       150       160
                     ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
mTMEM114       51    GPANRS---- ---QQEPLSS HSGLWRTC-- ---------RV QSSCTPLMNP FWQENVT--- --VSDSSRQL LTMHG----- 103
mTMEM235       43    ---------- ---GGVQLFS HSGLWRTC-- ---------EG QNSCVPLIDP FASAGLE--- --VSPSVQHL LSLHR-----  89
mgamma1        43    ---------- ---NETCEAA HFGLWRICTA RV---AVHNK DKSCEHVTPS GEKNCSYFRH FNPGESSEIF EFTTQKEYSI 106
mgamma6        73    ---------- ---SAVCEAA HLGLWKVCIK RLWQADVPAG RETCGPAELP GEANCTYFKF FTTGENARIF QRTTKKEVNL 139
mgamma5        48    ---------- ---TEVKMSL HSGLWRVCF- -----LAGEE RGRCFTIEYV MPMNSQM--- --TSESTVNV LKMIR-----  98
mgamma7        48    ---------- ---TEVKMAL HAGLWRVCF- -----FAGRE KGRCVASEYF LEPEINL--- --VTENTENI LKTVR-----  98
mgamma2        51    SK-------- ---KNEEVMT HSGLWRTCC- -----LEGNF KGLCKQIDH- FPEDADY--- --EADTAEYF LRAVR----- 102
mgamma3        51    SR-------- ---KNEEVMT HSGLWRTCC- -----LEGAF RGVCKKIDH- FPEDADY--- --EQDTAEYL LRAVR----- 102
mgamma4        51    GPPPR----- ---RARGDLT HSGLWRVCC- -----IEGIY RGHCFRINH- FPEDNDY--- --DHDSSEYL LRIVR----- 105
mgamma8        63    GPPHRGGSGS SEKKDPGGLT HSGLWRICC- -----LEGLK RGVCVKINH- FPEDTDY--- --DHDSAEYL LRVVR----- 125
mclaudin1      40    ---------- ---IVTAQAI YEGLWMSCV- -----SQSTG QIQCKVFDSL LNLNSTL--- ----QATRAL MV-------- 85
Clustal Consensus 2                      : ***  *             *                    .            8
```

```
                                     170       180        190       200        210        220        230        240
                                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
mTMEM114                    104 TFVILLPLSL IVMVFGGMTG -FLS-FLLRA HLLLLLTGIL FLFGAMVTLT GISIYIAYSA VAFREAVCLL EERALLDQVD  181
mTMEM235                     90 TVMVVLPLSL VLIVCGWVCG -LLS-SLSQS VPLLLATGCY FLLGGALTLA GLSIYISYSH LAFVEAAR-T YGVTHVQNVH  166
mgamma1                     107 SAAAIAIFSL GFIIVGSICA -FLS-FGNKR DYLLRPASMF YAFAGLCLIV SVEVMRQSVK RMI------D SED--TVWIE  176
mgamma6                     140 AAAVIAVLGL TAMALGCLCV -IMV-LSKGA ESLLRLGAVC FGLSGLLLFV SLEVFRHSVG ALLQGVNP-E TPP--APRLA  214
mgamma5                      99 SATPFPLVSL FFMFIGFILS -NIGHIRPHR TILAFVSGIF FILSGLSLVV GLVLYISSIN DEMLNRTK-D AET----YFN  172
mgamma7                      99 TATPFPMVSL FLVFTAFVIS -NIGHIRPQR TILAFVSGIF FILSGLSLVV GLVLYISSIN DEVMNRPS-S SEQ----YFH  172
mgamma2                     103 ASSIFPILSV ILLFMGGLCI -AASEFYKTR HNIILSAGIF FVSAGLSNII GIIVYISANA GDPSKS---D SKK-----NS  173
mgamma3                     103 ASSVFPILSV TLLFFGGLCV -AASEFHRSR HSVILSAGIF FVSAGLSNII GIIVYISANA GDPGQR---D SKK------S  172
mgamma4                     106 ASSVFPILST ILLLLGGLCI -GAGRIYSRK NNIVLSAGIL FVAAGLSNII GIIVYISSNT GDPSDKRD-E DKK-----NH  178
mgamma8                     126 ASSIFPILSA ILLLLGGVCV -AASRVYKSK RNIILGAGIL FVAAGLSNII GVIVYISANA GEPGPKRD-E EKK-----NH  198
mclaudin1                    86 IGILLGLIAI FVSTIGMKCM RCLEDDEVQK MWMAVIGGII FLISGLATLV ATAWYGNRIV QEFYDPLT-P INA------R  158
Clustal Consensus             8     .  ..      .               :    .  :  ..   .  .                            10

                                     250       260        270       280        290        300        310        320
                                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
mTMEM114                    182 IRFGWSLALG WISFVSELLT GVVFL----- ---------- ---------- ---------- ---------- -------AAA  209
mTMEM235                    167 ISFGWSLALA WASCASEVLS GALLL----- ---------- ---------- ---------- ---------- -------AAA  194
mgamma1                     177 HYYSWSFACA CAAFILLFLG GLFLL----- ---------- ---------- ---------- ---------- ----------  201
mgamma6                     215 YEYSWSLGCG VGAGLILLLG GVCFL----- ---------- ---------- ---------- ---------- ----------  239
mgamma5                     173 YKYGWSFAFA AISFLLTESA GVMSVYLFMK RYTA------ -EDMYRP--- ---------- ---HPGFYRP RLSNCSDYSG  229
mgamma7                     173 YRYGWSFAFA ASSFLLKEGA GVMSVYLFTK RYAE------ -EEMYRP--- ---------- ---HPAFYRP RLSDCSDYSG  229
mgamma2                     174 YSYGWSFYFG ALSFIIAEMV GVLAVHMFID RHKQLRATAR ATDYLQA--- ---------S AITRIPSYRY RYQRRSRSSS  241
mgamma3                     173 YSYGWSFYFG AFSFIIAEIV GVVAVHIYIE KHQQLRARSH -SELLKK--- ---------S TFARLPPYRY RFRR--RSSS  237
mgamma4                     179 YNYGWSFYFG ALSFIVAETV GVLAVNIYIE KNKELRFKTK -REFLKA--- ------SSSS PYARMPSYRY R-RRRSRSSS  247
mgamma8                     199 YSYGWSFYFG GLSFILAEVI GVLAVNIYIE RSREAHCQSR -SDLLKAGGG AGGSGGSGPS AILRLPSYRF RYRRRSRSSS  277
mclaudin1                   159 YEFGQALFTG WAAASLCLLG GVLLS----- ---------- ---------- ---------- ---------- ------CSCP  187
Clustal Consensus            10     :..::  .   :       *                                                        15
```

```
                                330        340        350        360        370        380        390       400
                           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
mTMEM114              210 R--------- ---------- ---------A LSLSQR---- ---------- ---------- ---------- ---------- 217
mTMEM235              195 R--------- ---------- ---------L LSLSQRPGV- ---------- ---------- ---------- ---------- 205
mgamma1               202 ---------- ---------- ---------L FSLPRMPQN- ---------- ---------- ---------- ---------- 211
mgamma6               240 ---------- ---------- ---------L LTLPS----- ---------- ---------- ---------- ---------- 245
mgamma5               230 Q-FLHP---- ---------- --------DA WIRGRSPSD- ---------- ---------- ---------- ---------- 245
mgamma7               230 Q-FLQP---- ---------- --------EA WRRGRSPSD- ---------- ---------- ---------- ---------- 245
mgamma2               242 R-STEPSHSR DASPVGVKGF NTLPSTEISM YTLSRDPLKA ATTPTATYNS ---------- ---------- ---------- 290
mgamma3               238 R-STEP-RSR DLSPIS-KGF HTIPSTDISM FTLSRDPSK- -LTMGTLLNS ---------- ---------- ---------- 282
mgamma4               248 R-STEASPSR DASPVGLKIT GAIPMGELSM YTLSREPLK- -VTTAASYSP ---------- ---------- ---------- 294
mgamma8               278 RGSSEASPSR DASPGGPGGP G-FASTDISM YTLSRDPSK- -GSVAAGLAS AGGGGSGAGV GAYGGAAGAA GGGGAGSERD 354
mclaudin1             188 R--------- ---------- ------KTTS YPTPRP---- ---------- ---------- ---------- ---------- 198
Clustal Consensus      15                                                                           15

                                410        420        430        440        450        460
                           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....
mTMEM114              218 ---------- ---------- -QDQAI---- ---------- ---------- ---------- --------- 222
mTMEM235              206 ---------- ---------- PHSVIL---- ---------- ---------- ---------- --------- 211
mgamma1               212 ---------- ----PWESCM DAEPEH---- ---------- ---------- ---------- --------- 223
mgamma6               246 ---------- ---WPWRSLC PKWGGP---- ---------- ---------- ---------- -------TA 260
mgamma5               246 -ISSDASLQM NSNYPALLKC PDYDQM---- ---------- ---------- ---------- ----SSSPC 275
mgamma7               246 -ISSDVSIQM TQNYPPAIKY PDHLHI---- ---------- ---------- ---------- ----STSPC 275
mgamma2               291 -DRDNSFLQV HNCIQKDSKD SLHANT---- ---------- ---------- ---------- -ANRRTTPV 323
mgamma3               283 -DRDHAFLQF HNSTPKEFKE SLHNNP---- ---------- ---------- ---------- -ANRRTTPV 315
mgamma4               295 -DQDAGFLQM HDFFQQDLKE GFHVSM---- ---------- ---------- ---------- -LNRRTTPV 327
mgamma8               355 RGSSAGFLTL HNAFPKEAAS GVTVTVTGPP AAPAPAPAPP APAAPAPGTL SKEAAASNTN TLNRKTTPV 423
mclaudin1             199 ---------- ---YPKPTPS SGKDYV---- ---------- ---------- ---------- --------- 211
Clustal Consensus      15                                                            15
```

N-myristoylation site
Protein kinase C phosphorylation site
N-glycosylation site
Leucine zipper pattern
Casein kinase II phosphorylation site
Amidation site
Tyrosine kinase phosphorylation
cAMP- and cGMP-dependent protein kinase phosphorylation
Cell attachment sequence
TTPV PDZ-binding motif
nPIST-binding motif

Appendix VI. Paralogons in the vicinity of CACNG genes.

VIa. Paralogons (homologous chromosome segments) in the vicinity of elephant shark CACNG genes on scaffolds.

scaffold NW_006890200.1		scaffold NW_006890063.1		scaffold NW_006890324.1		scaffold NW_006890083.1	
Gene	Location(Mb)	Gene	Location(Mb)	Gene	Location(Mb)	Gene	Location(Mb)
		GRIN2C	2.507-2.532				
		TMEM235	7.825-7.834			**TMEM114**	5.698-5.711
		SYNGR2	7.8606-7.867	SYNGR1	0.147-0.154		
		CACNG1	9.064-9.069				
CACNG3	1.031-1.048	**CACNG4**	9.078-9.081	**CACNG2**	0.891-0.916		
		CACNG5	9.097-9.1109				
PRKCB	0.974-1.023	PRKCA	9.119-9.209				
		CACNA1G	10.359-10.436				

VIb. Paralogons (homologous chromosome segments) in the vicinity of fugu CACNG genes on scaffolds.

Scaffold 40		Scaffold 112		Scaffold 115		Scaffold 3		Scaffold 345		Scaffold 61		scaffold_202	
Gene	Location(Mb)	Gene	Location(Mb)	Gene	Location(Mb)	Gene	Location(Mb)	Gene	Location(Mb)	Gene	Location(Mb)	Gene	Location(Mb)
				SSTR2 1of2	0.411-0.412	SSTR2 2of2	3.765-3.767						
CACNG3a	0.461-0.464	**CACNG3b**	0.283-0.292										
PRKCB1a	0.466-0.484	PRKCB1b	0.300-0.354										
SYNGR3a	0.403-0.406			SYNGR2 2of3	0.622-0.624								
				TMEM235 1of2	0.783-0.787	**TMEM235 2of2**	2.079-2.081						
						CACNA1G	4.004-4.109						
								CACNG1	0.081-0.088	**CACNG6a**	0.071-0.075	**CACNG6b**	0.209-0.212
								CACNG5a	0.102-0.106	**CACNG7a**	0.004-0.103	**CACNG7b**	0.185-0.189
								PRKCA	0.119-0.179			PRKCG	0.173-0.180
										CACNG8a	0.081-0.085	**CACNG8b**	0.200-0.205
						TBX6	0.350-0.351						

VIc. Paralogons (homologous chromosome segments) in the vicinity of zebra fish CACNG genes on chromosome 1, 3, 6, 12, 16 and 19.

Chromosome 1		Chromosome 3		Chromosome6		Chromosome12		Chromosome16		Chromosome 19	
Gene	Location(Mbp)	Gene	Location(Mbp)	Gene	Location(Mbp)	Gene	Location(Mbp)	Gene	Location(Mbp)	Gene	Location(Mbp)
		PRKCA(2 of 3)	10.84 - 10.89								
		GRIN2Ca	18.82 - 18.95			GRIN2Cb	33.59 - 33.70	GRIN2Db	14.93 - 15.14	GRIN2Da	10.53 - 10.70
SYNGR3b	7.49 - 7.50	SYNGR1a	21.46 - 21.49	SYNGR1b	0.3156 - 0.3186						
GRIN2Ab	7.83 - 7.99	GRIN2Aa	27.22 - 27.42								
		TMEM114	28.08 - 28.10			**TMEM235**	36.89 - 36.90				
CACNA1Hb	7.38 - 7.47	CACNA1I	29.23 - 29.52			CACNA1G	29.27 - 29.63				
		SSTR3	29.75 - 29.75								
		CACNG2a	29.86 - 29.99	**CACNG2b**	0.3011 - 0.3112						
		CACNA1Aa	33.87 - 33.98								
PRKCBa	8.083 - 8.14	PRKCBb	35.09 - 35.33								
CACNG3a	8.15 - 8.16	**CACNG3b**	35.38 - 35.46								
		CACNA1Ha	39.84 - 39.94								
		SYNGR3a	40.16 - 40.17								
		PRKCA(3 of 3)	56.72 - 56.76	PRKCA	22.91 - 23.07			PRKCG	14.77 -14.86		
		CACNG5b	56.77 - 56.80	**CACNG5a**	23.07 - 23.13			**CACNG7b**	14.71 - 14.75	**CACNG7a**	10.46 - 10.50
		SYNGR2a	58.58 - 58.59								
		CACNG1a	60.81 - 60.83	**CACNG1b**	23.15 - 23.17			**CACNG6b**	14.58 - 14.61	**CACNG6a**	10.40 - 10.42
		CACNG4a	61.55 - 61.57	**CACNG4b**	23.20 - 23.20			**CACNG8b**	14.63 - 14.70	**CACNG8a**	10.43 - 10.45
		SSTR2(2of 2)	63.07 - 63.07			STTR2(1 OF2)	1.73 - 1.734				
				PRKCDa	40.45 - 40.47						

VId. Paralogons (homologous chromosome segments) in the vicinity of frog CACNG genes on scaffold.

Scaffold NW_004668244.1		scaffold NW_004671306.1		scaffold NW_004668236.1		scaffold NW_004668240.1	
Gene	Location(Kb)	Gene	Location(Kb)	Gene	Location(Kb)	Gene	Location(Kb)
TMEM114	61,22-61,25						
GRIN2A	61,860-62,630					GRIN2D	108,070-108,910
SYNGR3	76,100-86,100			SYNGR1	73,745-73,783	SYNGR4	108,74-108,77
						CACNG7	109,082-109,222
		CACNG1	0.5-5.5			**CACNG6**	109,35-109,36
SSTR5	82,185-82,235			SSTR3	84,583-84,641		
		CACNA1A	210-620				
				CACNG2	84,900-85,035		

VIe. Paralogons (homologous chromosome segments) in the vicinity of chicken CACNG genes on chromosome 1, 14, and 18.

Chromosome 14		Chromosome 18		Chromosome 1	
Gene	location(Mb)	Gene	location(Mb)	Gene(Mb)	location(Mb)
SSTR5	5.61-5.61	SSTR2	9.25-9.25		
SYNGR3	6.17-6.19	SYNGR2	10.14-10.14	SYNGR1	50.53-50.54
GRIN2A	9.24-9.39	GRIN2C	10.76-10.77		
CACNA1H	5.23-5.33	CACNA1G	10.46-10.57	CACNA1I	50.26-50.35
		PRKCA	7.59-7.71		
		CACNG5(1 of 2)	7.57-7.59		
CACNG3	6.74-6.74	**CACNG4(1 of 2)**	7.51-7.55	**CACNG2**	51.49-51.54
		CACNG1(1 of 2)	7.50-7.51		
		CACNG5(2 of 2)	7.26-7.27		
		CACNG4(2 of 2)	7.20-7.24		
		CACNG1(2 of 2)	7.19-7.20		
TBX6	8.24-8.25				
TMEM114	10.11-10.13				

VIf. Paralogons (homologous chromosome segments) in the vicinity of dog CACNG genes on chromosomes 1, 6, 9, and 10.

Chromosome 6		Chromosome 9		Chromosome 10		Chromosome 1	
Gene	Location(Mb)	Gene	Location(Mb)	Gene	Location(Mb)	Gene	Location(Mb)
TMEM114	33.54-33.55	**TMEM235**	2.91-2.92				
		SYNGR2	2.97-2.97	SYNGR1	25.68-25.69	SYNGR4	107.87-107.87
GRIN2A	32.52-32.66	GRIN2C	5.58-5.59			GRIN2D	107.82-107.85
		SSTR2	6.98-7.00	SSTR3	27.32-27.33		
SSTR5	39.60-39.61	**CACNG1**	13.42-13.43			**CACNG6**	103.22-103.24
CACNG3	21.62-21.71			**CACNG2**	27.75-27.86	**CACNG8**	103.25-103.26
		CACNG5	13.57-13.58			**CACNG7(1 of 2)**	103.27-103.29
						CACNG7(2 of 2)	103.48-103.49
		PRKCA	13.64-14.03				
CACNA1H	39.50-39.53	CACNA1G	26.48-26.54	CACNA1I	25.37-25.49		
TBX6	18.09-18.09	TBX4	35.16-35.19	TBX1	29.52-29.53		
		TBX2	35.23-35.24				

VIg. Paralogons (homologous chromosome segments) in the vicinity of cattle CACNG genes on chromosomes 5, 18, 19, and 25.

Chromosome 25		Chromosome 19		Chromosome 5		Chromosome 18	
Gene	Location(Mb)	Gene	Location(Mb)	Gene	Location(Mb)	Gene	Location(Mb)
TBX6	26.456-26.460	TBX2	11.94-11.95				
		TBX4	11.87-11.89				
CACNA1H	0.959-0.984	CACNA1G	36.73-36.79	CACNA1I	111.52-111.61		
TMEM114	7.52-7.53	**TMEM235**	54.547-54.552				
SYNGR3	1.546-1.550	SYNGR2	54.611-54.615	SYNGR1	111.29-111.32	SYNGR4	55.51-55.52
GRIN2A	8.55-8.66	GRIN2C	57.19-57.20			GRIN2D	55.54-55.57
SSTR5	0.857-0.858	SSTR2	58.718-58.719	SSTR3	76.013-76.014		
		PRKCA	63.50-63.58			PRKCG	62.03-62.05
		CACNG5	63.621-63.627			**CACNG7**	62.05-62.07
CACNG3	22.17-22.27	**CACNG4**	63.68-63.73	**CACNG2**	75.34-75.46	**CACNG8**	62.08-62.10
		CACNG1	63.74-63.75			**CACNG6**	62.11-62.12

VIh. Paralogons (homologous chromosome segments) in the vicinity of rabbit CACNG genes on chromosomes 4, 6, 19, ? and ?.

Chromosome 6		Chromosome ?		Chromosome 19		Chromosome 4		Chromosome ?	
Gene	Location(Mb)	Gene	Location(Mb)	Gene	Location(Mb)	Gene	Location(Mb)	Gene	Location(Mb)
TBX6	18.45-18.46			TBX2	27.59-27.60				
				TBX4	27.64-27.67				
		SYNGR2	0.81-1.15			SYNGR1	84.15-84.15		
				CACNA1G	37.11-37.16			CACNA1A	14.05-25.86
				CACNG1	50.42-50.43			**CACNG6**	7.20-8.52
CACNG3	14.93-15.03			**CACNG4**	50.44-50.56	**CACNG2**	85.24-85.25		
								CACNG7	11.90-13.63
PRKCB	14.53-14.88			PRKCA	50.62-51.05				
TMEM114	2.56-2.57	**TMEM235**	5.02-5.56						
				SSTR2	56.72-56.72				

VIi. Paralogons (homologous chromosome segments) in the vicinity of mouse CACNG genes on chromosome 7, 11, 15, and 16.

Chromosome 7		chromosome 11		Chromosome 15		Chromosome 16	
Gene	Location	Gene	Location	Gene	Location	Gene	Location
PRKCG	7 A1						
CACNG7	7 A1						
CACNG8	7 A1						
CACNG6	7 A1						
GRIN2D	7 B4	GRIN2C	11 E2			GRIN2A	16 A1
SYNGR4	7 B4	SYNGR2	11 E2				
		TMEM235	11 E2			**TMEM114**	16 A3
PRKCB	7 F3	PRKCA	11 E1				
		CACNG5	11 E1				
CACNG3	7 F3	**CACNG4**	11 E1	**CACNG2**	15 E1		
		CACNG1	11 E1				
				SSTR3	15 E1		
				SYNGR1	15 E1		
		CACNA1G	11 D	CACNA1I	15 E1		
TBX6	7 F3	TBX4	11 C				
		TBX2	11 C				

VIj. Paralogons (homologous chromosome segments) in the vicinity of human CACNG genes on chromosome 16, 17, 19, and 22.

Chromosome 16		Chromosome 17		Chromosome 19		Chromosome 22	
Gene	Location	Gene	Location	Gene	Location	Gene	Location
SSTR5	16p13.3	SSTR2	17q24			SSTR3	22q13.1
CACNA1H	16p13.3	CACNA1G	17q21.33	CACNA1A	19p13.2	CACNA1I	22q13.1
SYNGR3	16p13.3	SYNGR2	17q25.3	SYNGR4	19q13.33	SYNGR1	22q13.1
TMEM114	16P13.2	**TMEM235**	17q25.3				
GRIN2A	16P13.2	GRIN2C	17q25.1	GRIN2D	19q13.33		
PRKCB1	16p12.2-p12.1	PRKCA	17q24.2	PRKCG	19q13.42		
		CACNG5	17q24.2	**CACNG7**	19q13.42		
CACNG3	16p12.1	**CACNG4**	17q24	**CACNG8**	19q13.42	**CACNG2**	22q12.3
		CACNG1	17q24.2	**CACNG6**	19q13.42		
TBX6	16p11.2	TBX2	17q23.2			TBX1	22q11.21
		TBX4	17q23.2				

Appendix VII. CACNG6 pairwise dN/dS

	D.rerio6a	D.rerio6b	T.rubripes6a	T.rubripes6b	O.niloticus6b	O.niloticus6a	L.chalumnae	C.bellii	C.familiaris	B.taurus	O.cuniculu	R.norvegicus	M.musculus	H.sapiens
D.rerio6a														
D.rerio6b	0.1087													
T.rubripes6a	0.0229	0.1170												
T.rubripes6b	0.0317	0.1099	0.2181											
O.niloticus6b	0.0484	0.1183	0.1505	0.0934										
O.niloticus6a	0.0167	0.1290	0.1626	0.1461	0.1097									
L.chalumnae	0.0111	0.0895	0.1991	0.3620	0.1995	0.1225								
C.bellii	0.1629	0.0056	0.1991	0.3275	0.1852	0.1427	0.2708							
C.familiaris	0.1458	0.3225	0.3293	0.4925	0.3789	0.2409	0.3880	0.3772						
B.taurus	0.0097	0.2161	0.2889	0.3379	0.2460	0.2098	0.2641	0.3721	0.2646					
O.cuniculu	0.0550	0.0084	0.3113	0.4082	0.0090	0.2002	0.2782	0.4337	0.3156	0.3638				
R.norvegicus	0.1102	0.2977	0.2282	0.2674	0.2511	0.1783	0.0058	0.2035	0.1775	0.2084	0.1634			
M.musculus	0.1421	0.2872	0.2877	0.3086	0.2339	0.2108	0.1685	0.2046	0.2113	0.2054	0.1725	0.0787		
H.sapiens	0.0089	0.2780	0.3205	0.4349	0.3031	0.2158	0.2811	0.3683	0.2743	0.2831	0.2637	0.1371	0.1510	

Appendix VIII. Results of branch model applied to additional copies of CACNG genes in osteichthyes

VIIIa. Two copies CACNG1 in osteichthyes LRT statistics of branch model

Two copies CACNG1	Number of sequences	Test 1	Test 2	2*(Lnl2-Lnl1)	p-value
after duplication	17	-7105.371431	-7101.014493	8.713876	0.003158
after speciation	17	-7096.448813	-7091.253677	10.39027	0.001267
between duplication and speciation	17	-7110.338209	-7110.213326	0.249766	0.61724

VIIIb. Two copies CACNG2 in osteichthyes LRT statistics of branch model

Two copies CACNG2	Number of sequences	Test 1	Test 2	2*(Lnl2-Lnl1)	p-value
after duplication	14	-6170.662206	-6167.394968	6.534476	0.01058
after speciation	14	-6171.854802	-6167.716557	8.27649	0.004016
between duplication and speciation	14	-6173.177464	-6173.176906	0.001116	0.97335

VIIIc. Two copies CACNG3 in osteichthyes LRT statistics of branch model

Two copies CACNG3	Number of sequences	Test 1	Test 2	2*(Lnl2-Lnl1)	p-value
after duplication	12	-7232.67	-7219.93	25.47447	4.48E-07
after speciation	12	-7241.573027	-7233.418811	16.30843	5.38E-05
between duplication and speciation	12	-7235.49	-7235.49	0	1

VIIId. Two copies CACNG4 in osteichthyes LRT statistics of branch model

Two copies CACNG4	Number of sequences	Test 1	Test 2	2*(Lnl2-Lnl1)	p-value
after duplication	13	-7878.380677	-7876.866615	3.028124	0.081833
after speciation	13	-7878.221063	-7876.788442	2.865242	0.090512
between duplication and speciation	13	-7876.823029	-7876.747478	0.151102	0.697484

VIIIe. Two copies CACNG5 in osteichthyes LRT statistics of branch model

Two copies CACNG5	Number of sequences	Test 1	Test 2	2*(Lnl2-Lnl1)	p-value
after duplication	15	-6258.022495	-6253.287101	9.470788	0.002088
after speciation	15	-6260.665712	-6255.773965	9.783494	0.001761
between duplication and speciation	15	-6250.916512	-6250.916512	0	1

VIIIf. Two copies CACNG6 in osteichthyes LRT statistics of branch model

Two copies CACNG6	Number of sequences	Test 1	Test 2	2*(Lnl2-Lnl1)	p-value
after duplication	14	-8288.845216	-8288.585788	0.518856	0.47133
after speciation	14	-8287.17655	-8287.016003	0.321094	0.570951
between duplication and speciation	14	-8304.041075	-8302.660627	2.760896	0.096594

VIIIg. Two copies CACNG7 in osteichthyes LRT statistics of branch model

Two copies CACNG7	Number of sequences	Test 1	Test 2	2*(Lnl2-Lnl1)	p-value
after duplication	15	-7269.150369	-7259.113058	20.07462	7.45E-06
after speciation	15	-7272.688768	-7261.80032	21.7769	3.06E-06
between duplication and speciation	15	-7260.161315	-7260.161315	0	1

VIIIh. Two copies CACNG8 in osteichthyes LRT statistics of branch model

Two copies CACNG 8	Number of sequences	Test 1	Test 2	2*(Lnl2-Lnl1)	p-value
after duplication	13	-10425.12055	-10424.98671	0.267688	0.604887
after speciation	13	-10425.12047	-10424.98662	0.267706	0.604875
between duplication and speciation	13	-10435.94662	-10435.94613	0.000982	0.975001

VIIIi. Two copies TMEM235 in osteichthyes LRT statistics of branch model

Two copies TMEM235	Number of sequences	Test 1	Test 2	2*(Lnl2-Lnl1)	p-value
before speciation	12	-7376.682703	-7376.496574	0.372258	0.54177603
after speciation	12	-7401.424358	-7400.639617	1.569482	0.210282256

Appendix IX CACNG sequences with suspected problems in annotation.

Sequences	Suspect problems
P. marinus γ_1 (ENSPMAT00000005800)	May lack exon 1 and 2
X. tropicalis γ_1 (ENSXETT00000033623)	No start codon
X. tropicalis γ_4 (ENSXETT00000055785)	Missing Roughly 52 nucleotides at beginning of exon 4
C.milii γ_3 (XP_007904006.1)	No start codon; exon 2 too divergent
D.rerio γ_2 (ENSDART00000013939)	According to amino acids and intron rule , delete 293 ~298 and replace as "TCAGAGTATTTTTTAC"
D.rerio γ_2 (ENSDART00000151099)	Comparing to the other gene structure , exon 2 might be not real