Redundancy In the Genetic Code: Selection Analysis and Its Implications for Reconstruction of Ancestral Protein Sequences

Ali Tehfe

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the degree of
Master of Science in Chemistry

Department of Chemistry and Biomolecular Sciences
Faculty of Sciences
University of Ottawa

Abstract

Ancestral Sequence Reconstruction is a technique used to statistically infer the most likely ancestor of a set of evolutionarily related sequences, but research which relies solely on protein data has the disadvantage of sequence information being lost upon translation of a protein from its gene transcript, due to the redundancy inherent in the genetic code. In this project, the amino acid sequences, and separately the corresponding codon sequences, of 184 homologous Acetylcholine receptor protein sequences were aligned, and phylogenetic analysis and ancestral sequence reconstruction was performed based on both alignments to infer several ancestral sequences representing important milestones in the evolutionary history of the homologous protein family. To further extract meaningful information from the nucleotide sequences, positive selection analysis was performed on the codon alignment using the Mixed Effects Model of Evolution method, which estimates and compares between the rates of synonymous and non-synonymous mutations across the alignment to detect the occurrence of positive selection events throughout their evolution. The Mixed Effects Model of Evolution can infer positive selection across both sites and evolutionary branches in a sequence alignment, thus highlighting residues along the evolutionary trajectory of the proteins which may have been functionally important in their evolution. Positive selection analysis detected positive selection at a multitude of sites and branches, and by mapping signatures at which selection is strongest with changes in the trajectory of ancestral states, several important sites were chosen as likely to be most valuable for future experimental testing. The implications of this study on the benefits of conducting ancestral sequence reconstruction with protein and codon sequences are discussed.

Acknowledgements

I would like to express my sincere appreciation to my research supervisor, Dr. Corrie daCosta, for his efforts, dedication, and patience in guiding me through this project. I would like to thank my fellow graduate students at the daCosta Lab, and especially Johnathon R. Emlaw, who collected the homologous AChR protein sequences used in this study. I am grateful to the graduate staff at the University of Ottawa for their assistance and understanding during the course of this program. I thank my brothers Ahmad and Reda for keeping me company during those many evenings spent typing away on my computer writing this paper. Finally, this work would not have been possible without love and motivation from my parents, Hussein and Batoul, to whom I dedicate my thesis.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

MEME	Mixed Effects Model of Evolution
EB Procedure	Empirical Bayes Procedure
AChR	Acetylcholine Receptor

Chapter I: Introduction

1.1 Ion Channels

Ion channels are porous protein molecules which span the plasma membranes of cells, allowing for the passage of ions through the membrane[1]. All ion channels contain an aqueous pore, which ions can selectively pass through once the ion channel is in its open conformation[2]. The conformational changes which result in channel gating and opening are controlled by different mechanisms. Voltage-gated ion channels are opened by changes in the electrical potential across the membrane, and this ability confers upon them the ability to play critical roles in controlling neuronal excitability[3]. Ligand-gated ion channels, on the other hand, open upon binding with a specific ligand, such as a neurotransmitter or a hormone, and have important roles within both the central and peripheral nervous systems[4]. Ion channels are typically assembled from multiple subunits which come together to form the pore-lining structure, and the number of subunits that comprise the complete protein varies between ion channel families[4,5].

1.2 Pentameric Ligand-Gated Ion Channels

Pentameric ligand-gated ion channels are the largest and most diverse family of ligand-gated ion channels[6]. As their name suggests, these ion channels are multimers formed from five subunits arranged around a central pore, and the vast number of combinations that can be made from individual subunits is an important reason behind why this protein family is so diverse[2]. Pentameric ligand-gated ion channels are allosteric receptors that transition between multiple conformations upon binding to their ligands, which may have excitatory or inhibitory effects on channel gating[7]. This family of proteins was previously known as Cys-loop receptors, as the Eukaryotic members of the family are notable for containing a 13-residue “Cys-loop”, which is sandwiched between two cysteines forming a disulfide bridge with one another[8]. Altogether, these proteins are

highly important for neuropharmacology, and further research can shed light on the mechanisms and potential treatments for numerous diseases that affect the nervous system[9], such as Alzheimer's disease, Parkinson's disease, and epilepsy.

1.3 Acetylcholine Receptors

Acetylcholine Receptors are pentameric ligand-gated ion channels that, upon binding to the ligand acetylcholine, are activated to allow for non-selective cation movement through its central pore[2]. AChRs play a major role in the central and peripheral nervous systems of many animal species, controlling muscle contraction at the neuromuscular junction and participating in neural communication in the sympathetic and parasympathetic nervous systems[2]. When an action potential reaches a motor nerve terminal, acetylcholine is released across the synaptic cleft and binds to acetylcholine receptors, where opening of the channels causes muscle contraction[2]. These acetylcholine receptors are often called "nicotinic" acetylcholine receptors to distinguish them from "muscarinic" acetylcholine receptors, which are G protein-coupled receptors that also play roles in the nervous system. However, it is important to note that these ion channels are not principally characterized by their binding to nicotine, and a large variety of ligands that can bind with AChRs exist[2].

Like other members of the superfamily, the overall structure of each AChR subunit consists of three domains: the extracellular ligand-binding domain (ECD), the transmembrane domain (TMD) consisting of four membrane-spanning helices (M1 to M4), where the M2 from each subunit lines the channel pore through the membrane, and finally the intracellular cytoplasmic domain (ICD) located between the M3 and M4 helices of each subunit[2] (Figure 1A). AChRs are formed from five (either identical or paralogous) AChR subunits (Figure 1B), with a wide variety of different subunit types having been found to exist[10]. The five muscle-type AChR subunits, associated with the neuromuscular junction, are the α, β, γ, δ, and ε subunits, with acetylcholine binding to muscle-type nAChR at both the interfaces of the α1 and γ subunits, and the α1 and either δ or ε subunits[11] (Figure 1C).

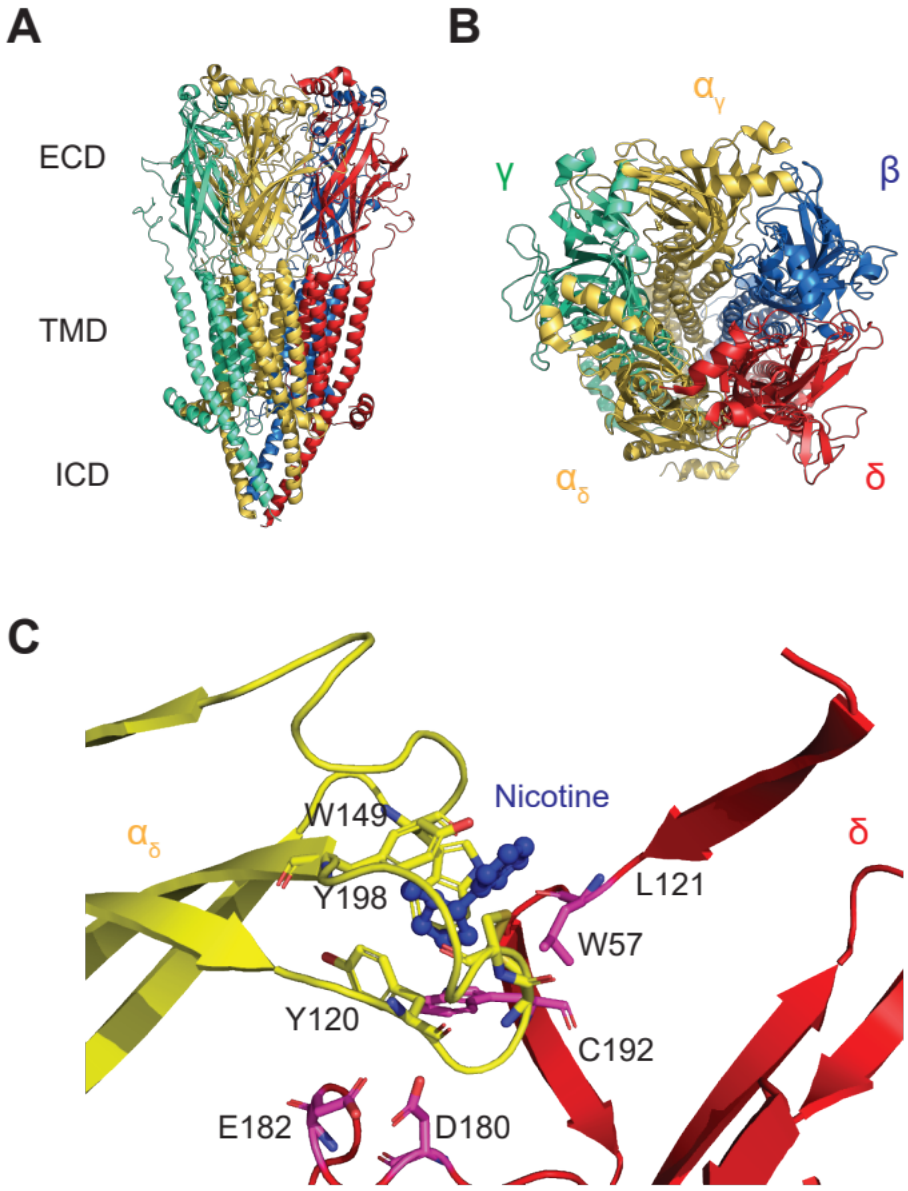


Figure 1: Structure of the Acetylcholine receptor from *Torpedo marmorata.* (A) Side-view and (B) top-view of the receptor structure in its resting state (PDB: 7QLO) with the extracellular domain (ECD), transmembrane domain (TMD), and intracellular domain (ICD) labelled in "A", and where each subunit is shown as a different colour. (C) The interface between α and δ subunits of the acetylcholine receptor bound to nicotine (PDB: 7QL5), with residues directly interacting with the nicotine ligand (blue, ball-and-sticks) shown as sticks and labelled with their one-letter amino acid codes and respective numbering. Figure made with PyMOL[71].

The other group of subunits are the neuronal-type AChR subunits[12], which are associated with inter-neuronal AChR activity and include at least twelve subunits, numbered α2 to α10 and β2 to β4, which are capable of either self-associating as homopentamers or interacting with each other as heteropentamers[13]. In each case, acetylcholine binds to the receptor in the extracellular domain near the N-terminus at the interface between two subunits, where three loops (A-C) forming the principal face of the first subunit captures and binds the ligand with the four loops (D-G) forming the complementary face of the second subunit[14]. As the principal and complementary faces are provided by different subunits, binding sites can be chemically distinct even within the same receptor (such as the muscle-type receptor, which possesses two binding sites), which illustrates the structural complexity of these proteins. The binding sites have evolved to bind with particular affinities and specificities to various ligands, and it is expected that certain sites participating in the binding site experienced selective pressures that contributed to their evolution.

The diversity of paralogous AChR subunits points to the receptor sequence having undergone functional specialization throughout its evolutionary history[15]. The specific details behind how mutations in the gene and protein sequence have fine-tuned the structure and function of its subunits will be further explored in the present study.

1.4 Ancestral Sequence Reconstruction

Ancestral Sequence Reconstruction (ASR) is a method for reconstructing the sequences of ancestral genes and proteins, and has emerged as a powerful method for characterizing the molecular evolution of proteins of interest[16]. In this technique, extant sequences of a protein family, either the nucleotide sequences of the genes or the amino-acid sequences of their protein products, are collected and aligned as a Multiple Sequence Alignment (MSA), which is then used to construct a phylogeny representing the relationship between the protein family members[17]. In the phylogenetic tree, each node represents an ancestor from which extant sequence lineages diverged, and ASR can use

evolutionary models to infer the sequences of these ancestors[10]. The evolutionary models are substitution models of DNA or protein evolution, which describe the relative evolutionary rates for substitutions between specific nucleotide or AA residues, coupled with a statistical approach - most often Maximum Likelihood (ML), although Maximum Parsimony (MP) and Bayesian methods may alternatively be implemented[18]- for inferring the optimal ancestral sequence[19].

Both ML and Bayesian-based methods are probabilistic and model evolutionary processes as per a time-reversible Markov process, whereas MP approaches optimize for ancestral sequences requiring a minimum number of substitutions from their descendants and are relatively unsophisticated, for that reason MP is often considered the least reliable method[18]. ML-based approaches score and calculate the residue that has the greatest statistical likelihood to occupy each particular position in the ancestral sequence, and have been shown to reconstruct more stable and active proteins than MP and Bayesian procedures, making it the most frequently used approach in ASR[20]. ML reconstruction methods may be categorized as either marginal or joint, with marginal reconstruction inferring character states independently for each node while joint reconstruction infers the most likely set of character states for all nodes taken together, and marginal reconstruction is thus most suitable for reconstructing particular ancestors[21]. As the ML framework uses the topology and branch lengths of the phylogenetic tree as parameters when inferring the ancestral sequences, the topology and construction of the phylogeny are by necessity crucial components to this method, and phylogenetic uncertainty can hamper the accuracy of reconstructions.

In the field of protein engineering, ASR can be used to engineer proteins with enhanced properties such as activity or stability, with some ancestral proteins having been found to be more theormostable than their modern descendants[22]. Furthermore, while ASR has most commonly been implemented for the study of protein sequences, it has also been found able to accurately reconstruct ancestral DNA sequences[23]. However, the primary value of ASR remains in its application towards molecular evolution, and particularly the study of the structure and function of modern-day proteins[24]. It may be used to better understand how a protein family diversified from their ancestors in stepwise fashion[25]. By comparing the evolution of ancestral sequences reconstructed by

ASR with the extant proteins descended from them, it is also possible to identify residues of interest that have most significantly contributed towards structural and functional changes within protein families[26]. In combination with molecular biology experiments, ASR has been used to reveal otherwise difficult to identify substitutions responsible for modifying function by impacting factors such as conformational dynamics[27], or even epistasis and evolutionary contingencies[28].

1.5 Positive Selection Analysis

The redundancy inherent to the genetic code has the consequence that a mutation of a nucleotide base in a gene sequence may be either synonymous, if the mutated codon codes for the same amino acid, or non-synonymous, if the resulting codon codes for a different amino acid (Figure 2A), or a stop codon. In positive selection analysis, each codon site is tested to determine the relative proportion of sequences that have experienced non-synonymous mutations at that site relative to synonymous mutations; as non-synonymous mutations frequently affect the fitness of the translated protein, it is assumed that a higher proportion of non-synonymous mutations indicate that the site is being positively selected for.

A simple illustration of positive selection analysis can be provided by comparing two codon sequences with each other (Figure 2B). The non-synonymous mutation rate can be calculated by counting the number of nucleotide differences in "sequence B" that have led to amino-acid changes compared to "sequence A", divided over the total number of nucleotide sites which, if mutated, would have caused an amino acid substitution. The calculation of the synonymous mutation rate follows the same logic, and comparing the two mutation rates allows for the ratio of synonymous to non-synonymous mutations, or "ω", to be easily calculated (Figure 2B).

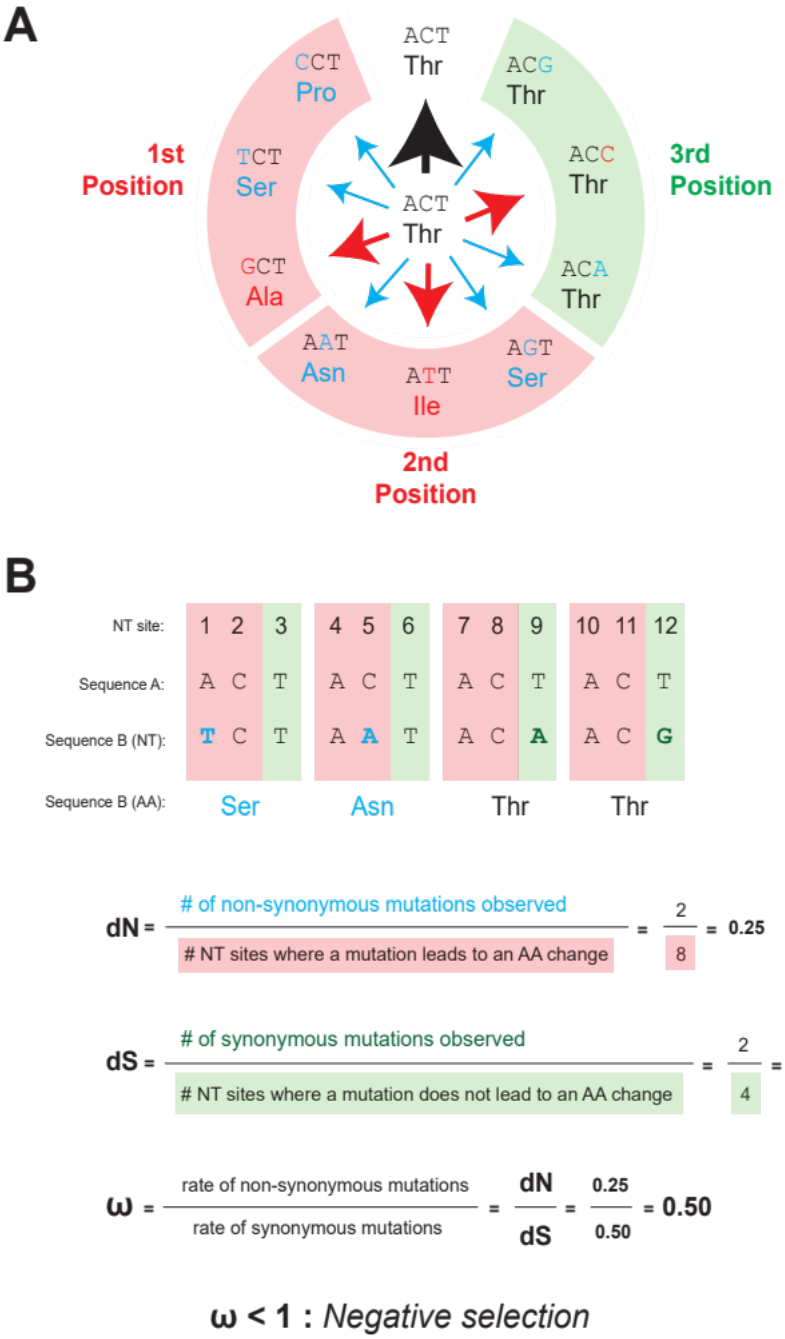


Figure 2: An illustration of the redundancy of the genetic code and its use in positive selection analysis. (A) Mutations in the ACT codon may be non-synonymous (1st and 2nd positions, red) or synonymous (3rd position, green) depending on whether they result in a mutation in the expressed amino acid. The arrows are weighted according to their likelihood of occurrence, with transitions (red) more likely than transversions (blue). (B) A simple example showing how the relationship between synonymous and non-synonymous mutations can be used to identify positive selection in codon sequences. The calculation reveals that the ω ratio is smaller than 1 for these two sequences, suggesting negative selection.

In practice, however, such calculations are too simple to be used in positive selection analysis tests, and the substitution rates are instead determined using codon substitution models coupled with an ML/Bayesian optimization approach to find ω (refer to the appendix for further information)[19].

The ratio of synonymous to non-synonymous mutations, or "ω", is the backbone of the branch-site test for detecting positive selection[29]. The branch-site test is a likelihood-ratio test that estimates the synonymous and non-synonymous mutation rates and compares them using at least two models: a "constrained" model that assumes neutral or negative selection, and an "unrestricted" model that allows for positive selection. If using the unrestricted model at some sites significantly improves the overall model fit relative to the constrained model, the branch-site test detects positive selection as likely to be occuring. The branch-site test can be readily applied to determine ω at different branches within a phylogeny (Figure 3), or to determine ω at individual sites within an alignment (Figure 4).

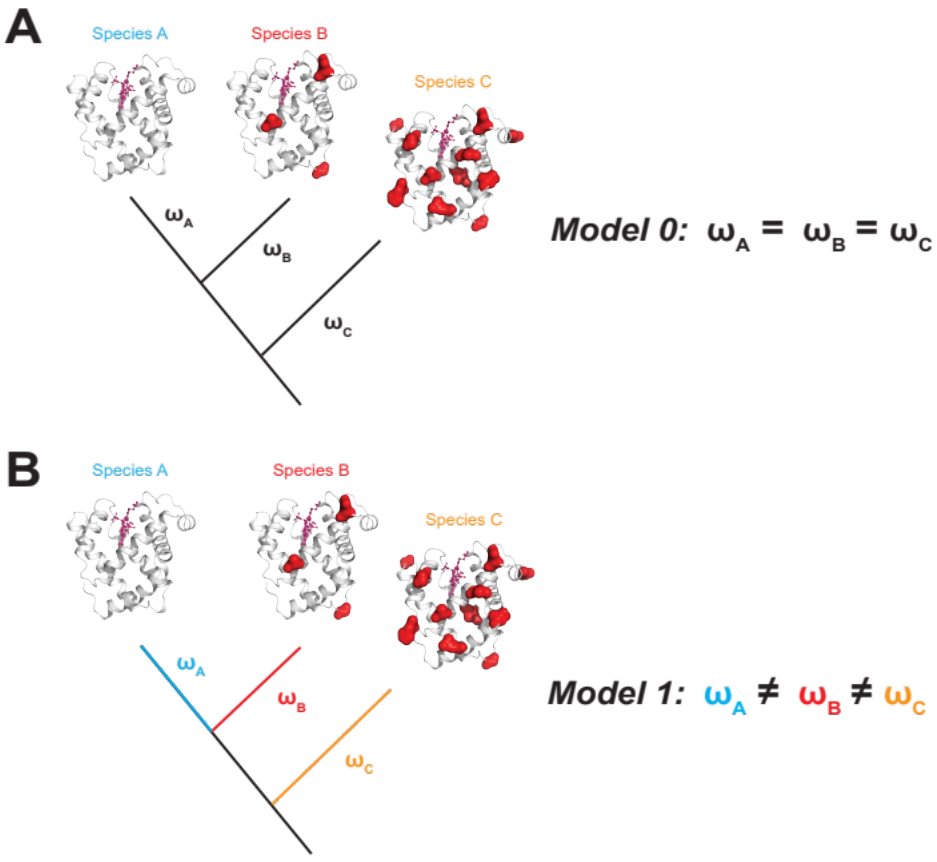


Figure 3: Models for comparison of ω in different branches of a phylogeny. (A) In model 0, the branches leading to the "species A", "species B", and "species C" nodes are assumed to have the same ω value. (B) In model B, the branches leading to "species A" (blue), "species B" (red), and "species C" (yellow) all have their own ω value.

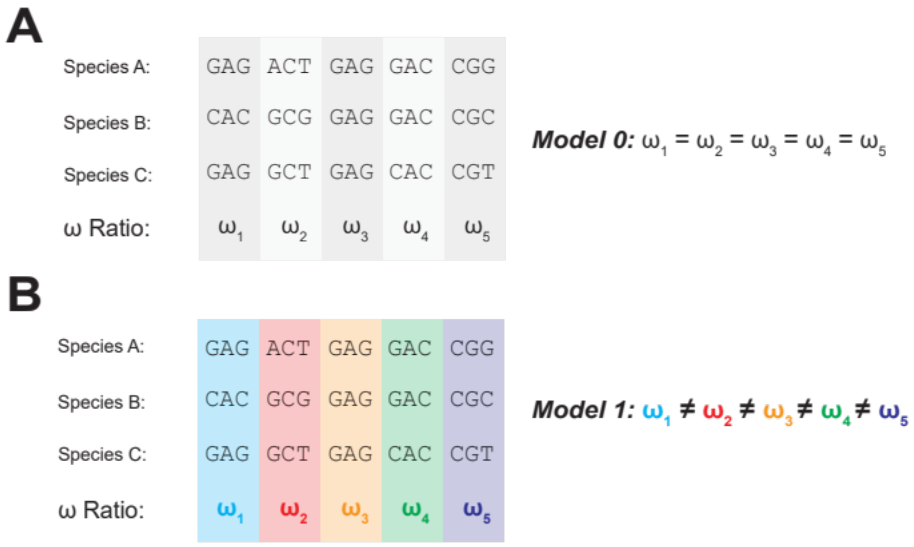


Figure 4: Models for comparison of ω at different sites within an alignment (A) In model 0 for an alignment of three sequences with five codon sites, each codon site is assumed to shares the same ω value. (B) In model 1, each of the five codon sites in the alignment is assigned its own ω value (depicted as their own colour).

The original branch-site test for detecting positive selection in sequences would keep the ω ratio constant for all sites in an alignment (as in Figure 3), but was succeeded by "random effects" branch-site tests[29] that used codon substitution models capable of estimating substition rates while simultaneously permitting the ω ratio to vary from site to site (as in Figure 4). This made it possible to identify instances where only a small proportion of sites in a sequence are undergoing positive selection. However, these random effects branch-site tests do not allow ω to vary between branches, which has the disadvantage of assuming that most branches for that site in the phylogeny are undergoing positive selection, which is rarely the case[30]. To control for this, a new random effects branch-site test was introduced, the Mixed Effects Model of Evolution (MEME)[31]. While keeping the "fixed effect" of varying ω from site to site, MEME also allows the ω ratio to vary from branch to branch at a site, called the "random effect" because it allows for all possible assignments of branch rates to sites to be considered (Figure 5). This is especially useful in detecting positive selection at sites in very large sequence alignments, as it has been noted that including too many sequences in an alignment may result in sites no longer being detected as significant, because purifying selection occurring on some lineages may mask the positive selection on other lineages[31].

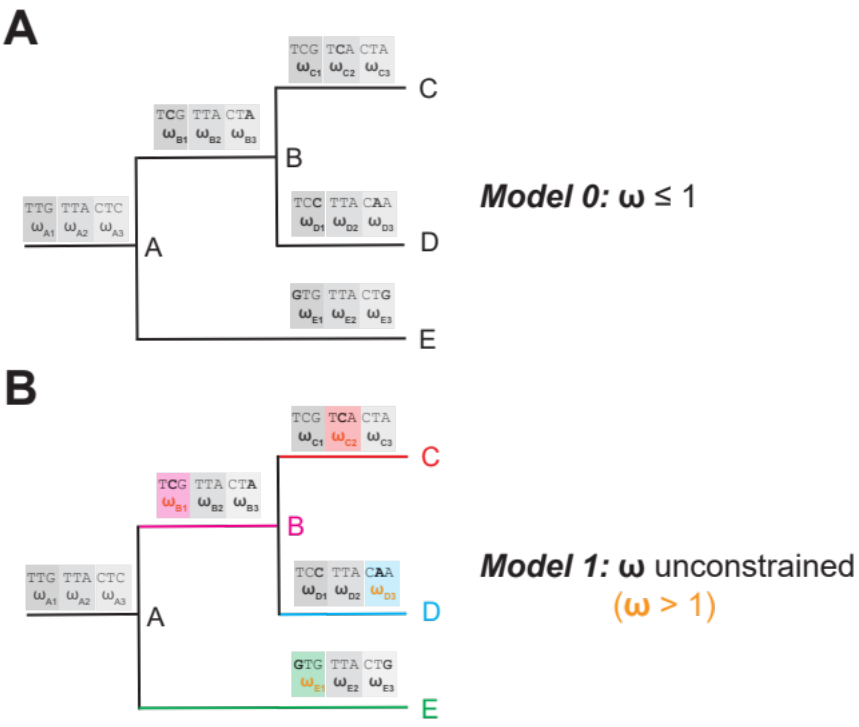


Figure 5: The Mixed Effects Model of Evolution (MEME) is a random effects approach to positive selection analysis which allows ω to vary both between sites and between branches. (A) In model 0, each site at each node in the phylogeny is assumed to be under neutral or negative selection. (B) In model 1, ω is unconstrained, and certain codon sites (coloured) across the phylogeny have been detected to be under positive selection.

MEME was found to be superior in detecting positive selection relative to the older random effects branch-site tests. However, it is incorrect to assume, as branch-site tests do, that all synonymous mutations are neutral. Synonymous mutations may influence many processes, such as selecting for codon bias during translation[32], and studies have shown synonymous mutations can in many cases even influence the overall fitness of an organism to a similar extent to non-synonymous mutations[33]. This presents a potential limitation to our confidence in positive selection detection testing and MEME. Furthermore, it is important to note that while MEME may be able to detect if a proportion of branches at a site are under episodic selection, it cannot accurately identify which of the individual branches at that site are undergoing positive selection because of the uncertainty inherent to attempting to infer both simultaneously[31]. Nevertheless, one Bayesian method, the empirical Bayes (EB) procedure[34], was suggested as having potential to shed light on whether a specific branch is likely to be experiencing selection for a site found to be significant by MEME, but because such inferences would by necessity be based on limited evidence[31], the EB procedure was not recommended and is not implemented in this study.

1.6 Thesis Objectives

Acetylcholine receptors are important for many physiological processes, and their dysfunction is implicated in neuropathic diseases such as Alzheimer's, Parkinson's, schizophrenia, and autism spectrum disorder[35]. Acetylcholine receptors are formed from a variety of subunits, each with their own unique properties, which determines their pharmacological profile and physiological function[36]. Understanding how specific subunits impart different properties will help us understand disease, and allow for the design of more effective therapeutics[37]. Unfortunately, aspects of AChR function remain elusive, and the relative contributions of individual residues towards receptor functions such as ligand binding, ion conductance, conformational transitions, and protein structure remains unclear, as well as how these residues have evolved to differentially interact with

other subunits. Through the following three specific objectives (I, II, and III), this study aims to address these gaps in our knowledge by taking advantage of the large amount of amino acid and nucleotide data presently available in sequence databases:

I. Perform phylogenetic analysis of a diverse set of acetylcholine receptor subunit sequences from different animal species, using both the amino acid and complementary nucleotide sequences, and compare the amino acid-based phylogeny with the nucleotide-based phylogeny.
   - ***Hypothesis I: That the nucleotide-based phylogeny will outperform[38] the amino acid-based phylogeny in accuracy and branch support.***

II. Reconstruct and compare key ancestral acetylcholine receptor subunit sequences, using both the amino acid and nucleotide sequence data.
   - ***Hypothesis II: That the nucleotide-based ancestral sequence reconstructions will have higher confidence than the respective amino acid-based ancestral protein reconstructions.***

III. Exploit the extensive nucleotide-based multiple sequence alignment to perform positive selection analysis of the acetylcholine receptor subunits, using the Mixed Effects Model of Evolution (MEME) method, to detect codon sites that have signatures of positive selection.
   - ***Hypothesis III: That those codon sites detected to be undergoing positive selection will have documented contributions to the structure and function of the acetylcholine receptor.***

In summary, this research project, focusing on the acetylcholine receptors, aims to provide a proof of concept for how nucleotide data can be used to gain insight into proteins by highlighting the residues most likely to be contributing to protein function, providing new opportunities for uncovering the mysteries of protein structure, function, and evolution.

Chapter II: Methods

2.1 Retrieving the AChR sequences

The protein sequence alignment used in this thesis was collected and prepared by Johnathon R. Emlaw. All acetylcholine receptor protein sequences, both paralogues and orthologues, were retrieved from the NCBI GenBank database[39]. From this initial collection, certain sequences, such as those from teleost fish species (Table 1), were excluded from the alignment so that the resulting phylogeny would match with the consensus species phylogeny fully accessible and curated in the Open Tree of Life[40]. The signal peptides of the sequences chosen were also trimmed prior to alignment to further improve the accuracy of the resulting phylogeny. The final alignment contained 184 homologous AChR sequences, along with two $5HT_3A$ (5-hydroxytryptamine receptor 3A) sequences chosen as the outgroup.

To collect the complementary DNA sequences, each protein sequence was searched as a query sequence using the NCBI TBLASTN tool[41] for identifying the complementary DNA sequence from which they were translated (Figure 6A). The NCBI accession code for both the protein and cDNA sequences used in creating the multiple sequence alignments were recorded (Table 1 in the appendix). The complementary DNA sequence was found for each protein sequence, and no residue identity was changed, with two notable exceptions discussed further below. Since the signal peptide, and certain other residues, had previously been removed from the protein sequences in order to provide for a tighter protein alignment, it was necessary that the cDNA sequences be trimmed in an identical manner. To accomplish this, each cDNA sequence was input into the ANTICALIgN program[42], which is capable of automatically translating DNA sequences so that they can be easily and individually modified to correspond to each protein sequence.

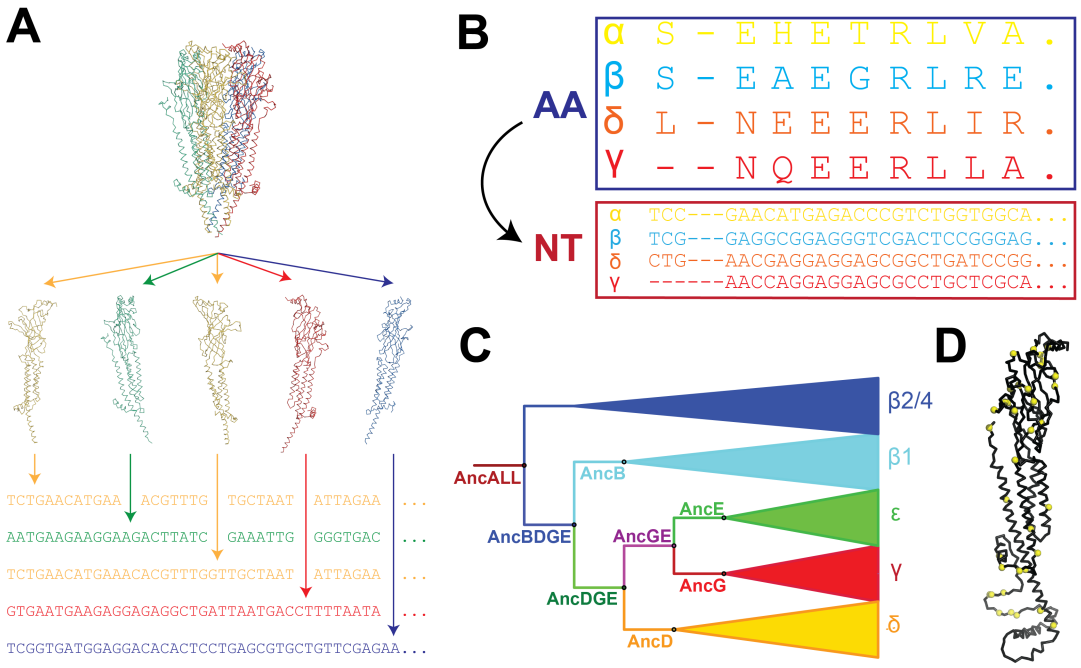


Figure 6: Methods used in this research project. (A) Collection of both protein (PDB: 7QLO) and complementary nucleotide sequences from the NCBI database. (B) Alignment of the nucleotide sequences (red box) to perfectly match the amino-acid alignment (blue box) of their respective protein sequences. (C) Phylogenetic analysis and ancestral sequence reconstruction is performed to reconstruct representative ancestral subunit sequences (labelled nodes shown in bold). (D) Detection of sites undergoing positive selection (shown as yellow spheres for AncALL) using MEME and correspondance to ancestral subunit sequences. Structural model of the ancestral sequence was generated by AlphaFold2.[48,52,69,70.]

2.2 Dense taxon alignment

The protein sequences were aligned with PRANK[43], using 5-hydroxytryptamine sequences as an outgroup (see end of Table 1 in the appendix). To map the DNA sequences onto the protein alignment, it was necessary that they match perfectly with the alignment, and as such, two sequences had to be modified. Firstly, the protein sequence for the α1 subunit of *Torpedo marmorata* was modified with the V323L and S418C substitutions so that they matched with the codons in the respective DNA sequence, which were likely more accurate as the protein sequence had been sequenced earlier. Secondly, in the δ subunit of *Vicugna pacos*, there is an unresolved residue in both the protein (X83) and cDNA (n247) sequences that are recorded in the NCBI database. As it was necessary that all sequence positions be known, the unknown residue was predicted based on homology by using TBLASTN for pairwise alignment of the *Vicugna pacos* δ subunit sequence with the δ sequences of its closest relatives, the camelids of genus *Camelus* (*Camelus dromedarius*, *Camelus bactrianus*, and *Camelus ferus*), where the unresolved residue aligned with a proline. The X83P and n247c substitutions were thus made in the protein and DNA sequences, respectively.

The bioinformatics program DAMBE[44] was used to map the collected cDNA sequences onto the protein alignment (Figure 7). The names of the protein and DNA sequences were set to exactly match one another, so that DAMBE could align the nucleotide sequences by mapping them to the amino acids which they translated to in the protein alignment. The cDNA alignment was then translated into a protein alignment using DAMBE, and likelihood analysis of the two alignments using PhyML-SMS[45] found the likelihood scores and predicted best model (JTT+G+I) to match each other, proving that they are identical.

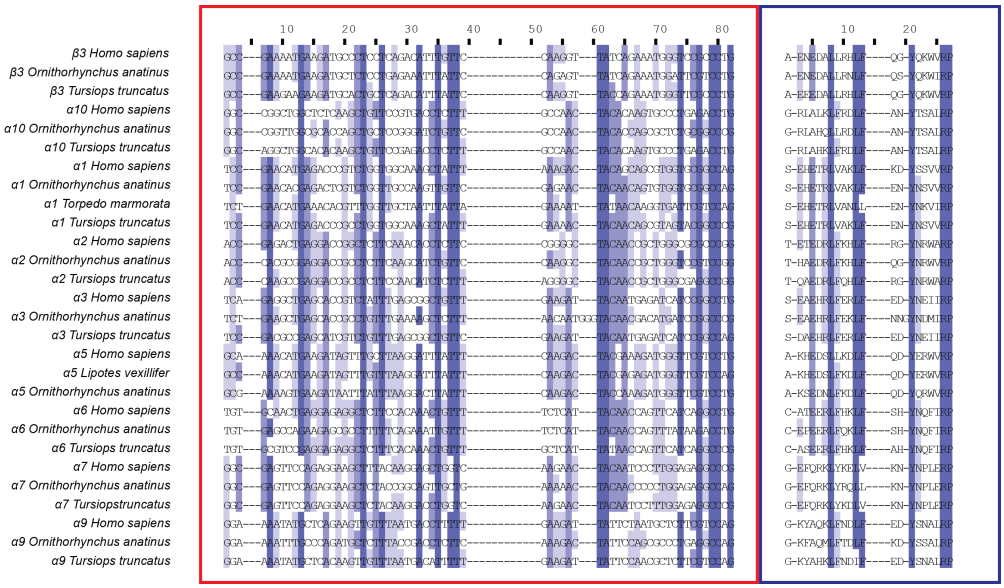


Figure 7: Select region from the multiple sequence alignment of the homologous acetylcholine receptor sequences displaying both the nucleotide (red box) and respective protein (blue box) sequences. Both alignments are coloured based on their relative conservation at each site, with darker blue indicating the site is more conserved. Conservation analysis done based on percentage identity. Figure made with JalView[52].

2.3 Phylogenetics and ASR

Phylogenetic trees were inferred from the cDNA and protein alignments using PhyML-SMS, with SH-like aLRT values also recorded. The best model for the protein alignment was JTT+G+I, and the best model predicted for the cDNA alignment was GTR+G+I. The $5HT_3A$ sequences were culled from both alignments, such that the corresponding $5HT_3A$ tips were dropped from the resulting trees. The phylogenetic trees, minus these $5HT_3A$ outgroups, were then rooted to the last universal common ancestor using the APE package in R[46]. The APE package was also used to construct dendograms of both phylogenetic trees, and to compare them by creating a tanglegram to compare their topologies[46].

Important ancestral sequences proximal to extant AChR sequences were selected to be reconstructed (Figure 6C): these are the ancestors of all β1 (AncB), δ (AncD), ε (AncE), and γ (AncG) subunits, as well as AncGE (ancestor of AncG and AncE), AncDGE (ancestor of AncGE and AncD), AncBDGE (ancestor of AncDGE and AncB), and AncALL (ancestor of AncBDGE and β2 and β4 sequences).

Ancestral sequence reconstruction of both the protein and DNA sequences was performed using the program Lazarus[47] which is based on another program, PAML[19]. The substitution matrices were generated using baseml (DNA) or codeml (protein) in PAML, and GTR and JTT+G+I selected as the substitution model in Lazarus for the DNA and protein models, respectively. The posterior probability distributions were also characterized using the plot-pp-distribution script within Lazarus. For the ancestral DNA sequences, this was done by characterizing the posterior probability distributions seperately for the first, second, and third position nucleotide sites in each codon, and summing the calculated posterior probability for each codon that translates into the amino acid with the highest posterior probability at that site. For example, suppose that the amino acid with the highest posterior probability at a site was leucine. In this case, the posterior probabilities for the TTA, TTG, CTT, CTC, CTA, and CTG would be individually calculated, and then these would be summed together to get the total value.

2.4 Positive Selection Analysis

Positive selection analysis with MEME was performed using the phylogenetics suite HyPhy (QuickSelectionDetection.bf batch file method) on the cDNA alignment. The bootstrap values were first manually removed from the phylogenetic tree file as they are not accepted in HyPhy. The sites detected to be significant ($p \leq 0.05$) as undergoing positive selection were compared with the ancestral subunit sequences (Figure 6D). Predicted structures of the reconstructed ancestral subunits (from both amino acid reconstructions and translated cDNA reconstructions) were generated by the artificial intelligence program AlphaFold[48].

Chapter III: Results

3.1 Comparison of Phylogenetic Trees

The phylogenetic trees obtained from the cDNA (Figure 8) and amino acid (Figure 9) sequence alignments were similar, but nevertheless different. In a tanglegram, which depicts side-by-side comparison of the two trees (Figure 10), where differences in the topology of the two trees are reflected in the number of lines crossing one another and distinct branches are shown as dotted lines, it can be seen that the α2/α4 subunits are more closely related to the α3/α6 subunits in the nucleotide phylogeny, while the α2/α4 subunits are more closely related to the β3/α5 subunits in the protein phylogeny. Furthermore, in the α2 and α7 clades, which are composed of three sequences each, the three sequences differ between the two trees in their relationships to one another, as also indicated by the dotted lines. As these paralogous sequences do not descend from any of the keystone ancestral subunits in this study, they should not impact the sequence reconstructions of the identified ancestors of interest. On the other hand, a close inspection of the ε subunits does show that the *Latimeria chalumnae* branches are slightly distinct, being more closely related to the chondrichthyes (*Torpedo marmorata*, *Hypnos monopterygius,* and *Scyliorhinus torazame*) in the nucleotide phylogeny relative to the protein phylogeny. In terms of the number of sequences, there is ample coverage for the subunits descended from AncALL: β1 (32), β2/β4 (33), γ (25), δ (36), and ε (27). However, relative to the other muscle-type subunits, α1 has very low coverage (4), and the remaining accessory subunits consist of 3 sequences each.

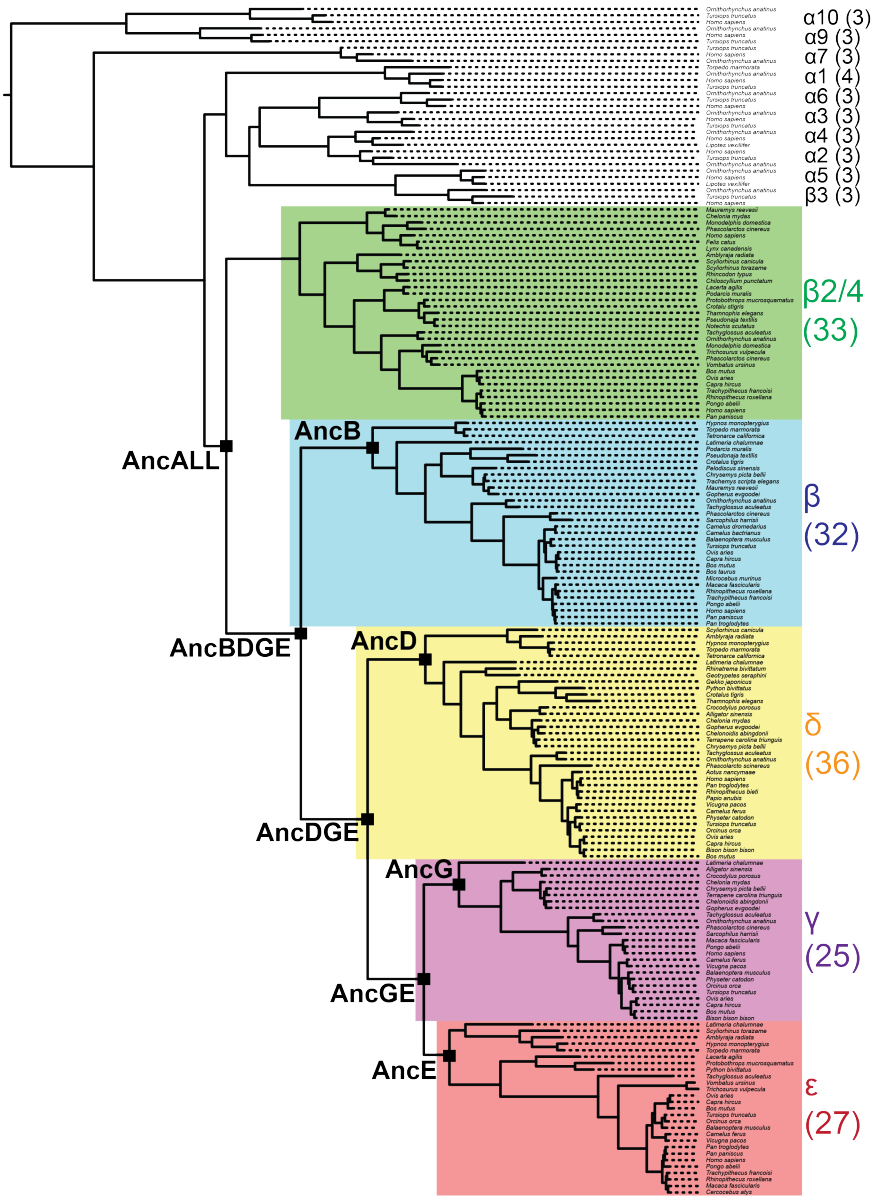


Figure 8: The inferred phylogenetic relationship between nucleotide sequences (n=184). The acetylcholine receptor subunits are highlighted by their type: β (blue), γ (purple), ε (red), δ (yellow), β2/β4 (green), and remaining paralogous subunits (grey). Ancestral subunits of interest (AncALL, AncBDGE, AncB, AncDGE, AncD, AncGE, AncG, and AncE) are also indicated on the tree (bold). Figure made with FigTree[70].

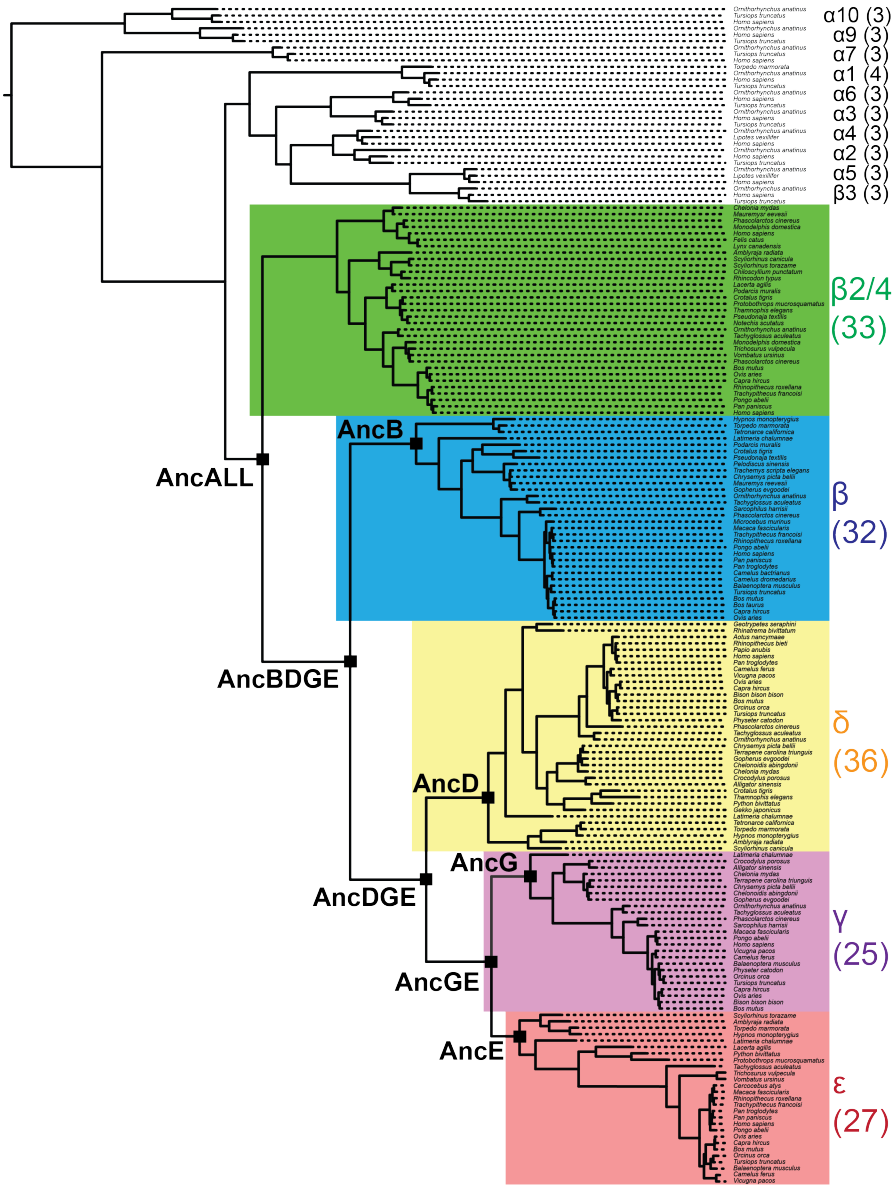


Figure 9: The inferred phylogenetic relationship between amino acid sequences (n=184). The acetylcholine receptor subunits are highlighted by their type: β (blue), γ (purple), ε (red), δ (yellow), β2/β4 (green), and remaining paralogous subunits (grey). Ancestral subunits of interest (AncALL, AncBDGE, AncB, AncDGE, AncD, AncGE, AncG, and AncE) are also indicated on the tree (bold). Figure made with FigTree[70].

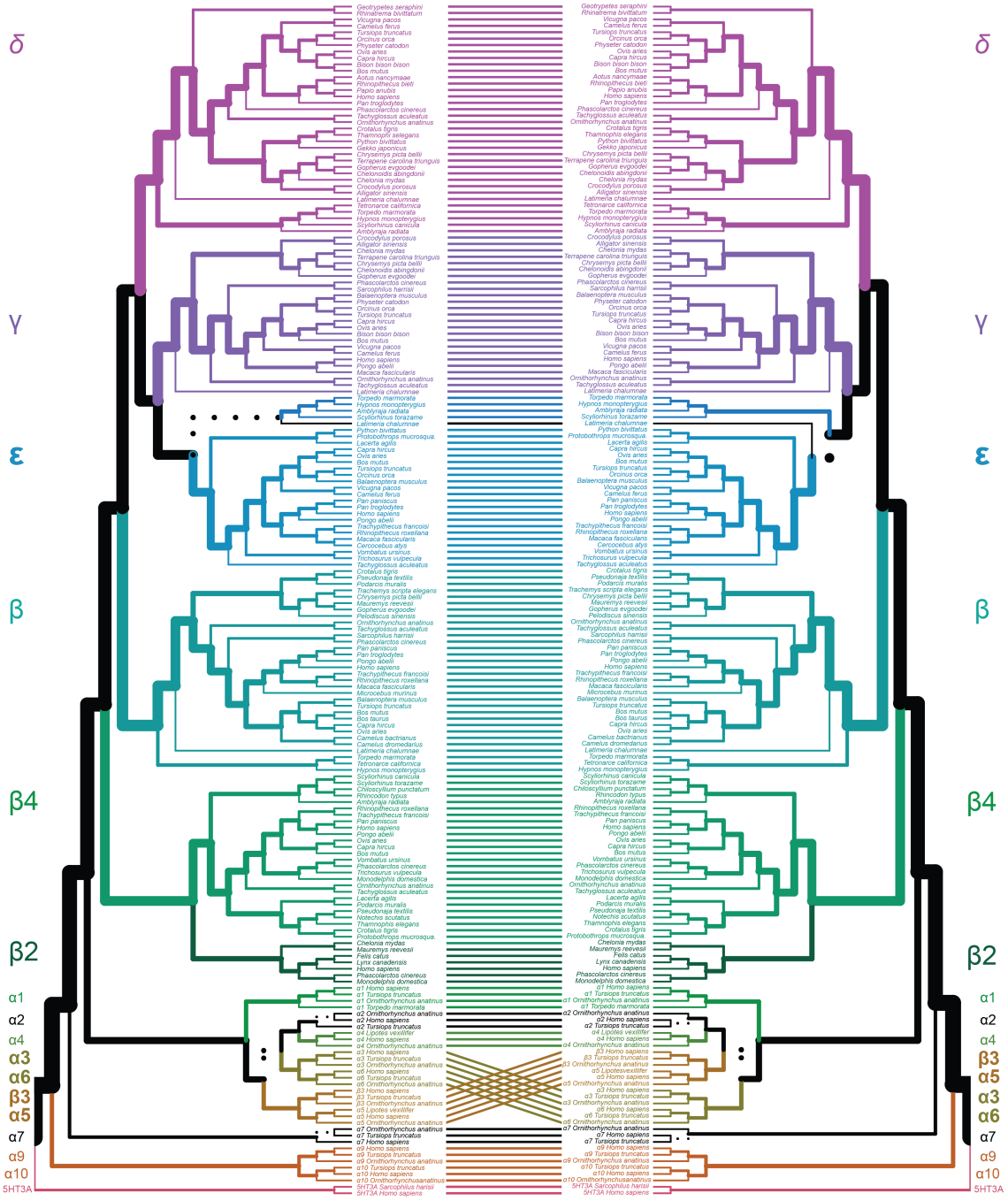


Figure 10: Tanglegram comparing the topology of phylogenies generated using the nucleotide (left) and amino acid (right) alignments. Lines crossing in the center indicates a discrepancy in topology of the two inferred alignments, and discrepant subunits are in bold. Distinct branches are shown as dotted lines. Figure constructed with R[72].

The relative branch-lengths in the cDNA and protein phylogenies (averaged at each branch for the α, β, γ, δ, and ε subunit branches) were compared (Figure 11), by comparing each branch from the tree root to the *Homo sapiens* sequence of each subunit. Moving forward in time from the tree root, the first six branches are longer in the protein phylogeny compared to the cDNA phylogeny, while the latter eight branches are longer in the cDNA phylogeny (Figure 11A). In both phylogenetic trees, the branch-lengths tend to become shorter moving further in time along the tree (Figure 11B), with the exception of branch 3 where the branch-lengths are much shorter compared to branches 2 and 4.

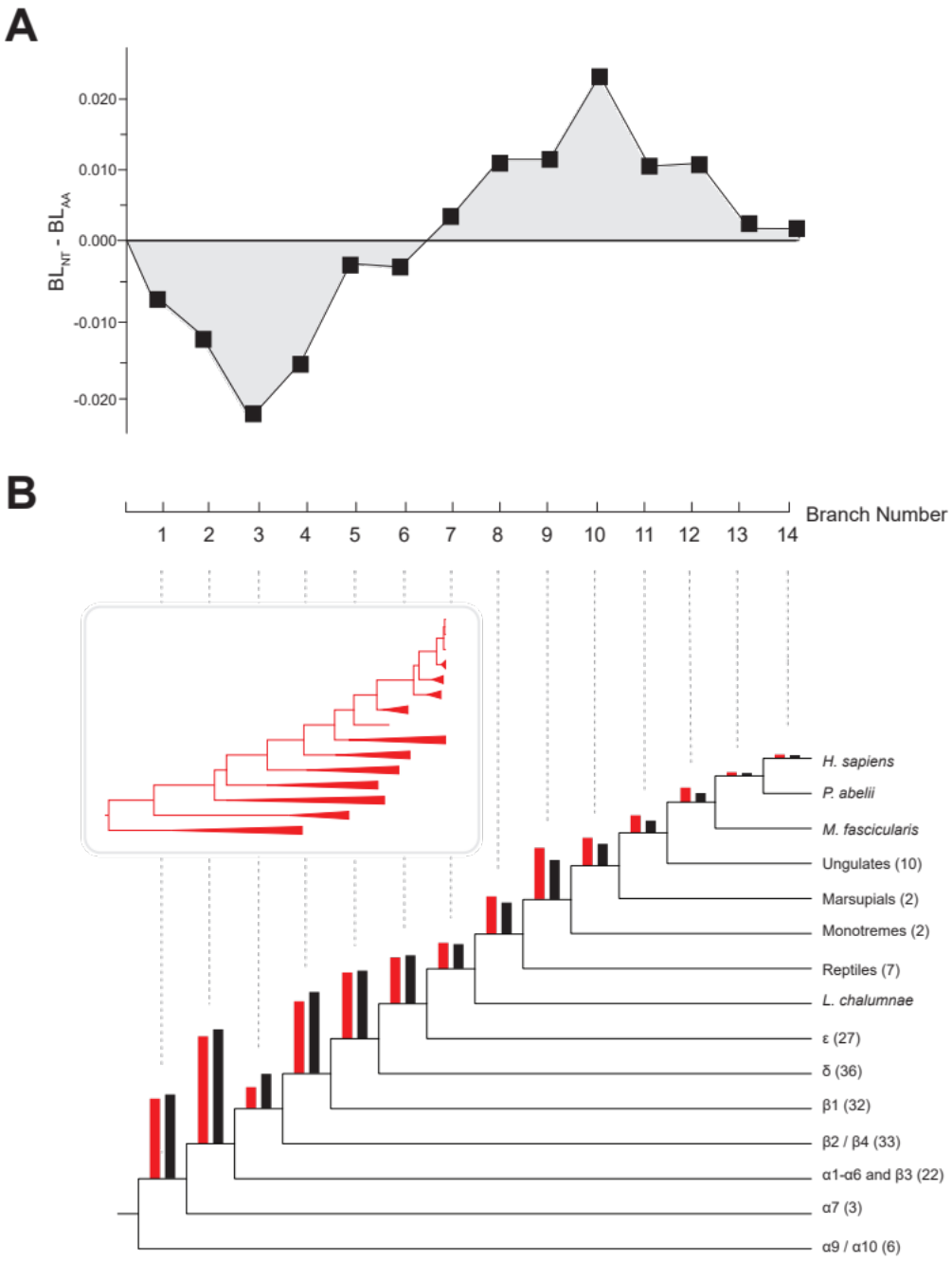


Figure 11: Comparison of branch-lengths in the phylogenies inferred from nucleotide and amino acid sequences. (A) A line chart comparing the relative branch-lengths of the two phylogenies. Branches at which the branch-length is larger in the phylogeny inferred from amino acid sequences are under the X-axis, while branches at which the branch-length is larger in the phylogeny inferred from nucleotide sequences are above the X-axis. (B) A cladogram of the phylogeny inferred from amino acid sequences, where above each branch are two bars indicating the indicated branches relative length in the nucleotide (red) and amino acid (black) phylogenies. The nucleotide-based phylogenetic tree (red) is shown in the inset.

The aLRT scores of the two phylogenetic trees were compared to each other. In the aLRT-SH method used by PhyML, the aLRT score reflects the confidence level in any putative branch in a phylogeny, making it useful for interpreting the statistical confidence of a particular inferred phylogenetic tree[45]. However, as this method is dependent on the topology of the tree[49], it is important that the two phylogenetic trees share the same topology as one another to control for such a comparison; as the tree topologies do match one another with only the minor exceptions previously mentioned, this should not significantly interfere with the results. A comparison of the two phylogenetic trees (Figure 12) shows that most branches have very similar aLRT scores (the scores are not shown for branches with aLRT score differences smaller than 0.1). Of those branches with score differences of 0.1 or higher, averaging the scores together reveals that the aLRT scores of the nucleotide tree branches are higher by an average of 0.12, indicating a higher level of overall confidence for the nucleotide phylogeny. However, there are some exceptions, such as early in the tree at the branch ancestral to the β2/4, β, γ, δ, and ε subunits, where the aLRT score is 0.90 in the amino acid phylogeny but only 0.23 in the nucleotide phylogeny.

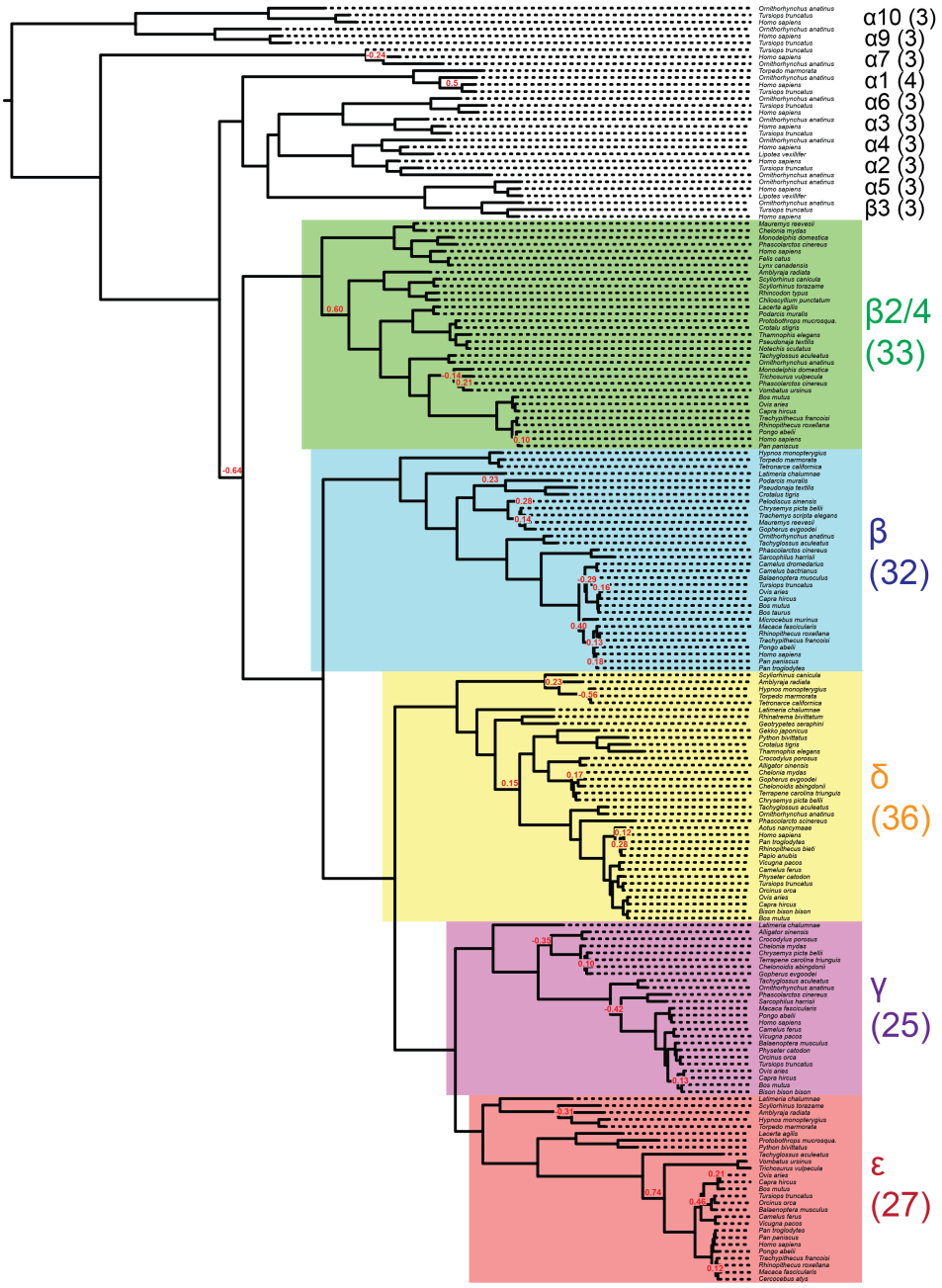


Figure 12: Comparison of the approximate likelihood ration test (aLRT) scores for each branch from the phylogenies inferred from nucleotide and amino acid sequences. The aLRT score of the phylogeny inferred from nucleotide sequences was subtracted by the aLRT score from the phylogeny inferred from amino acid sequences, and the resulting values are shown in red above their respective branches in this phylogentic tree inferred from the nucleotide sequences; however, if the aLRT difference is below 0.1, or if there is no corresponding branch in the phylogenetic tree inferred from amino acid sequences, no value above the branch is shown.

3.2 Comparison of Ancestral Sequence Reconstructions

The ancestral subunit sequences reconstructed by ASR are presented in the appendix (Tables 3 and 4) for a select number of significant sites (columns in the alignment, not nodes) in the alignments from which they were generated. The ancestral subunit sequences reconstructed from the protein dataset, and the ancestral subunit sequences reconstructed from the cDNA dataset and then translated to their corresponding amino acid sequences, differed from one another in percentage identity. The percentage identity differences were close to one another (a range of less than 10%), varying from 8.85% in AncG to a maximum of 17.88% for AncBDGE.

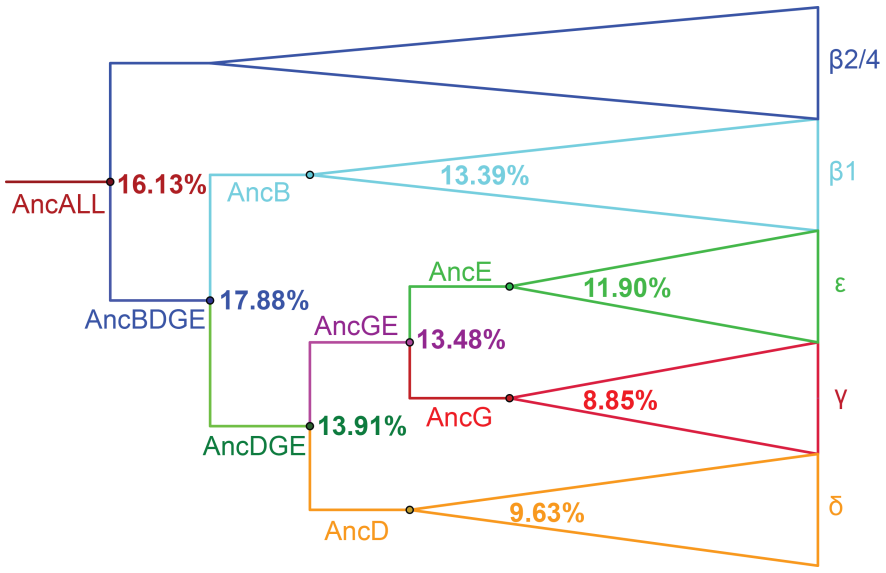


Figure 13: Percentage identity difference between the reconstructed ancestral amino acid sequences from the amino acid and translated nucleotide sequences. The percentage identity difference for each ancestral subunit is shown in bold to the right of its respective node in the phylogenetic tree. Percentage identities were calculated by pairwise alignment with Jalview[52].

To compare the relative confidence in the ancestral state inferences, the posterior probabilities were compared to identify differences between the two sets of reconstructions for AncALL, and the resulting posterior probability distributions graphed as histograms (Figure 14). In general, most sites have a higher posterior probability score in the protein-based reconstruction relative to the nucleotide-based reconstructions (Figure 14C), and this is true for most of the sites identified to be undergoing positive selection (shown in red) as well. Nevertheless, a number of sites do have higher posterior probability scores in the nucleotide-based reconstruction, and many also have similar posterior probability scores (close or equal to 0 in Figure 14C).

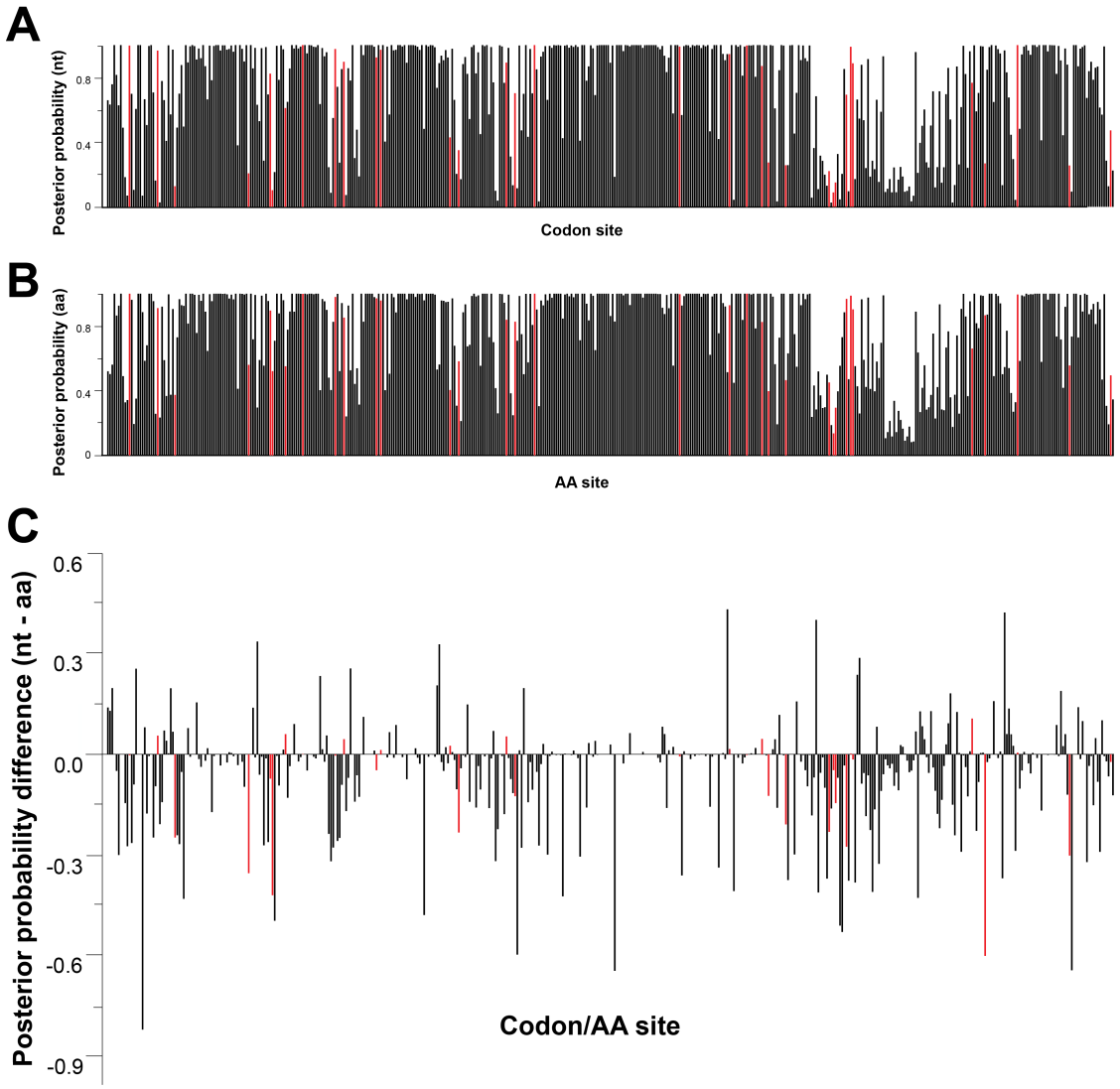


Figure 14: Histograms comparing the posterior probability scores of the reconstructed AncALL amino acid sequences reconstructed from the nuceotide (nt) and amino acid (aa) data-sets at each site. Sites detected to be undergoing positive selection are shown in red. (A) Histogram of posterior probability scores of the translated AncALL amino acid sequence reconstructed from the nucleotide data-set. (B) Histogram of posterior probability scores of the AncALL amino acid sequence reconstructed from the amino acid data-set. (C) Histogram comparing the posterior probability scores in the nucleotide and amino acid data-sets. Posterior probability scores larger in the nucleotide-based reconstruction are above the X-axis, while posterior probability scores larger in the protein-based reconstruction are below the X-axis.

3.3 Detection of Significant Sites by MEME

Positive selection analysis with MEME detected 47 codon/AA sites (Tables 2-4) that have been significantly positively selected for throughout the evolution of the AChR sequences. These sites occurred at various parts of the receptor, including within the extracellular domain, binding-site loops, transmembrane domain, and intracellular domain (Figure 16). After predicting the protein structures of the reconstructed ancestral sequences using AlphaFold2[48], these sites were mapped onto the 3D structures (Figures 17, 18) so that residues substituted along the evolutionary trajectory can be discerned. As any sequence may have a gap in the alignment at a significant site, not all subunits contained all 47 significant sites, and as such, a change from a gap to an amino acid (as an insertion or deletion) is not marked on the 3D structures. Although many substitutions are identical in both the protein and translated DNA ancestral sequence sets, a large number of them also differed, with substitutions at sites in the protein set not occuring in the respective sites of the DNA set, or vice versa.

Sites detected to be undergoing positive selection generally have ω values from between 1 to 100 (Figure 15, shown in red), which are larger than the ω values of most sites (less than 1). Some sites undergoing positive selection, particularly towards the end of the sequence, also have very high ω values of 600 and above. It is notable that some sites that have ω values between 1 to 100 were not detected to be undergoing positive selection, which reflects the statistical methods used by MEME (refer to Appendix II).

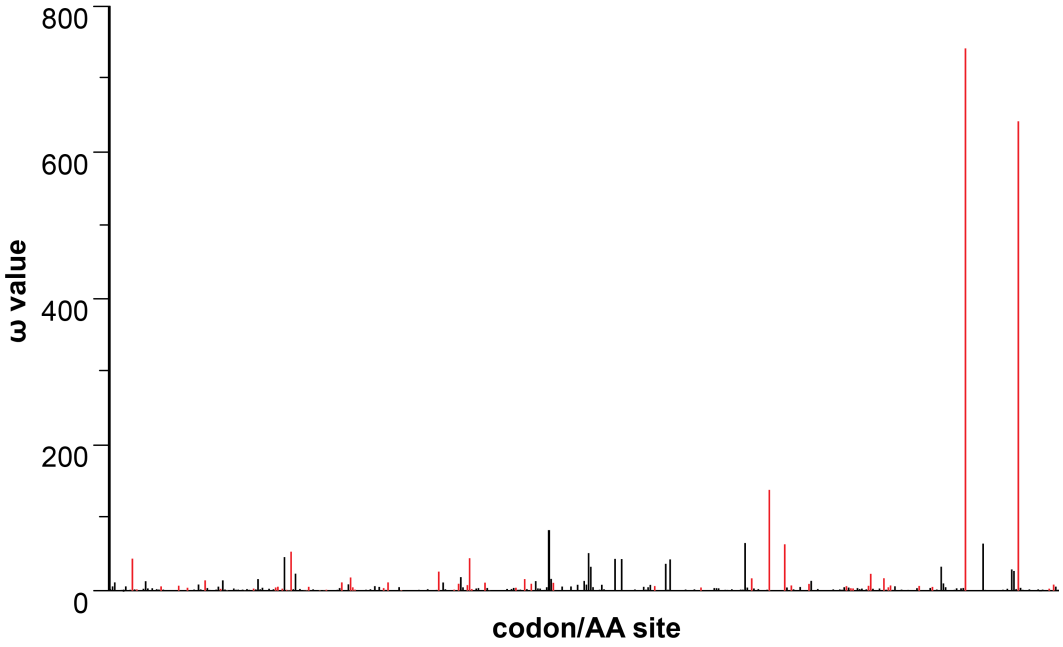


Figure 15: Bar graph showing the ω value at each site in the alignment for the *Homo sapiens* α1 subunit sequence. The ω value was calculated as the β+ (unrestricted rate of non-synonymous mutations) score over the α (rate of synonymous mutations) score for each site in the alignment.

Chapter IV: Discussion

4.1. Comparing the Nucleotide and Amino Acid-based Phylogenies and Sequence Reconstructions

Inspection of the phylogenies inferred from the nucleotide (Figure 8) and amino acid-based (Figure 9) phylogenetic trees, as well as the tanglegram comparing them (Figure 10), reveals that the two phylogenetic trees are largely similar to one another. The nucleotide tree shows the α2/α4 subunit clade to be more closely related to the α3/6 clade than the β3/α5 clade, which is accurate according to the literature[50], unlike the amino acid-based tree, which has the α2/α4 receptors most closely related to the β3/α5 clade. On the other hand, for the α2 and α7 subunit sequences, the nucleotide-based tree groups the *Homo sapiens* sequence more closely to the *Ornithorhyncus anatinus* sequence than to that of *Tursiops truncatus*, which appears to be inaccurate as *Homo sapiens* is more closely related to *Tursiops truncatus* than *Ornithorhyncus anatinus*[40], as reflected in the amino acid-based phylogenetic tree.

The nucleotide-based phylogeny does appear to outperform the amino acid-based phylogeny in terms of branch support, with branches in the nucleotide-based phylogeny having higher aLRT scores (Figure relative to the amino acid-based phylogeny on average (by 0.12). However, there are some exceptions, such as the branch leading to the AncALL node, which is higher by 0.67 in the amino acid-based phylogeny, and it should be pointed out that most branches have very similar aLRT scores with differences smaller than 0.1. Therefore, the evidence does suggest that, overall, the nucleotide-based phylogeny outperformed the amino acid-based phylogeny, and thus demonstrates an advantage of using nucleotide sequence data in phylogenetic analysis.

Comparison of the relative branch-lengths (Figure 11) in the two phylogenies shows that the branch-lengths are generally longer in the amino acid-based phylogeny early in the tree (first six branches onwards from the root), but become shorter compared

to the nucleotide phylogeny in the remaining latter branches of the phylogenetic trees. This would suggest that the reconstructed ancestral sequences that occur later within the phylogenetic (such as AncG and AncE) tree may be more accurate in the amino acid-based phylogeny, because short branches help mitigate issues related to substitution saturation, as there is less time for multiple substitutions to accumulate at any single site[51]. Conversely, branches earlier in the tree (such as for AncALL and AncBDGE) would be implied to be more accurate for the nucleotide-based data set. However, when the relative posterior probabilities (Figure 14) were calculated for AncALL, it was seen that more sites have a higher posterior probability score in the amino acid-based reconstruction relative to the nucleotide-based reconstructions, which would suggest that this pattern is consistent across all of the other reconstructed sequences as well, and it would be worth confirming this in future research.

The percentage identity differences for the reconstructed ancestral subunit sequences were close to one another at a range of less than 10%, with the average percentage difference being 11.5% (Figure 13). In general, pairwise alignment of ancestral sequences predicted from the amino acid-based data-set with the respective sequences predicted from the nucleotide-based data-set shows that most sites are either identical or chemically similar (i.e. with positive PAM250 scores[52]) - of the minority of sites that differ, most appear to lie within the intracellular domain, possibly because this region has been less conserved among the extant AChR sequences used in the ancestral sequence reconstruction. The pLDDT scores for all 3D structures (Supplemental Figure 1 in the appendix) predicted from the reconstructed sequences also shows a high degree of confidence at all regions except for the intracellular domain, although this is likely because intracellular domain structures are often flexible due to being exposed to an aqueous cytoplasmic environment and being involved in conformational transitions[53], rather than reflecting inaccuracy in the reconstructed sites at that region.

In summary, the comparison suggests that both nucleotide-based and amino acid-based sequence reconstructions have their strengths and weaknesses, and so the hypothesis that the nucleotide-based ancestral sequence reconstructions would outperform the amino acid-based reconstructions is not supported. The choice of which method to use may depend on specific objectives and regions of interest within the

sequences, and the findings highlight the importance of considering multiple factors and utilizing both nucleotide and amino acid-based data sets when reconstructing ancestral sequences with ASR.

4.2. Functional Analysis of Significant Sites

The redundancy inherent to the genetic code means that sequence information is lost upon translation of a gene transcript into a protein. In this study, protein and complementary DNA alignments and phylogenies were constructed for 184 existing homologous AChR sequences, and positive selection analysis and ASR were conducted to extract meaningful information from the data. The molecular structure of AChR subunits as a general whole are put into perspective by numbering sites according to their respective aligned residue positions within the nAChR α1 subunit from *Homo sapiens* (Figure 16B), Table 2, full-length amino-acid sequence taken from NCBI accession: XP_016858746.1), and the positively selected sites were also mapped onto the 3D structure of the *Torpedo marmorata* α1 subunit (Figure 16A) for illustrative purposes.

As expected from previous studies[54], analysis of the calculated ω value at each position along the alignment showed that the AChR sequences were predominantly conserved at the amino acid level along their evolution (Figure 15), with most sites having the low ω values (most being under 1) indicative of neutral and purifying selection, and most sites having large ω values being detected to be undergoing positive selection. The transmembrane domain was, consistent with previous research confirming the critical importance of the pore region to receptor function[55], particularly well-conserved, with only a few residue sites (275 in M2, 319 in M3, and 422 and 446 in M4, as shown for the α1 subunit in Figure 16) found to be undergoing positive selection in at least a minority of branches, with the M4 sites having particularly high ω values. The coupling region, which connects the transmembrane domain to the extracellular domain and is considered an important link connecting the agonist-binding site to gating of the channel pore, and thus in the allosteric mechanisms behind the conformational transitions of the receptor[2], was also mostly conserved across the alignment, although, for example,

site 296 in the M2-M3 loop was detected as undergoing positive selection. Overall, the majority of the sites detected to be undergoing positive selection reside within the extracellular domain, with the intracellular domain also containing a large number of significant sites.

The AChR subunits are diverse in function, and their coassembly into either homo- or hetero-pentamers results in receptors with highly variable properties[56]. Using the codon alignment, selection analysis with MEME identified 47 sites (Figure 16, Table 2) that were undergoing positive selection at a proportion of their branches across the alignment, and further analysis of these sites can shed information on the functional differentiation of these subunits.

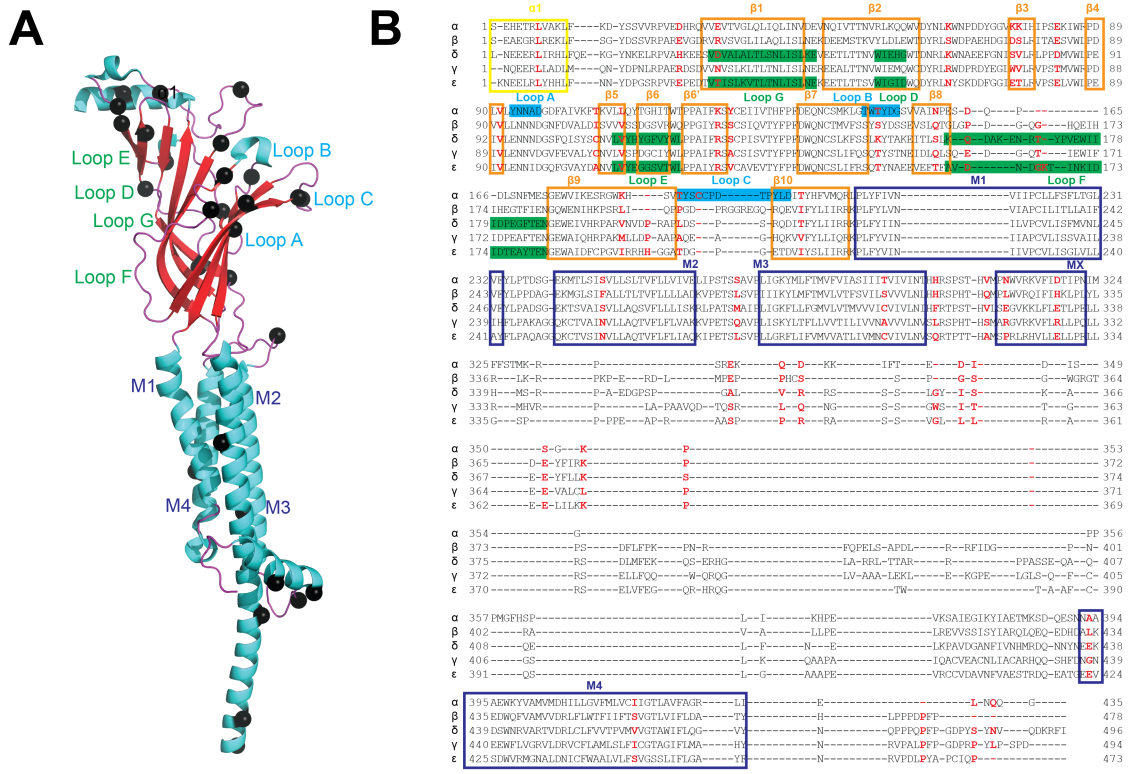


Figure 16: Mapping sites detected to be undergoing positive selection by MEME onto the acetylcholine receptor structure. (A) AChR α1 subunit structure of *Torpedo marmorata* (PDB: 7QLO) shown in PyMOL. The transmembrane helices and acetylcholine receptor binding site loops are labelled, and the residues detected to be significant by MEME are shown as black spheres. (B) Alignment of the muscle-type subunits of the *Homo sapiens* acetylcholine receptor. The alpha-helix (α1, yellow box) and beta-sheet strands (β1-10, orange boxes) of the ECD, and the transmembrane helices (M1-4, dark blue boxes) of the TMD are labelled, and residues forming the loops in the principal (loops A-C, light blue) and complementary (loops D-G, green) faces of the binding site are labelled and highlighted within the alignment. Sites identified to be significant by positive selection analysis with MEME are in red. Figure made with JalView[52].

Positive selection at the intracellular domain may be contributing to the adaptation of acetylcholine receptors to possess distinct conformational transitions[53] involved with transduction of ligand binding to channel gating and other processes[57]. For example, functional diversification at this region may lead to subunits evolving to have distinct activation or desensitization kinetics. At the N-terminus of the extracellular domain, the α1 helix and the α1-β1 linker also had significant sites, which may be relevant to allosteric modulation in the neuronal subunits because this region can experience conformational changes in response to ligand binding[58]. The several binding-site loops that make up the ligand-binding site, which are loops A-C in the subunit contributing the principal face and loops D-G in the subunit contributing the complementary face, also contained sites undergoing positive selection, which can lead to different binding affinities and specificities for particular ligands across diverse acetylcholine receptors.

The following residues are numbered according to their position in the *Homo sapiens* α1 subunit sequence (Figure 16B). The mutation of residue 177 in loop B may result in the alteration of the specificity and affinity of agonist binding[59], and residues 216 and 219 of loop C in the β9-β10 linker could contribute to differential ligand binding through their influence on electrostatic interactions[2]. At the complementary face, residue 57 (a glutamate in the α1 subunit) is found in loop G of the β1 strand, where it may participate in acetylcholine or other agonist binding[2]. Three significant sites were detected in loop F, where it is known that in addition to participating in agonist binding, the β8-β9 loop region is also responsible for calcium potentiation of acetylcholine receptors[60]. Loop E is generally more variable and less conserved among the binding loops[55], and the significant site in the β5 strand might be contributing to altered pharmacology of agonist binding among a subset of the AChR sequences. Another two sites, including K212 in the *Homo sapiens* α1 subunit, that were identified as undergoing positive selection are located in β9 strand, which, while not directly participating in binding, is involved in regulating sensitivity and specificity to a diverse selection of ligands[61], and has been found to influence the desensitization kinetics of certain subunits[62].

It is important to keep in mind that the acetylcholine receptor subunits do not function as independent entities, but rather work together as a combined pentamer. As such, it is worth observing whether any of the sites detected to be undergoing positive selection exist at the subunit interface, where they could be contributing to the overall receptor structure. Looking at the *Homo sapiens* α1 subunit (of which there are two in the heteropentamer), it can be seen that a number of the detected sites undergoing positive selection are present at the interface with neighbouring subunits. In the extracellular domain, residues 24, 91, 127, and 150 appear to interact with residues in the rightmost subunit, while residues 76 and 77 are located adjacent to residues from the leftmost subunit. In addition, residues 248, 269, and 300 in the transmembrane domain are positioned close to adjacent subunits, with residue 269 being of particular interest because it is positioned adjacent to a conserved serine residue in the modern-day subunit sequences that is known to participate in channel conductance[28].

Altogether, it is clear that many of the positive selection-detected sites could be playing important roles in AChR function, by influencing receptor properties such as ligand affinity and specificity, as well as the kinetics of conformational transitions leading to desensitization and other allosteric mechanisms[36] such as channel gating.

4.4 Influence of Positive Selection on AChR Subunit Evolution

All AChR subunits ultimately descend from a single common ancestor (Figures 8, 9), and the ancestral AChR receptor is believed to have been a single subunit capable of self-assembling into a homomer, similar to the modern-day a7 homopentamer[12]. From this single ancestor, gene duplications followed by both natural selection and functional differentiation led to the evolution of new subunit sequences with unique properties. The muscle-type subunits of the AChR receptor occupy a notable role in their function at the neuromuscular junction, and as such, the most important ancestors of these subunits were selected for reconstruction using ASR. The positively selected sites identified by MEME were mapped to the reconstructed ancestral sequences to shed light on how positive selection has influenced the evolutionary trajectory of muscle-type AChR (Figures 17, 18, and Tables 3, 4).

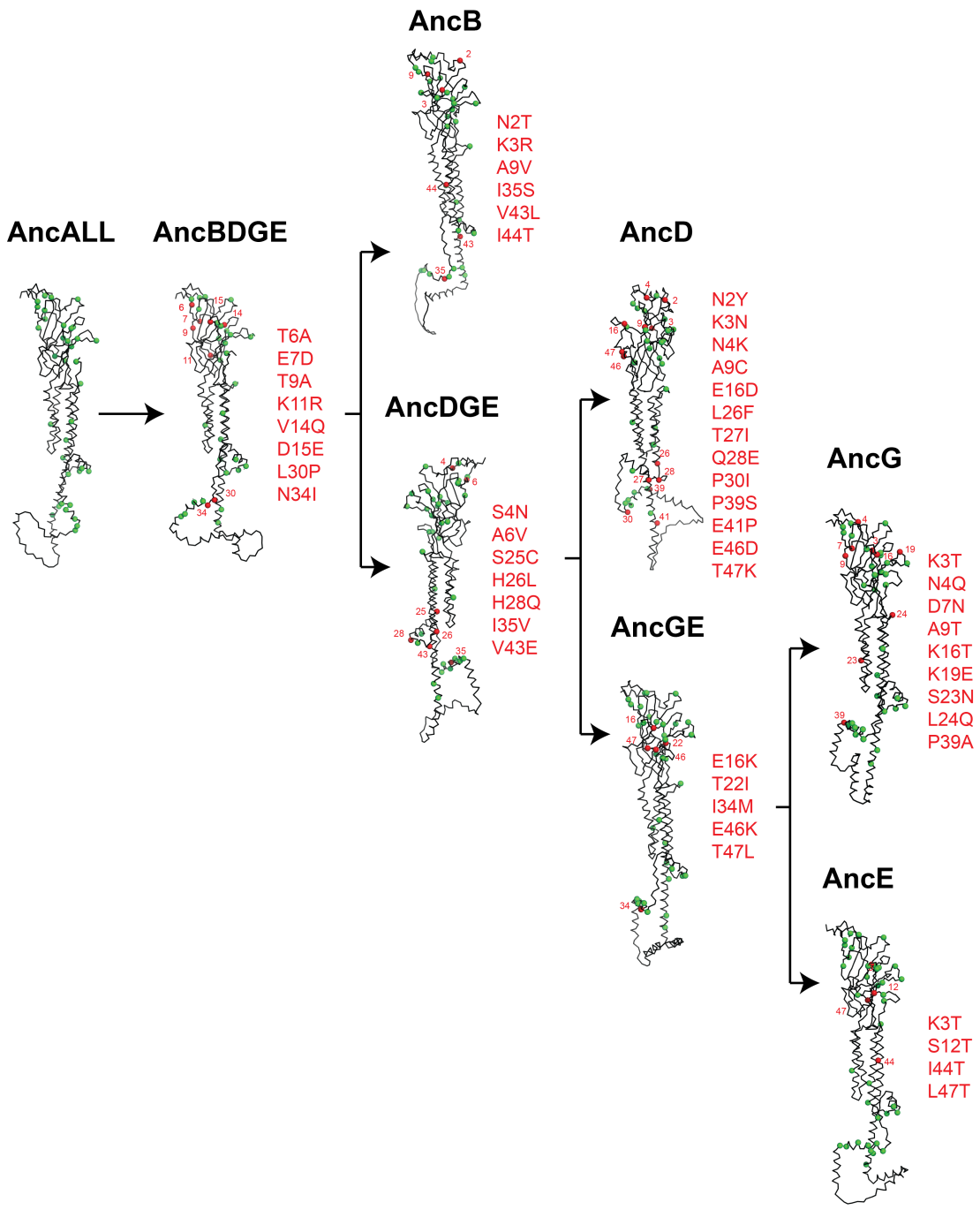


Figure 17: AlphaFold2 modelled Cα backbone structures of reconstructed ancestral subunits inferred from the nucleotide sequence alignment and phylogeny. In AncALL, the residues (shown as spheres) that were detected to be undergoing positive selection by MEME are highlighted in green; in each subsequent subunit, residues that have been substituted during the last divergence are instead coloured in red, and listed to the right of the structure. The substitution notation shows the previous residue symbol first, the new residue symbol last, and the respective number of the significant site (which number between 1-47) in the middle. Figure made with PyMOL[71].

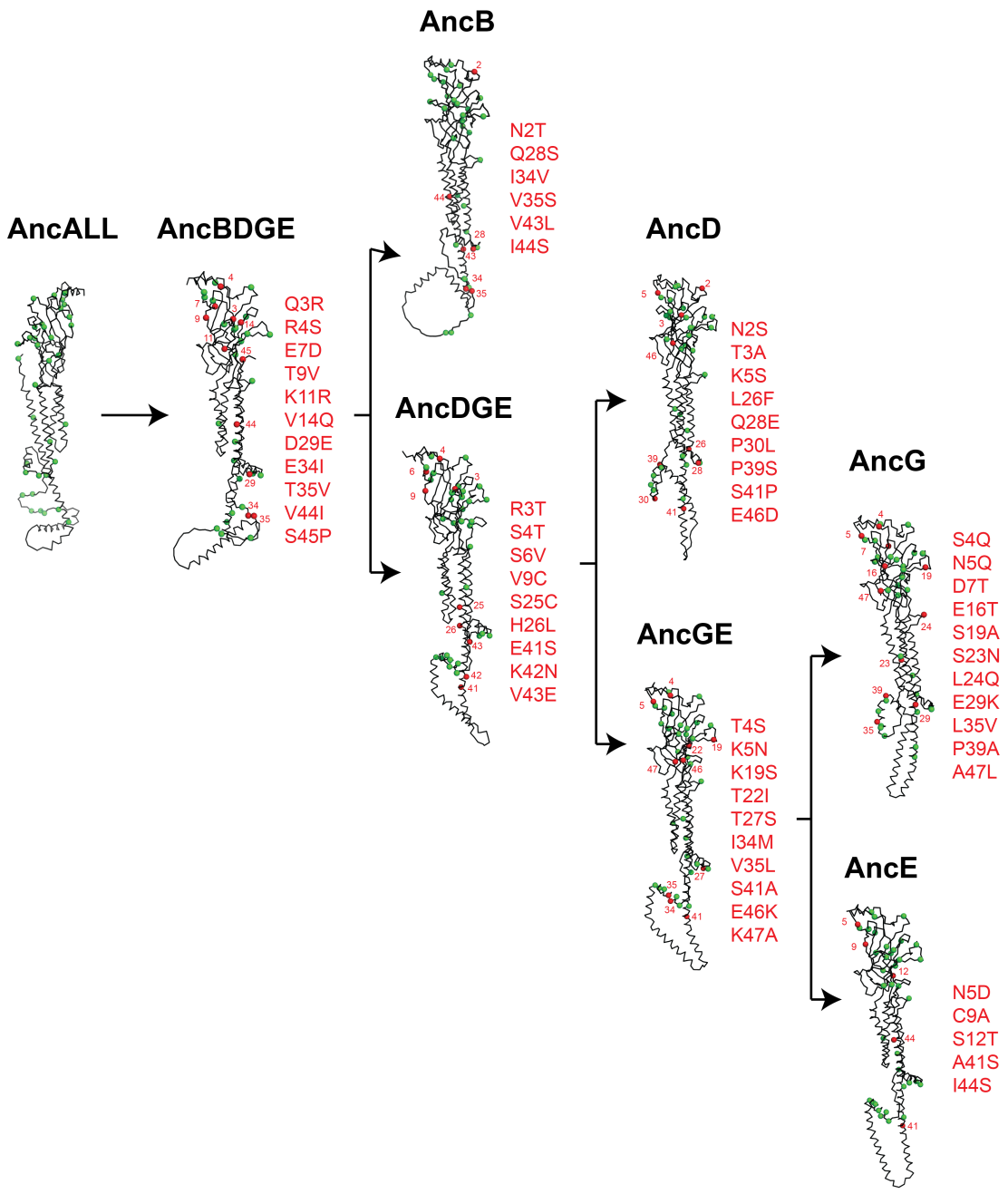


Figure 18: AlphaFold2 modelled Cα backbone structures of reconstructed ancestral subunits inferred from the amino acid sequence alignment and phylogeny.. In AncALL, the residues (shown as spheres) that were detected to be undergoing positive selection by MEME are highlighted in green; in each subsequent subunit, residues that have been substituted during the last divergence are instead coloured in red, and listed to the right of the structure. The substitution notation shows the previous residue symbol first, the new residue symbol last, and the respective number of the significant site (which number between 1-47) in the middle. Figure made with PyMOL[71].

The sites detected to be undergoing positive selection were numbered from 1-47 (as shown in Figures 17, 18, and Tables 3, 4, in the Appendix), and the remainder of the discussion section uses this numbering system to describe these sites. Comparison of the mutations from one ancestral subunit to the next, in the amino acid and cDNA-based data, shows that the two are largely the same or conserved; for example, many mutations are shared between them (like E7D between AncALL and AncBDGE, or N2T between AncBDGE and AncB), and even when there is a discrepancy, the mutated residues often share similar chemical properties (such as T9V and T9A in the AncALL to AncBDGE transition, with valine and alanine belonging to the same non-polar class) or have become equivalent at the successive ancestral node (e.g. R4S in the amino acid-based AncALL/AncBDGE sequences is S4S in the respective cDNA sequences, with the latter thus not shown in Figure 17). However, when a mutation does appear to differ between the two sequence sets (T3A vs K3N in the AncDGE to AncD transition), this may reflect on the comparative accuracies of amino acid vs. cDNA methods, and could be useful in highlighting an important mutation that would have been missed using only a single sequence-based method.

It is known that among the muscle-type subunits, the α1 subunit plays the role of the principal binding face, the γ/δ/ε subunits evolved to provide the complementary binding face with differing affinities and specificities to nicotinic ligands, and the β subunit emerged as a structural subunit essential for receptor formation and transport to the cell surface[63], with other residues also changing to occupy specific niches within the muscle-type receptor[54]. Figures 17 and 18 show that site 3, located within loop G, changed during the evolutionary trajectory, with a Q3R mutation from AncALL to AncBDGE, an R3T mutation from AncBDGE to AncDGE, T3A (AA) or K3N (cDNA) from AncDGE to AncD, and finally mutating to K3T in AncG and AncE. During the AncALL to AncBDGE transition, the following sites of interest were mutated, which may have differentiated these muscle-type subunits: Q3R, T6A, T9V (AA) or T9A (cDNA), K11R, V14Q, E34I (AA) or N34I (cDNA), and T35V. The AncBDGE to AncB transition, which could have contributed to AncB taking a greater role as a structural subunit, was marked by the following mutations: N2T, V35S (AA) or I35S (cDNA), and

I44S (AA) or I44T (cDNA). Assuming that the AncBDGE subunit was capable of acting in some capacity as a complementary subunit, the AncBDGE to AncDGE transition also experienced a number of mutations potentially contributing to the γ, δ, and ε subunits roles as complementary binding face subunits: R3T, S6V (AA) or A6V (cDNA), V9C, S25C, H26L, H28Q, E41S, and V43E.

The remaining transitions are within the ancestral complementary subunits, and so relate to the evolution of those receptors to have different pharmacological properties towards the various ligands that are capable of binding at the AChR LBD. The AncDGE to AncGE transition involves mutations K5N, E16K, K19S, T22I, S41A, E46K, and K47A (AA) or T47L. The δ subunit has lower affinity in binding *d*-tubocurarine compared to the γ and ε subunits[64], and the evolution of AncDGE to AncD occurred with the significant site mutations N2Y, T3A (AA) or K3N (cDNA), N4K, K5S, A9C, Q28E, P39S, S41P (AA) or E41P (cDNA), and T47K. The AncGE to AncG significant mutations are K3T, D7T (AA) or D7N (cDNA), A9T, E16T (AA) or K16T (cDNA), S19A (AA) or K19E (cDNA), L24Q, E29K, and P39A, with the γ subunit having relatively low affinity to the agonist carbamylcholine[64]. On the other hand, the notable mutations between AncGE and AncE are K3T, N5D, C9A, A41S, I44S (AA) or I44T (cDNA), and L47T.

To further elucidate the nature of the residues being substituted, it is useful to look at the chemical structure of their side-chains. Amino-acid side-chains can be broadly divided into four main classes depending on their chemical properties: positively-charged (R, K, H), negatively-charged (D, E), uncharged polar (S, T, N, Q, Y, C), and uncharged non-polar (remaining residues). A mutation from one side-chain class to another (such as the substitution of a positively-charged side-chain with a negatively-charged side-chain, or the substitution of a negatively-charged side-chain with a non-polar side-chain) would be of particular interest as the chemical sequence and structure of the acetylcholine receptor is closely tied to its function[57]. Of the substitutions previously discussed, it can be seen that a number of them involved transitioning between the amino acid side-chain classes:

- Substitutions between positively-charged and negatively-charged amino acids:
    - AncDGE to AncGE: E16K, E46K.

    - AncGE to AncG: K19E, E29K.
- Substitutions between charged and polar amino acids:
    - AncALL to AncBDGE: Q3R.
    - AncBDGE to AncDGE: R3T, H28Q, E41S.
    - AncDGE to AncGE: K5N, K19S.
    - AncDGE to AncD: K3N (cDNA), N4K, Q28E.
    - AncGE to AncG: D7N (cDNA).
    - AncGE and AncE: N5D.
- Substitutions between charged and non-polar amino acids:
    - AncALL to AncBDGE: E34I (AA).
    - AncBDGE to AncDGE: H26L, V43E.
    - AncDGE to AncGE: K47A (AA).
    - AncDGE to AncD: K5S, E41P (cDNA), T47K.
    - AncGE to AncG: K3T, D7T (AA), E16T (AA) or K16T (cDNA).
    - AncGE and AncE: K3T.
- Substitutions between polar and non-polar uncharged amino acids:
    - AncALL to AncBDGE: T6A, T9V (AA) or T9A (cDNA), V14Q, N34I (cDNA), T35V.
    - AncBDGE to AncB: V35S (AA) or I35S (cDNA), I44S (AA) or I44T (cDNA).
    - AncBDGE to AncDGE: S6V (AA).
    - AncDGE to AncGE: T22I, S41A, T47L.
    - AncDGE to AncD: T3A (AA), P39S, S41P (AA).
    - AncGE to AncG: A9T, S19A (AA), L24Q.
    - AncGE and AncE: A41S, I44S (AA) or I44T (cDNA), L47T.

These mutations are therefore proposed to be the most likely to have influenced the functions of the ancestral subunits during their evolution into the modern-day muscle-type subunits, and are hence likely to be the most informative sites for further studies of these receptors. As the codon sites detected to be undergoing positive selection have documented contributions to the structure and function of the acetylcholine receptor, this

hypothesis is supported, although further experimental studies must still be done to confirm the importance of these mutations in the evolution of the acetylcholine receptor subunits.

4.5 Future Experiments Following from this Study

By combining ancestral subunit sequence reconstruction with positive selection analysis, it is possible to identify substitutions likely to have influenced the evolutionary trajectory of the acetylcholine receptor. However, to prove that the codon sites detected to be undergoing positive selection correspond to mutations responsible for functional specialization of the subunits, it is necessary to conduct further experimental studies. For example, in order to study the contribution of a substitution to channel gating and conductance, a combination of ancestral sequence reconstruction and patch clamp experiments can be conducted, as done by Emlaw et al[28] to investigate how two substitutions in the transmembrane domain have contributed to increased conductance in the human β subunit relative to AncB. In brief, this method used an alignment of 60 orthologous muscle-type β subunits to generate a maximum likelihood molecular phylogeny, used ASR to reconstruct AncB from this phylogeny and inverse PCR to introduce the point mutations into the AncB and human β sequence constructs, and transfected them into cells along with wild-type human α, δ, and ε subunits. Single-channel patch clamp recordings were then performed and analyzed to determine unitary current amplitude, thus permitting detection of alterations in channel conductance. Patch clamp experiments can similarly be conducted to study how mutations can alter ligand binding[65] and through electrical fingerprinting studies to determine alterations in subunit composition[15]. While this method relied on ancestral sequence reconstruction and emperical observation to pinpoint the importance of these two mutations, exploiting positive selection analysis allows for many more substitutions of interest to be highlighted, as discussed in the previous sections.

4.6 Strengths and Weaknesses to this Study

This study possesses both strengths and weaknesses that should be taken into consideration. This project took advantage of the vast sequence information now publicly available on the NCBI database to create a very large alignment consisting of 184 sequences across many different vertebrate species, and furthermore was able to align them in a very dense alignment that generated phylogenetic trees closely matching the consensus phylogeny (Figures 8, 9). As MEME has been shown capable of detecting positive selection at sites in a minority of branches even in large alignments[24], this makes the results even more likely to be accurate compared to studies utilising other branch-site tests.

Conversely, as it is of interest to differentiate between selective pressures in different subunit types with the goal of making specific inferences about mutations along the evolutionary trajectory, MEME does have a drawback in not being able to identify which branches at a site are undergoing positive selection. Rather, all it can conclude is that some branches at that site are experiencing positive selection, and while it remains useful in narrowing down sites of interest, especially in combination with ASR results about how residues have been mutated from one ancestral node to its descendents, any inferences made will remain uncertain until experimentally verified. While this uncertainty is fundamentally intrinsic to positive selection analysis[31], it could be useful to consider other branch-site models; for example, a similar study[66] examining positive selection in the neuronal AChR subunits grouped the subunits into their constitutive subunit families, and manipulated the branch-site model proposed by Yang et al[67] so that each subunit family was set as the foreground branch, and the remaining sequences were set as the background branch, allowing them to draw more specific inferences on the functional specialization of each subunit. Furthermore, recent research has found that these branch-site models, including MEME, are prone towards false inferences of positive selection due to the misinterpretation of multinucleotide mutations that occur naturally during DNA replication, and a branch-site model that incorporates multinucleotide mutations may be perferable to use as it would help alleviate this bias[68].

Ultimately, the results of this study should be verified by other methods such as molecular biology and electrophysiological techniques to confirm or refute the hypotheses postulated in this work. As positive selection can contribute to the evolution of receptors to possess a diverse range of pharmacological properties, it is hoped that the combination of cDNA/codon and amino acid-based methods employed and the results obtained from this project may provide new insights into AChR structure, function, and evolution.

Appendix

Appendix I - Tables

Table 1: The homologous acetylcholine receptor and outgroup sequences retrieved from the NCBI database. The NCBI accession numbers for both the amino acid and nucleotide sequences are provided. An asterisk marks the names of the two sequences which differ between the original amino acid and nucleotide sequences (α1 subunit of *Torpedo marmorata* and δ subunit of *Vicugna pacos*). The acetylcholine receptor sequences which were not used in the alignment and following analyses are in red.

Species Name	AA Sequence Accession	cDNA Sequence Accession
α10 Homo sapiens	NP_065135.2	NM_020402.4
α10 Ornithorhynchus anatinus	XP_039770641.1	XM_039914707.1
α10 Tursiops truncatus	XP_033718286.1	XM_033862395.1
α1 Homo sapiens	XP_016858746.1	XM_017003257.1
α1 Lepisosteus oculatus	XP_015214325.1	XM_015358839.1
α1 Ornithorhynchus anatinus	XP_001514882.1	XM_001514832.5
*α1 Torpedo marmorata**	P02711.1	M25893.1
α1 Tursiops truncatus	XP_019778162.1	XM_019922603.2
α2 Homo sapiens	AAB40109.1	U62431.1
α2 Ornithorhynchus anatinus	XP_028908521.1	XM_029052688.2
α2 Tursiops truncatus	XP_033713511.1	XM_033857620.1
α3 Homo sapiens	NP_000734.2	NM_000743.5

α3 Lepisosteus oculatus	XP_015198894.1	XM_015343408.1
α3 Ornithorhynchus anatinus	XP_039768124.1	XM_039912190.1
α3 Tursiops truncatus	XP_019805902.1	XM_019950343.2
α5 Homo sapiens	NP_000736.2	NM_000745.4
α5 Lepisosteus oculatus	XP_015198885.1	XM_015343399.1
α5 Lipotes vexillifer	XP_007471904.1	XM_007471842.1
α5 Ornithorhynchus anatinus	XP_028921637.2	XM_029065804.2
α6 Homo sapiens	NP_004189.1	NM_004198.3
α6 Nanorana parkeri	XP_018413413.1	XM_018557911.1
α6 Ornithorhynchus anatinus	XP_028922041.1	XM_029066208.2
α6 Tursiops truncatus	XP_004327124.1	XM_004327076.3
α7 Erpetoichthys calabaricus	XP_028661667.1	XM_028805834.1
α7 Homo sapiens	CAA49778.1	Y08420.1
α7 Lepisosteus oculatus	XP_015201366.1	XM_015345880.1
α7 Ornithorhynchus anatinus	XP_028927153.1	XM_029071320.2
α7 Tursiops truncatus	XP_019806811.1	XM_019951252.1
α9 Homo sapiens	AAI13550.1	BC113575.1

α9 Lepisosteus oculatus	XP_006630105.1	XM_006630042.2
α9 Ornithorhynchus anatinus	XP_028938906.1	XM_029083073.2
α9 Tursiops truncatus	XP_004312383.2	XM_004312335.3
α4 Homo sapiens	NP_000735.1	NM_000744.7
α4 Lipotes vexillifer	XP_007462309.1	XM_007462247.1
α4 Ornithorhynchus anatinus	XP_001506577.1	XM_001506527.4
α4 Scyliorhinus canicula	XP_038659724.1	XM_038803796.1
β3 Homo sapiens	NP_000740.1	NM_000749.5
β3 Lepisosteus oculatus	XP_015201369.1	XM_015345883.1
β3 Ornithorhynchus anatinus	XP_028921557.1	XM_029065724.1
β3 Tursiops truncatus	XP_019776801.1	XM_019921242.1
β2 Scyliorhinus canicula	XP_038643165.1	XM_038787237.1
β2 Latimeria chalumnae	XP_005996559.1	XM_005996497.1
β2 Crotalus tigris	XP_039219545.1	XM_039363611.1
β2 Protobothrops mucrosquamatus	XP_029139447.1	XM_029283614.1
β2 Thamnophis elegans	XP_032090171.1	XM_032234280.1
β2 Chelonia mydas	XP_037739529.1	XM_037883601.1
β2 Mauremys reevesii	XP_039368075.1	XM_039512141.1

β2 Phascolarctos cinereus	XP_020847304.1	XM_020991645.1
β2 Monodelphis domestica	XP_001373153.1	XM_001373116.3
β2 Felis catus	XP_023103325.1	XM_023247557.1
β2 Lynx canadensis	XP_030158890.2	XM_030303030.2
β2 Pan troglodytes	PNI90705.1	XM_001152392.5
β2 Homo sapiens	NP_000739.1	NM_000748.3
β4 Chiloscyllium punctatum	GCC33331.1	BEZZ01000507.1
β4 Scyliorhinus torazame	GCB75362.1	BFAA01015039.1
β4 Rhincodon typus	XP_020387477.1	XM_020531888.1
β4 Scyliorhinus canicula	XP_038640406.1	XM_038784478.1
β4 Amblyraja radiata	XP_032870801.1	XM_033014910.1
β4 Podarcis muralis	XP_028600253.1	XM_028744420.1
β4 Pseudonaja textilis	XP_026565611.1	XM_026709826.1
β4 Notechis scutatus	XP_026540336.1	XM_026684551.1
β4 Crotalus tigris	XP_039193436.1	XM_039337502.1
β4 Lacerta agilis	XP_033016736.1	XM_033160845.1

β4 Protobothrops mucrosquamatus	XP_015685807.1	XM_015830321.2
β4 Thamnophis elegans	XP_032089304.1	XM_032233413.1
β4 Ornithorhynchus anatinus	XP_028920406.1	XM_029064573.2
β4 Tachyglossus aculeatus	XP_038623209.1	XM_038767281.1
β4 Phascolarctos cinereus	XP_020831651.1	XM_020975992.1
β4 Monodelphis domestica	XP_007478061.1	XM_007477999.2
β4 Trichosurus vulpecula	XP_036591397.1	XM_036735502.1
β4 Vombatus ursinus	XP_027696403.1	XM_027840602.1
β4 Bos mutus	XP_005887683.2	XM_005887621.2
β4 Capra hircus	XP_013828545.1	XM_013973091.2
β4 Ovis aries	XP_004017871.2	XM_004017822.5
β4 Trachypithecus francoisi	XP_033045178.1	XM_033189287.1
β4 Rhinopithecus roxellana	XP_030787168.1	XM_030931308.1
β4 Pongo abelii	XP_024088464.1	XM_024232696.1
β4 Pan paniscus	XP_003813864.1	XM_003813816.4
β4 Homo sapiens	NP_000741.1	NM_000750.5
β Tetronarce californica	P02712.1	J00964.1
β Torpedo marmorata	4AQ5_B	AY472103.1

β Hypnos monopterygius	AAR29365.1	AY472107.1
β Esox lucius	XP_034146453.1	XM_034290562.1
β Esox lucius	XP_010886031.2	XM_010887729.3
β Danio rerio	NP_001240739.1	NM_001253810.1
β Danio rerio 2	NP_001233249.1	NM_001246320.1
β Lepisosteus oculatus	XP_006627492.1	XM_006627429.2
β Lepisosteus oculatus 2	XP_015195943.1	XM_015340457.1
β Acipenser ruthenus	XP_034770737.1	XM_034914846.1
β Acipenser ruthenus 2	XP_034770770.1	XM_034914879.1
β Scleropages formosus	XP_018599910.1	XM_018744394.2
β Scleropages formosus 2	XP_018599870.2	XM_018744354.2
β Betta splendens 2	XP_028987844.1	XM_029132011.2
β Betta splendens	XP_028998705.1	XM_029142872.2
β Gasterosteus aculeatus aculeatus	XP_040038956.1	XM_040183022.1
β Gasterosteus aculeatus aculeatus2	XP_040039172.1	XM_040183238.1
β Pungitius pungitius	XP_037336277.1	XM_037480380.1

β Pungitius pungitius 2	XP_037336278.1	XM_037480381.1
β Oncorhynchus tshawytscha	XP_024229341.1	XM_024373573.1
β Oncorhynchus tshawytscha 2	XP_024292164.1	XM_024436396.2
β Carassius auratus	XP_026123015.1	XM_026267230.1
β Carassius auratus2	XP_026107116.1	XM_026251331.1
β Latimeria chalumnae	XP_014349798.1	XM_014494312.1
β Podarcis muralis	XP_028559587.1	XM_028703754.1
β Crotalus tigris	XP_039193790.1	XM_039337856.1
β Pseudonaja textilis	XP_026578993.1	XM_026723208.1
β Gopherus evgoodei	XP_030404056.1	XM_030548196.1
β Chrysemys picta bellii	XP_005313767.3	XM_005313710.4
β Trachemys scripta elegans	XP_034612815.1	XM_034756924.1
β Pelodiscus sinensis	XP_006112165.1	XM_006112103.3
β Mauremys reevesii	XP_039355828.1	XM_039499894.1
β Ornithorhynchus anatinus	XP_028911396.1	XM_029055563.1
β Tachyglossus aculeatus	XP_038624148.1	XM_038768220.1
β Phascolarctos cinereus	XP_020822356.1	XM_020966697.1
β Sarcophilus harrisii	XP_031821294.1	XM_031965434.1

β Camelus bactrianus	XP_010964205.1	XM_010965903.1
β Ovis aries	XP_004012703.1	XM_004012654.5
β Camelus dromedarius	XP_010994836.1	XM_010996534.2
β Bos mutus	XP_005889998.1	XM_005889936.2
β Tursiops truncatus	XP_033704040.1	XM_033848149.1
β Bos taurus	NP_0776941.1	NM_174516.2
β Capra hircus	XP_005693560.2	XM_005693503.2
β Balaenoptera musculus	XP_036691576.1	XM_036835681.1
β Microcebus murinus	XP_012631830.1	XM_012776376.2
β Macaca fascicularis	XP_005582811.1	XM_005582754.2
β Rhinopithecus roxellana	XP_01036065.1	XM_010362349.2
β Trachypithecus francoisi	XP_033040654.1	XM_033184763.1
β Pongo abelii	XP_002827007.4	XM_002826961.4
β Pan paniscus	XP_003810132.1	XM_003810084.3
β Pan troglodytes	XP_511298.4	XM_511298.7
β Homo sapiens	CAA32939.1	X14830.1
δ *Amblyraja radiata*	XP_032888019.1	XM_033032128.1
δ *Scyliorhinus canicula*	XP_038672806.1	XM_038816878.1

δ *Callorhinchus milii*	NP_001279747.1	NM_001292818.1
δ *Hypnos monopterygius*	AAR29367.1	AY472109.1
δ *Tetronarce californica*	P02718.1	J00965.1
δ *Torpedo marmorata*	4AQ5_C	AY472105.1
δ *Danio rerio*	AAH90405.1	BC090405.1
δ *Electrophorus electricus*	XP_026863880.2	XM_027008079.2
δ *Esox lucius*	XP_034145360.1	XM_034289469.1
δ *Maylandia zebra*	XP_014268561.2	XM_014413075.3
δ *Paramormyrops kingsleyae*	XP_023654102.1	XM_023798334.1
δ *Pungitius pungitius*	XP_037311827.1	XM_037455930.1
δ *Takifugu rubripes*	XP_029698942.1	XM_029843082.1
δ *Cyprinus carpio*	XP_018923129.1	XM_019067584.1
δ *Acipenser ruthenus*	XP_034784722.1	XM_034928831.1
δ *Polyodon spathula*	MBN3274219.1	XM_041269708.1
δ *Scleropages formosus*	XP_018616387.1	XM_018760871.2
δ *Oncorhynchus tshawytscha*	XP_024230260.1	XM_024374492.1
δ *Carassius auratus*	XP_026093699.1	XM_026237914.1
δ *Lepisosteus oculatus*	XP_015216562.1	XM_015361076.1

δ *Geotrypetes seraphini*	XP_033815039.1	XM_033959148.1
δ *Rhinatrema bivittatum*	XP_029471528.1	XM_029615668.1
δ *Thamnophis elegans*	XP_032081741.1	XM_032225850.1
δ *Python bivittatus*	XP_007429429.1	XM_007429367.2
δ *Crotalus tigris*	XP_039183334.1	XM_039327400.1
δ *Gekko japonicus*	XP_015263317.1	XM_015407831.1
δ *Crocodylus porosus*	XP_019388679.1	XM_019533134.1
δ *Alligator sinensis*	XP_006017156.1	XM_006017094.2
δ *Chrysemys picta bellii*	XP_023966042.1	XM_024110274.2
δ *Gopherus evgoodei*	XP_030431043.1	XM_030575183.1
δ *Terrapene carolina triunguis*	XP_024073134.2	XM_024217366.2
δ *Chelonoidis abingdonii*	XP_032659764.1	XM_032803873.1
δ *Chelonia mydas*	XP_037764876.1	XM_037908948.1
δ *Tachyglossus aculeatus*	XP_038604933.1	XM_038749005.1
δ *Ornithorhynchus anatinus*	XP_039770221.1	XM_039914287.1
δ *Phascolarctos cinereus*	XP_020861479.1	XM_021005820.1
δ *Camelus ferus*	XP_006191204.1	XM_006191142.3
δ *Physeter catodon*	XP_007125086.2	XM_007125024.3

δ *Vicugna pacos**	XP_031533709.1	XM_031677849.1
δ *Capra hircus*	XP_017913645.1	XM_018058156.1
δ *Orcinus orca*	XP_033278985.1	XM_033423094.1
δ *Ovis aries*	XP_027821117.1	XM_027965316.1
δ *Bison bison bison*	XP_010835257.1	XM_010836955.1
δ *Bos mutus*	XP_014338943.1	XM_014483457.1
δ *Tursiops truncatus*	XP_033716265.1	XM_033860374.1
δ *Aotus nancymaae*	XP_012289463.1	XM_012434040.1
δ *Rhinopithecus bieti*	XP_017741699.1	XM_017886210.1
δ *Papio anubis*	XP_031507469.1	XM_031651609.1
δ *Pan troglodytes*	XP_016806195.1	XM_016950706.2
δ *Homo sapiens*	NP_000742.1	NM_000751.3
γ *Acipenser ruthenus*	XP_034780700.1	XM_034924809.1
γ *Lepisosteus oculatus*	XP_015216581.1	XM_015361095.1
γ *Danio rerio*	AEL23188.1	JN242070.1
γ *Electrophorus electricus*	XP_035384279.1	XM_035528386.1
γ *Bettas plendens*	XP_028987446.1	XM_029131613.2
γ *Paramormyrops kingsleyae*	XP_023674242.1	XM_023818474.1

γ Scleropages formosus	XP_018608735.1	XM_018753219.2
γ Takifugu rubripes	XP_011614733.2	XM_011616431.2
γ Salvelinus alpinus	XP_023833413.1	XM_023977645.1
γ Oncorhynchus tshawytscha	XP_024292002.1	XM_024436234.1
γ Carassius auratus	XP_026104300.1	XM_026248515.1
γ Anolis carolinensis	XP_016851795.1	XM_016996306.1
γ Pogona vitticeps	XP_020667154.1	XM_020811495.1
γ Crotalus tigris	XP_039183335.1	XM_039327401.1
γ Python bivittatus	XP_025023150.1	XM_025167382.1
γ Lacerta agilis	XP_033006749.1	XM_033150858.1
γ Gekko japonicus	XP_015263318.1	XM_015407832.1
γ Crocodylus porosus	XP_019388678.1	XM_019533133.1
γ Alligator sinensis	XP_006017158.1	XM_006017096.2
γ Chelonoidis abingdonii	XP_032659779.1	XM_032803888.1
γ Chelonia mydas	XP_037764962.1	XM_037909034.1
γ Chrysemys picta bellii	XP_005292657.2	XM_005292600.3
γ Gopherus evgoodei	XP_030431465.1	XM_030575605.1
γ Terrapene carolina triunguis	XP_024073127.2	XM_024217359.2

γ Ornithorhynchus anatinus	XP_028918519.1	XM_029062686.1
γ Tachyglossus aculeatus	XP_038601156.1	XM_038745228.1
γ Phascolarctos cinereus	XP_020861481.1	XM_021005822.1
γ Sarcophilus harrisii	XP_003770322.1	XM_003770274.3
γ Camelus ferus	XP_006191202.1	XM_006191140.2
γ Vicugna pacos	XP_006209177.1	XM_006209115.2
γ Bison bison bison	XP_010835256.1	XM_010836954.1
γ Bos mutus	XP_005911142.1	XM_005911080.1
γ Capra hircus	XP_017922709.1	XM_018067220.1
γ Physeter catodon	XP_023980501.1	XM_024124733.2
γ Ovis aries	XP_027821120.1	XM_027965319.2
γ Orcinus orca	XP_004262635.1	XM_004262587.2
γ Tursiops truncatus	XP_033716262.1	XM_033860371.1
γ Balaenoptera musculus	XP_036714458.1	XM_036858563.1
γ Aotus nancymaae	XP_012289470.1	XM_012434047.1
γ Macaca fascicularis	XP_005574682.1	XM_005574625.2
γ Pongo abelii	XP_024098943.1	XM_024243175.1
γ Homo sapiens	AAY24103.1	NM_005199.5

ε *Scyliorhinus torazame*	GCB67712.1	BFAA01003046.1
ε *Torpedo marmorata*	4BOG_3	AY472104.1
ε *Amblyraja radiata*	XP_032872204.1	XM_033016313.1
ε *Hypnos monopterygius*	AAR29366.1	AY472108.1
ε *Cyprinus carpio*	XP_018921300.1	XM_019065755.1
ε *Danio rerio*	NP_001314812.1	NM_001327883.1
ε *Esox lucius*	XP_010891762.2	XM_010893460.3
ε *Gasterosteus aculeatus aculeatus*	XP_040037026.1	XM_040181092.1
ε *Polypterus senegalus*	XP_039603092.1	XM_039747158.1
ε *Maylandia zebra*	XP_004563946.1	XM_004563889.3
ε *Paramormyrops kingsleyae*	XP_023700625.1	XM_023844857.1
ε *Pungitius pungitius*	XP_037322865.1	XM_037466968.1
ε *Carassius auratus*	XP_026098120.1	XM_026242335.1
ε *Protobothrops mucrosquamatus*	XP_015668648.1	XM_015813162.1
ε *Lacerta agilis*	XP_033025005.1	XM_033169114.1
ε *Python bivittatus*	XP_007441665.1	XM_007441603.2
ε *Thamnophis elegans*	XP_032092536.1	XM_032236645.1

ε *Tachyglossus aculeatus*	XP_038625000.1	XM_038769072.1
ε *Trichosurus vulpecula*	XP_036624896.1	XM_036769001.1
ε *Vombatus ursinus*	XP_027712861.1	XM_027857060.1
ε *Orcinus orca*	XP_033278994.1	XM_033423103.1
ε *Ovis aries*	XP_027830863.1	XM_027975062.1
ε *Tursiops truncatus*	XP_033704285.1	XM_033848394.1
ε *Vicugna pacos*	XP_031530788.1	XM_031674928.1
ε *Balaenoptera musculus*	XP_036692530.1	XM_036836635.1
ε *Bos mutus*	XP_005906122.1	XM_005906060.2
ε *Camelus ferus*	XP_032354439.1	XM_032498548.1
ε *Capra hircus*	XP_017920003.1	XM_018064514.1
ε *Rhinopithecus roxellana*	XP_010362526.1	XM_010364224.2
ε *Trachypithecus francoisi*	XP_033040884.1	XM_033184993.1
ε *Cercocebus atys*	XP_011910638.1	XM_012055248.1
ε *Macaca fascicularis*	XP_005582649.1	XM_005582592.2
ε *Pongo abelii*	XP_024090773.1	XM_024235005.1
ε *Pan paniscus*	XP_034797235.1	XM_034941344.1
ε *Pan troglodytes*	XP_016786791.1	XM_016931302.1

ε Homo sapiens	NP_000071.1	NM_000080.4
5HT$_3$A Homo sapiens	NP_998786.3	NM_213621.4
5HT$_3$A Sarcophilus harrisii	XP_003764254.1	XM_003764206.2

Table 2: Sites detected to be undergoing positive selection by MEME. Their position in the acetylcholine receptor alignment and corresponding to the sequence of the *Homo sapiens* α1 subunit, and their location within the structures of the acetylcholine receptor subunits, are provided.

Significant Site	Position in Alignment	Position in full-length *Hs* α1 subunit sequence	Location in Receptor
1	12	38	α1
2	30	51	α1-β1 linker
3	38	57	β1, Loop G
4	72	93	β2-β3 linker
5	82	103	β3
6	83	104	β3
7	89	110	β3-β4 linker
8	97	118	β4
9	112	133	Immediately preceding β5
10	116	137	β5, Loop E
11	131	152	β6'
12	133	154	β6' (adjacent to the Cys loop)
13	156	177	β7-β8 linker, Loop B
14	165	186	β8

15	170	190	β8-β9 linker, Loop F
16	181	-	β8-β9 linker, Loop F
17	182	-	β8-β9 linker, Loop F
18	211	212	β9
19	215	-	β9
20	220	216	β9-β10 linker, Loop C
21	223	219	β9-β10 linker, Loop C
22	239	229	β10
23	303	275	M2
24	324	296	M2-M3 loop
25	347	319	M3
26	355	327	M3-MX linker
27	363	334	M3-MX linker
28	366	337	MX
29	374	345	MX
30	418	361	ICD
31	426	364	ICD
32	429	365	ICD

33	450	-	ICD
34	454	372	ICD
35	456	373	ICD
36	457	-	ICD
37	484	379	ICD
38	490	381	ICD
39	506	382	ICD
40	560	-	ICD
41	737	401	ICD
42	743	407	ICD
43	759	422	M4
44	783	446	M4
45	828	-	C-terminus of the ECD
46	836	460	C-terminus of the ECD
47	839	462	C-terminus of the ECD

Table 3: Alignment of ancestral acetylcholine receptor sequences reconstructed from amino acid sequence data. Only the 47 residues which have been detected to be undergoing significant positive selection are displayed, with the respective position of the site in the protein alignment also shown. If a site has undergone a substitution following divergence from the ancestral subunit immediately before it in the trajectory, it is indicated with an asterisk. A second asterisk indicates that the site was also found to be undergoing positive selection at that particular node.

Significant Site #	Position in Alignment	AncALL	AncBDGE	AncB	AncDGE	AncD	AncGE	AncG	AncE
1	12	L	L	L	L	L	L	L	L
2	30	N	N	T*	N	S*	N	N	N
3	38	Q	R*	R	T*	A*	T	T	T
4	72	R	S*	S	T*	T	S*	Q*	S
5	82	K	K	K	K	S*	N*	Q*	D*
6	83	S	S	S	V*	V	V	V	V
7	89	E	D*	D	D	D	D	T*	D
8	97	V	V	V	V	V	V	V	V
9	112	T	V*	V	C*	C	C	C	A*
10	116	V	V	V	V	V	V	V	V
11	131	K	R**	R	R	R	R	R	R
12	133	S	S	S	S	S	S	S	T*
13	156	T	T	T	T	T	T	T	T

14	165	V	Q*	Q	Q	Q	Q	Q	Q
15	170	E	E	E	E	E	E	E	E
16	181	-	E	E	E	E	E	T*	E
17	182	-	-	-	-	-	-	-	-
18	211	K	K	K	K	K	K	K	K
19	215	-	-	-	K	K	S*	A*	S
20	220	P	P	P	P	P	P	P	P
21	223	-	-	-	-	-	-	-	-
22	239	T	T	T	T	T	I*	I	I
23	303	S	S	S	S	S	S	N*	S
24	324	L	L	L	L	L	L	Q*	L
25	347	S	S	S	C*	C	C	C	C
26	355	H	H	H	L*	F*	L	L	L
27	363	T	T	T	T	T	S*	S	S
28	366	Q	Q	S*	Q	E*	Q	Q	Q
29	374	D	E*	E	E	E	E	K*	E
30	418	P	P	P	P	L*	P	P	P
31	426	-	-	-	-	-	-	-	-

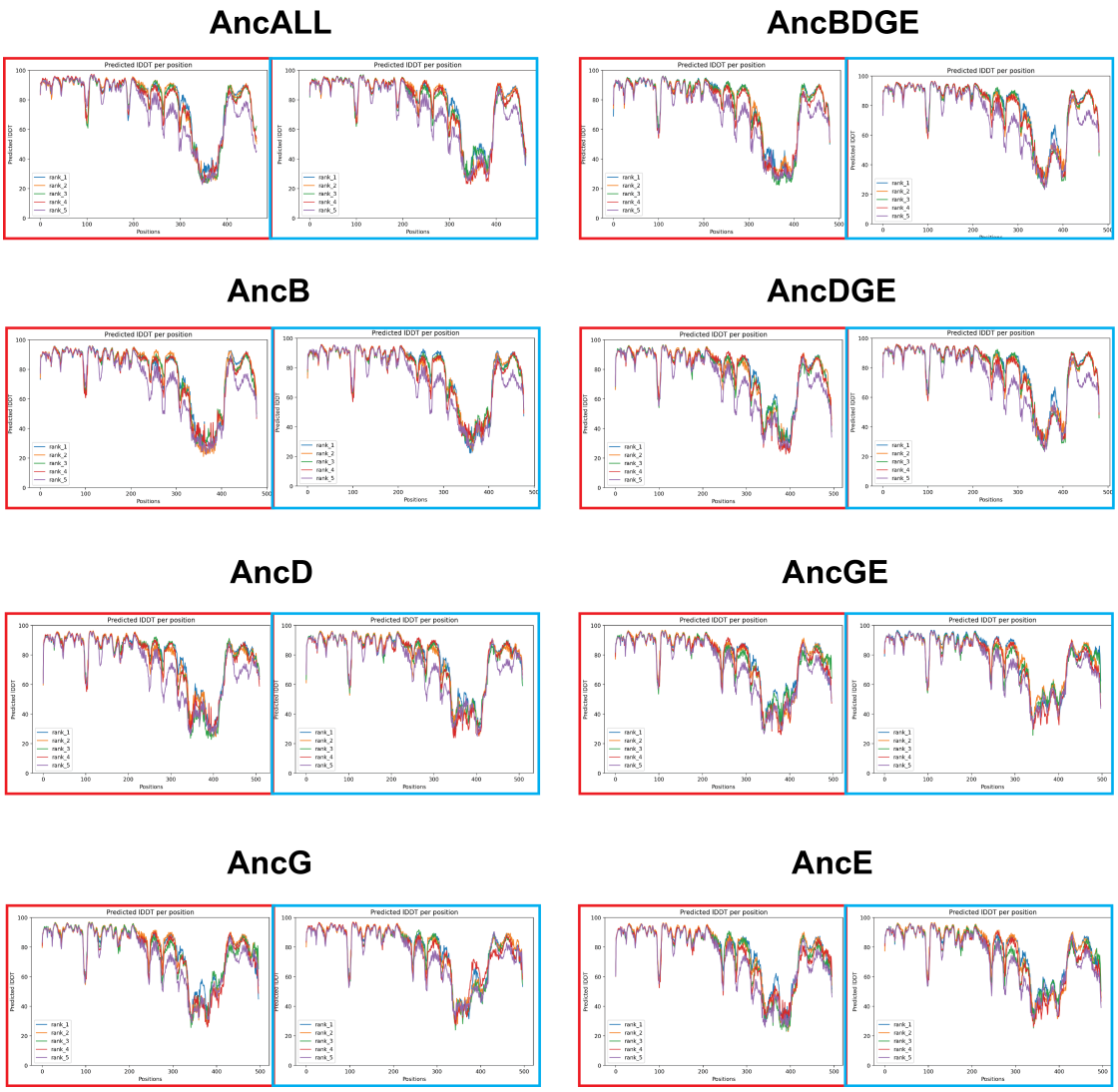
32	429	-	-	-	-	-	-	-	-
33	450	-	-	-	G	G	G	G	G
34	454	E	I*	V*	I	I	M*	M	M
35	456	T	V*	S*	V	V	L*	V*	L
36	457	-	-	-	-	-	-	-	-
37	484	E	E	E	E	E	E	E	E
38	490	K	K	K	K	K	K	K	K
39	506	P	P	P	P	S*	P	A*	P
40	560	-	-	-	-	-	-	-	-
41	737	E	E	E	S*	P*	A*	A	S*
42	743	K	K	K	N*	N	N	N	N
43	759	V	V	L*	E*	E	E	E	E
44	783	V	I*	S*	I	I	I	I	S*
45	828	S	P*	P	P	P	P	P	P
46	836	-	-	-	E	D*	K*	K	K
47	839	-	-	-	K	K	A*	L*	A

Table 4: Alignment of ancestral acetylcholine receptor sequences reconstructed from the amino acid sequences translated from nucleotide sequences. Only the 47 residues which have been detected to be undergoing significant positive selection are displayed, with the respective position of the site in the codon alignment also shown. If a site has undergone a substitution mutation following divergence from the ancestral subunit prior to it, it is indicated with an asterisk. A second asterisk indicates that the site was furthermore found to be significantly undergoing positive selection at that particular node.

Significant Site	Position in Alignment	AncALL	AncBDGE	AncB	AncDGE	AncD	AncGE	AncG	AncE
1	12	L	L	L	L	L	L	L	L
2	30	N	N	T*	N	Y*	N	N	N
3	38	K	K	R*	K	N*	K	T*	T*
4	72	S	S	S	N*	K*	N	Q*	N
5	82	K	K	K	K	K	K	K	K
6	83	T	A*	A	V*	V	V	V	V
7	89	E	D*	D	D	D	D	N*	D
8	97	V	V	V	V	V	V	V	V
9	112	T	A*	V*	A	C*	A	T**	A
10	116	V	V	V	V	V	V	V	V
11	131	K	R**	R	R	R	R	R	R
12	133	S	S	S	S	S	S	S	T*
13	156	T	T	T	T	T	T	T	T
14	165	V	Q*	Q	Q	Q	Q	Q	Q

15	170	D	E*	E	E	E	E	E	E
16	181	-	E	E	E	D*	K*	T*	K
17	182	-	-	-	-	-	-	-	-
18	211	K	K	K	K	K	K	K	K
19	215	-	-	-	K	K	K	E*	K
20	220	P	P	P	P	P	P	P	P
21	223	-	-	-	-	-	-	-	-
22	239	T	T	T	T	T	I*	I	I
23	303	S	S	S	S	S	S	N*	S
24	324	L	L	L	L	L	L	Q*	L
25	347	S	S	S	C*	C	C	C	C
26	355	H	H	H	L*	F*	L	L	L
27	363	T	T	T	T	I*	T	T	T
28	366	H	H	H	Q*	E*	Q	Q	Q
29	374	E	E	E	E	E	E	E	E
30	418	L	P*	P	P	I*	P	P	P
31	426	-	-	-	-	-	-	-	-

32	429	-	-	-	-	-	-	-	-
33	450	-	-	-	G	G	G	G	G
34	454	N	I*	I	I	I	M*	M	M
35	456	I	I	S*	V*	V	V	V	V
36	457	-	-	-	-	-	-	-	-
37	484	E	E	E	E	E	E	E	E
38	490	K	K	K	K	K	K	K	K
39	506	P	P	P	P	S*	P	A	P
40	560	-	-	-	-	-	-	-	-
41	737	E	E	E	E	P*	E	E	E
42	743	N	N	N	N	N	N	N	N
43	759	V	V	L*	E*	E	E	E	E
44	783	I	I	T*	I	I	I	I	T*
45	828	P	P	P	P	P	P	P	P
46	836	-	-	-	E	D*	K*	K	K
47	839	-	-	-	T	K*	L*	L	T*



Supplemental Figure 1: Predicted Local Distance Difference Test (pLDDT) scores for the acetylcholine receptor ancestral subunit structures predicted by AlphaFold2. The charts show the pLDDT score (percentages out of 100) at each residue in the sequence for the five top-ranked structures (ranks 1 to 5) predicted by AlphaFold2. The charts predicted for the translated amino acid sequences translated from nucleotide sequences are bordered in red, while the charts predicted for the amino acid sequences are bordered in blue.

Appendix II - Positive Selection Analysis

The examples used in the introduction (section 1.5) to illustrate the concept of positive selection analysis are conceptual, and by necessity overly simplistic, and thus do not reflect the actual methods used in this thesis. As such, a more detailed explanation is provided here (for further information, one may consult the original Mixed Effects Model of Evolution paper[31]).

All maximum likelihood-based branch-site tests, including the Mixed Effects Model of Evolution (MEME), employ codon substitution models for positive sequence analysis. In codon substitution models, a Markov chain is used to describe substitutions from one codon to another (as stop codons are not allowed inside proteins, they are not included in the model). The instantaneous rate of substitution (q) from codon I to codon J can be determined and is specified as q_{IJ}, with the rate matrix (Q) describing the rates for all possible codon substitutions. The substitution model incorporates three primary features of sequence evolution for determining the codon substitution rates:

1. Nucleotide mutational biases (θ).
2. The non-synonymous (β) and synonymous (α) mutation rates.
3. Codon frequencies (π_J for codon "*J*").

MEME uses a codon sequence model to describe the evolution of codons along branches in a phylogenetic tree. Given a phylogenetic tree τ, with B branches and total branch length T, the probability of substitution from codon I to codon J at a site along branch b in time t_b is recorded in the transition matrix $\boldsymbol{M_b(t_b)}=\mathbf{e}^{Q_{IJ}(t_b)}$, with the instantaneous substitution rates and rate matrix Q determined based on the features previously described (θ, β, α, and π variables). Unlike in previous models, in order for MEME to allow ω to vary across both sites and branches, the non-synonymous mutation rate (β) at each branch b is treated as a random draw from one of K selective categories, where each branch is assigned an independent category c_b that can take values from 1 ... K, with the configuration vector $C = (c_{1,\ldots,} c_B)$ describing all branch categories.

The goal is to detect sites where a proportion of lineages are evolving with ω>1. To accomplish this, each site is provided a set of free parameters that govern the strength of selection at that site for two discrete β categories, and these parameters are shared for all branches at that site. These two categories are:

1. Constrained (β^-): where β is either smaller or equal to α.
2. Unrestricted (β^+): where β may take on any value, including values larger than α.

The probability that branch *b* (for all *B* branches) is evolving with $\beta^b = \beta^-$ is $0 \leq q^- \leq 1$, and the complementary probability that $\beta^b = \beta^+$ is $q^+ = 1 - q^-$. Thus, the phylogenetic likelihood at a site can be found by computing the standard likelihood function with the following transition matrix for each branch *b*: $\boldsymbol{M_b(\alpha,\beta^-,\beta^+,q^-) = q^- e^{Q(\alpha,\beta^-;\theta,\pi)t_b} + (1-q^-)e^{Q(\alpha,\beta^+;\theta,\pi)t_b}}$. Therefore, four parameters are inferred (jointly from all branches) at each site: α, β^-, β^+, and q^-, and because α is estimated independently at each site (in previous branch-site models, α was normally kept constant at all sites), MEME has the added advantage of accounting for site-to-site variability in synonymous mutation rates[69].

Finally, at every site, the model from the previous step is fitted with $\beta^+ \leq \alpha$ (the null hypothesis, where the site is not under positive selection). MEME uses a test statistic for the likelihood ratio test in which the worst-case null of $\beta^+ = \beta^- = \alpha$ is a 0.33:0.3:0.37 mixture of χ^2_0 , χ^2_1, and χ^2_2[31]. In this way, MEME is able to test for which sites are undergoing positive selection at a proportion of its branches[31].

If the likelihood ratio tests detects that a site *s* is undergoing positive selection at a proportion of its branches, MEME also includes an emperical Bayes (EB) procedure for exploring which branches at that site are experiencing diversifying evolution. However, because such inferences would by necessity be based on limited evidence[31], it is not implemented in this study.

References

1. Hille, B. Ionic channels in excitable membranes. Current problems and biophysical approaches. *Biophys. J.* **22**, 283–294 (1978).
2. Sine, S. M. End-Plate Acetylcholine Receptor: Structure, Mechanism, Pharmacology, and Disease. *Physiol. Rev.* **92**, 1189–1234 (2012).
3. Catterall, W. A. Voltage-gated sodium channels at 60: structure, function and pathophysiology. *J. Physiol.* **590**, 2577–2589 (2012).
4. Sivilotti, L. & Nistri, A. GABA receptor mechanisms in the central nervous system. *Prog. Neurobiol.* **36**, 35–92 (1991).
5. Jiang, Y. *et al.* X-ray structure of a voltage-dependent K+ channel. *Nature* **423**, 33–41 (2003).
6. Jaiteh, M., Taly, A. & Hénin, J. Evolution of Pentameric Ligand-Gated Ion Channels: Pro-Loop Receptors. *PLOS ONE* **11**, e0151934 (2016).
7. Miller, P. S., Topf, M. & Smart, T. G. Mapping a molecular link between allosteric inhibition and activation of the glycine receptor. *Nat. Struct. Mol. Biol.* **15**, 1084–1093 (2008).
8. Thompson, A. J., Lester, H. A. & Lummis, S. C. R. The structural basis of function in Cys-loop receptors. *Q. Rev. Biophys.* **43**, 449–499 (2010).
9. Dajas-Bailador, F. & Wonnacott, S. Nicotinic acetylcholine receptors and the regulation of neuronal signalling. *Trends Pharmacol. Sci.* **25**, 317–324 (2004).
10. Improgo, Ma. R. D., Scofield, M. D., Tapper, A. R. & Gardner, P. D. The nicotinic acetylcholine receptor CHRNA5/A3/B4 gene cluster: Dual role in nicotine addiction and lung cancer. *Prog. Neurobiol.* **92**, 212–226 (2010).
11. Raftery, M. A., Hunkapiller, M. W., Strader, C. D. & Hood, L. E. Acetylcholine receptor: complex of homologous subunits. *Science* **208**, 1454–1456 (1980).
12. Le Novere, N. & Changeux, J.-P. Molecular evolution of the nicotinic acetylcholine receptor: An example of multigene family in excitable cells. *J. Mol. Evol.* **40**, 155–172 (1995).

13. Tammimäki, A., Horton, W. J. & Stitzel, J. A. Recent advances in gene manipulation and nicotinic acetylcholine receptor biology. *Biochem. Pharmacol.* **82**, 808–819 (2011).

14. Lee, W. Y. & Sine, S. M. Principal pathway coupling agonist binding to channel gating in nicotinic receptors. *Nature* **438**, 243–247 (2005).

15. Emlaw, J. R. *et al.* A single historical substitution drives an increase in acetylcholine receptor complexity. *Proc. Natl. Acad. Sci.* **118**, e2018731118 (2021).

16. Garcia, A. K. & Kaçar, B. How to resurrect ancestral proteins as proxies for ancient biogeochemistry. *Free Radic. Biol. Med.* **140**, 260–269 (2019).

17. Thornton, J. W. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**, 366–375 (2004).

18. Zhang, J. & Nei, M. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. *J. Mol. Evol.* **44 Suppl 1**, S139-146 (1997).

19. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

20. Wheeler, L. C., Lim, S. A., Marqusee, S. & Harms, M. J. The thermostability and specificity of ancient proteins. *Curr. Opin. Struct. Biol.* **38**, 37–43 (2016).

21. Yang, Z., Kumar, S. & Nei, M. A New Method of Inference of Ancestral Nucleotide and Amino Acid Sequences. *Genetics* **141**, 1641–1650 (1995).

22. Trudeau, D. L., Kaltenbach, M. & Tawfik, D. S. On the Potential Origins of the High Stability of Reconstructed Ancestral Proteins. *Mol. Biol. Evol.* **33**, 2633–2641 (2016).

23. Blanchette, M., Diallo, A. B., Green, E. D., Miller, W. & Haussler, D. Computational reconstruction of ancestral DNA sequences. *Methods Mol. Biol. Clifton NJ* **422**, 171–184 (2008).

24. Harms, M. J. & Thornton, J. W. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat. Rev. Genet.* **14**, 559–571 (2013).

25. Ugalde, J. A., Chang, B. S. W. & Matz, M. V. Evolution of coral pigments recreated. *Science* **305**, 1433 (2004).

26. Siddiq, M. A., Hochberg, G. K. A. & Thornton, J. W. Evolution of protein specificity: Insights from ancestral protein reconstruction. *Curr. Opin. Struct. Biol.* **47**, 113–122 (2017).

27. Clifton, B. E. *et al.* Evolution of cyclohexadienyl dehydratase from an ancestral solute-binding protein. *Nat. Chem. Biol.* **14**, 542–547 (2018).

28. Emlaw, J. R., Burkett, K. M. & daCosta, C. J. B. Contingency between Historical Substitutions in the Acetylcholine Receptor Pore. *ACS Chem. Neurosci.* **11**, 2861–2868 (2020).

29. Yang, Z. & dos Reis, M. Statistical properties of the branch-site test of positive selection. *Mol. Biol. Evol.* **28**, 1217–1228 (2011).

30. Bielawski, J. P. & Yang, Z. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J. Mol. Evol.* **59**, 121–132 (2004).

31. Murrell, B. *et al.* Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLoS Genet.* **8**, e1002764 (2012).

32. Plotkin, J. B. & Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* **12**, 32–42 (2011).

33. Shen, X., Song, S., Li, C. & Zhang, J. Synonymous mutations in representative yeast genes are mostly strongly non-neutral. *Nature* **606**, 725–731 (2022).

34. Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A. M. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**, 431–449 (2000).

35. Dineley, K. T., Pandya, A. A. & Yakel, J. L. Nicotinic ACh receptors as therapeutic targets in CNS disorders. *Trends Pharmacol. Sci.* **36**, 96–108 (2015).

36. Role, L. W. & Berg, D. K. Nicotinic receptors in the development and modulation of CNS synapses. *Neuron* **16**, 1077–1085 (1996).

37. Bertrand, D. & Terry, A. V. The wonderland of neuronal nicotinic acetylcholine receptors. *Biochem. Pharmacol.* **151**, 214–225 (2018).

38. Simmons, M. P., Ochoterena, H. & Freudenstein, J. V. Amino acid vs. nucleotide characters: challenging preconceived notions. *Mol. Phylogenet. Evol.* **24**, 78–90 (2002).

39. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **41**, D36–D42 (2013).

40. OpenTreeOfLife *et al.* Open Tree of Life Synthetic Tree. (2019) doi:10.5281/zenodo.3937742.

41. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

42. Jarasch, A. *et al.* ANTICALIgN: visualizing, editing and analyzing combined nucleotide and amino acid sequence alignments for combinatorial protein engineering. *Protein Eng. Des. Sel.* **29**, 263–270 (2016).

43. Löytynoja, A. Phylogeny-aware alignment with PRANK. in *Multiple Sequence Alignment Methods* (ed. Russell, D. J.) 155–170 (Humana Press, 2014). doi:10.1007/978-1-62703-646-7_10.

44. Xia, X. DAMBE7: New and Improved Tools for Data Analysis in Molecular Biology and Evolution. *Mol. Biol. Evol.* **35**, 1550–1552 (2018).

45. Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

46. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).

47. Finnigan, G. C., Hanson-Smith, V., Stevens, T. H. & Thornton, J. W. Evolution of increased complexity in a molecular machine. *Nature* **481**, 360–364 (2012).

48. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

49. Anisimova, M. & Gascuel, O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **55**, 539–552 (2006).

50. Changeux, J. P. *et al.* Brain nicotinic receptors: structure and regulation, role in learning and reinforcement. *Brain Res. Brain Res. Rev.* **26**, 198–216 (1998).

51. Duchêne, D. A., Mather, N., Van Der Wal, C. & Ho, S. Y. W. Excluding Loci With Substitution Saturation Improves Inferences From Phylogenomic Data. *Syst. Biol.* **71**, 676–689 (2021).

52. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

53. Cheng, X., Ivanov, I., Wang, H., Sine, S. M. & McCammon, J. A. Nanosecond-timescale conformational dynamics of the human alpha7 nicotinic acetylcholine receptor. *Biophys. J.* **93**, 2622–2634 (2007).

54. Tsunoyama, K. & Gojobori, T. Evolution of nicotinic acetylcholine receptor subunits. *Mol. Biol. Evol.* **15**, 518–527 (1998).

55. Corringer, P.-J., Novère, N. L. & Changeux, J.-P. Nicotinic Receptors at the Amino Acid Level. *Annu. Rev. Pharmacol. Toxicol.* **40**, 431–458 (2000).

56. Millar, N. S. & Gotti, C. Diversity of vertebrate nicotinic acetylcholine receptors. *Neuropharmacology* **56**, 237–246 (2009).

57. Lee, W. Y., Free, C. R. & Sine, S. M. Nicotinic receptor interloop proline anchors beta1-beta2 and Cys loops in coupling agonist binding to channel gating. *J. Gen. Physiol.* **132**, 265–278 (2008).

58. Spurny, R. *et al.* Molecular blueprint of allosteric binding sites in a homologue of the agonist-binding domain of the α7 nicotinic acetylcholine receptor. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E2543-2552 (2015).

59. Corringer, P.-J. *et al.* Critical Elements Determining Diversity in Agonist Binding and Desensitization of Neuronal Nicotinic Acetylcholine Receptors. *J. Neurosci.* **18**, 648–657 (1998).

60. Galzi, J. L., Bertrand, S., Corringer, P. J., Changeux, J. P. & Bertrand, D. Identification of calcium binding sites that regulate potentiation of a neuronal nicotinic acetylcholine receptor. *EMBO J.* **15**, 5824–5832 (1996).

61. Harvey, S. C., McIntosh, J. M., Cartier, G. E., Maddox, F. N. & Luetje, C. W. Determinants of specificity for alpha-conotoxin MII on alpha3beta2 neuronal nicotinic receptors. *Mol. Pharmacol.* **51**, 336–342 (1997).

62. McCormack, T. J. *et al.* Rapid desensitization of the rat α7 nAChR is facilitated by the presence of a proline residue in the outer β-sheet. *J. Physiol.* **588**, 4415–4429 (2010).

63. Stokes, C., Treinin, M. & Papke, R. L. Looking below the surface of nicotinic acetylcholine receptors. *Trends Pharmacol. Sci.* **36**, 514–523 (2015).

64. P, B. & Jp, M. Molecular basis of the two nonequivalent ligand binding sites of the muscle nicotinic acetylcholine receptor. *Neuron* **3**, (1989).

65. daCosta, C. J. B. *et al.* Ancestral Reconstruction Approach to Acetylcholine Receptor Structure and Function. *Biophys. J.* **114**, 296a (2018).

66. Zhao, M., Ma, Y., Xin, J., Cao, C. & Wang, J. Detection of differential selection pressure and functional-specific sites in subunits of vertebrate neuronal nicotinic acetylcholine receptors. *J. Biomol. Struct. Dyn.* **0**, 1–10 (2021).

67. Zhang, J., Nielsen, R. & Yang, Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **22**, 2472–2479 (2005).

68. Venkat, A., Hahn, M. W. & Thornton, J. W. Multinucleotide mutations cause false inferences of lineage-specific positive selection. *Nat. Ecol. Evol.* **2**, 1280–1288 (2018).

69. Pond, S. K. & Muse, S. V. Site-to-site variation of synonymous substitution rates. *Mol. Biol. Evol.* **22**, 2375–2385 (2005).

70. Rambaut, A. FigTree. (2018).

71. The PyMOL Molecular Graphics System.

72. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).