

# Phylogenetic and Structural Analysis of BC200 and Hominoidea Homologs

Derek Robinson, Mina Emadi, Mazyar Ghezelji

University of Victoria

Victoria, Canada

{drobinson, minaemadi, mazyarghezelji}@uvic.ca

**Abstract**—Alzheimer's disease (AD) is a neurodegenerative disorder, resulting from synaptic plasticity failure in neurons, which greatly affects memory and thinking skills of patients. This disease involves multiple cell types and signalling pathways which makes it complex and difficult to analyze, but understanding it is the main step towards diagnosing and preventing AD. There are some previous studies suggesting that AD is uniquely human disease, but as there are some similarities between *Homo sapiens* and other species of the hominidae superfamily, we try to answer the question of “is there any possibility that other species have AD”? In order to answer this question we first identify the causes of AD; accumulation of beta-amyloid protein and tau protein. Abnormal levels of both these proteins can disrupt cell functions and eventually cause diseases such as AD. We break down our problem to two crucial parts; performing phylogenetic analysis of BC200 and investigating the structural differences between BC200 and its four most closely related hominoidea homologs. We start by the most related sequences to BC200 by multiple sequence alignment, then building phylogenetic tree of the sequences so that we can determine which homologs are closely related. Observing the most similar sequences and building their secondary structure, which then can be used to compare functions of BC200 and its homologs. Although we found some similarities between these sequences, these comparisons reveal some structural differences between human BC200 and its homologs which means that their functions in body may be different. As a result, our research supports the hypothesis that AD is uniquely human.

## I. INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disease which greatly affects the lives of those who are diagnosed and those who care for the diagnosed. In the United States, 6.2 million patients are estimated to be living with AD and it is the sixth leading cause of death for American adults [1]. In Canada, over 747,000 patients are living with AD, or another form of dementia

[2]. Symptoms of AD range from memory loss and poor judgement in mild cases to the inability to communicate and seizures in severe cases [3].

AD involves multiple cell types and signaling pathways [4], as such, the collective knowledge of AD is spread across many different domains. This spread of knowledge means that fully understanding AD in humans is difficult, let alone understanding the disease in other species of the hominoidea superfamily. Finch and Austad argue that with our current understanding of AD, it is not possible to determine if AD is uniquely human [5]. As such, this paper explores one facet of AD pathogenesis, the long non-coding RNA (lncRNA) BC200. BC200 has been implicated in AD as it upregulates the expression of b-site APP-cleaving enzyme1 (BACE1) [4], [6]. The upregulation of BACE1 in turn leads to higher levels of beta-amyloid ( $A\beta$ ) in the brain, thus, disrupting cell function [4], [6].

In hopes of better understanding how AD may affect other species in the hominoidea superfamily, specifically, the role that hominoidea homologs of BC200 play in AD, we answer the following research questions:

- RQ1. What does the phylogenetic tree of BC200 look like?
- RQ2. What structural differences exist between BC200 and its four most closely related hominoidea homologs?

The following paper is structured as follows: section II gives background into BC200 and the role it plays in AD, section III lays out the methods used for selecting BC200 homologs and performing both the phylogenetic structural analysis, section IV presents the results of our analysis, section V discusses the relevancy of the results, and section VI concludes the paper.

## II. BACKGROUND

### A. What is BC200?

Brain Cytoplasmic 200 lncRNA (BC200), sometimes referred to as BCYRN1, is a 200 nucleotide long RNA transcript which is found mostly in the brain [7]. As a non-coding RNA, BC200 is not translated into protein but can be used as a potential therapeutic target and biomarker due to its regulatory role in biological processes involved in disease development [4], [8]. This lncRNA has recently been studied extensively because of its role in regulating translation and inhibiting its initiation, as well as its impacts in the pathogenesis of Alzheimer's disease and cancer [4], [7]. Because of the crucial role of BC200 in translation control, it impacts the synthesis of dendritic proteins which facilitates long-term plastic changes at the synapse [8].

### B. The Relation Between AD and BC200

Alzheimer's disease (AD) is a neurodegenerative disease resulting from synaptic plasticity failure in neurons [8]. It is a complex disease, meaning that it involves multiple cell types and signaling pathways [4].

AD is thought to occur due to the accumulation of two proteins in the brain. One of them is beta-amyloid ( $A\beta$ ) which accumulates in neurons, forms plaques, and disrupts cell functions. The other one is hyperphosphorylated tau protein which in abnormal levels can form neurofibrillary tangles in neurons and block synaptic transmissions [4].

$A\beta$ , a cleavage product of the amyloid precursor protein (APP), is generated by b-site APP-cleaving enzyme1 (BACE1) and  $\gamma$ -secretase complex, and it strongly influences the pathogenesis of AD. Inhibition of BACE1 activity and the subsequent reduction in  $A\beta$  levels may cure or prevent AD [4], [6].

BC200 facilitates AD pathogenesis by upregulating  $A\beta$  production through the modulation of BACE1 expression. The inhibition of BC200 significantly suppresses BACE1 expression, increases cell viability and reduces cell apoptosis in an AD model, and these effects are reversed by BC200 over-expression [4], [6].

Many researches have demonstrated the important role of BC200 in AD. El Mus *et al.* show that there are steady decline in BC200 level from age 49 to 86, but, in AD brain its level was substantially higher [8]. They

also observe that BC200 expression is increased in brain areas that are involved in AD and it is parallel with severity of disease. Huanyen Li *et al.* establish an AD cell model over expressing  $A\beta$ 1-42 to observe the effects of BC200 on the cell viability and apoptosis and to investigate the associated underlying mechanisms [6]. They observe that BC200 and BACE1 were increased upon treatment with  $A\beta$ 1-42, and inhibition of BC200 rescued this  $A\beta$ 1-42-mediated dysfunction, as indicated by the interaction of BC200 directly targeting BACE1. Moreover, inhibition of BC200 increased AD cell growth and reduced cells apoptosis. They demonstrate that BC200 is a potent positive regulator of BACE1 in AD cells and in conclusion, lncRNA BC200 facilitates AD pathogenesis by up-regulating  $A\beta$  through BACE1.

## III. MATERIALS AND METHODS

### A. Selection of lncRNAs

The lncRNA BC200 was selected for phylogenetic and structural analysis due to the role it plays in AD as discussed in section II. The homologs of BC200 were selected as a result of an NCBI Blast [9] search. Specifically, Megablast [10] with default parameters was used as it is able to compare closely related sequences [11]. From the Blast results, the top six sequences which are known BC200 homologs as indicated by the inclusion of BC200 in their name were chosen. Two of the top hits in the Blast results were complete bacterial artificial chromosome sequences for pan troglodytes. Inclusion of these two sequences (accession numbers: AC185986.3, AC183594.3) caused errors in the analysis software described in section III-B and section III-C, as such we did not consider these full chromosome sequences in our analysis. Table I outlines the sequence name including the organism and the accession number of each of the chosen sequences.

**TABLE I:** lncRNA Sequence Information

Organism Name	Sequence Accession Number
Homo sapiens	NR_001568.1
Pongo pygmaeus	AF067780.1
Pan paniscus	AF067778.1
Gorilla gorilla	AF067779.1
Macaca mulatta	AF067784.1
Hylobates lar	AF067781.1
Papio hamadryas	AF067782.1

### B. Phylogenetic Analysis

The lncRNA sequences in table I were subject to phylogenetic analysis. The phylogenetic tree was built with the MEGA11 [12] bioinformatics software. First the sequences outlined in table I were aligned using the MEGA11 alignment module. The multiple sequence alignment was performed using the MUSCLE algorithm with default parameters. Once the multiple sequence alignment had been performed in MEGA11, it was then possible to build the phylogenetic tree.

The phylogenetic tree was built using the maximum likelihood statistical method and Tamura-Nei substitution model [13], [14]. The bootstrap method was used as a test of phylogeny with 500 replications. In order to further support the phylogenetic tree, estimates of evolutionary divergence were also calculated by MEGA11 [12]. Analyses were conducted using Maximum Composite Likelihood and the Tamura-Nei model [13], [14]. All ambiguous positions were removed for each sequence pair (pairwise deletion option).

### C. Structural Analysis

In order to find RNA secondary structure of BC200 and three of closest homologs of it, we used FORNA, a platform for drawing the secondary RNA structure [15]. FORNA is able to draw the RNA secondary structure from the dot bracket notation provided by the user or from the minimum free energy (MFE) structure as calculated by RNAFold [16]. As this analysis aims at computationally determining a secondary structure, the latter option of using the MFE structure was employed. Each of the five structures present in section IV-B was drawn individually. First, we analyzed BC200 structure to get a better understanding of its function in body. Knowing that BC200 structure consist of three main domains, we also look at its homologs from this view point. Comparing each homolog structure to BC200, we divided them to three main domains like BC200 structure and based on each part function we analyzed the structural differences.

### D. Finding Alu Domains

Alu elements are elements of DNA that can excise and then get re-introduced into the genome. They called retrotransposons, also they are very common in human and primate genomes [17]. In order to find them, we used Dfam website (<https://www.dfam.org/>). The Dfam

database is an open collection of Transposable Element DNA sequence alignments, hidden Markov Models (HMMs), consensus sequences, and genome annotations. By using the sequence search we find the Alu domains of the BC200 and its homologs. Comparing the Alu domain structure gives us a suitable result for the RNA function.

### E. Finding Conserved Parts of BC200

Over many generations, random mutations and deletions can change nucleic acid sequences in the genome of an evolutionary lineage. Chromosomal rearrangement may also recombine or delete sequences. Conserved sequences are sequences which persist in the genome despite such forces, and have slower rates of mutation than the background mutation rate [18]. One of the best ways for visualize conserved sequences during evolution, is multiple sequence alignment. Here, we aligned BC200 and the four most related homologs with each other to find the conserved parts of the sequence. We used Clustal Omega for aligning the sequences [19]. The CLUSTAL format includes a plain-text key to annotate conserved columns of the alignment, denoting conserved sequence (\*), conservative mutations (:), semi-conservative mutations (.), and non-conservative mutations ( ). The output format is in ClustalW, and the guide tree is shown below.

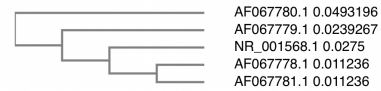


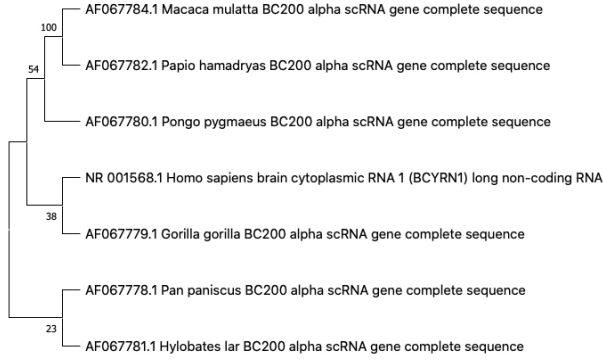
Fig. 1: Multiple Sequence Alignment Guide Tree

## IV. RESULTS

### A. Phylogenetic Tree of BC200

The results of the phylogenetic analysis can be found in figure 2. The tree shows that the Homo sapiens BC200 lncRNA and the Gorilla gorilla BC200 lncRNA are likely to share a common ancestor. Similarly, the BC200 homologs from Pan paniscus and Hylobates lar are likely to share a common ancestor. This is also the case for the BC200 homologs from Macaca mulatta, Papio hamadryas, and Pongo pygmaeus.

Additionally, table II shows the pairwise distances between sequences which estimate the evolutionary divergence between sequences. The evolutionary divergence estimates show that the four most closely related homologs of BC200 found in other hominoidea

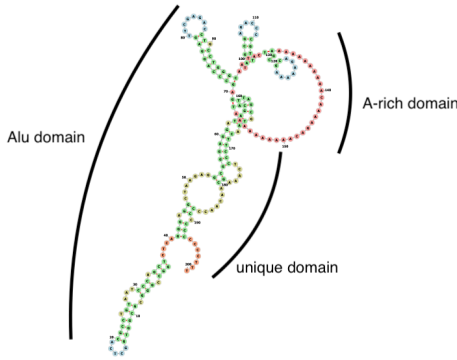


**Fig. 2:** Phylogenetic tree of BC200 and hominoidea homologs. The numbers at each node are the bootstrap support values obtained by maximum likelihood.

come from *Pan paniscus* (AF067778.1), *Pongo pygmaeus* (AF067780.1), *Hylobates lar* (AF067781.1), and *Gorilla gorilla* (AF067779.1), in order of evolutionary distances. Thus, the aforementioned BC200 homologs (AF067778.1, AF067780.1, AF067781.1, AF067779.1) are the closest relatives of BC200 and have their secondary structure analyzed in section IV-B.

### B. Structural difference between BC200 and its homologs

The BC200 structure consist of three main parts: A-rich domain, Alu domain and unique domain [20].



**Fig. 3:** BC200 RNA Secondary Structure

Finding the Alu domain of BC200 and its homologs (supplementary info), we concluded that the Alu domain between human BC200 and its Select seq *Pan paniscus* and *Pongo pygmaeus* homolog are completely the same. On the other hand the alu domain between human BC200 and its gorilla gorilla and *Hylobates lar* and other

homologs have strong differences. We didn't even find any similarity between alu domains of human BC200 and gorilla gorilla BC200.

The Alu domain of the mammalian signal recognition particle (SRP) comprises the heterodimer of proteins SRP9 and SRP14 bound to the 5 and 3 terminal sequences of SRP RNA [21]. So, their function in brain may be same with each other. It's worth mentioning that, all of the results that we get are based on simipilifying assumptions and in the organsim body, conditions may be different, so each long non-coding RNA function can be different too. Figures 4a, 4b, and 4c show the RNA secondary structure of the three most closely related BC200 homologs from great apes (hominidae). Additionally, we have chosen to depict the RNA secondary structure of the *Hylobates lar* BC200 homolog in Figure 4d. While *Hylobates lar* is not a great ape, it is still part of the hominoidea superfamily, and Phylogenetic analysis revealed that it was closely related to BC200.

### C. Finding Conserved Parts of the Sequence

Using multiple sequence alignment we aligned BC200 with its four mostly related homologs to find the conserved parts of the sequence. Here is the alignment result:

It can be concluded from the results, most parts of the BC200 sequence are conserved. Non-coding sequences important for gene regulation, such as the binding or recognition sites of ribosomes and transcription factors. However, sequence conservation in non-coding RNAs is generally poor compared to protein-coding sequences, and base pairs that contribute to structure or function are often conserved instead [22]. So, the result shows that there should be a high similarity between BC200 RNA function in human body and in other species. So, the following result does not support our hypothesis, and they suggest that Alzheimers can be shared between human and great apes.

## V. DISCUSSION

In this study we first began by gathering known homologs of *Homo sapiens* BC200 via NCBI Blast [9], [23]. Then, using these homologs, we built the phylogenetic tree in order to determine which homologs were the most evolutionarily related. Once we had determined that the BC200 homologs from *Pan paniscus*, *Pongo pygmaeus*, *Hylobates lar*, and *Gorilla gorilla* were the

**TABLE II:** Estimates of Evolutionary Divergence between Sequences

Accession Number						
NR_001568.1						
AF067780.1	0.01023					
AF067778.1	0.00503	0.02169				
AF067779.1	0.03070	0.02890	0.01709			
AF067784.1	0.03088	0.04573	0.03802	0.05235		
AF067781.1	0.02561	0.02763	0.01562	0.02439	0.04353	
AF067782.1	0.03120	0.04478	0.03947	0.04731	0.00852	0.04259

most closely related to Homo sapiens BC200 we individually built their RNA secondary structure. We began with Homo sapiens BC200, in order to identify the differing domains present in the secondary structure. By determining the domains present in Homo sapiens BC200, shown in figure 3, and comparing to the secondary structures of the homologs we observe some differences. The main differences occur in the Alu domain of Homo sapiens BC200 and that of Gorilla gorilla and Hylobates lar. However, we do observe similarities between the Alu domains of Homo sapiens BC200 and Pan paniscus and Pongo pygmaeus BC200. Thus, based on the results we present here, we conclude that the function of BC200 in the brain of Homo sapiens versus other hominoidea **may** be different.

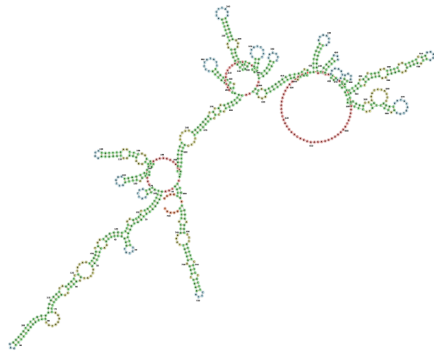
The two main characteristics of a brain which has developed AD are the presence of beta-amyloid ( $A\beta$ ) in cerebral vessels and abundance of neurofibrillary tangles (NFTs) [24]. The beta-amyloid protein involved in Alzheimers comes in several different molecular forms that collect between neurons. It is formed from the breakdown of a larger protein, called amyloid precursor protein [4]. One form, beta-amyloid 42, is thought to be especially toxic. In the Alzheimers brain, abnormal levels of this naturally occurring protein clump together to form plaques that collect between neurons and disrupt cell function. Neurofibrillary tangles are abnormal accumulations of the tau protein, which collect inside neurons. Healthy neurons, in part, are supported internally by structures called microtubules, which help guide nutrients and molecules from the cell body to the axon and dendrites. In healthy neurons, tau normally binds to and stabilizes microtubules. In Alzheimers disease, however, abnormal chemical changes cause tau to detach from microtubules and stick to other tau molecules, forming threads that eventually join to form tangles inside neurons. These tangles block the neurons transport system, which harms the synaptic communication between neurons [25]. Mus *et al.* showed that long non-

coding RNAs may be involved in AD [8]. It was shown that the expression of the BC200 lncRNA, which is a translational repressor, is differentially regulated in normal aging and AD [8].

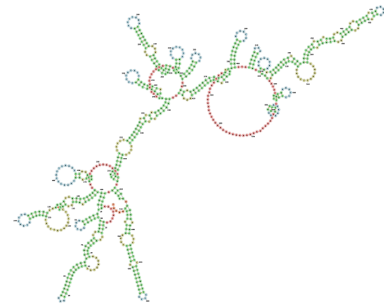
Here we showed that several hominoidea homologs of BC200 share evolutionary ancestors as shown in figure 2. Additionally, we have shown that the four most closely related hominoidea homologs of Homo Sapiens BC200 have differing RNA secondary structures and that they may act differently amongst each host organism. On the other hand, there are many papers suggest that presence of beta-amyloid can be the cause of Alzheimer disease, we cannot be certain that it is the real cause. There are biological evidence that contradict this hypothesis [26]. As mentioned above, overexpression of BC200 will increase expression of BACE1, which in turn generates beta-amyloid. Although there is high similarity between BC200 lncRNA in Homo sapiens and in Pan paniscus and Pongo pygmaeus, this does not definitively answer the questions of whether or not AD is uniquely human. Alzheimers disease is extremely complex and this work is only aimed determining one facet of how AD may proceed in hominoidea other than homo sapiens. In a recent review paper, Drummond *et al.* discuss the many models of AD pathology including transgenic mice/rats, invertebrate animals, and *in vitro* human cell culture models [27]. After reviewing the many models of AD pathology they conclude that AD is a uniquely human disease [27]. The conclusion of Drummond *et al.* is supported by our observations that the function of BC200 across different hominoidea is likely different.

## VI. CONCLUSION

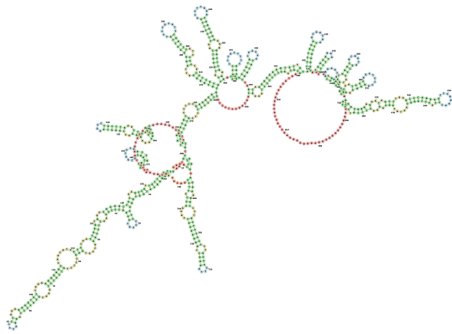
In this paper our goal was to determine if there was evidence which supports the hypothesis that Alzheimers disease is uniquely human. In order to determine if such evidence exists we first performed phylogenetic analysis to determine candidate species for RNA secondary



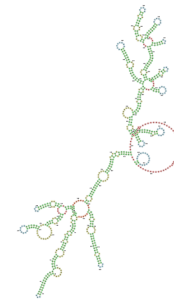
(a) BC200 RNA Secondary Structure in Gorilla gorilla



(b) BC200 RNA Secondary Structure in Pan paniscus



(c) BC200 RNA Secondary Structure in Pongo pygmaeus



(d) BC200 RNA Secondary Structure in Hylobates lar

**Fig. 4:** RNA Secondary Structures of Hominoidea BC200 Homologs, generated by FORNA [15] using default parameters

structure analysis. Once we had determined the RNA secondary structure for the four most closely related hominoidea homologs of BC200, we compared their secondary structures to the secondary structure of Homo sapiens BC200. We determined that the Alu domain of Homo sapiens, Pan paniscus, and Pongo pygmaeus BC200 are similar. However, this was not the case for the Alu domain of Gorilla gorilla or Hylobates lar. Further analysis determined that more than 90% of Homo sapiens BC200 is conserved during evolution, as shown in figure 5. While much of the Homo sapiens BC200 sequence is conserved, all of the homologs which we analyzed contain many extra nucleotides. The result of a different Alu domain and sequence length may suggest that BC200 has a different regulatory function each host organism. In conclusion, our analysis of BC200 homologs found in hominoidea supports the conclusion

presented by Drummond *et al.* that Alzheimers Disease is uniquely human.

## REFERENCES

- [1] "Alzheimer's disease — disease or condition of the week — cdc," <https://www.cdc.gov/dotw/alzheimers/index.html#:~:text=Alzheimer's%20disease%20is%20the%20most,of%20death%20for%20all%20adults.>, June 2021, (Accessed on 10/19/2021).
- [2] "Alzheimer's and dementia in canada," <https://www.alz.org/ca/dementia-alzheimers-canada.asp>, 2021, (Accessed on 10/19/2021).
- [3] "What are the signs of alzheimer's disease? — national institute on aging," <https://www.nia.nih.gov/health/what-are-signs-alzheimers-disease>, May 2017, (Accessed on 11/21/2021).

CLUSTAL O(1.2.4) multiple sequence alignment

```

AF067780.1      GAAGCCAGCAGAGAAGACATGGTGGACTCTCTGTGAGGCCCAACCCCAATCATGAGC
AF067779.1      GAAGCCAGCAGAGAAGACATGGTGGACTCTCTGTGAGGCCCAACCCCAATCATGAGC
NR_001568.1      -----CCAAACCCCAATCATGAGC
AF067778.1      -----CCAAACCCCAATCATGAGC
AF067781.1      -----CCAAACCCCAATCATGAGC

AF067780.1      CTCGCCCCCTGAACGATATTCAGGAATCTCCAGGGAG-TATTGGAGGGGGCCCAAGTC
AF067779.1      CTCGCCCCCTGAACGATATTCAGGAATCTCCAGGGAGTTTGGAGGGGGCCCAAGTC
NR_001568.1      CTCGCCCCCTGAACGATATTCAGGAATCTCTGAGGGAGTTTGGAGGGGGCCCAAGTC
AF067778.1      CTCGCCCCCTGAACGATATTCAGGAATCTCTGAGGGAGTTTGGAGGGGGCCCAAGTC
AF067781.1      CTCGCCCCCTGAACGATATTCAGGAATCTCCAGGGAGTTTGGAGGGGGCCCAAGTC

AF067780.1      CTCCAATTGGGTGGTGGGGCTCCAGCCATTGTTCTTGCACATTCATGCCAATGTGAC
AF067779.1      CTCCAATTGGGTGGTGGGGCTCCAGCCATTGTTCTTGCACATTCATGCCAATGTGAC
NR_001568.1      CTCCAATTGGGTGGTGGGGCTCCAGCCATTGTTCTTGCACATTCATGCCAATGTGAC
AF067778.1      CTCCAATTGGGTGGTGGGGCTCCAGCCATTGTTCTTGCACATTCATGCCAATGTGAC
AF067781.1      CTCCAATTGGGTGGTGGGGCTCCAGCCATTGTTCTTGCACATTCATGCCAATGTGAC

AF067780.1      CTAATTGGCTTAAGCAGAGCTGGAGAAGCTGGATAACCTATTTAAATTCAGAGCCAAA
AF067779.1      TTAATTGGCTTAAGCAGAGCTGGAGAAGCTAGATAACCTATTTAAATTCAGAGCCAAA
NR_001568.1      CTAATTGGCTTAAGCAGAGCTGGAGAAGCTAGATAACCTATTTAAATTCAGAGCCAAA
AF067778.1      CTAATTGGCTTAAGCAGAGCTGGAGAAGCTAGATAACCTATTTAAATTCAGAGCCAAA
AF067781.1      CTAATTGGCTTAAGCAGAGCTGGAGAAGCTAGATAACCTATTTAAATTCAGAGCCAAA

AF067780.1      AACACTCAAGAGTGTTTTCTGAGGGGGCTGAGGGTTGAAAAATATCTATATTTGATT
AF067779.1      GACACTCAAGAGTGTTTTCTGAGGGGGCTGAGGGTTGAAAAATATCTATATTTGATT
NR_001568.1      -----GACACTCAAGAGTGTTTTCTGAGGGGGCTGAGGGTTGAAAAATATCTATATTTGATT
AF067778.1      GACACTCAAGAGTGTTTTCTGAGGGGGCTGAGGGTTGAAAAATATCTATATTTGATT
AF067781.1      GACACTCAAGAGTGTTTTCTGAGGGGGCTGAGGGTTGAAAAATATCTATATTTGATT

AF067780.1      TATCTCCCACTCATGTCTGAGTCAATTCCTTAGTCTTAGATGAGGAGAGGACTTGGG
AF067779.1      TATCTCCCACTCATGTCTGAGTCAATTCCTTAGTCTTAGATGAGGAGAGGACTTGGG
NR_001568.1      -----TATCTCCCACTCATGTCTGAGTCAATTCCTTAGTCTTAGATGAGGAGAGGACTTGGG
AF067778.1      TATCTCCCACTCATGTCTGAGTCAATTCCTTAGTCTTAGATGAGGAGAGGACTTGGG
AF067781.1      TATCTCCCACTCATGTCTGAGTCAATTCCTTAGTCTTAGATGAGGAGAGGACTTGGG

AF067780.1      AGTCATCCTTGTTTTGTAGTGAATATCAACCCATGCGCAGCAGAGGGAATCTGCAT
AF067779.1      AGTCATCCTTGTTTTGTAGTGAATATCAACCCATGCGCAGCAGAGGGAATCTGCAT
NR_001568.1      -----AGTCATCCTTGTTTTGTAGTGAATATCAACCCATGCGCAGCAGAGGGAATCTGCAT
AF067778.1      AGTCATCCTTGTTTTGTAGTGAATATCAACCCATGCGCAGCAGAGGGAATCTGCAT
AF067781.1      AGTCATCCTTGTTTTGTAGTGAATATCAACCCATGCGCAGCAGAGGGAATCTGCAT

AF067780.1      TTCAGAGCGCAATTTGAGATCTATGAAGAATTTCAATCAGAAATAGAGGCCGGGGCGC
AF067779.1      TTCAGAGCGCAATTTGAGATCTATGAAGAATTTCAATCAGAAATAGAGGCCGGGGCGC
NR_001568.1      -----TTCAGAGCGCAATTTGAGATCTATGAAGAATTTCAATCAGAAATAGAGGCCGGGGCGC
AF067778.1      TTCAGAGCGCAATTTGAGATCTATGAAGAATTTCAATCAGAAATAGAGGCCGGGGCGC
AF067781.1      TTCAGAGCGCAATTTGAGATCTATGAAGAATTTCAATCAGAAATAGAGGCCGGGGCGC

AF067780.1      GGTGGCTCAGCGCTGTAATCCAGCTCTCAGGGAGGC-AAGAGCGGGAGGATAGCTTGA
AF067779.1      GGTGGCTCAGCGCTGTAATCCAGCTCTCAGGGAGGCTAAGAGCGGGAGGATAGCTTGA
NR_001568.1      GGTGGCTCAGCGCTGTAATCCAGCTCTCAGGGAGGCTAAGAGCGGGAGGATAGCTTGA
AF067778.1      GGTGGCTCAGCGCTGTAATCCAGCTCTCAGGGAGGCTAAGAGCGGGAGGATAGCTTGA
AF067781.1      GGTGGCTCAGCGCTGTAATCCAGCTCTCAGGGAGGCTAAGAGCGGGAGGATAGCTTGA

AF067780.1      GCCCAGGAGTTCGAGACCTGCCTGGCAATATAGCGAGCCCGTCTCCAGAAAAAGGA
AF067779.1      GCCCTGGAGTTCGAGACCTACCGGGCAATATAGTGAGCCCGTCTCCAGAAAAATGA
NR_001568.1      GCCCAGGAGTTCGAGACCTGCCTGGCAATATAGCGAGCCCGTCTCCAGAAAAATGA
AF067778.1      GCCCAGGAGTTCGAGACCTGCCTGGCAATATAGCGAGCCCGTCTCCAGAAAAAGGA
AF067781.1      GCCCAGGAGTTCGAGACCTGCCTGGCAATATAGCGAGCCCGTCTCCAGAAAAAGGA

AF067780.1      AAAAA-----AAAAAAGAGCAAAAAAATAAGCGTAACCTCCCTCAAGCAACA
AF067779.1      AAAAA-----AAAAAAGAGCAAAAAAATAAGCGTAACCTCCCTCAAGCAACA
NR_001568.1      AAAAA-----AAAAAAGAGCAAAAAAATAAGCGTAACCTCCCTCAAGCAACA
AF067778.1      AAAAA-----AAAAAAGAGCAAAAAAATAAGCGTAACCTCCCTCAAGCAACA
AF067781.1      AAAAA-----AAAAAAGAGCAAAAAAATAAGCGTAACCTCCCTCAAGCAACA

AF067780.1      CCCCCCCC-----CCCTTTTCTATATTTCTGCTGGAAGAAAGCGCTTCTCTCTCTC
AF067779.1      CCCCCCCC-----CCCTTTTCTATATTTCTGCTGGAAGAAAGCGCTTCTCTCTCTC
NR_001568.1      CCCCCC-----CCCTTTTCTATATTTCTGCTGGAAGAAAGCGCTTCTCTCTCTC
AF067778.1      CCCCCC-----CCCTTTTCTATATTTCTGCTGGAAGAAAGCGCTTCTCTCTCTC
AF067781.1      CCCCCC-----CCCTTTTCTATATTTCTGCTGGAAGAAAGCGCTTCTCTCTCTC

AF067780.1      TTAAGCTGGGA-AGTATAAGCATCTGAAAT
AF067779.1      TTAAGCTGGGAAGATTAAGCATCTGAAAT
NR_001568.1      TTAAGCTGGGAAGATTAAGCATCTGAAAT
AF067778.1      TTAAGCTGGGAAGATTAAGCATCTGAAAT
AF067781.1      TTAAGCTGGGAAGATTAAGCATCTGAAAT

```

**Fig. 5:** BC200 and it's homologs ClustalW alignment Result

- [4] Y. Zhang, Y. Zhao, X. Ao, W. Yu, L. Zhang, Y. Wang, and W. Chang, "The role of non-coding rnas in alzheimers disease: From regulated mechanism to therapeutic targets and diagnostic biomarkers," *Frontiers in Aging Neuroscience*, p. 384, 2021.
- [5] C. E. Finch and S. N. Austad, "Commentary: is alzheimer's disease uniquely human?" *Neurobiology of aging*, vol. 36, no. 2, pp. 553–555, 2015.
- [6] H. Li, L. Zheng, A. Jiang, Y. Mo, and Q. Gong, "Identification of the biological affection of long noncoding rna bc200 in alzheimers disease," *Neuroreport*, vol. 29, no. 13, pp. 1061–1067, 2018.
- [7] H. Tiedge, W. Chen, and J. Brosius, "Primary structure, neural-

specific expression, and dendritic location of human bc200 rna," *Journal of Neuroscience*, vol. 13, no. 6, pp. 2382–2390, 1993.

- [8] E. Mus, P. R. Hof, and H. Tiedge, "Dendritic bc200 rna in aging and in alzheimer's disease," *Proceedings of the National Academy of Sciences*, vol. 104, no. 25, pp. 10 679–10 684, 2007.
- [9] "Nucleotide blast: Search nucleotide databases using a nucleotide query," [https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE\\_TYPE=BlastSearch&LINK\\_LOC=blasthome](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome), (Accessed on 10/20/2021).
- [10] A. Morgulis, G. Coulouris, Y. Raytselis, T. L. Madden, R. Agarwala, and A. A. Schäffer, "Database indexing for production megablast searches," *Bioinformatics*, vol. 24, no. 16, pp. 1757–1764, 2008.
- [11] F. Amirmahani and K. Jamshidi Goharrizi, "Phylogenetic analysis of three long non-coding rna genes: Ak082072, ak043754 and ak082467," *Journal of Genetic Resources*, vol. 4, no. 1, pp. 56–64, 2018.
- [12] K. Tamura, G. Stecher, and S. Kumar, "Mega11: molecular evolutionary genetics analysis version 11," *Molecular Biology and Evolution*, vol. 38, no. 7, pp. 3022–3027, 2021.
- [13] K. Tamura and M. Nei, "Estimation of the number of nucleotide substitutions in the control region of mitochondrial dna in humans and chimpanzees," *Molecular biology and evolution*, vol. 10, no. 3, pp. 512–526, 1993.
- [14] K. Tamura, M. Nei, and S. Kumar, "Prospects for inferring very large phylogenies by using the neighbor-joining method," *Proceedings of the National Academy of Sciences*, vol. 101, no. 30, pp. 11 030–11 035, 2004.
- [15] P. Kerpedjiev, S. Hammer, and I. L. Hofacker, "Forna (force-directed rna): Simple and effective online rna secondary structure diagrams," *Bioinformatics*, vol. 31, no. 20, pp. 3377–3379, 2015.
- [16] R. Lorenz, S. H. Bernhart, C. Höner zu Siederdisen, H. Tafer, C. Flamm, P. F. Stadler, and I. L. Hofacker, "Viennarna package 2.0," *Algorithms for molecular biology*, vol. 6, no. 1, pp. 1–14, 2011.
- [17] B. A. Dombroski, Q. Feng, S. L. Mathias, D. M. Sassaman, A. F. Scott, H. H. Kazazian Jr, and J. D. Boeke, "An in vivo assay for the reverse transcriptase of human retrotransposon 11 in saccharomyces cerevisiae," *Molecular and cellular biology*, vol. 14, no. 7, pp. 4485–4492, 1994.
- [18] M. Kimura and T. Ohta, "On some principles governing molecular evolution," *Proceedings of the National Academy of Sciences*, vol. 71, no. 7, pp. 2848–2852, 1974.
- [19] F. Madeira, Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. Tivey, S. C. Potter, R. D. Finn *et al.*, "The embl-ebi search and sequence analysis tools apis in 2019," *Nucleic acids research*, vol. 47, no. W1, pp. W636–W641, 2019.
- [20] E. Jung, J. Lee, H. J. Hong, I. Park, and Y. Lee, "Rna recognition by a human antibody against brain cytoplasmic 200 rna," *Rna*, vol. 20, no. 6, pp. 805–814, 2014.

- [21] O. Weichenrieder, K. Wild, K. Strub, and S. Cusack, "Structure and assembly of the alu domain of the mammalian signal recognition particle," *Nature*, vol. 408, no. 6809, pp. 167–173, 2000.
- [22] P. Johnsson, L. Lipovich, D. Grandér, and K. V. Morris, "Evolutionary conservation of long non-coding rnas; sequence, structure, function," *Biochimica et Biophysica Acta (BBA)-General Subjects*, vol. 1840, no. 3, pp. 1063–1071, 2014.
- [23] T. Madden, "Chapter 16: The blast sequence analysis tool, the ncbi handbook (2002)," 2012.
- [24] M. A. DeTure and D. W. Dickson, "The neuropathological diagnosis of alzheimers disease," *Molecular neurodegeneration*, vol. 14, no. 1, pp. 1–18, 2019.
- [25] R. H. Swerdlow, "Brain aging, alzheimer's disease, and mitochondria," *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, vol. 1812, no. 12, pp. 1630–1639, 2011.
- [26] D. J. Selkoe and J. Hardy, "The amyloid hypothesis of alzheimer's disease at 25 years," *EMBO molecular medicine*, vol. 8, no. 6, pp. 595–608, 2016.
- [27] E. Drummond and T. Wisniewski, "Alzheimers disease: experimental models and reality," *Acta neuropathologica*, vol. 133, no. 2, pp. 155–175, 2017.