On the origin of Human speech and intelligence

By Joseph Kang and Anirudh Jandhayala

Abstract

The emergence of human cognitive development and speech has been traced to a transcription factor gene, FOXP2. Presumed to originate from a common ancestor 400 million years ago, the FOXP2 gene is utilized by multicellular eukaryotes for various functions, most notably speech, cognitive development, and tissue repair. Furthermore, new data was recently released detailing the potential benefits of FOXP2 aiding in synaptic plasticity[1] as well as the ability to build and use complex tools, which is exclusive to humans. Due of this, many scientists believe that the development of human intelligence would have been impossible if not for this gene, because it gave humans the ability to cooperate with one another and learn from experiences much better than other species. However, much of this data has not been verified and conflicting opinions on the importance of this gene still differ in the scientific community. Thus, an experiment was conducted analyzing 12 different species' FOXP2 tract to determine when it evolved into the form we see in humans today and why it is significant. All FOXP2 amino acid strands were compared to the human form using a longest common subsequence algorithm in the software Python. The percent difference between all amino acid strands to humans were determined based

[1] Synaptic plasticity - the construction of neural networks in the brain allowing for memories and experiences to be remembered and repeated.

off of the data collected from Python. A phylogenetic tree was then constructed utilizing the distance method to provide a visual representation of the deviances that FOXP2 orthologs have between species. This FOXP2 phylogenetic tree was then compared to a phylogenetic tree containing the 12 species' entire genome to determine where this gene was conserved and where rapid change occurred. This data can also be compared to phenotypic differences in speech between animals and whether there is a correlation. Concluding the experiment, it was discovered that the human FOXP2 form differs at two distinct amino acid placements compared to chimpanzees at positions 301 and 325, which were recurring substitutions that also took place across all species when compared to humans. These two nonsynonymous[2] substitutions could be in part responsible for human intelligence, but it is safe to conclude that this is only one in numerous factors that drove human intelligence to form. This discovery could, however, give some insight on our linguistic heritage as a species. Another interesting fact about the data is that the FOXP2 of hummingbirds and alligators were the same and only contained 5.4% of the human form even though these two species are known to be able to make sounds for simple communication. Further research is needed to explain this phenomenon.

[2] Nonsynonymous - a mutation that results in a change in at least one coded amino acid

Introduction

The objective of this experiment was to investigate the development of speech in humans (*homo sapiens*) in the form of the gene FOXP2 when compared to speech-like homologs in other relative species. Conflicting inferences on the origin of human speech and cognitive development has presented a dichotomy over the past century, within the scientific community. This altercation was first addressed in the late 1980s when geneticists at Oxford University analyzed a family with severe speech impairments known as verbal dyspraxia. A pedigree spanning three generation showed a discernible trend between individuals with and without the disorder. The pattern of an offspring's genotype in relation to his/her parents followed an dominant, autosomal monogenic inheritance (a trait controlled by a single gene). Further research into the family confirmed that this disorder was caused by a single dominant gene, as a result of a gene mutation at position 31 of the 7th chromosome (Spinelli). Around this time, new data was released detailing the potential benefits of FOXP2 aiding in synaptic plasticity and the buildings of connections between neurons. Synaptic plasticity is the ability for neurons to form permanent connections that allow for long term memory and the ability to repeat experiences (muscle memory) (Huang). A PNAS paper published in 2014 discussed an experiment in which certain mice were genetically modified to make their FOXP2 code for the human protein form. The mice were significantly better at navigating through mazes and performing certain oral actions as compared to their normal counterparts. However, the mice were still unable to be

taught to understand human language which suggests that FOXP2 is not the only factor that helps in speech development.

The mutated gene would later be named FOXP2 because it codes for regulatory proteins within the Forkhead box protein domain which regulate the transcription of speech related genes. Proteins that are within this domain possess a special three dimensional property that allows binding interactions on nucleic acids to form (Marcus). This is one of the strategies that regulatory proteins use to aid in the essential parts of learning and development of an organism. As a regulatory protein, FOXP2's protein, usually referred to as SPCH1 (for 'speech 1'), holds true to this precedent. Besides language development, it has been hypothesized that this protein may have the capability to regulate hundreds of other proteins as a result of this unique property. Though SPCH1 is active in all parts of the brain, it's mainly focused on the planum temporale, one of the most important areas for language. The FOXP2 gene is also essential for speech related activities in all multicellular organisms which include human complex language, echolocation in bats and songs in birds (Staes).

Nascent human evolution began in the African Savannah 7-8 million years ago where bipedalism[3] was advantageous over other species (Binder). Obtaining more food and natural resources, in an efficient manner, were driving factors to develop manual action and tool use in humans. A higher caloric/protein intake (proportionally, humans consume 400, 635, and 800 more daily calories compared to chimps, gorillas, and orangutans respectively), led to human brains nearly tripling in the past 3.5 million years (from 450 cm3 to 1350 cm3). Perhaps the most significant change happened about half a million years ago when the appearance and

[3] Bipedalism - the ability to stand upright using only two limbs for transportation.

stabilization of the human FOXP2 allowed for cooperation between populations and much more complex tool use (Wilson). Furthermore, human abundancy coincided with this appearance. Although all animals contain a variation of the FOXP2 gene, the human form differs at two distinct amino acids from their most similar ancestor, the common chimpanzee. Due to this, it has been hypothesized by many scientists that the appearance of FOXP2 is what led to human population and evolution skyrocketing in the past 100,000 years. Most likely, the intelligence that humans possess today would be impossible without this gene (Ideggyogy).

Studies from renowned scientists at MIT constructed a phylogenetic tree, based on the FOXP2 divergence in relation to other species that contain the gene (Marcus) and claimed that this gene could provide a pathway to tracking our linguistic heritage. This research was conducted using nucleotide sequences with respect to zebrafish. To build on this research, a phylogenetic tree of the relationship between humans and other species as well as zebrafish and other species needed to be developed. However, this data has not been verified and no study has been done comparing the human FOXP2 form to animals that are not primates. Furthermore, no research has compared the derivation of the human FOXP2 to our entire genome's derivation. Thus, further research is warranted to not only confirm this new found evidence, but also to expand knowledge as to how this gene was altered through generations and if it is as significant as its made out to be.

An experiment was conducted analyzing 12 different species' FOXP2 differences when compared to humans. The differences in FOXP2 were compared to differences in the 12 species' entire genome, to see if we can determine when each of the species' FOXP2 form changed until

the human form was brought up. With all of this information in mind, it was hypothesized that If two phylogenetic trees are made (one containing a selected amount of species' entire genomes, the other containing the same species' FOXP2 tract), then both will be similar because FOXP2 should be mostly evolutionarily conserved because of how important it is.

Procedure

1. A phylogenetic tree of the FOXP2 gene was constructed which included humans, primates, amphibians, birds, and fish
    a. The tree showed the relative divergence for each of the species
2. Each FOXP2 gene was researched for the corresponding amino acid sequence in 10 different species ()
    a. The species were chosen based on the ancestral trace of the human FOXP2 genome, with primates being the closest relatives extending to birds and amphibians
    b. Data was collected from the genome database UniProt
    c. Data was collected from amino acid sequences (as opposed to nucleotide sequences) of the FOXP2 gene as the amount of data provided by the nucleotide sequences was infeasible to process
    d. Using this data, an amino acid sequence for each species was created
        i. Each letter denotes a specific amino acids
            1. Example: "W" corresponds to "Tryptophan"

ii. The amino acids were concatenated into one continuous sequence with no spaces or special characters

iii. Amino acids generally corresponded to the letter that it starts with, however certain amino acids start with the same letter and those are assigned a random one

3. Every gene was copied, pasted, and saved into its own text file, which was labeled by the respective species' scientific name as shown in Figure 1.1

```
  10         20         30         40         50
MMQESANETI SNSSMSQNGM SSLSSQLDAG SRDGRSSGET SSEVSAVELL
        60         70         80         90        100
HLQQQQALQA ARQLLLQQPG SGLKSPKNND KQRPLQVPVS VAMMSPQVIT
       110        120        130        140        150
PQQMQQILQQ QVLSPQQLQA LLQQQQAVML QQQHLQEFYK KQQEQLHLQL
       160        170        180        190        200
LQQQHPGKQA KEQQQQQQQL AAQQLVFQQQ LLQMQQLQQQ QHLLNMQRQG
       210        220        230        240        250
LLSMPPGPGQ PTLPGQTLPP AGLSPAELQQ LWKDVTASHT MEDNGMKHSG
       260        270        280        290        300
LDLSTNNNTS TTSTSNPKAS PPITHHSMSN GQSPALNNNR RESSLHEETA
       310        320        330        340        350
VSHSLYGHGV CKWPGCESIC DDFGQFLKHL NNEHALDDRS TAQCRVQMQV
       360        370        380        390        400
VQQLEIQLSK ERERLQAMMA HLHMRPSEPK PSPKPLNLVS SVTMSKNLPS
       410        420        430        440        450
ISPPNLPQTP TTPTAPVTPL SQMPQVPNVL SPANVPSMGA MRRRHTDKYS
       460        470        480        490        500
MALSSEIAPN YEFYKNADVR PPFTYATLIR QAIMESSDMQ LTLNEIYSWF
       510        520        530        540        550
TRTFAYFRRN AATWKNAVRH NLSLHKCFVR VENVKGAVWT VDEMEYQKRR
       560        570        580        590        600
SQKITGSPTL VKNLPSSLGY GAALNASLQA ALAETTLPLL GNPGLMNSAS
       610        620        630        640        650
AMMGASPPVM MSGSPTGLLQ GTTHEELNGT LDHLDTNGHS SPGYSPHTHL
       660        670        680        690
PPIHVKEEPL NMEDEDCPMS LVTTANHSPE LDDDRELEEG NLSEDLE
```

Figure 1.1: The list of all amino acids in the FOP2 gene in homo sapiens separated into series of 10 amino acids. To find an amino acid, find the associated multiple of 10 and count the required units.

4. Python, a computer software programming language (https://www.python.org/), was used to develop a program to find the longest common subsequence (LCS) between pairwise amino acid sequences and determine the percent difference and location/type of differences between the two sequences. The program works as follows:

Genomic Sequence Comparison Program

5. A dictionary was created with a variable being assigned to a the entire library of amino acids and their corresponding letter
    a. For example, the variable name would be “Amino_Acid” and it would be set = to the contents it contained
        i. amino_acid = {“W” : “Tryptophan}
6. The parameters that were passed into the function in the code were text files of the animals amino acids genome that were previously download as text files.
    a. The text files needed to be modified, so no spaces and numbers were in the file as they would count as characters and the code would count any characters within the file, therefore affecting the result of the data collected. (Insert screenshot here)
    b. The code essentially finds the longest common subsequence between the two text files, from this data, the length of the characters that the code outputs, is the longest common subsequence, and therefore shows how long the common sequence is between two animals genome
    c. Each species’ sequence was converted into a list of amino acids
7. Each amino acid sequence was compared to the amino acid sequence for homo sapiens through an implementation of the longest common subsequence (LCS) algorithm
    a. The LCS algorithm constructs a (n x m) matrix containing the two amino acid sequences under comparison. Valuations are given at each position in the matrix by comparing the two elements based on if the elements are the same and the valuations of the previous vertical position and previous horizontal position (Figure 1.2).

b. Two pointers traverse horizontally, vertically, and diagonally along the matrix determining to get the elements of the longest common subsequence in reverse.

   i. Diagonally - If the two elements match, decrement both the x pointer and the y pointer by 1.

   ii. Horizontally - If the valuation of the previous element in the x direction is greater than the valuation of the previous element in the y direction, decrement the x pointer by 1 (moving closer to 0).

   iii. Vertically - If the valuation of the previous element in the y direction is greater than the valuation of the previous element in the x direction, decrement the y pointer by 1 (moving closer to 0).

   iv. After one of the pointers (or both) has reached 0, all of the elements of the longest common subsequence would have been retrieved.

$$c[i,j] = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0, \\ c[i-1, j-1] + 1 & \text{if } i, j > 0 \text{ and } x_i = y_j, \\ \max(c[i, j-1], c[i-1, j]) & \text{if } i, j > 0 \text{ and } x_i \neq y_j. \end{cases}$$

Figure 1.2: Rule for creating the longest common subsequence (LCS) algorithm using a (n x m) matrix

c. The output would contain the elements and the length of the longest common subsequence.

   i. For example, the longest subsequence in "**buchannon**p" and

   ii. "**b**p**u**b**channon**g," would be "**buchannon**" The longest common subsequence does not necessarily have to be a consecutive subsequence of letters as long as it can be read in order

Reading the Data:

1. Similar to the procedure above, a function was utilized to determine the differences (insertions/deletions) between two different amino acid sequences
    a. The modifications were shown as "-" for one sequence and "+" for the other sequence indicating a point mutation for the amino acid. The program will also include which amino acids are different within the sequence.
    b. In case the of a frameshift mutation, the LCS function will handle the mispositioned amino acids by considering them with respect to their relative position, not absolute position
2. For calculating percent difference between the two amino acid sequences, the following formula was used:

    (amino acids in common sequence) / (amino acids in human sequence) X 100%

    a. The formula yields a result that is always less than or equal to the amino acids in a human sequence.
    b. These percentage of similarities were mapped onto the phylogenetic tree, along with locations of the differences between the human sequence and other species will also be shown for each species.

Creating a Phylogenetic Tree for FOXP2 Gene Using the Unweighted Pair-Group Method with Arithmetic Mean (UPGMA):

1. The differences in each amino acid was found between each animal
    a. 12 Choose 2, 66 different results were collected

2. After the data was collected, the distance method was utilized to determine how close certain animals were to each other.
    a. The sequences of all the species were aligned
    b. All the sequences were compared pairwise to get the number of differences
        i. The differences were recorded in a 12x12 table
        ii. Comparisons in the vertical direction have already been performed were ignored
    c. Finding the smallest weighted difference of all differences, the two associated species can be grouped together
        i. For example, in Figure 1.3 the differences between all the species were found, to determine the pairings of species.

Differences between sequences

	A	B	C	D	E	F
A		9	2	4	9	10
B			9	6	2	10
C				5	9	10
D					6	10
E						10
F						

Figure 1.3

Since Species A and C and species B and E have the smallest weighted difference, they were selected to be grouped together. The distance method is then applied with these groupings.

    d. The weighted difference was found between the species in the grouping (namly A-C and B-D) and each other species that have not been grouped yet

Differences between sequences

	A / C	B / E	D	F
A / C		9	4.5	10
B / E			6	10
D				10
F				

Figure 1.4

The averages have been found between the groupings (A-C and B-D) and the species being compared, this will allow the viewer to see how the species that have not been grouped together relates with species that have been grouped together.

e. Eventually, by repeating this process, a phylogenetic tree based unweighted difference was generated

Data

Table 2 - An analysis comparing the time of divergence, percent similarity, and specific deviances between species when compared with humans.

Species	**Time of divergence (from humans)**	**% of FOXP2 similarity (to humans)**	**% of FOXP2 similarity (to zebrafish)**	**Types and location of differences (from humans)**
Humans	0 years ago	100%	81.79%	N/A
Common Chimpanzees	5-6 million years ago	99.7%	81.98%	- Position 223, histidine to glutamine - Position 303, asparagine to threonine - Position 325,

				Serine to asparagine
Bonobo	6.5 million years ago	99.3%	81.84%	- Position 154, addition of Glutamine - Position 201, addition of glutamine - Position 223, Histidine to Glutamine - Position 303, threonine to asparagine - Position 325, serine to asparagine
Gorilla	7-8 million years ago	99.4%	82.18%	- Position 163, Deletion of Glutamine - Position 164, Deletion of Glutamine - Position 198, Insertion of Glutamine - Position 223, Histidine to Glutamine - Position 303, asparagine to threonine - Position 324, serine to asparagine
Orangutan	8.5 million years ago	99.5%	82.04%	- Position 6, A to V - Position 153, deletion of Q - Position 374, S to N - Position 223, H

				to Q - Position 303, N to T
Gibbon	17 million years ago	99.4%	82.18%	- Position 153, deletion of Q - Position 223, H to Q - Position 303, N to T
House Mouse	80 million years ago	99.2%	81.93%	- Position 80, d to e - Position 223 h to q - Position 303 n to t -Position 325
Dog	90-100 million years ago	95%	79.86%	Greater than 10 amino acid changes
Hummingbird	150 million years ago	5.4%	5.9%	Greater than 10 amino acid changes
American Alligator	275-310 million years ago	5.4%	5.9%	Greater than 10 amino acid changes
Western Clawed Frog	340 million years ago	95%	77.51%	Greater than 10 amino acid changes
Zebrafish	410 million years ago	82%	100%	Greater than 10 amino acid changes

Table 2 - The number of amino acid differences between the all permutations of selected species. This table aided in the construction of our second phylogenetic tree below using the distance method.

	Gibbon	Gorilla	Human	House Mouse	Bonobo	Chimp	Orangutan	Zebra fish	Humming bird	American Alligator	Western Clawed Frog	Dog
Gibbon	0	1	4	2	3	2	1	127	699	699	60	32
Gorilla		0	5	3	3	3	2	127	699	699	61	32
Human			0	4	5	4	5	139	700	700	61	33
House Mouse				0	3	0	3	129	699	699	59	33
Bonobo					0	3	4	130	699	699	60	32
Chimp						0	3	129	699	699	59	33
Orangutan							0	128	699	699	61	33
Zebrafish								0	407	407	166	147
Hummingbird									0	0	676	386
American Alligator										0	676	386
Western-Clawed Frog											0	67
Dog												0

Figure 2.1 - Comparison of selected species' evolutionary track generated based on BLAST data viewer.

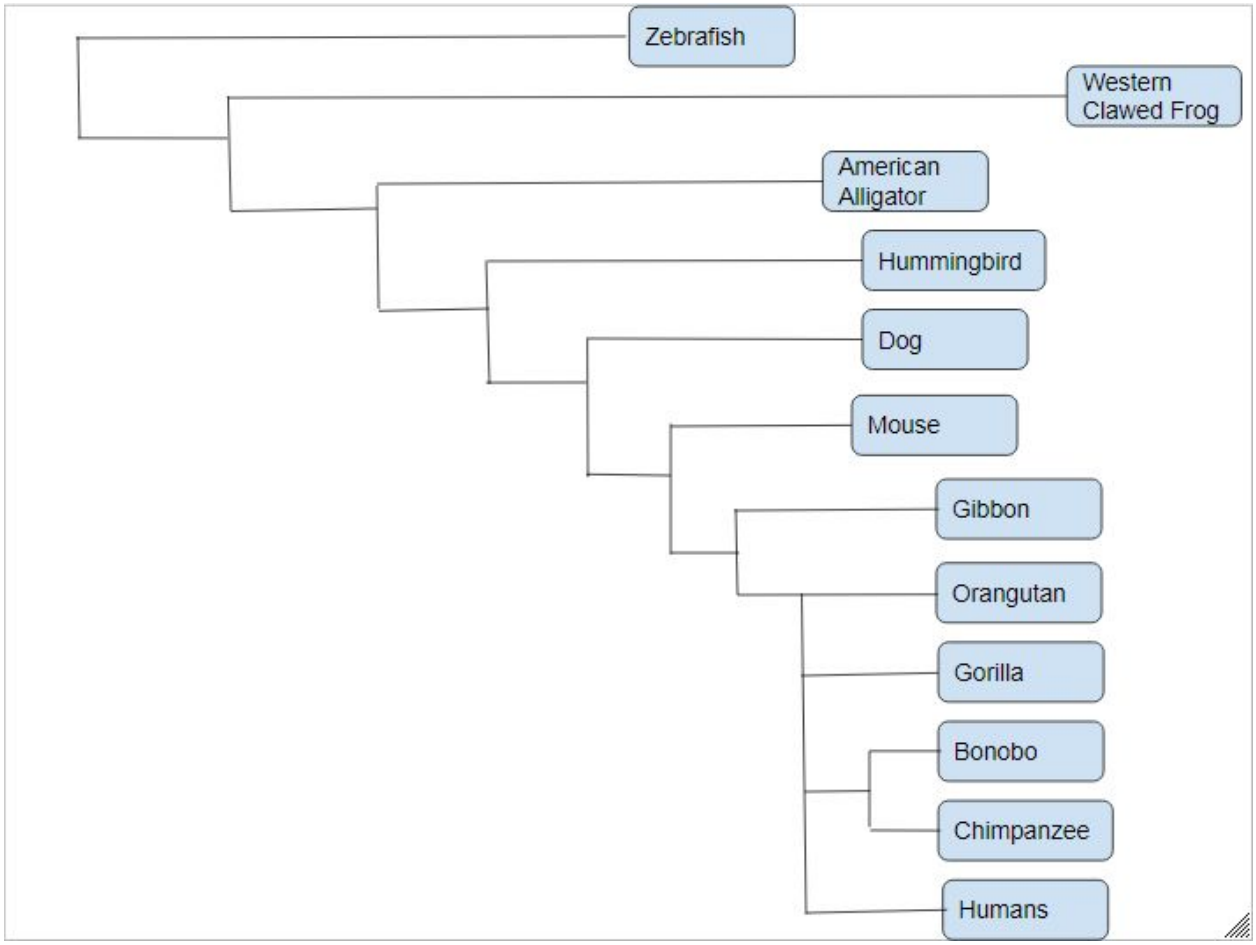


Figure 2.2 - The evolutionary track of FOXP2 between the selected species made using the distance method. The top percentage shows the percent similarity of each species as compared to humans and the bottom number shows a percent similarity of each species as compared to zebrafish.

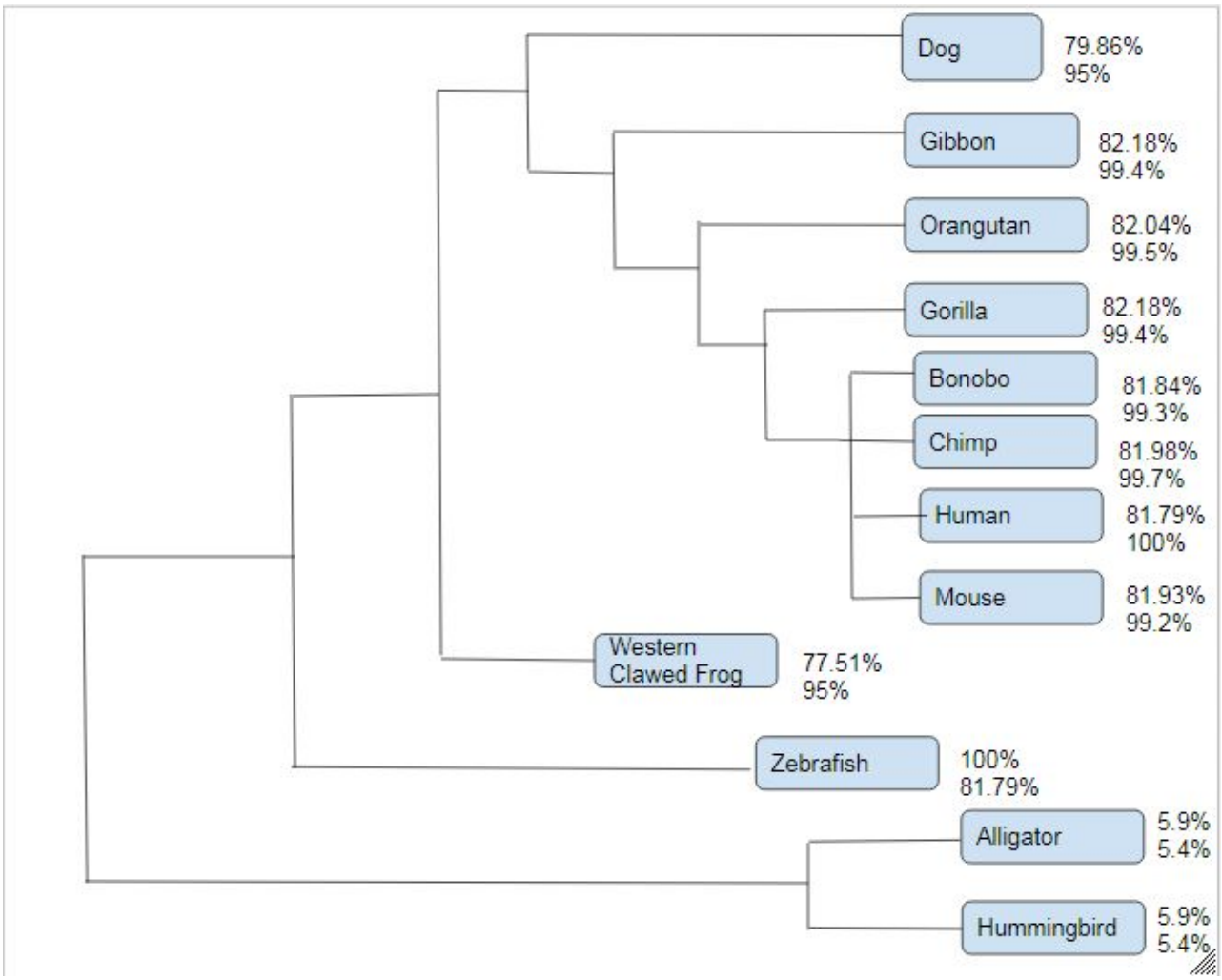
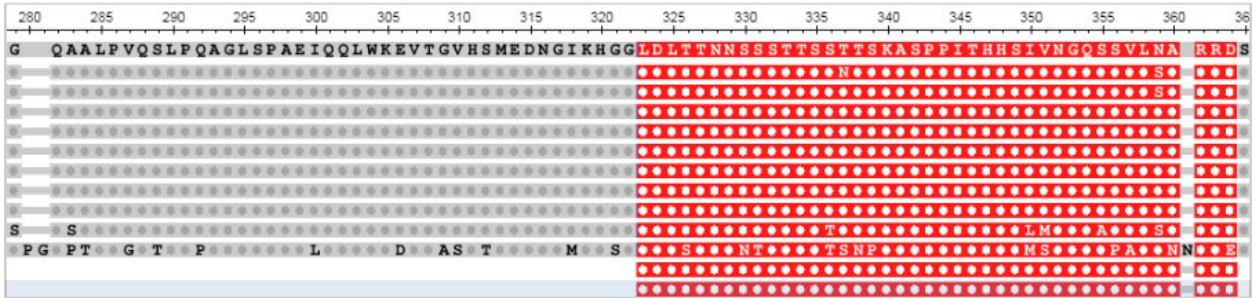


Figure 2.3 - A portion of all 12 species' FOXP2 genes as compared with humans which make up the first row. A blank space indicates that the amino acid is the same as in humans. A Shading of read indicates that all species are the same.



Data Analysis

All the primates had the closest similarity in their FOXP2 gene to the human FOXP2 as originally expected because they are two species that digressed the closest in relative years when compared to other species. Species further away in divergence than humans had less similarity in their FOXP2 gene. For instance, American alligator (diverged ~275 million years ago) and hummingbirds (diverged ~150 million years ago) showed a miniscule similarity of 5.4%. An outlier was the Western Clawed Frog (diverged ~340 million years ago), while it had little to no similarities with humans, it still had 95% similarity with the human FOXP2 gene. In almost all comparisons, a trend was observed where histidine was present in humans 303 while the other species had glutamine in those areas. Another outlier was the house mouse, containing over 99% of amino acids within the human FOXP2 genome, even though humans and mouse digressed over 80 million years ago. Dogs also showed a relatively large amount of similarities in their

FOXP2 gene to the human FOXP2 gene, even though they digressed from humans around 100 million years ago. There was some sort of point change within the differences between humans and some sort of insertion or deletion that also occurred. There was some correlation between the time of species divergence and how similar the two genomes are of the respective species, with the outliers being the Western-Clawed Frog, Humming Birds and Dogs. It seemed as if there was no amino acid sequence change past 400 and before 100(with some notable exceptions such as the Orangutan and the Common Chimpanzee). It was also observed that American Alligators and Hummingbirds have the same exact sequence.

While comparing other species to zebrafish (due to how genetically similar it is with humans), it was observed that primates all revolved around the 80 percent mark in similarity to zebrafish, while other species showed sub 80 percentage similarity to zebrafish. Primates when being compared to zebrafish all had near the same percentage of around 80 percent. Some of the comparisons to humans, namly Western-Clawed Frog, Dog, Hummingbird, American Alligator and Zebrafish, were too large to list all the specific deletions, point changes or insertions. When looking at the two phylogenetic trees, it is evident that the FOXP2 gene didn't follow the tract of the rest of the genomes closely. For starters, the most distant animal to humans including the entire genome was zebrafish while the largest difference between humans in terms of FOXP2 was obviously the alligator or hummingbird with only a 5% similarity. These two were obvious outliers as they interestingly have the exact same FOXP2 gene both only containing 49 amino acids. The house mouse and humans in a normal phylogenetic tree digressed over 80 million years ago(diverged a long time ago in terms of evolution and change), but when implementing the distance method in accordance to the FOXP2 gene, the house mouse is as close to humans as chimpanzees are. All the other primates are as expected with the distance method however.

Conclusion

Specific modifications within the FOXP2 gene of the human genome could have led to human speech and cognitive development. Some other factors, such as environment, structure of the brain within humans and higher protein intake could have led to speech within humans.

FOXP2 orthologs[4] and their impacts have been fields of interest for a long time. A prominent part of FOXP2 codes for a protein that is mostly active in the left planum temporale, a region of the brain. Posterior to the part of the brain that controls auditory learning, the left planum temporale is controls almost all functions of speech comprehension. The left planum temporale is also a pristine example of brain asymmetry as it is much larger in the left hemisphere as compared to the right. Brain asymmetry is a precondition of human intelligence as many parts of the brain have become more specialized as compared to some of our most common ancestors. However, a surprising amount of other great gapes have brain asymmetry as well.

This may explain why certain chimpanzees have been able to not only use certain tools, but also make them. This missing link in their left planum temporale could explain why they were unable to communicate and work together to make these tools. There are a few exceptions of non-human primates being able to be taught to understand human language through sign-language. These individuals however, still lack the auditory and oral motor skills that humans possess. The fact that non-human primates can be taught to understand human language suggest that they may not be as inferior in terms of intelligence as humans have thought.

[4] Orthologs- the same gene in different organisms.

Recent information shows that neanderthals likely had similar speech capabilities as humans do. This raises the question of why they died out and were inferior in terms of intelligence as compared to humans. It is known that neanderthals' main diet composed of animals from the ice age era. Their shivas were specialized in hunting large animals and were adapted to colder temperatures. As the temperatures changed, they were unable to adapt to the new environment. They didn't evolve fast enough and coevolved with their fellow ice age animals. They relied on them to get food and without them they slowly died of starvation. Furthermore, humans had access to primitive technology such as simple tools and food networks which neanderthals did not.

The FOXP2 gene is essential for speech related activities in all multicellular organism which include, human complex language, echolocation in bats and, songs in birds,
Within these parameters, it has been hypothesized for a long time that this single gene was responsible for the development in human languages. One article writes that FOXP2 has no relevance to the evolution of speech humans, and that this development of complex human language was almost solely responsible for the growth of human civilization since humans were now able to cooperate and use tools. This gene also heavily aids in synaptic plasticity which is the ability to think and remember experience with the ability to repeat those same experiences. The group sought out to see the exact differences that gave humans such a big advantage in terms of speech and ability to remember.

The hypothesis that the FOXP2 genome divergence would be similar to the general phylogenetic that has been constructed to explain the divergence of other species, was partially proven. In the data collected, there were strong similarities between the human FOXP2 gene and all the other primates. As expected, there was a correlation between the time of divergence from a species to humans and the similarity of the FOXP2 gene between a certain species and humans, but as mentioned previously, there were some outliers so a conclusion based on divergence cannot fully be derived. Primates and humans

shared around 99.2-99.7% of their FOXP2 gene, leading to a predisposed assumption that humans and primates share the FOXP2 gene in close similarity.

As this amino acid chain was also contained in humans with the exception of amino acid 325 that replaced “n” with “s.” This was one of the few differences that only humans had along with the difference at 303 changing an “h” with a “q.” However, mice and humans shared over 99% of their FOXP2 with each other, so a conclusion founded on the idea of divergence of FOXP2 gene cannot be fully derived due to the mouse being a notable outlier in this experiment. An inference can be drawn that the FOXP2 gene has been passed down for many generations with little no mutations occuring this time period, however this is also falsified based on the fact that birds and alligators have little to no similarities to the FOXP2 genome that the primates in general possess.

In relevance, the research into FOXP2 could be utilized to help with other speech impaired diseases and other developmental diseases. The FOXP2 gene could also help explain and support the Theory of Evolution since humans are able to communicate while other species are speech impaired or are only restricted to basic verbal expressions. Further research into the FOXP2 gene could be utilized to understand why certain individual inherit speech and developmental disorders rather than an error occuring mitosis or a chromosomal mutation. FOXP2 is also involved in the development of the central neural system and extended research into the area could explain brain-related diseases such as Alzheimer's.

Some errors within the experiment was using amino acids instead of nucleotides, this could result in multiple genomes being created because a more than one codon can be attributed to one amino acid, leading to multiple variations in the FOXP2 genome of the species being compared. A logistical error in the coding may have occurred which could have yielded inaccurate results, such as the number of differences between two genomes and the location of the difference between the two species. Another

area of the experiment that involved coding was phylogenetic trees. When coding the distance method mentioned in the procedure, a computational error could have occurred, which may show an error in the digression of the FOXP2 gene. While parsing the sequences, an error could have occured while modifying the text files to make it compatible for the code, which may have created a syntax error or the wrong animals may have been compared which could yield inaccurate data collection. Another error that may have occurred is pinpointing the exact location of the change between the two amino acid sequences, due to multiple letters of repetition. Utilizing the distance method may have been proven to be ineffective because it takes into consideration the number of difference between two genomes and groups them based on the relative similarity, rather than evolutionary divergence and homologous structures that were passed down.

Citations

Gibbons, John. “Why Did Neanderthals Go Extinct?” Smithsonian Insider, 21 Mar. 2016, insider.si.edu/2015/08/why-did-neanderthals-go-extinct/.

Lakatos, L, and Z Janka. “[Evolution of Human Brain and Intelligence].” Ideggyogyaszati Szemle., U.S. National Library of Medicine, 30 July 2008, https://news.nationalgeographic.com/news/2001/10/1004_TVlanguagegene.html

Staes, Nicky, et al. “FOXP2 Variation in Great Ape Populations Offers Insight into the Evolution of Communication Skills.” Scientific Reports, Nature Publishing Group UK, 2017, www.ncbi.nlm.nih.gov/pmc/articles/PMC5715162/.

Binder, Jeffrey R. “The Wernicke Area: Modern Evidence and a Reinterpretation.” Neurology, Lippincott Williams & Wilkins, 15 Dec. 2015, www.ncbi.nlm.nih.gov/pmc/articles/PMC4691684/.

Spinelli, Edorado G. “KE Family.” KE Family - an Overview | ScienceDirect Topics, Science Direct, www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/ke-family.

http://www.ai.mit.edu/projects/dm/foxp2.pdf - A research project conducted by MIT on foxp2
Staes, Nicky, et al. “FOXP2 Variation in Great Ape Populations Offers Insight into the Evolution of Communication Skills.” Scientific Reports, Nature Publishing Group UK, 2017, www.ncbi.nlm.nih.gov/pmc/articles/PMC5715162/.

Shah, R, et al. “Expression of FoxP2 during Zebrafish Development and in the Adult Brain.”The International Journal of Developmental Biology., U.S. National Library of Medicine, 31 Oct. 2006, www.ncbi.nlm.nih.gov/pubmed/16525940.

harvey-girard, Erik. Research Gate, www.researchgate.net/figure/221723021_fig7_Figure-2-Phylogenetic-comparison-of-apteronotid-FoxP2-sequence-A-Sequence-comparison.

Wilson , Clare. “Human ‘Language Gene’ Makes Mice Smarter.” New Scientist, 15 Sept. 2014, www.newscientist.com/article/dn26216-human-language-gene-makes-mice-smarter/.

Marcus, Gary F., and Simon E. Fisher. “FOXP2 In Focus: What Can Genes Tell Us about Speech and Language?” MIT, 3 Feb. 2003, www.ai.mit.edu/projects/dm/foxp2.pdf+.

Wade, Nicholas. “Speech Gene Shows Its Bossy Nature.” The New York Times, The New York Times, 11 Nov. 2009, www.nytimes.com/2009/11/12/science/12gene.html.

Huang, Shiyong, et al. “Associative Hebbian Synaptic Plasticity in Primate Visual Cortex.” The Journal of Neuroscience, Society for Neuroscience, 28 May 2014, www.ncbi.nlm.nih.gov/pmc/articles/PMC4035519/.

Schreiweisab1, Christiane, et al. “Christiane Schreiweis.” *Proceedings of the National Academy of Sciences*, National Acad Sciences, 29 Apr. 2014, www.pnas.org/content/111/39/14253.abstract.

European Bioinformatics InstituteProtein Information ResourceSIB Swiss Institute of Bioinformatics. "European Bioinformatics Institute." UniProt, 4 Feb. 2014, www.uniprot.org/uniprot/?query=Foxp2&sort=score.

Creating a phylogenetic tree https://www.youtube.com/watch?v=09eD4A_HxVQ&t=308s