EMATM0004 Final Project
Is snake closer to lizard or eel?

Jeonghyun Kim
University of Bristol
Student ID: 1871686
Candidate ID: 97485
cc18316@bristol.ac.uk

ABSTRACT

Snakes seem to have not legs and move like a marine organism like eel. However, there are some arguments that snakes originally have their legs. They argues that ancestor of snakes' fossil has legs, so some researches suggest that they lose their legs during the evolution. To find out evidences of which animals, lizard and eel, are closer to snake we will use some methods and compute distance between those animals. For comparison of each animals, this paper selects cytochrome b (CYTB) and cytochrome c oxidase subunit I (COX1) sequences as target proteins. The result shows lizard is closer to snake. However, our method of calculating distance and designing tree is not a way for increasing accuracy, because those are simple and fast. To handle the issue, we refer some credential articles whether our result is right or not. In conclusion, lizard is closer than eel proved by our experiments and comparison with credential articles.

1 INTRODUCTION

Ancestor of snake has legs according to their fossil. It can assume that lizards and snakes are closed to each other because they are reptile and have legs. In truth, lizard and snakes are classified by reptile, not a fish. However, there is one other animal who have similar features with snake, eel. Interestingly, one of eel has almost same name called snake eel. Their action of movement is similar, so Caldwell and Lee in 1997 [2] asserted that snakes are from sea to land with legs. Here is a research question of this research. Is snake closer to lizard or eel?

In this paper, we will try to find evidence that snakes are closer to lizard or eel. To prove that which animals are closer, this research will compare some of proteins such as CYTB and COX1 from the three animals. Then, we can generate distance based on sequence of the protein with both cases of amino acid and nucleotide sequences. As a result, this paper have main hypothesis.

* Lizard will be closer to snake as they are reptile and have legs.

If this hypothesis not right, then eel is closer than lizard. It is totally related to answer of the research question.

The results from this study could be one evidence that snakes are evolved from sea to land, or snakes originally lived on land with legs.

2 METHODOLOGY AND RESULT

To compare the three animals precisely, methodology is divided by 8 tasks to do step by step.

* **Task1**: Investigate snakes detailed with statistical method.
* **Task2**: Find 5 closest animals through comparison of CYTB sequence.
* **Task3**: Generate a rooted phylogenetic tree.
* **Task4**: Add lizard and eel, and calculate distance of their sequences.
* **Task5**: Compare trees with CYTB and CYTC cases, and further with amino acid and nucleotide sequences.
* **Task6**: Compare relationship of trees with result of Taxbrowser.
* **Task7**: Investigate with result of multiple alignments.
* **Task8**: review with precious articles related to evolution of snakes to get final result.

Task1

In this task, python regious called ball python is a representative of a snake.

Taxonomy ID: 51751
Scientific name: Python regius (Shaw, 1802)
syntype of Python regius: BMNH:IV:3. 3a
homotypic synonym: Boa regia
homotypic synonym: Shireenhoserus regius
NCBI BLAST® name: snakes
Rank: species
Genetic code: Translation table 1 (Standard)
Mitochondrial genetic code: Translation table 2 (Vertebrate Mitochondrial)
Other names:
    common name: ball python
    common name: royal python
Taxonomy name: Serpentes (Snake)
Accession number: NC_007399

Protein code:
CYTB:
translation="MPHHYILTLFGLLPVATNISTWWNFGSMLLTC
LMLQVLTGFFLAVHYTANINLAFSSIIHITRDVPYGWLMQNL
HAIGASMFFICIYIHIARGLYYGSHLNKETWVSGITLLITLMA
TAFFGYVLPWGQMSFWAATVITNLLTAVPYLGATMTTWL
WGGFAINDPTLTRFFALHFILPFAIISLSSLHIILLHEEGSSNPL
GTNPDIDKIPFHPYHSYKDLLLLTLMLLTLMITVSFFPDIFND
PDNFSKANPLITPQHIKPEWYFLFAYGILRSIPNKLGGALALV
MSIMILLTAPLTHTAHLRPMTFRPLSQLMFWTLISTFITITWA
ATKPVEPPYIIISQATATLYFTFFISTPILGWIENKMMNS"

Cytochrome c oxidase subunit II:
translation="MPYATQLSLQEGTGPAMEEVVFLHDHVLLLTFL
MSLVILLFATTAITATVTHNDPTEEVEQLEAAWTAAPIMILIL
TALPSVRSLYLMEEVFDPYVTIKATGHQWYWNYEYTDGTN
VSFDSYMIQTQDLPNGAPRLLEVDHRMVMPANLQTRIVVT
AEDVLHSWALPSLGIKVDAVPGRLNQLPLATSRTGVFFGQC
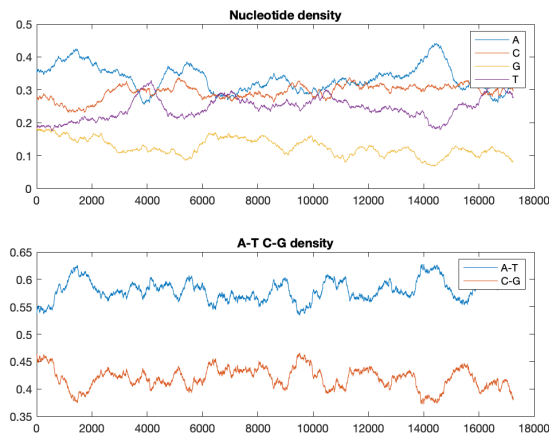SEICGANHSFMPIVVEAVPLTYFEQWLLTTKQ"

Basic Statistics:

Figure 1: Density of A-T and G-C Graph

Task2

Table 1: Comparison of identities (CYTB)

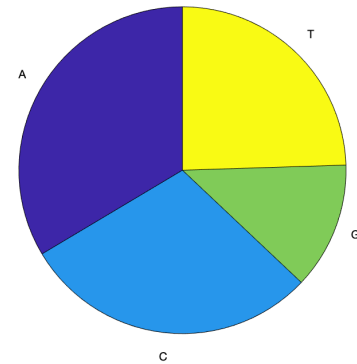| Species	Identities	Accession
Python regius	100%	YP_313706.1
Python anchietae	94.18%	AIL52626.1
Python bivittatus	92.16%	AGL12235.1
Xenopeltis unicolor	89.07%	AAM83132.1
Cylindrophis ruffus	85.68%	AAL83566.1

Figure 2: Pie chart of distribution of nucleotides

Table 1 is a top 5 results of Python regius' CYTB protein from BLAST. Python regius, Ball python, is one of famous snakes. As the table shows, closest three are same species called 'Python' and they have similar identities more than 90%.

Interestingly, Xenopeltis unicolor, sunbeam snake, and Cylindrophis ruffus named 'Red tail pipe snake' have high level of identities although it is not python and live different habitats, because they are all snakes. This result is one of evidences that if their species same or similar, then identities value will be high.
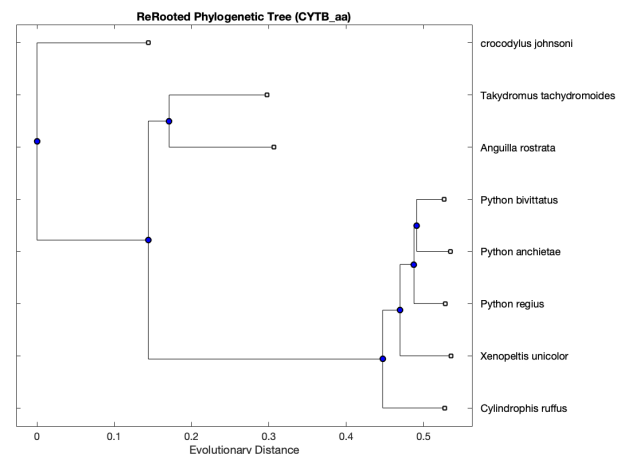
Task3

Figure 3: Rooted Phylogenetic Tree

Figure 3 is a phylogenetic tree from Task2 extracted by BLAST. Furthermore, the tree shows two more animals to

compare with snakes: Takydromus tachydromoides named 'Japanese grass lizard' and Anguilla rostrata named 'American eel'. As an outgoup for this study, crocodylus johnsoni called Freshwater crocodile is chosen because this animal is reptile but has different features. We assume that the Freshwater crocodile is a descent outgroup animal between snakes, lizard and eel. Those added animals are related to the research question to compare each other. Then, re-root the tree with crocodylus johnsoni root.

Through the phylogenetic tree, result of Task2 can be proved. X axis of the tree is distance, which is calculated by sequence of amino acid, and it presents that all three pythons has closer relationship than others. Other snakes are little bit far away from python, but two of added species have even more gap from snakes.

Above explanation is based on computing results. We can also think about evolution of the species. Snakes have different name and features, but they have similar sequence of a protein than added animals such as the lizard and the eel. It means that snakes evolve from same kinds of ancestor. Furthermore, lizard, eel and snakes might have one kind of ancestors as the tree shows roots.

Task4

Table 2: Distance Matrix (CYTB)

	P.Reg	Cylin	Eel	Lizard	Croc
P. Reg	0	0.1607	0.5395	0.5496	0.6619
Cylin	0.1607	0	0.5685	0.5292	0.6644
Eel	0.5395	0.5685	0	0.2620	0.4417
Lizard	0.5496	0.5292	0.2620	0	0.4500
Croc	0.6619	0.6644	0.4417	0.4500	0

Table 3: Distance Matrix (CYTC)

	P. Reg	Cylin	Eel	Lizard	Croc
P.. Reg	0	0.0782	0.3146	0.2863	0.3337
Cylin	0.0782	0	0.3088	0.2714	0.3112
Eel	0.3146	0.3088	0	0.1444	0.1940
Lizard	0.2863	0.2714	0.1444	0	0.1801
Croc	0.3337	0.3112	0.1940	0.1801	0

In order to check an species who closed to a snake, we added two animals to distance matrix: Japanese grass lizard and American eel. In addition, to see concise results, other pythons without ball python and xenopeltis unicolor will be removed from the table.

If a snake is closer to a lizard, which means that they are evolved from closer ancestor than eel, distance between a snake and a lizard should be closer than eel.

The method of calculating distance is Jukes-Cantor distance in this research. CYTB and COX1 proteins are used for computing distance in amino acid level.

Table 1, 2 are the results based on above method. Two tables show distance of each pair. From the result, CYTB has big differences than CYTC. It means CYTB can be easier to recognise gaps. However, distance of CYTB and CYTC are different on some points. For instance, python is closer to eel than lizard in case of CYTB, but the case of CYTC is contrast. It means that it is hard to decide which one is closer to snakes thought distance of two proteins only.
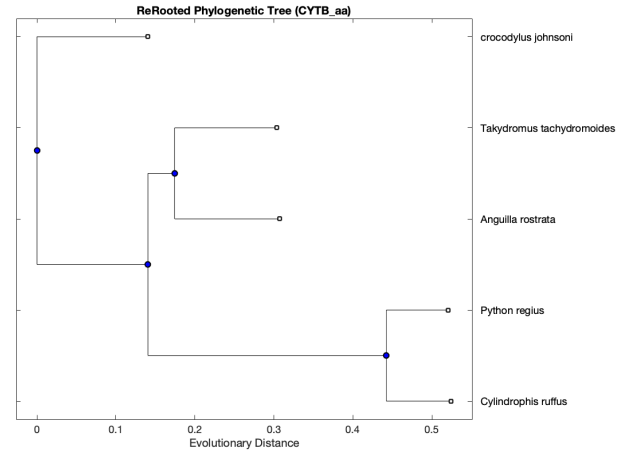
Task5

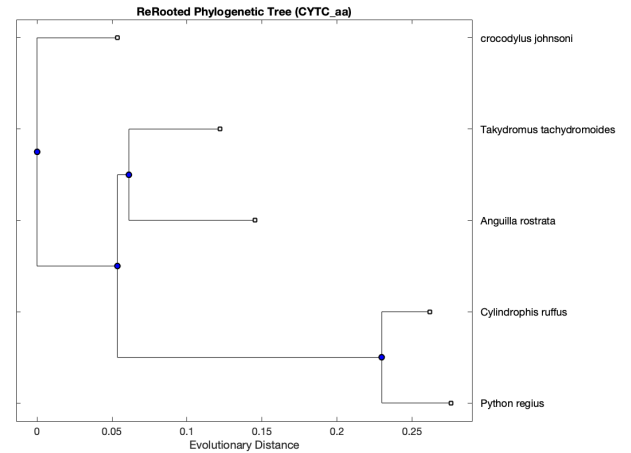**Figure 4: Rooted Phylogenetic Tree (CYTB_aa)**
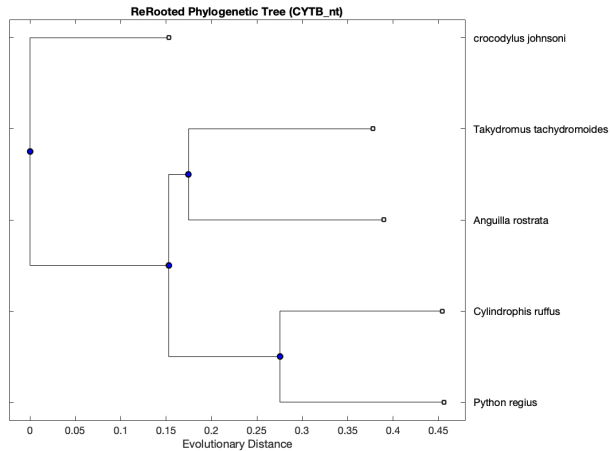

Figure 5: Rooted Phylogenetic Tree (CYTC_aa)
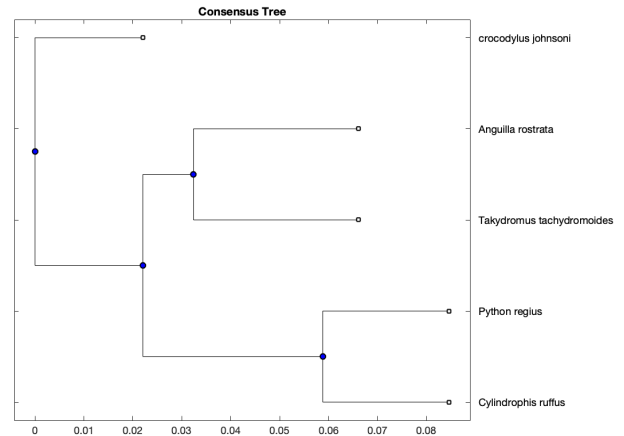
Figure 6: Rooted Phylogenetic Tree (CYTB_nt)

Figure 8: Consensus Tree (Average)

In the result of consensus tree, Figure 8, eel and lizard has similar distance against a snake, but the result of distance shows lizard is slightly closer than eel.
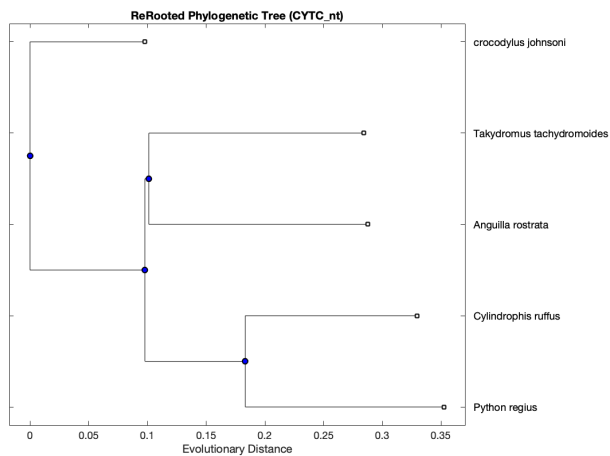
Task6

Figure 9: Result of Taxbrowser

Figure 7: Rooted Phylogenetic Tree (CYTC_nt)

Figure 4,5,6,7 are rooted phylogenetic trees. Before comparing it detail, here is brief explanation about the tree. A node is a branch that represents an individual or gene in a tree, and connects the gene to the gene. In relation to the tree, we call the node at the bottom of the tree as leaf. Ancestor, which traces back to the top of a tree, is called root.

In those results, we can find that amino acid case has lower distance value, while nucleotide case has larger distance. It is because that one character of amino acid is changed into three character of nucleotide. It means that amino acid case has larger distance than nucleotide.

If the partition distance indicates how the two trees are different, the consensus tree is a tree representing common points in several trees.

Result from Taxbrowser presents lizard is closer than eel because they have the closest organism named Squamata. the result is similar to our result, which is lizard's distance of sequence is closer than eel's one in average. Overall tree is similar, but one different point is that our node of tree shows Lizard and eel has same node directly while Taxbrowser present lizard and snake are same node.
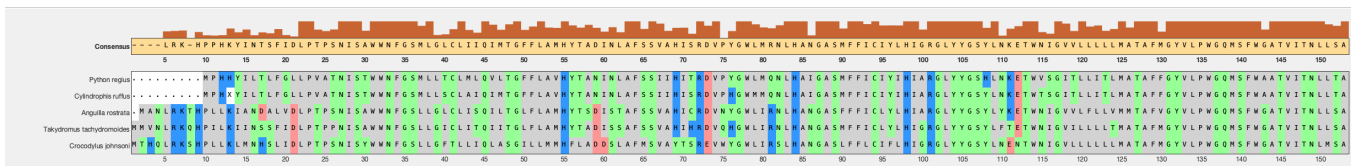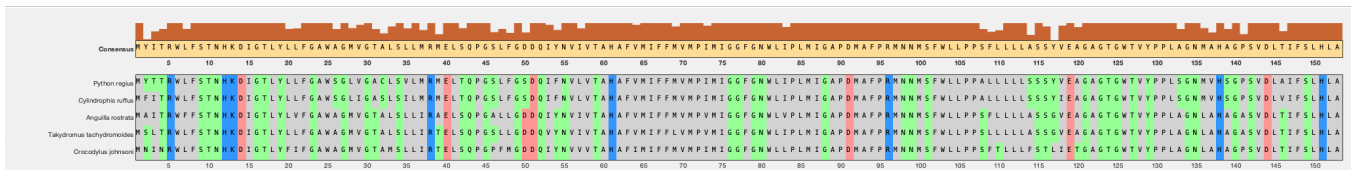
Figure 10: Multi Aligned with CYTB

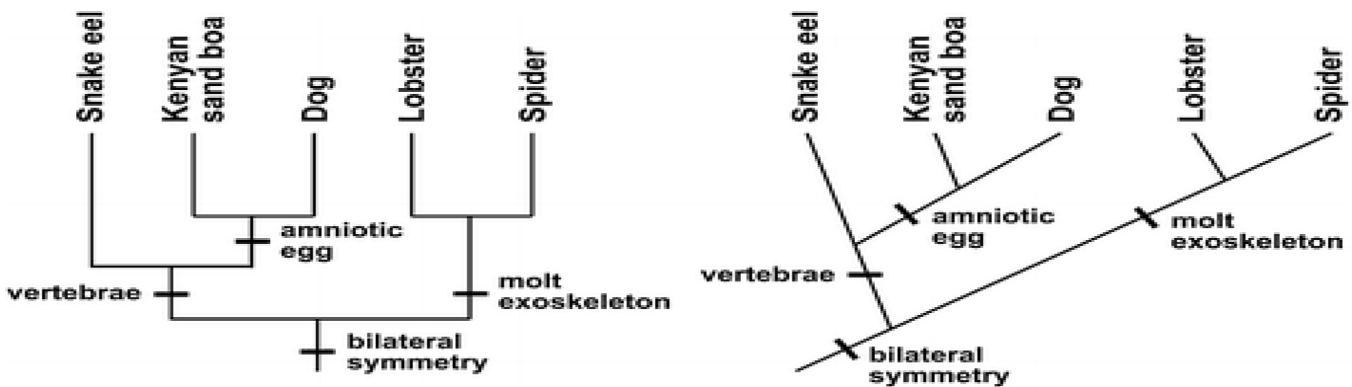**Figure 11: Multi Aligned with CYTC**

Figure 12: Tree and ladder cladograms adapted from [3]

Here are our thoughts why it is different. We only compared CYTB and COX1 proteins, and it is not enough to decide which animal is closer to snake. Furthermore, Jukes-Cantor distance is fast, but its accuracy is lower than other algorithms. Those reasons may influence to the difference.

Task7

Figure 10, 11 are multiple alignment results on amino acid level. There are two types of alignments: CYTB and CYTC (COX1). In the figures, there are bar chart on the top, and it means that amino acid is different at each regions. Alignment of CYTB shows slopes of each bars are imbalanced while case of COX1 is stable. It means that COX is harder to find differences than CYTB. As shown in task 4, CYTB is more clear to recognise differences of distance. In this case, it shows same as the results.

Task8

From the above results, hypothesis in this paper is accepted through distance of sequences and Taxbrowser. It means that

lizard is closer to snake than eel. However, those approaches are a simple way to find closest animal. Therefore, it needs to compare with related articles which are trustworthy. To prove our hypothesis is correct, we mention one article found from Pubmed that can support the hypothesis. According to Bonnet et al. in 2017 [1], oviparity have evolved into viviparity and it makes progenies to compete each other for access to water reserves. It is one history about snakes' evolution, but we can find one of lizards and python regius in oviparity in their explanation. They employ ESF as a method to find out relationship between target animals. In here they refer to many articles for their accuracy. It shows that lizard is really closed to snakes.

Furthermore, there are one more article that can support the hypothesis. Even though Catley et al. in 2013 [3] suggest educational ways to improve understanding of cladogram tree, its examples are related to the hypothesis. Figure 12 shows snake-eel and kenyan sand boa, which is one type of snakes, has same organism with vertebrae. Vertebrae is

5

high level of organism, so we can notice that eel and snake distance is far than lizard.

3 CONCLUSION

In methodology and result section, we found that lizard is closer to snake than eel, which is acceptable hypothesis, through comparing distances of some protein's sequences and Taxbrowser. It is almost correct because there are some supporting articles from credential sites. Although the way we did in this study is not exactly correct, but it is descent method if a purpose is finding relationships since it is fast and not required deep knowledge of bioinformatics. It means that our method has some weakness in accuracy. Furthermore, it has more limitations. We only use two proteins to compare each other, but it is not always correct. This topic needs further study with better algorithm to compute distance and method of designing trees.

REFERENCES

[1] Xavier Bonnet, Guy Naulleau, and Richard Shine. 2017. The evolutionary economics of embryonic-sac fluids in squamate reptiles. *The American Naturalist* 189, 3 (2017), 333–344.

[2] Michael W Caldwell and Michael SY Lee. 1997. A snake with legs from the marine Cretaceous of the Middle East. *Nature* 386, 6626 (1997), 705.

[3] Kefyn M Catley, Brenda C Phillips, and Laura R Novick. 2013. Snakes and eels and dogs! Oh, my! Evaluating high school students' tree-thinking skills: An entry point to understanding evolution. *Research in Science Education* 43, 6 (2013), 2327–2348.