3.0 Chapter 3 Methods and Materials

3.1 Collection of Data

I obtained the sequences from NCBI database genome.Alignment-based and alignment-free methods applied to mitochondrial of 5 primates which are *Homo sapiens* (human, NC_012920.1), *Pan troglodytes* (chimp, NC_001643.1), *Gorilla gorilla* (gorilla, NC_001645.1), *Pongo pygmaeus* (orangutan, D38115.1), *Hylobates lar* (gibbon, NC_002082.1) and three species of bird family, *Dromaius novaehollandiae* (emu, NC_002784.1), *Gallus gallus* (chicken, AY235570.1), *Struthio camelus* (ostrich, NC_002785.1) and also *Terebratulina retusa* (brachiopoda, NC_000941.1). The complete mitochondria genes were saved in fasta format.

3.2 Software and Tools

This project used Linux as the operating system environment. I used PyCogent (Knight, 2007) package in Python scripts to run ClustalW and Muscle. For working with T-Coffee, I used the T-Coffee libraries in Linux. The D2 and D2z implementations were obtained from http://veda.cs.uiuc.edu/d2z. I used the Kr method via a webserver because the standalone package given by author was not able to work properly and also MplusD software was an executable file in terminal. The results in alignment-based methods were saved in .aln format which was later converted to phylogeny tree by R. This was also done for current alignment-free methods.

I used Neighbor-Joining (Saitou & Nei, 1987) tree for alignment-based methods in which two sequences that have the smallest evolutionary distance are joined (that is, their most recent common ancestor is added to the tree), and then the third sequence or a set of sequences, which are closest to them, are added to the tree, and so on.

Hardware specifications used in this project was an Intel® Core™ i5-2450M CPU with 2.50 GHz clock speed, 4 GB installed memory (RAM) and 64 bit operating system.

3.3 Genome Comparison

One of references that can assist in verifying the correctness of results in this project is full mitochondrial genome comparison. I used the Organellar Genome Retrieval (OGRe), (Jameson *et al*, 2003) website to compare the nine species of all mitochondrial genomes in this research. Within the primate and bird taxa, gene order is conserved.
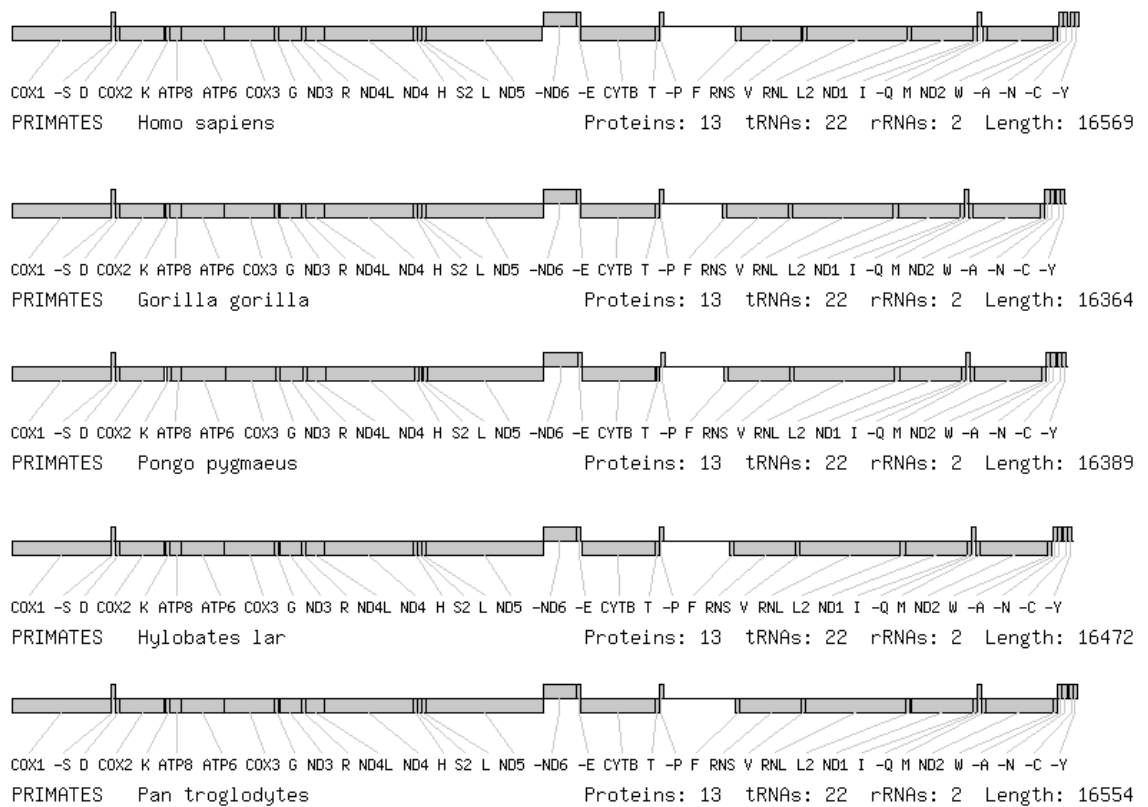

Figure 2.6. Mitochondrial genome comparison among primates. In this figure it is noticeable that the length of sequences in five species are different but there is not break point among them.
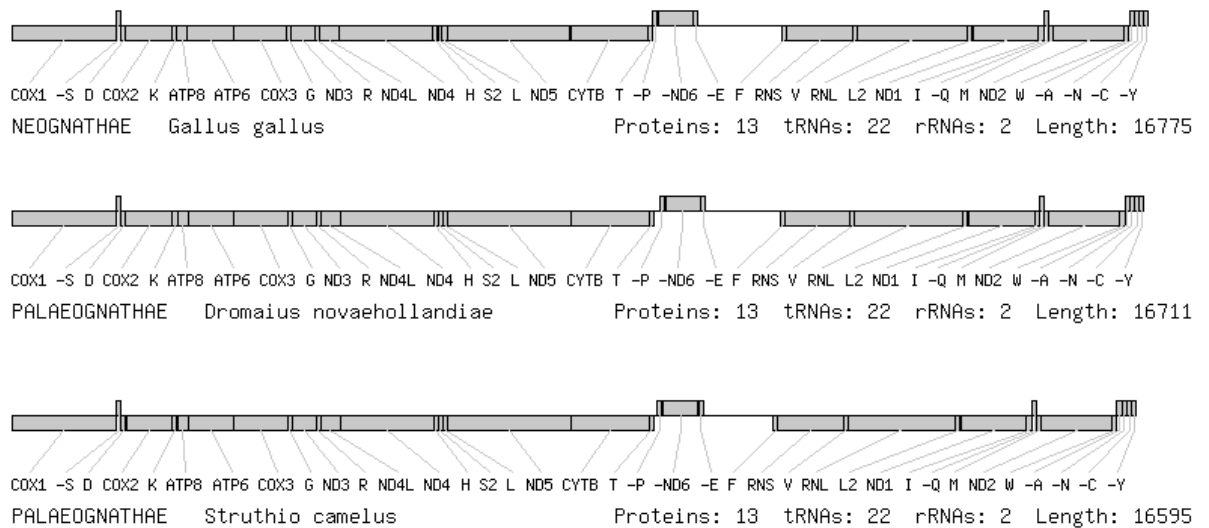
COX1 -S D COX2 K ATP8 ATP6 COX3 G ND3 R ND4L ND4 H S2 L ND5 CYTB T -P -ND6 -E F RNS V RNL L2 ND1 I -Q M ND2 W -A -N -C -Y

NEOGNATHAE   Gallus gallus   Proteins: 13   tRNAs: 22   rRNAs: 2   Length: 16775

COX1 -S D COX2 K ATP8 ATP6 COX3 G ND3 R ND4L ND4 H S2 L ND5 CYTB T -P -ND6 -E F RNS V RNL L2 ND1 I -Q M ND2 W -A -N -C -Y

PALAEOGNATHAE   Dromaius novaehollandiae   Proteins: 13   tRNAs: 22   rRNAs: 2   Length: 16711

COX1 -S D COX2 K ATP8 ATP6 COX3 G ND3 R ND4L ND4 H S2 L ND5 CYTB T -P -ND6 -E F RNS V RNL L2 ND1 I -Q M ND2 W -A -N -C -Y

PALAEOGNATHAE   Struthio camelus   Proteins: 13   tRNAs: 22   rRNAs: 2   Length: 16595

Figure 2.7. Mitochondrial genome comparison among birds.

Between the primate and bird taxa, I used *Homo sapiens* and *Gallus gallus* as an example which is illustrated in figure 2.8.
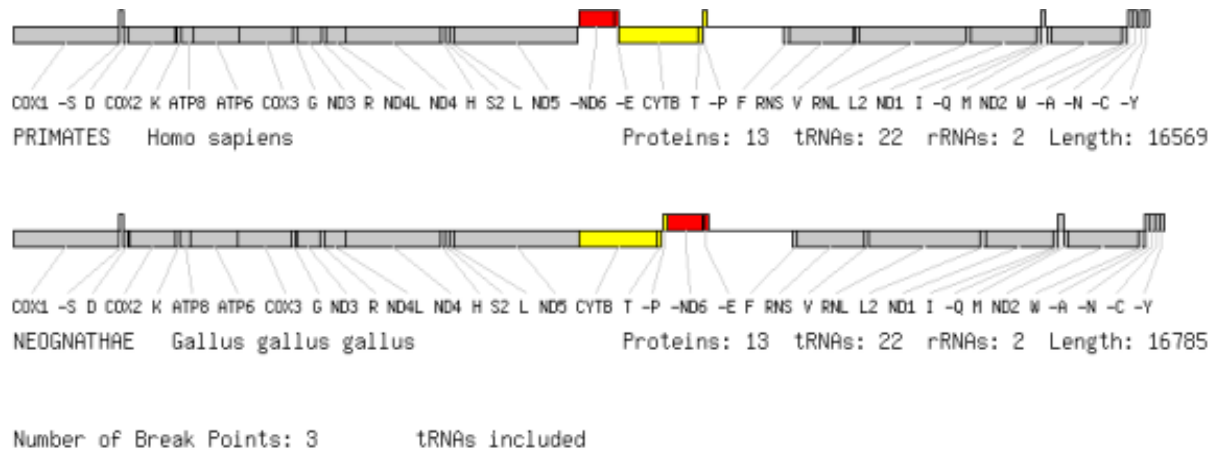

COX1 -S D COX2 K ATP8 ATP6 COX3 G ND3 R ND4L ND4 H S2 L ND5 -ND6 -E CYTB T -P F RNS V RNL L2 ND1 I -Q M ND2 W -A -N -C -Y

PRIMATES   Homo sapiens   Proteins: 13   tRNAs: 22   rRNAs: 2   Length: 16569

COX1 -S D COX2 K ATP8 ATP6 COX3 G ND3 R ND4L ND4 H S2 L ND5 CYTB T -P -ND6 -E F RNS V RNL L2 ND1 I -Q M ND2 W -A -N -C -Y

NEOGNATHAE   Gallus gallus gallus   Proteins: 13   tRNAs: 22   rRNAs: 2   Length: 16785

Number of Break Points: 3   tRNAs included

Figure 2.8. Mitochondria genome comparison between human and chicken.

There are three break points between these two taxa, shown in red and yellow colors. Break points identify the number of excisions that have to take place to replace the misplaced partitions. There were also 31 break points between *Terebratulina* and *Homo sapiens*.
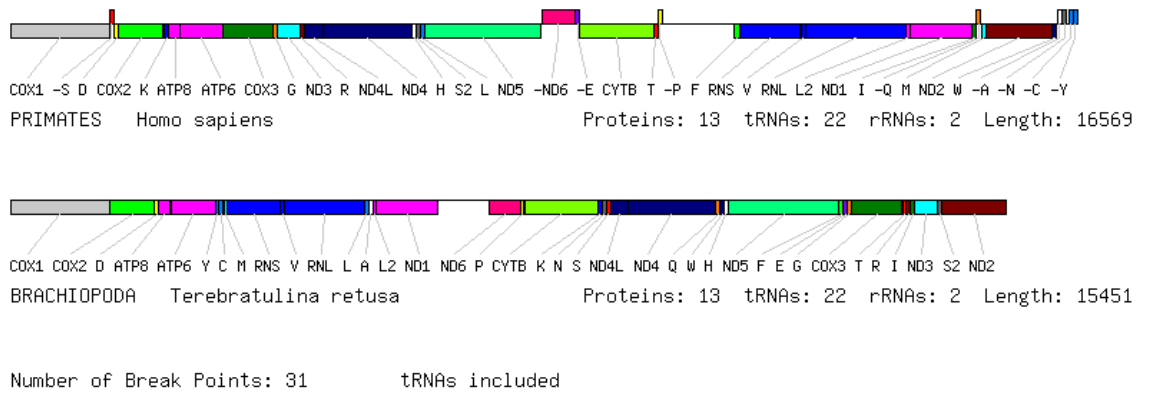
```
COX1 -S D COX2 K ATP8 ATP6 COX3 G ND3 R ND4L ND4 H S2 L ND5 -ND6 -E CYTB T -P F RNS V RNL L2 ND1 I -Q M ND2 W -A -N -C -Y
PRIMATES    Homo sapiens                              Proteins: 13  tRNAs: 22  rRNAs: 2  Length: 16569
```

```
COX1 COX2 D ATP8 ATP6 Y C M RNS V RNL L A L2 ND1 ND6 P CYTB K N S ND4L ND4 Q W H ND5 F E G COX3 T R I ND3 S2 ND2
BRACHIOPODA   Terebratulina retusa                   Proteins: 13  tRNAs: 22  rRNAs: 2  Length: 15451
```

```
Number of Break Points: 31        tRNAs included
```

Figure 2.9. Mitochondria genome comparison between human and brachiopoda

Also there were 30 break points between brachiopoda and chicken.


```
COX1 -S D COX2 K ATP8 ATP6 COX3 G ND3 R ND4L ND4 H S2 L ND5 CYTB T -P -ND6 -E F RNS V RNL L2 ND1 I -Q M ND2 W -A -N -C -Y
NEOGNATHAE    Gallus gallus                           Proteins: 13  tRNAs: 22  rRNAs: 2  Length: 16775
```

```
COX1 COX2 D ATP8 ATP6 Y C M RNS V RNL L A L2 ND1 ND6 P CYTB K N S ND4L ND4 Q W H ND5 F E G COX3 T R I ND3 S2 ND2
BRACHIOPODA   Terebratulina retusa                    Proteins: 13  tRNAs: 22  rRNAs: 2  Length: 15451
```

```
Number of Break Points: 30        tRNAs included
```
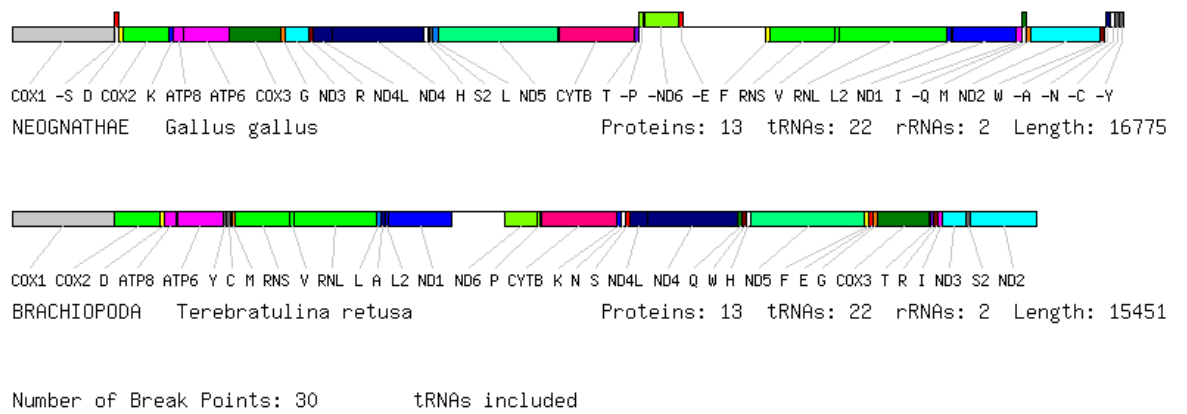
Figure 2.10. Mitochondria genome comparison between chicken and brachiopoda

The large number of break points between Terebratulina with the primate and bird taxa justifies its use as an outgroup.