

## Exploring Evolution: Unveiling the History of Flightless Birds Through Phylogenetic Trees

Workshop developed by Carla H Finn<sup>1</sup>

Acknowledgement and thanks to:

Wilda Laux<sup>2</sup> for thoughtful edits and feedback.

Peter Ritchie<sup>1</sup> for feedback and help with concept.

<sup>1</sup> Te Herenga Waka – Victoria University of Wellington

<sup>2</sup> Genomics Aotearoa

In this activity, you will explore the evolutionary relationships among flightless birds by performing two complementary activities:

1. **Manual DNA Sequence Alignment:** In the first part, you will complete a hands-on activity where you will manually align simplified nucleotide sequences from different ratite species. By identifying similarities and differences in the sequences, you will gain an understanding of how sequence alignments work.
2. **Computer-Based Sequence Alignment:** In the second part, you will use a computational tool (**Clustal Omega**) to align actual nucleotide sequences from more ratite species. This tool allows for the analysis of much longer and more complex sequences quickly and accurately. From the computer-generated alignment, you will build a phylogenetic tree that visually represents the evolutionary history of the ratites.

Together, these exercises will give you a deeper appreciation of how genetic data can be used to map the evolutionary paths of species and the benefits of using bioinformatics tools in research.

### Part- 1 – Manual DNA Sequence Alignment

#### Materials:

A3 sheet of paper with portions of a nucleotide sequence ranging from 122-132bp from Ostrich, Rhea, Kiwi, Moa, and Cassowary.

Tape

Scissors

#### Aims:

To manually align simplified nucleotide sequences AND understand how sequence alignments work and how differences in DNA can be used to build a basic phylogenetic tree.

## Methods:

### Step 1: Align the DNA Sequences Manually

#### 1. Prepare the Sequences:

- Cut the strips of DNA sequences.
- Spread out the paper strips for each species so the strips are side by side.

#### 2. Align the Sequences:

- Begin aligning the sequences by matching identical nucleotides (A, T, G, C) across the species.
- Look for regions where the sequences are the same; these are called conserved regions.

### Example:

Consider the sequences for Ostrich and Rhea:

- **Ostrich:**     G A T **G** T
- **Rhea:**        G A T **C** T

Notice that the first three nucleotides (G, A, T) are the same, but the fourth nucleotide differs (G for Ostrich and C for Rhea). This is a single-nucleotide polymorphism, or 'SNP'.

**Note:** The nucleotide sequences provided are simplified and curated for this exercise to reflect real phylogenetic relationships. In actual research, longer sequences or entire genomes are used for greater accuracy, but manual alignment of such sequences would be impractical.

### Step 2: Count the Number of Differences and Fill in the Matrix of Differences

#### 1. Pairwise Comparison:

- For each pair of species, count how many nucleotides are different between the pairs as indicated in the matrix.

#### 2. Record the Differences:

- Use the matrix below to record the number of differences between each pair of species.

### Example:

As an example, here we are comparing a simple sequence for the Ostrich and Rhea:

- **Ostrich:**     G A T **G** T
- **Rhea:**        G A T **C** T

There is **1 difference** at the fourth position. This counts for 1. Count how many more differences there are along their DNA sequences and enter the final count in the matrix.

### 3. Continue with the other species.

#### Matrix of Differences:

	Ostrich	Rhea	Kiwi	Moa	Cassowary
Ostrich	0				
Rhea		0			
Kiwi			0		
Moa				0	
Cassowary					0

### Step 3: Interpret Your Matrix

#### 1. Analyse the Results:

- Look at the numbers in your matrix.
- Identify which species pairs have the fewest differences and which have the most.

#### 2. Consider Evolutionary Implications:

- Species with fewer differences are likely more closely related.
- Species with more differences likely diverged earlier in evolutionary history.

### Questions to Consider:

- **a.** Which species pairs have the **fewest differences**, indicating they are closely related?
- **b.** Which species has the **most differences** from the others, suggesting it diverged earlier?
- **c.** Do the results align with your expectations based on the species' physical characteristics or geographical distributions?

### Step 4: Building a Basic Phylogenetic Tree

#### Instructions:

#### 1. Create a Diagram:

- Using the data from your matrix, draw a phylogenetic tree.
- Start with the species pair that has the least difference, and draw a bracket connecting them

#### 2. Determine Branching Order:

- Then, connect the next pair with fewest differences

Your teacher will demonstrate how you can start drawing your phylogenetic tree.

**Note** that fewer differences represent a closer relationship, meaning the species share a more recent common ancestor, while more differences suggest a more distant relationship and an older common ancestor.

## Step 5: Reflection and Discussion

### Questions to Reflect On:

1. **Most Divergent Species:**
  - Which species has the most differences from the others? What might this suggest about its evolutionary history?
2. **Limitations of Manual Alignment:**
  - What limitations arise from manually aligning sequences and building a tree?
  - How might computer-assisted methods differ in accuracy and ease?

## Conclusion and what's next?

### Conclusion and what's next?

The above manual alignment process mirrors the approach used by a popular computational algorithm known as **Neighbor-Joining (NJ)**. This algorithm is widely used in bioinformatics for constructing phylogenetic trees because it provides a **fast and efficient method** for inferring evolutionary relationships between species.

### What is Neighbor-Joining?

**Neighbor-Joining (NJ)** is an algorithm that estimates the evolutionary distances between species based on **genetic similarities and differences** in the sequences you align. The NJ method starts with the assumption that all species are equally related and then iteratively joins the "closest" neighbors—those with the smallest genetic distances—until a complete tree is constructed.

### Why Use Computer Algorithms for Phylogeny?

In reality, scientists work with DNA sequences which are often thousands to millions of nucleotides long. Comparing and aligning these longer sequences by hand would be impractical, if not impossible. That's where bioinformatics tools like **Clustal Omega** and algorithms like **Neighbor-Joining** come in.

- **Speed and Accuracy:** Computer algorithms quickly calculate genetic distances and align large datasets, producing accurate trees that reflect millions of years of evolutionary history in minutes.
- **Handling Complex Data:** We can account for a variety of factors, such as **insertions**, **deletions**, and **mutations**, which happen over evolutionary time and complicate the relationships between species.

- **Statistical Support:** Statistical support indicates the level of confidence in the relationships being proposed.

**Table: pros and cons of manual vs computer alignment.**

Manual alignment	Computer alignment (Clustal Omega)
Provides a tactile understanding of sequence alignment and phylogeny construction.	Handles much larger and more complex sequences in a fraction of the time.
Limited to simplified, shorter sequences.	Generates a more accurate and comprehensive phylogenetic tree.
Useful for grasping fundamental concepts but impractical for real-world genetic data.	Reflects the real power of bioinformatics in understanding evolutionary relationships at the molecular level.

## Part 2: Computer-Based Sequence Alignment and Phylogeny Construction

In this part, you will use **Clustal Omega**, a bioinformatics tool designed to align DNA sequences from multiple species, helping with quickly and efficiently observing the evolutionary relationships between them.

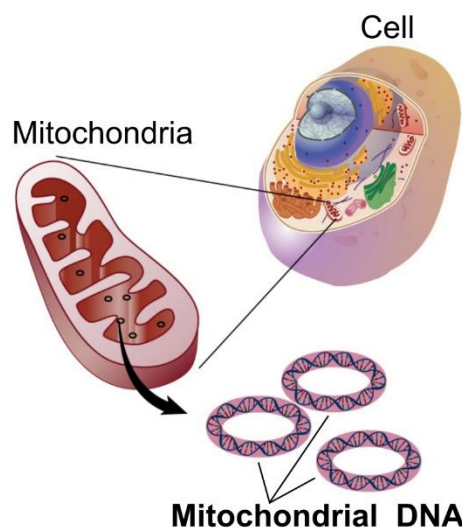
### Why Use Clustal Omega?

Clustal Omega can analyse real-world datasets that are often too large or complex to handle by hand. The computer algorithm can efficiently align sequences, account for evolutionary events (e.g., insertions, deletions), and generate a detailed phylogenetic tree that reflects the evolutionary history of the species you're studying.

### The sequence we're using: Cytochrome B

Cytochrome B gene is part of the mitochondrial genome. Mitochondria are organelles ("small organs") that are found in plant and animal cells. They are where glucose is changed into energy using the process of cellular respiration.

Cytochrome B is particularly useful for phylogenetic studies because:



Cytochrome B is a protein-coding gene found in the mitochondria of all animals. While its function is conserved across species, the gene accumulates mutations at a steady rate over evolutionary time, which provides the variation necessary to distinguish between different species.

This balance between conservation (because it's part of the essential mitochondrial function) and variability (due to mutations over time) makes Cytochrome B an excellent marker for studying evolutionary relationships. It allows us to track both recent divergences (as seen in ratites) and older evolutionary splits.

Image: By National Human Genome Research Institute - National Institutes of Health. National Human Genome Research Institute. "Talking Glossary of Genetic Terms". Retrieved November 17, 2016, from, Public Domain, <https://commons.wikimedia.org/w/index.php?curid=53240683>

## Why Cytochrome B for Ratites?

Ratites (flightless birds) are a relatively young evolutionary group, and Cytochrome B evolves at a rate that is particularly useful for resolving relationships between species that diverged more recently. The variation in this gene between species provides just the right level of genetic difference to reveal how these species are related. For example, it helps us explore the evolutionary connections between the kiwi, ostrich, rhea, and other ratites.

### The ratite species we're analysing today:

- *Ostrich*
- *Rhea*
- *Kiwi* (Great Spotted, North Island Brown, and Okarito)
- *Moa*
- *Elephant Bird*
- *Brushland Tinamou*



**World map demonstrating the geographical distribution of the ratite species used in this activity.**

Image accreditations:

Moa: "Euryapteryx," by [FunkMonk](#) (Michael B. H.) / Wikimedia Commons, licensed under CC BY-SA 3.0

(<https://creativecommons.org/licenses/by-sa/3.0/>). No modifications were made. Original Moa

Elephant bird: By Acrocynus - Acrocynus, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=3830345>

image available at: <https://commons.wikimedia.org/wiki/File:Euryapteryx.jpg>

Ostrich: By MathKnight - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=5010729>

Rhea: By Rufus46 - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=36109606>

Kiwi: By Maungatautari Ecological Island Trust - <http://www.maungatrust.org/news/default.asp>, Public Domain,

<https://commons.wikimedia.org/w/index.php?curid=2450957>

Tinamou: By Allan Drewitt - <https://www.flickr.com/photos/31003918@N05/6437912125/in/photostream>, CC BY 2.0,

<https://commons.wikimedia.org/w/index.php?curid=18145163>

**Information for each of the ratite species included in this activity.**

<b>Species</b>	<b>Origin</b>	<b>Appeared (Estimate in Mya)</b>	<b>Extinction (If Relevant)</b>
Aepyornis (Elephant Bird)	Madagascar	Quaternary (~2.6)	17th century
Anomalopteryx didiformis (Little Bush Moa)	New Zealand	Miocene (~23-5)	14th century
Apteryx haastii (Great Spotted Kiwi)	New Zealand	Pleistocene (~2.6)	Extant
Apteryx rowi (Okarito Kiwi)	New Zealand	Pleistocene (~2.6)	Extant
Apteryx mantelli (North Island Brown Kiwi)	New Zealand	Pleistocene (~2.6)	Extant
Nothoprocta cinerascens (Brushland Tinamou)	South America	Cenozoic (~65)	Extant
Ostrich	Africa	Pliocene (~5)	Extant
Rhea americana (Greater Rhea)	South America	Pleistocene (~2.6)	Extant



## Methods:

### Step 1: Align DNA Sequences Using Clustal Omega

1. In your web browser, go to <https://www.ebi.ac.uk/jdispatcher/msa/clustalo>
2. Copy and paste the provided DNA sequences provided in the text file, 'ratite\_sequences.fasta' into the input box.
  - Note: FASTA file stores DNA or protein sequences. Each entry starts with a header line that begins with ">", giving the name or ID of the sequence. Below the header, you'll see the sequence itself.
3. **Set Parameters:**
  - **Sequence Type:** Select **DNA**.
  - **Output Format:** Choose **ClustalW with character counts** (this helps visualize the positions better).
  - **Optional:** add a title to your alignment (e.g., "Ratite Phylogeny").
4. Click **Submit** to start the alignment process. This may take a few moments, depending on the number of sequences.

### Step 2: Analyse the Alignment

1. Click on 'Alignment'.

### Understanding the Output:




Each nucleotide is represented by a different colour, making it easy to spot similarities (conserved regions) and differences.

- **Conserved regions** (positions where all species have the same nucleotide) are key indicators of evolutionary conservation.
- **Variable regions** highlight differences, potentially reflecting evolutionary changes since the species diverged.

### Step 3: Analyse a Phylogenetic Tree

Once you have the aligned sequences, Clustal Omega will automatically generate a phylogenetic tree based on the alignment. **Click on 'Phylogenetic Tree' to view this.** The tree shows evolutionary relationships, with the scale representing genetic distances between species.



- Click on  the to make the tree easier to read.
- Click on the  a few times to make the tree easier to read.
- Click on  the to sort the tree by the oldest/deepest clade.
- 

### Understanding the Phylogenetic Tree:

#### 1. Interpret the Tree:

- **Branches:** Each branch represents a lineage (a species or group of species). The **length of the branches** indicates the genetic distance between species; longer branches suggest more genetic differences and earlier divergence.
- **Nodes:** A node is where two branches meet and represents a common ancestor of those species.
- **Scale Bar:** The scale bar at the bottom (e.g., 0.10) represents genetic distance. It shows the number of nucleotide differences per sequence length. For example, a distance of 0.1 means that, on average, 10% of nucleotides differ between the species.

### How to Read the Tree:

- **Closely Related Species:** Species with short branches that meet at a node are closely related. For example, **Kiwi species (Great Spotted and Okarito)** should have shorter branches between them, indicating a more recent common ancestor.
- **Geographical Considerations:** Given the geographic distribution of ratites, you can hypothesize how continental drift and land connections (such as the ancient Gondwana supercontinent) might explain the patterns in the tree.

### Questions to Consider:

- Which species appear to be the most closely related, based on the tree?
- How does the tree reflect the geographic distribution of species? For example, do the Kiwis from New Zealand group together, and how do they relate to the Moa?
- Do the molecular data support what you expected based on geography? How could this be explained?