

BioNusantara: Interactive Map of Indonesian Biodiversity

Shafiq Irvansyah
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jalan
Ganesa 10 Bandung
13522003@std.stei.itb.ac.id

Ahmad Naufal Ramadan
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jalan
Ganesa 10 Bandung
13522005@std.stei.itb.ac.id

Elijah Darrellshane Suryanegara
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jalan
Ganesa 10 Bandung
13522097@std.stei.itb.ac.id

Abstract—BioNusantara is an interactive Python-based tool designed to explore Indonesia's rich biodiversity, focusing on species distribution patterns influenced by the Wallace and Weber Lines. Its multi-stage pipeline integrates pygbif for GBIF data acquisition, geopandas and Biopython for data processing and genetic sequence fetching, and folium for interactive web mapping. The tool performs phylogenetic analysis, demonstrating relationships like the close kinship of *Crocodylus porosus* and *Dermochelys coriacea* within the transitional zone, contrasting with their distant relation to *Chelonia mydas* from the western zone. BioNusantara provides a robust framework for managing, analyzing, and interactively presenting complex biodiversity data, supporting biogeographical research and conservation efforts in Indonesia.

Keywords—Biodiversity, Indonesia, Biogeography, Phylogenetics, Interactive Mapping

I. INTRODUCTION

Indonesia, a vast archipelago nation, stands as a global epicenter of biodiversity, boasting an unparalleled wealth of flora and fauna. Its strategic geographical position between the Indian and Pacific Oceans has fostered an extraordinary array of terrestrial and marine ecosystems, including expansive tropical rainforests, critical mangrove habitats, and some of the world's most diverse coral reefs. This ecological richness is evident in the sheer number of species it hosts: approximately 1,849 bird species, 498 amphibian species, 791 reptile species, 750 mammal species, and an estimated 30,000 plant species, many of which are endemic to the archipelago. The nation's marine environments are equally vibrant, supporting over 4,523 species of reef fish, 1,300 mollusk species, and approximately 600 coral species, accounting for about 70% of global coral diversity. From the majestic orangutans of Borneo and Sumatra to the Komodo dragons of Nusa Tenggara and the countless marine species inhabiting its extensive coral triangle, Indonesia's natural heritage is truly exceptional and vital for global ecological balance.

The remarkable distribution of species across the Indonesian archipelago is famously delineated by the Wallace Line, a profound biogeographical boundary proposed by the 19th-century naturalist Alfred Russel Wallace. This invisible line, running between Borneo and Sulawesi, and through the Lombok Strait, marks a distinct separation between Asian and Australasian faunal realms. To the west of the Wallace Line, species exhibit strong affinities with those found on the Asian mainland, including large mammals like tigers, elephants, and rhinoceroses, as well as various primates and Asian-type birds.

Conversely, to the east of the Wallace Line, the fauna begins to resemble that of Australia, featuring marsupials such as cuscus, and distinct bird species like cockatoos and birds of paradise. This sharp divide is attributed to deep ocean trenches that persisted even during periods of lower sea levels, preventing land bridges from forming and thus isolating the two biotas. Further east, the less distinct Weber Line identifies a more subtle boundary, representing the point where Australasian species begin to significantly outnumber Asian species, highlighting a transitional zone where the influence of both continental biotas gradually shifts.

Together, the Wallace and Weber Lines serve as powerful illustrations of evolutionary processes and geological history shaping biodiversity. They underscore Indonesia's pivotal role as a natural laboratory for understanding biogeography and evolution. Given its immense biodiversity and the unique evolutionary patterns observed within its boundaries, Indonesia is a critical area for global conservation efforts, facing the ongoing challenge of balancing development with the urgent need to protect its irreplaceable natural heritage.

II. THEORETICAL BASIS

A. Biodiversity and the Theory of Evolution

Biodiversity, the astonishing variety of life on Earth at all levels—from the genetic makeup within a single species to the vast array of ecosystems—is fundamentally a product of evolution by natural selection. This foundational scientific theory, independently conceived by Charles Darwin and Alfred Russel Wallace, posits that species is not static but change over geological timescales. The core mechanism involves individuals within a population possessing heritable traits that confer a survival or reproductive advantage in a given environment. These advantageous traits are then more likely to be passed on to subsequent generations, leading to a gradual shift in the characteristics of the population over time. This continuous process of adaptation and change is the engine driving the proliferation of life forms from common ancestral origins (Darwin, 1859; Wallace, 1858).

A key process in the generation of biodiversity is speciation, the evolutionary mechanism by which new and distinct species arise. One of the most common modes of speciation is allopatric speciation, which occurs when populations of a single species become geographically

separated by a physical barrier. This barrier could be a mountain range, a large body of water, a desert, or even a newly formed river. Once isolated, gene flow—the exchange of genetic material between populations—ceases. This interruption is crucial because it allows the separated populations to evolve independently, free from the homogenizing effects of interbreeding (Mayr, 1963).

Over vast periods, these isolated populations accumulate genetic differences through several mechanisms. Random mutations introduce new genetic variations, serving as the ultimate source of all genetic diversity upon which evolutionary forces can act (Luria & Delbrück, 1943). Genetic drift, the random fluctuation of gene frequencies, can have a more pronounced effect in smaller, isolated populations, leading to changes in allele frequencies and potentially the loss or fixation of certain alleles purely by chance (Frankham et al., 2017). Crucially, each isolated population will experience unique selective pressures from its local environment. For example, one population might face a different climate, a new set of predators, or a distinct food source compared to its ancestral or sister population. These differing pressures favor different traits, leading to divergent adaptations in each isolated group.

As these genetic and phenotypic differences accumulate, the two populations may eventually reach a point of reproductive isolation. This means that even if the geographical barrier were to disappear, individuals from the two populations would no longer be able to interbreed successfully to produce fertile offspring (Coyne & Orr, 2004). This inability to exchange genes defines them as distinct species. The cumulative effect of countless such speciation events, occurring across diverse environments and over millions of years, has resulted in the breathtaking array of species that populate our planet.

B. Wallace Line and Weber Line

The Wallace Line stands as one of the most compelling biogeographical boundaries on Earth, a testament to the profound influence of geological history on species distribution. Named after the British naturalist Alfred Russel Wallace, who meticulously documented this faunal divide during his extensive expeditions through the Malay Archipelago (1854–1862), the line runs primarily through the Lombok Strait (between Bali and Lombok) and the Makassar Strait (between Borneo and Sulawesi). To the west of this line, the islands of Sundaland (including Sumatra, Java, Borneo, and Bali) are characterized by a fauna strongly akin to mainland Asia, featuring placental mammals such as tigers (*Panthera tigris*), elephants (*Elephas maximus*), rhinoceroses (*Rhinoceros sondaicus*, *Dicerorhinus sumatrensis*), and various primates. Bird species here often include pheasants, woodpeckers, and sunbirds, typical of the Oriental realm.

The striking abruptness of the Wallace Line, particularly evident across the narrow Lombok Strait (approximately 35 km wide), puzzled early naturalists. The explanation lies in deep ocean trenches that have persisted for millions of years. Unlike the relatively shallow Sunda Shelf to the west

(which connected to mainland Asia during Pleistocene glacial periods when sea levels were significantly lower), the waters along the Wallace Line remained deep. This acted as an impenetrable barrier for most terrestrial species, preventing their dispersal between the Asian and Australasian continental plates. Consequently, the distinct faunas on either side evolved largely in isolation, leading to divergent evolutionary trajectories. For instance, while Bali hosts Asian-type birds like weaver birds, Lombok, just a short distance away, is home to Australasian species such as cockatoos and megapodes. This geological isolation over extended periods has been a primary driver of the unique biodiversity patterns observed today (Wallace, 1869; EarthDate, n.d.).

In contrast to the sharp demarcation of the Wallace Line, the Weber Line describes a more gradual and nuanced biogeographical transition within the eastern parts of the Indonesian archipelago. Proposed by Max Carl Wilhelm Weber in the early 20th century, this line generally passes through the central Moluccas (Maluku Islands) and is typically situated to the east of the Wallace Line. It identifies a "line of faunal balance" where the influence of Australasian species begins to equal, and then increasingly dominate, that of Asian species. While the Wallace Line marks a clear separation, the Weber Line highlights a complex zone of intermingling and shifting species composition (Weber, 1902; IASZoology.com, n.d.).

The area between the Wallace Line and the Weber Line is collectively known as Wallacea. This vast and geologically complex region, including islands like Sulawesi, the Lesser Sundas (excluding Bali), and the Moluccas, was never fully connected to either the Asian or Australian continental shelves during periods of low sea levels. Instead, these islands remained isolated by deep ocean waters, becoming "oceanic islands" in the biogeographical sense. Consequently, Wallacea exhibits a fascinating mix of species with both Asian and Australasian affinities, alongside a very high proportion of endemic species that have evolved in isolation on individual islands. For example, Sulawesi, a key island in Wallacea, boasts unique macaque species found nowhere else in the world. The Weber Line, therefore, provides a valuable framework for understanding the complex dispersal and evolutionary patterns within this unique transitional zone, where limited inter-continental exchange and long-term island isolation have fostered extraordinary levels of biodiversity and endemism (Michaux, 2010; Wikipedia, n.d. "Wallacea").

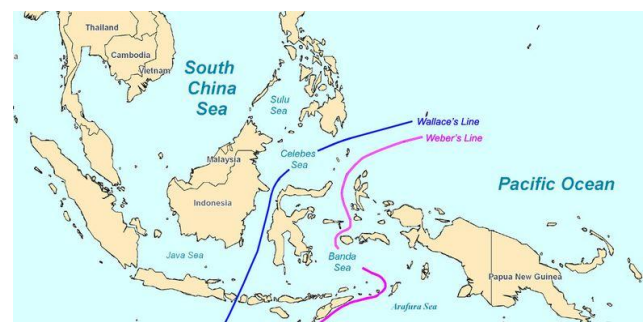


Image 1. Visualization of Wallace and Weber Line

C. Phylogenetics

Phylogenetics is a rigorous scientific discipline dedicated to inferring and reconstructing the evolutionary history of life. At its core, it seeks to elucidate the genealogical relationships among organisms, populations, genes, or even cells, typically visualized as a phylogenetic tree or cladogram. These branching diagrams are hypotheses representing the inferred patterns of divergence from common ancestors. Each node on the tree signifies a hypothetical common ancestor, while the branches denote evolutionary lineages leading to descendant taxa (operational taxonomic units, or OTUs). The lengths of these branches can sometimes represent the amount of evolutionary change (in phylograms) or simply the order of divergence (in cladograms), providing a visual narrative of diversification over geological time (Baum & Smith, 2013; Yang & Rannala, 2012).

The construction of phylogenetic trees relies on the analysis of shared characteristics, with molecular data—specifically DNA, RNA, or protein sequences—being the predominant choice due to its quantifiable and abundant nature (Yang & Rannala, 2012). The process typically begins with multiple sequence alignment, where homologous (shared due to common ancestry) positions in genetic sequences are aligned to identify similarities and differences. Subsequent phylogenetic inference employs sophisticated computational algorithms, broadly categorized into distance-based methods (e.g., Neighbor-Joining, UPGMA) and character-based methods (e.g., Maximum Parsimony, Maximum Likelihood, Bayesian Inference). Character-based methods, particularly Maximum Likelihood and Bayesian Inference, are statistically robust as they explicitly model the evolutionary process, accounting for factors like nucleotide substitution rates and variations across sites (Felsenstein, 2004; Huelsenbeck & Ronquist, 2001). These methods aim to find the tree topology that best explains the observed sequence data under a given evolutionary model, often requiring extensive computational searches across a vast tree space.

A critical challenge in phylogenetic inference is distinguishing between homology and homoplasy. Homologous traits are those shared by two or more taxa because they were inherited from a common ancestor. In contrast, homoplasy refers to similarities that arise independently in different lineages, often due to convergent evolution (similar selective pressures leading to similar adaptations), parallel evolution, or evolutionary reversals (loss of a derived trait, reverting to an ancestral state) (Hall, 2007). While homologous characters provide reliable evidence for shared ancestry, homoplastic characters can mislead phylogenetic analyses if not properly accounted for by appropriate evolutionary models. Robust phylogenetic methods and careful data selection are essential to minimize the impact of homoplasy and accurately reconstruct evolutionary relationships.

The insights derived from phylogenetics have profound implications across numerous biological disciplines. In systematics and taxonomy, phylogenies provide a natural,

evolutionary basis for classifying organisms. In conservation biology, understanding phylogenetic diversity helps prioritize species or lineages for protection, ensuring the preservation of evolutionary distinctness. Epidemiology heavily utilizes phylogenetics to trace the origins, spread, and evolution of pathogens, informing public health responses and vaccine development (e.g., tracking SARS-CoV-2 variants). Furthermore, phylogenetics aids in comparative genomics, drug discovery, and paleontology, by providing a framework for understanding trait evolution, predicting biochemical properties, and interpreting the fossil record within an evolutionary context (Yang & Rannala, 2012; Microbe Notes, n.d.). By unraveling the intricate web of life's evolutionary history, phylogenetics continues to be an indispensable tool for biological inquiry and application.

III. ANALYSIS AND IMPLEMENTATION

The implementation for this project is structured into several Python scripts, each handling a distinct phase of the biodiversity data processing and visualization pipeline. These scripts leverage various libraries for data manipulation, geospatial operations, and bioinformatics.

Several specialized Python libraries are used to build a comprehensive biodiversity analysis and visualization pipeline. Pygbif serves as the initial data acquisition component, enabling the direct download of species occurrence records from the GBIF database. Once raw data is obtained, geopandas extends pandas' data manipulation capabilities with geospatial functionalities, allowing the conversion of tabular occurrence data into rich geometric objects and facilitating spatial operations, such as identifying species within national park boundaries. Biopython is crucial for bioinformatics tasks, specifically utilizing its Entrez module to fetch genetic sequences from NCBI for phylogenetic analysis and its Phylo module to construct and visualize phylogenetic trees based on these sequences. Finally, folium powers the interactive visualization aspect, generating dynamic web maps that display the processed species occurrences, overlay biogeographic lines, highlight national parks, and incorporate interactive elements like popups for species details and embedded phylogenetic tree images. Together, these libraries form a robust framework for managing, analyzing, and interactively presenting complex biodiversity data.

A. Scripts

GBIF downloader (gbif_downloader.py)

This script is responsible for programmatically downloading species occurrence data from the Global Biodiversity Information Facility (GBIF) database.

Purpose: To fetch raw species occurrence records, categorized by taxonomic class, for Indonesia.

Key Functions:

- `download_gbif_data()`:
 - Iterates through a predefined dictionary of taxa (e.g., Mammalia, Aves, Amphibia, and specific reptilian orders) and their corresponding GBIF classKey identifiers.

- For each taxonomic group, it uses the pygbif.occurrences.search API to query GBIF for records from Indonesia (country="ID"), ensuring they have coordinate data (hasCoordinate=True).
- It implements a pagination mechanism to fetch records in chunks (per_page) up to a specified max_records limit to avoid overwhelming the API and to manage memory.
- Downloaded data for each taxonomic group is saved as a separate CSV file in the ../data/raw/ directory (e.g., gbif_mammalia_indonesia.csv).
- combine_reptilian_data():
 - After individual reptilian orders (Crocodylia, Squamata, Testudines, Sphenodontia) are downloaded, this function reads their respective CSVs.
 - It concatenates these dataframes into a single, combined reptilian dataset.
 - The combined data is then saved as ../data/gbif_reptilia_combined_indonesia.csv, providing a consolidated file for further processing.

Implementation Notes: The script sets an Entrez.email as required by NCBI for sequence fetching, though the actual fetching is performed in data_processing.py. It also sets specific record limits for different taxonomic groups (e.g., 1500 for reptiles, 6000 for others).

Data processing (data_processing.py)

Purpose: To take the raw GBIF data, clean it, validate coordinates, classify species into broad taxonomic groups, and, optionally, fetch genetic sequences from NCBI. It also integrates national park data.

Main Class: DataProcessor

- Initialization (__init__): Sets up empty attributes for species_data (GeoDataFrame for occurrences) and parks_data (GeoDataFrame for national parks).
- Data Loading and Cleaning:
 - load_and_clean_gbif_data(file_paths, fetch_sequences, progress_file_path_override):
 - Loads multiple GBIF CSV files (output from gbif_downloader.py).
 - Calls clean_gbif_data for detailed cleaning.
 - Transforms the cleaned Pandas DataFrame into a GeoPandas GeoDataFrame using create_geodataframe.
 - clean_gbif_data(df, fetch_sequences, progress_file_path_override):

- Removes duplicate records based on the 'key' column.
- Drops records with missing or invalid decimalLatitude or decimalLongitude.
- Filters out records with high coordinate uncertainty (coordinateUncertaintyInMeters < 1000m).
- Applies geographical bounding box filters to retain only records within Indonesia's approximate coordinates.
- Selects a subset of essential columns.
- Adds a taxonomic_group column using classify_taxonomic_group.
- Genetic Sequence Fetching (Conditional): If fetch_sequences=True, it iterates through unique scientific names and calls _fetch_ncbi_sequence. It includes robust error handling, retries, and a progress saving/loading mechanism using JSON files to resume interrupted fetches, minimizing repeated API calls to NCBI.

- Taxonomic Classification: classify_taxonomic_group(class_name):
 - Categorizes species into broader groups (Mammals, Birds, Amphibians, Reptiles, Fish, Arthropods, Molluscs, Plants, Other Invertebrates, Unknown) based on their 'class' name.
- Geospatial Data Preparation:
 - create_geodataframe(df): Converts a Pandas DataFrame containing decimalLatitude and decimalLongitude into a GeoPandas GeoDataFrame with Point geometries, setting the CRS to EPSG:4326.
 - load_parks_data(shapefile_path): Loads national park boundaries from a shapefile into a GeoPandas GeoDataFrame, reprojecting to EPSG:4326 if necessary.
- Spatial Analysis:
 - spatial_join_species_parks(): Performs a spatial inner join between the species_data (occurrences) and parks_data (national parks) GeoDataFrames. This identifies which species occurrences fall within the boundaries of national parks.
 - calculate_park_statistics(species_in_parks): Computes summary statistics for each national park based on the species occurrences found within them. This

includes total records, unique species counts, taxonomic group distribution, and counts of threatened species (based on IUCN Red List categories).

- `identify_endemic_candidates(park_specific_species, all_species_in_all_parks)`: Identifies species that are present in a specific park but not recorded in any other park within the dataset, marking them as potential endemic candidates.

Output: Saves the cleaned species data as a GeoJSON file in the `../data/processed/` directory.

Geospatial analysis (geospatial_analysis.py)

This script is responsible for creating an interactive geographical map to visualize the processed biodiversity data.

Purpose: To generate an interactive Folium map displaying species occurrences, biogeographic lines, national parks, and various biodiversity statistics.

Main Class: BiogeographicAnalyzer

- Initialization (`__init__`): Sets up attributes for Wallace and Weber lines and the Folium map object.
- Biogeographic Line Creation: `create_wallace_weber_lines()`: Defines the coordinates for the Wallace and Weber biogeographic lines and creates GeoDataFrames for them. These lines are crucial for understanding faunal distributions in the Indonesian archipelago.
- Biogeographic Zone Classification: `classify_biogeographic_zones(species_gdf)`: Assigns each species occurrence to one of three biogeographic zones (Western/Asian, Central/Wallacea, Eastern/Australasian) based on its longitude relative to the Wallace and Weber lines.
- Map Creation (`create_biodiversity_map`):
 - Initializes a Folium map centered on Indonesia.
 - Adds a custom HTML title to the map.
 - `_add_biogeographic_lines()`: Overlays the Wallace and Weber lines on the map.
 - `_add_species_points(species_gdf)`: Adds species occurrence points using MarkerCluster for efficient rendering of many points. Each point has a popup displaying species details and a button to view its phylogenetic tree. Markers are color-coded by taxonomic group.
 - `_add_species_heatmap(species_gdf)`: Adds a heatmap layer to show areas of high species density.
 - `_add_national_parks_layer(national_parks_data_path)`: Adds national park locations

(from a JSON file) as pulsing markers with tooltips displaying park information.

- `_add_biodiversity_statistics(species_gdf)`: Adds an interactive HTML panel displaying overall biodiversity statistics (total species, records, threatened species, taxonomic group counts, biogeographic zone counts) directly on the map.
- `_add_taxonomic_group_layers(species_gdf)`: Creates separate toggleable layers for each taxonomic group, allowing users to view specific groups independently.
- `_add_modal_html_css_js()`: Injects HTML, CSS, and JavaScript for a modal window that pops up to display the phylogenetic tree image when its corresponding button is clicked. This allows for interactive viewing of the phylogenetic trees generated by `phylogenetic_analysis.py`.
- `_add_pulsing_dot_css()`: Injects CSS for the pulsing effect on national park markers.
- `_add_custom_legend()`: Adds a custom legend to explain map symbols, lines, and taxonomic group colors.
- Adds a `folium.LayerControl` to allow users to toggle different map layers.
- Map Output:
 - `save_map(filename)`: Saves the generated interactive map as an HTML file (e.g., `bionusantara_map.html`) in the `../output/` directory.
 - `display_map()`: Returns the Folium map object for direct display in environments like Jupyter notebooks.

Interoperability: This script relies on the GeoJSON output from `data_processing.py` for species occurrence data and expects phylogenetic tree images generated by `phylogenetic_analysis.py` to be available in the `phylo_trees` subdirectory.

Phylogenetic analysis (phylogenetic_analysis.py)

This script is responsible for generating phylogenetic trees for individual species based on their fetched genetic sequences.

Purpose: To perform Multiple Sequence Alignment (MSA) and construct Neighbor-Joining phylogenetic trees for selected species and their closest relatives, then save these trees as image files.

Key Functions:

- `standardize_species_name_for_id(original_name)`: Cleans and standardizes species names (e.g., "Genus species" to "Genus_species") for use as unique identifiers in sequence records and tree leaves. This is crucial for consistent naming across files.

- `sanitize_filename_for_display(original_name_for_file)`: A utility to sanitize names for filenames, though `standardize_species_name_for_id` is primarily used for consistency.
- `run_msa(seq_records_list, mafft_exe)`:
 - Takes a list of SeqRecord objects (containing species names and their genetic sequences).
 - Writes these sequences to a temporary FASTA file.
 - Invokes the MAFFT command-line tool (`mafft.bat` for Windows) to perform Multiple Sequence Alignment. MAFFT (MAFFT_EXECUTABLE) is an external program used for aligning DNA or protein sequences.
 - Reads the aligned sequences back into a MultipleSeqAlignment object.
- `generate_dummy_newick_for_subset(species_name_std_subset)`: Creates a basic Newick string (a standard phylogenetic tree format) for a given subset of standardized species names. This is used when a full tree reconstruction might be too complex or when only a simple representation is needed for a subset.

- `create_and_save_image(tree_to_draw, species_to_highlight_original, output_path, all_leaf_names_in_tree_std)`:
 - Uses Biopython's Phylo.draw to render the phylogenetic tree.
 - Highlights the target species in red on the plot for easy identification.
 - Saves the generated tree visualization as a PNG image in the specified `output_path`.

- `main()`:
 - Loads the `sequence_fetch_progress.json` file (generated by `data_processing.py`) to get the genetic sequences.
 - Prepares SeqRecord objects for valid sequences.
 - Attempts to load an existing master phylogenetic tree (`master_phylo_tree.nwk`). If it doesn't exist or is out of sync, it constructs a new master tree:
 - Performs MSA on all valid sequences using `run_msa`.
 - Calculates a distance matrix (identity-based) from the alignment.
 - Builds the master tree using the Neighbor-Joining method.
 - Saves the master tree in Newick format.

- For each target species, it identifies its `MAX_SPECIES_IN_SUBSET_TREE - 1` closest relatives based on evolutionary distance from the master tree.
- Generates a subset phylogenetic tree for the target species and its closest relatives.
- Saves this subset tree as a PNG image in the `../output/phylo_trees/` directory, named after the standardized species name (e.g., `Genus_species_phylo.png`).

The full code can be accessed on <https://github.com/SandWithCheese/bio-nusantara>

B. Features

Taxonomic Group Analysis

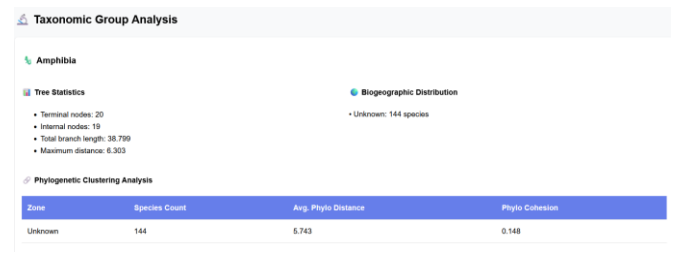


Image 2. Taxonomic Group Analysis Feature

Biodiversity map

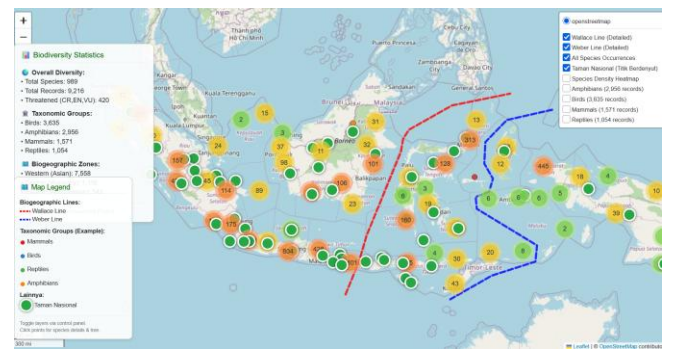


Image 3. Overview of Biodiversity Map

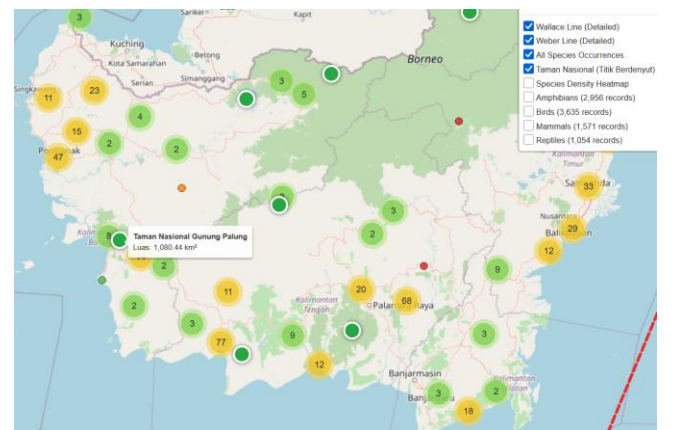


Image 4. View of National Park Detail on the Biodiversity Map

Species Details



Image 5. View of Species Details

Phylogenetic Tree

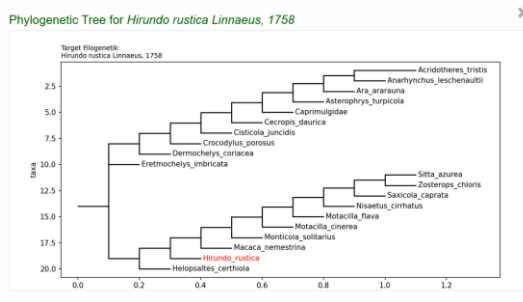


Image 6. An example of a phylogenetic tree detail

C. Analysis

If we compare two reptilian species from the class Testudines in the Transitional zone, such as *Crocodylus porosus* and *Dermochelys coriacea*, they will appear to be closely related on the phylogenetic tree.

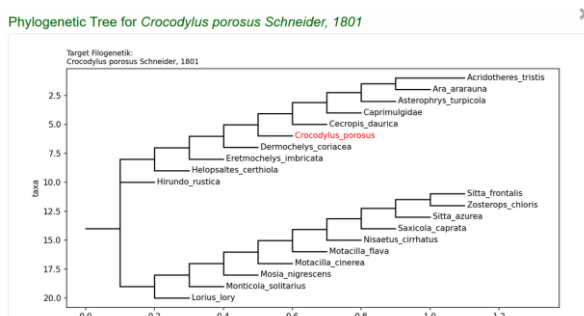


Image 7. Phylogenetic tree for *Crocodylus porosus* Schneider

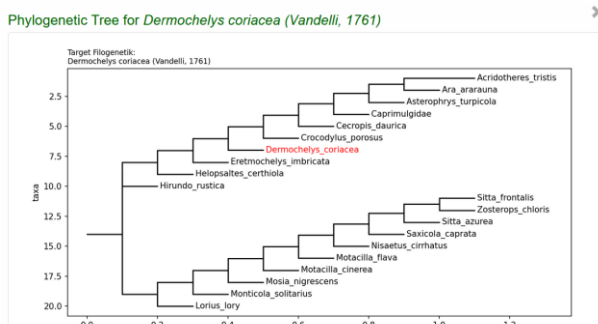


Image 8. Phylogenetic tree for *Dermochelys coriacea*

However, if both are compared to another reptilian species from the class Testudines but from the western zone, such as *Chelonia mydas*, we can see that they would be much further related

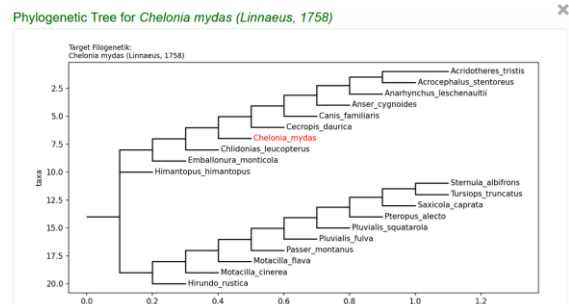


Image 9. Phylogenetic tree for *Chelonia mydas*

A similar phenomenon is also observed if we try to compare it to another reptilian species from the class Testudines in the Australasian zone, such as *Lepidochelys olivacea*

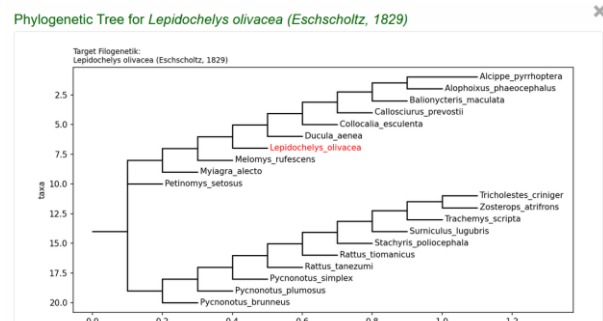


Image 10. Phylogenetic tree for *Lepidochelys olivacea*

A similar phenomenon is also observed on birds which belong to the order Passeriformes. Two species from the Asian zone, such as *Lonchura leucogastroides* and *Lonchura maja*, are very closely related.

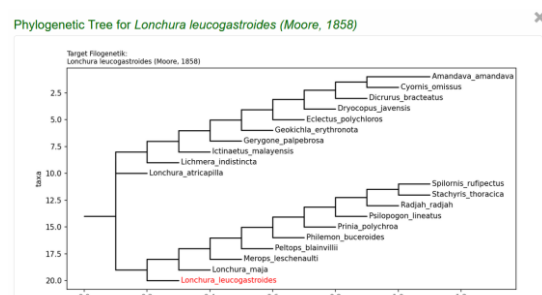


Image 11. Phylogenetic tree for *Lonchura leucogastroides*

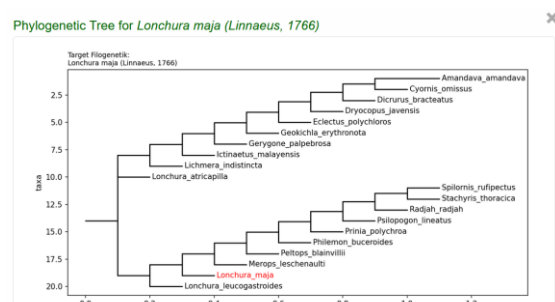


Image 12. Phylogenetic tree for *Lonchura maja*

However, if we then try to compare it to *Lonchura atricapilla* which also belong to the order Passeriformes but from the Transition zone, the relationship is observed to be more distant on the phylogenetic tree.

Phylogenetic Tree for *Lonchura atricapilla* (Vieillot, 1807)

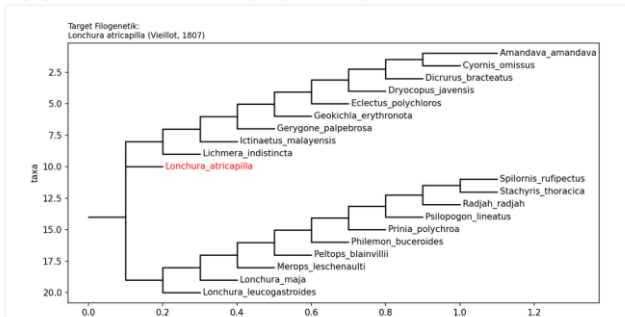


Image 13. Phylogenetic tree for *Lonchura atricapilla*

A similar conclusion is found when comparing it to *Oriolus boroensis*, another species belonging to the order Passeriformes but from the Australasian zone.

Phylogenetic Tree for *Oriolus boroensis* (Quoy & Gaimard, 1830)

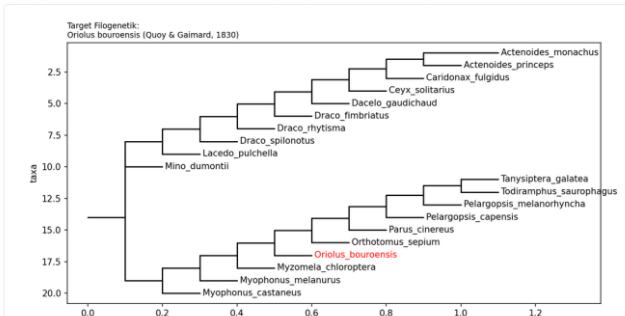


Image 14. Phylogenetic tree for *Oriolus boroensis*

IV. CONCLUSION

This project successfully developed and implemented BioNusantara, a comprehensive and interactive platform for analyzing and visualizing Indonesia's immense biodiversity. By integrating data acquisition from GBIF using pygbif, advanced data processing and genetic sequence fetching with geopandas and Biopython, and dynamic geospatial visualization via folium, the tool provides a powerful means to explore complex ecological and evolutionary patterns. The platform effectively delineates biogeographic zones marked by the Wallace and Weber Lines, showcasing the unique faunal transitions across the archipelago. Through its phylogenetic analysis component, BioNusantara allows for the visual comparison of evolutionary relationships, such as the observed close kinship between *Crocodylus porosus* and *Dermochelys coriacea* within the transitional zone, contrasting with their more distant relation to a species like *Chelonia mydas* from the western zone. This integrated approach not only enhances understanding of Indonesian biodiversity but also provides a valuable resource for researchers and conservationists in prioritizing protection efforts and studying speciation influenced by geological history.

V. REFERENCES

- [1] Jacobs, I. S., & Bean, C. P. (1963). Fine particles, thin films and exchange anisotropy. In G. T. Rado & H. Suhl (Eds.), *Magnetism* (Vol. III, pp. 271–). New York: Academic.
- [2] Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray. I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271
- [3] Wallace, A. R. (1858). On the Tendency of Varieties to Depart Indefinitely from the Original Type. *Proceedings of the Linnean Society of London*, 3, 53–62.
- [4] ASEAN Centre for Biodiversity. (2023). *ASEAN Biodiversity Outlook 3*. ASEAN Centre for Biodiversity. Retrieved from <https://asean.org/publication/asean-biodiversity-outlook-3/>
- [5] Mayr, E. (1963). *Animal Species and Evolution*. Harvard University Press.