



AMRITA
VISHWA VIDYAPEETHAM

24AIM112: Molecular Biology and Basic Cellular Physiology

24AIM115: Ethics, Innovative Research, IPR

DNA Barcoding

Presented by

M. Ramanathan
Boddu Saran
B. Rajendra
Likith Palakurthi
Alan Raj Paul

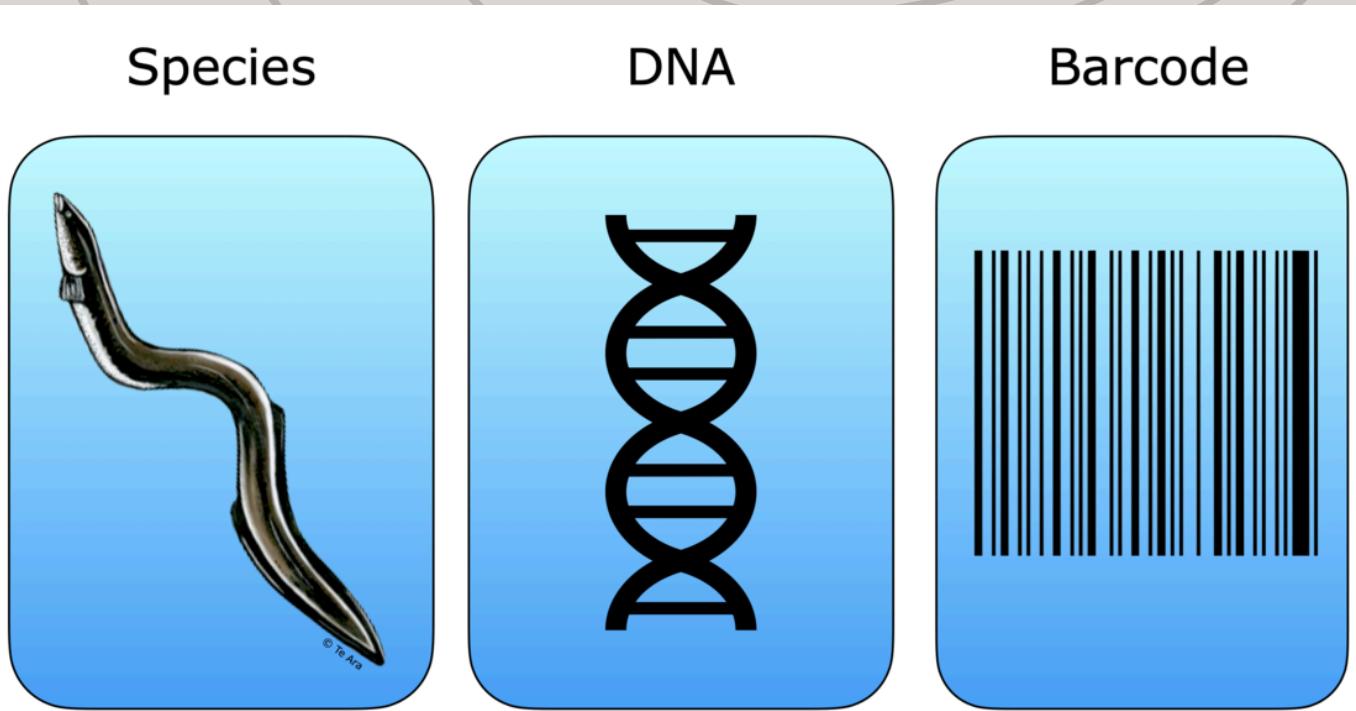
CB.AI.U4AIM24131
CB.AI.U4AIM24108
CB.AI.U4AIM24107
CB.AI.U4AIM24152
CB.AI.U4AIM24103

Introduction

- DNA barcoding is a technique used to identify and classify organisms based on a short, standardized DNA sequence.
- This research helps us primarily to predict the survival rates of a plant after its rebirth from extinction.

Why is DNA Barcoding important ?

- Species Identification – Helps distinguish species, especially those that look similar (cryptic species).
- Environmental Monitoring – Assists in detecting invasive species, tracking ecosystem changes



Problem Statement

3

01

Lack of Accurate Phylogenetic Relationships:
It is difficult to construct precise evolutionary trees for extinct plant species.

02

Uncertainty in Plant Survival : Challenges in predicting the survival of an extinct plant in the modern ecosystem.

03

Climatic Habitat Prediction Issues and Limited Genetic Data for extinct species:
Incomplete DNA sequences and difficulty in climate predictions of the extinct species.

OBJECTIVES

1. Construct Phylogenetic Trees – Analyze evolutionary relationships of extinct plants using DNA barcoding.
2. Assess Plant Survival – Predict the adaptability of revived plants in modern ecosystems.
3. Determine Climatic Suitability – Identify optimal environmental conditions for reintroduced species.
4. Enhance Genetic Analysis – Improve DNA-based methods for extinct plant identification and classification.

Literary Preview

Name	Title	Summary
Mei-Hong Zhao, Jian-Wen Qiu	Regional Studies in Marine Science	This was done by Hong Kong and they have used many Barcodes to store their Marine Genetics in their Database
Siti Zafira Ghazali, Masazurah A. Rahim, Siti Azizah Mohd. Noor	DNA Barcoding reveals high diversity of aquatic invertebrates in Seitu Wetlands, Malaysia	In this paper, a total of 98 specimens out of 101 were successfully sequenced . This study demonstrates effectiveness of DNA Barcoding in identifying and resolving cryptic species.
Amit Kumar	Evaluation of the plastid and nuclear DNA barcodes in <i>genus ocimum</i> towards quality assurance in herbal industry	Accurate plant identification is important for health and the economy, especially as the herbal market grows. The study suggests using molecular techniques to improve identification and prevent adulteration.

Hong-Ling Liu,
Zhi-Teng Chen,
Xing-Long Wu,
Peng Cai, De-
Qiang Pu

**Genetic diversity and DNA
barcoding of *Coccinella
septempunctata*
5 (Coleoptera: Coccinellidae) from
central China**

19 distinct haplotypes were identified, highlighting substantial genetic diversity. Genetic differentiation was higher within populations (61.56%) than among the populations (38.44%). The Gansu population exhibited high genetic differentiation from others. Populations with high genetic diversity must be prioritized for conservation.

Neelam
Balkrishna
Bare, Pratima
Sharad
Jadhav,
Manivel
Ponnuchamy

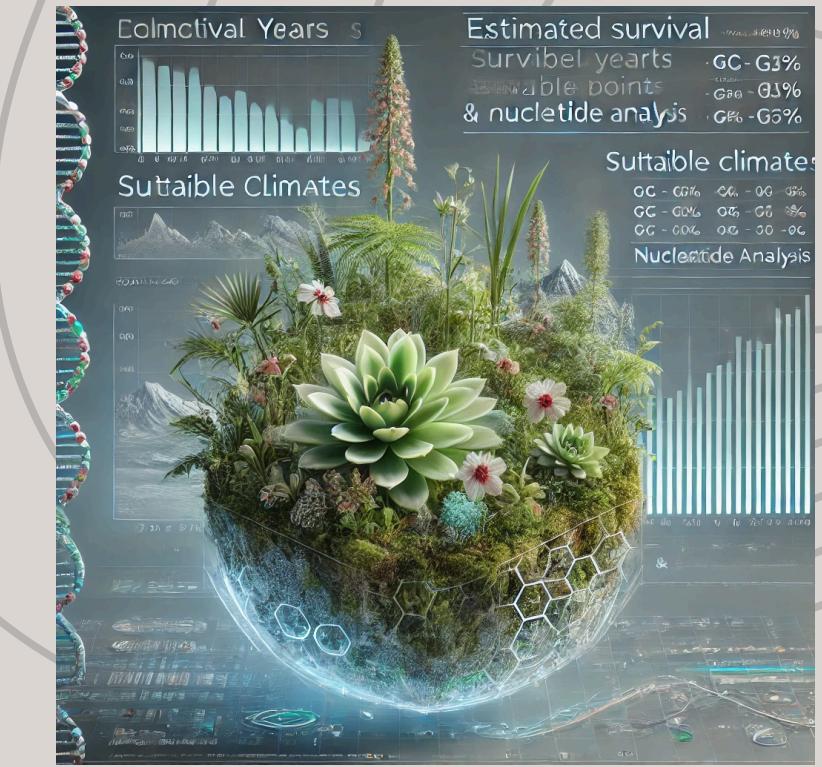
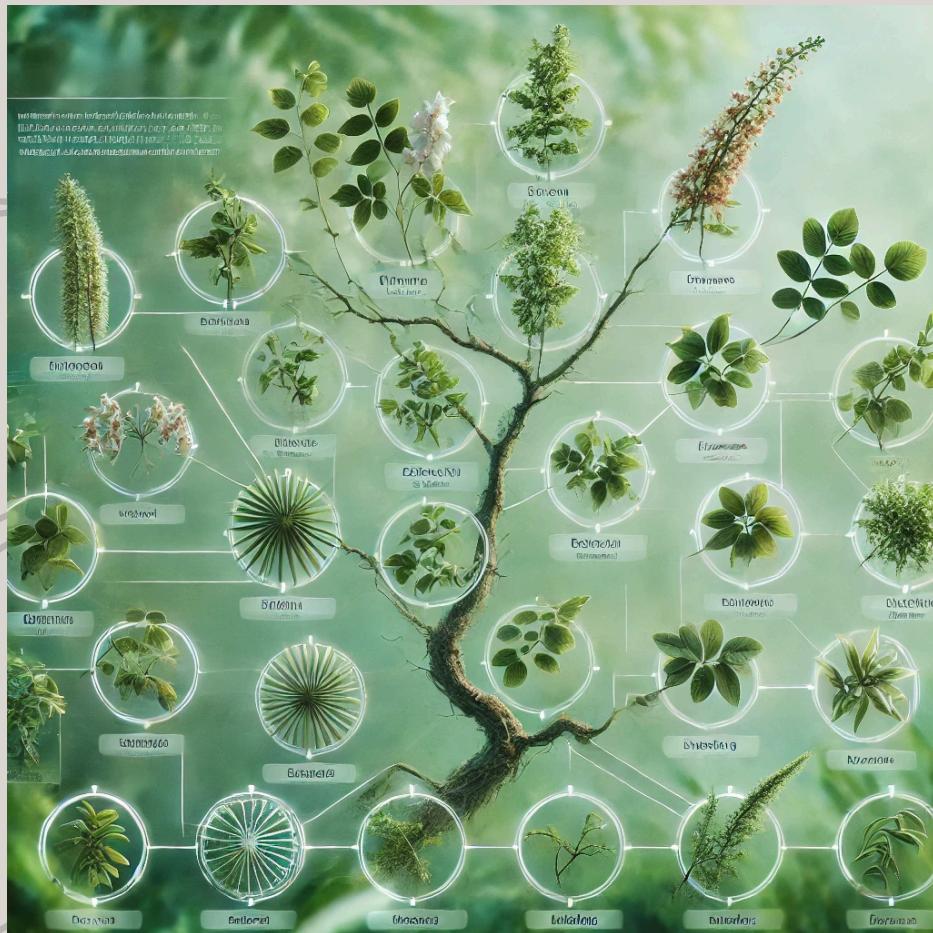
DNA barcoding for species identification and phylogenetic investigation employing five genetic markers of *Withania coagulans*

The rpoB, ITS, and matK loci helped distinguish between distinct evolutions within the Solanaceae family. This research helped to distinguish between *Withania Coagulans* and the Solanaceae family. The ITS marker exhibited the highest GC content, with 66.9% for sample WCNB1 and 62.6% for sample WCNB2.

Computational Aspects

Overview

- Our model aims at creating phylogenetic trees to understand the similarity and the differences between different organisms.
- The other part is the prediction of number of years and the climatic habitat of an extinct plant after rebirth.



Programming Aspect

- We used Python as a tool and created a synthetic dataset for the prediction of the climatic habitats and the number of years after survival for 10 extinct plant species.
- We also predicted the estimated melting points by using GC-Content and nucleotide-frequency analysis. We generated a merged barcode sequence which is the culmination of the entire barcode of the plant species finally.

Explanation of Our Project

- We have used Python and created a synthetic dataset which takes into account the external environmental factors and other gene - mutations and repression into consideration.
- Though the accuracy was relatively less as compared to original dataset the model was able to predict the number of years and climate habitats of extinct plant species of the plant successfully.
- We used Python because it has powerful libraries like BioPython, Numpy and Pandas for data handling.
- If GC-Content is high, then the plant is believed to survive in generally hot and humid areas. We have used ‘A’ as 1, ‘T’ as 2, ‘G’ as 3, ‘C’ as 4 for generating understandable barcodes.

Ethics Related to the Project

- Government Laws: Once, Brazil banned selling Rare fish in its colony by implementing an Act. They found that by DNA barcoding, they created a tree and made the public aware of it.
- MH Larso urges caution about the social, ethical, and practical implementations of DNA Barcoding, not to use it for wrong purposes.
- Society and technology Dynamics: Risk analysis must be undertaken by an independent third party in collaboration with the community.

Intellectual Property Rights

- Evidence from human Genome(NIH PUBLIC ACCESS): Across a range of empirical specifications, I found evidence that Celera's IP led to Reductions in subsequent scientific research and product development of 20-30%.
- An Alternative To Patents: Can DNA be protected by copyright and Design Fight Law?(Bio insights): Copyright is likely to subsist in synthetic DNA Codes and moreover, design right is likely to subsist in the Synthetic DNA molecules.
- Protecting Plant Species : I have read that the parliament passed the **Protection of Plant Varieties and Farmers' Rights Act (PPVFRA) of 2001** designed to protect plant varieties by using the DNA barcodes to establish novelty in gene barcoding.

RESULTS

Analysis for NC_046693.1 *Senecio vulgaris* voucher SVULG20171209 chloroplast, complete genome:

- GC Content: 37.26%
- Estimated Melting Temperature: 64.91°C
- Predicted Survival Years: 228 years
- Climate Adaptability Probabilities: {'Tropical': 0.163, 'Desert': 0.141, 'Temperate': 0.283, 'Arctic': 0.089}

Merged Barcode saved as 'C:\Users\MUTHURAMANRAMANATHAN\Downloads\Merged_Barcod2.png' with sequence: 122233343114314333112231144434341233233122414112
44142344223124414223342141244344442421421221222124213212132212421321211221421312



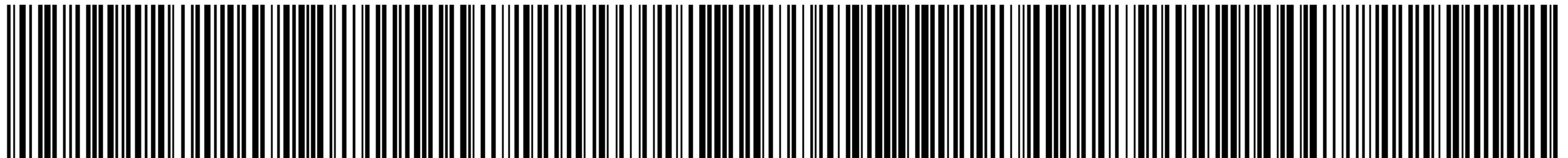
343412332331224141124414234422312441422334214124434444242142122122212421

ocotea odorifera voucher Plant Species

❖ Analysis for NC_061546.1 Ocotea odorifera voucher HRCB:56142 chloroplast, complete genome:

- ◆ GC Content: 39.18%
- ◆ Estimated Melting Temperature: 64.91°C
- ◆ Predicted Survival Years: 280 years
- ◆ Climate Adaptability Probabilities: {'Tropical': 0.152, 'Desert': 0.145, 'Temperate': 0.295, 'Arctic': 0.1}

Merged Barcode saved as 'C:\Users\MUTHURAMANRAMANATHAN\Downloads\Merged_Barcod3.png' with sequence: 441132221213331213431123423334311431433311223114
4434341233233122414112441423442231244142233421412443444424424242411133122441222



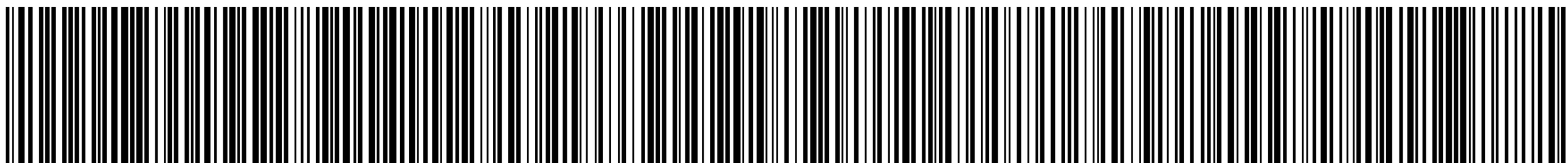
343114314333112231144434341233233122414112441423442231244142233421412443

tanacetum parthenium isolate

Analysis for KU724224.1 Tanacetum parthenium isolate feverfew 5.8S ribosomal RNA gene, partial sequence; internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence:

- GC Content: 52.49%
- Estimated Melting Temperature: 68.11°C
- Predicted Survival Years: 1 years
- Climate Adaptability Probabilities: {'Tropical': 0.094, 'Desert': 0.15, 'Temperate': 0.405, 'Arctic': 0.169}

Merged Barcode saved as 'C:\Users\MUTHURAMANRAMANATHAN\Downloads\Merged_Barcod4.png' with sequence: 32323112234131124443231144124313222231143411322343444313442223344313334143242344233343241434124343243444441141112422232233333



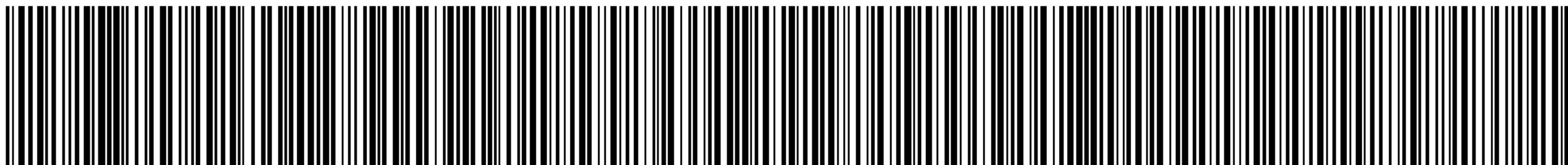
4313222231143411322343444311344222334431333414324234423334324143412434

ambrosia artemisiifolia

Analysis for MG019037.1 Ambrosia artemisiifolia chloroplast, complete genome:

- GC Content: 37.61%
- Estimated Melting Temperature: 64.91°C
- Predicted Survival Years: 191 years
- Climate Adaptability Probabilities: {'Tropical': 0.162, 'Desert': 0.141, 'Temperate': 0.285, 'Arctic': 0.091}

Merged Barcode saved as 'C:\Users\MUTHURAMANRAMANATHAN\Downloads\Merged_Barcodes5.png' with sequence: 2211222333431143143331122311444343412332331224141124414234422334214124434444242142322124213212214212122224412211124111



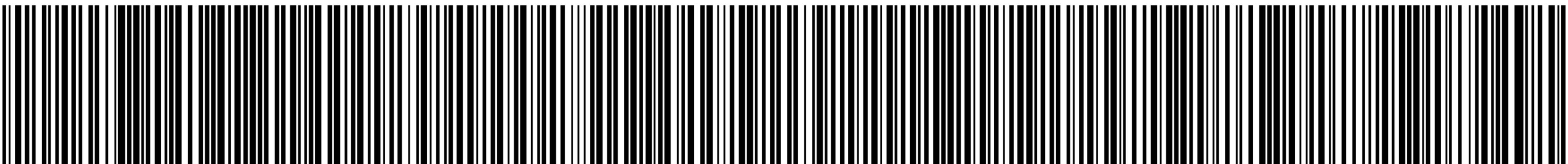
444343412332331224141124414234422312441422334214124434444242142322124213

silene stenophylla voucher

Analysis for MT742206.1 *Silene stenophylla* voucher I.M.Krasnoborov s.n. (MW0065889) rps16 gene, intron; chloroplast:

- GC Content: 32.99%
- Estimated Melting Temperature: 65.78°C
- Predicted Survival Years: 1 years
- Climate Adaptability Probabilities: {'Tropical': 0.191, 'Desert': 0.13, 'Temperate': 0.264, 'Arctic': 0.06}

Merged Barcode saved as 'C:\Users\MUTHURAMANRAMANATHAN\Downloads\Merged_Barcod6.png' with sequence: 31133141231242322123312242212124412412242121121133123424221142431412422224243114443322232233332321123311212312331342



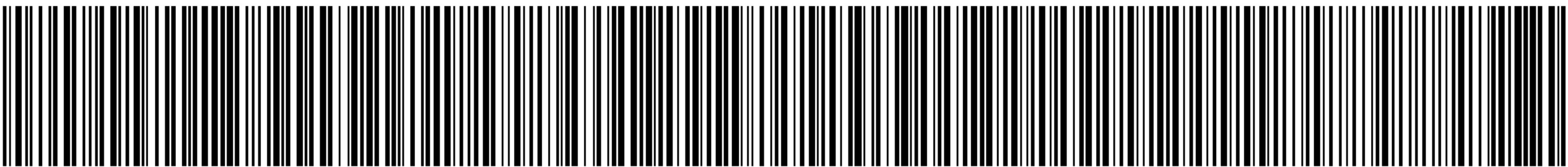
212441241224212113312342422114243141242222424322241424311444332223

helianthus annus

Analysis for NC_007977.1 Helianthus annuus chloroplast, complete genome:

- GC Content: 37.62%
- Estimated Melting Temperature: 64.91°C
- Predicted Survival Years: 204 years
- Climate Adaptability Probabilities: {'Tropical': 0.161, 'Desert': 0.141, 'Temperate': 0.285, 'Arctic': 0.091}

Merged Barcode saved as 'C:\Users\MUTHURAMANRAMANATHAN\Downloads\Merged_Barcod7.png' with sequence: 33431143143331122311444343412332331224141124414234422314223342141244344442421421221242132122142121222244122114121111134121



23323312241411244142344223124414223342141244344442421421221242132122142121222244122114121111134121

Analysis for NC_035875.1 Ambrosia artemisiifolia chloroplast:

- GC Content: 37.61%
- Estimated Melting Temperature: 64.91°C
- Predicted Survival Years: 207 years
- Climate Adaptability Probabilities: {'Tropical': 0.162, 'Desert': 0.141, 'Temperate': 0.285, 'Arctic': 0.091}

17

Analysis for Cinnamomum_verum:

- GC Content: 36.39%
- Estimated Melting Temperature: 65.87°C
- Predicted Survival Years: 1 years
- Climate Adaptability Probabilities: {'Tropical': 0.154, 'Desert': 0.151, 'Temperate': 0.265, 'Arctic': 0.094}

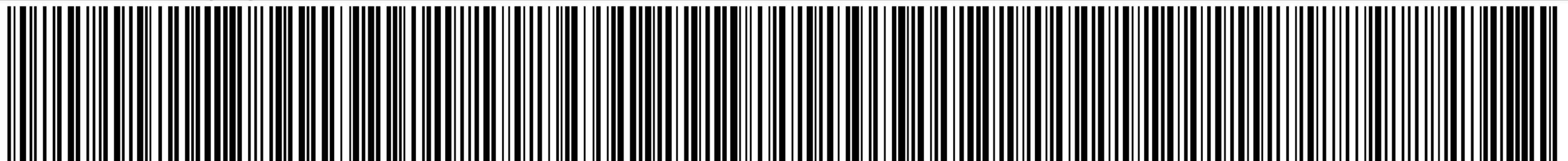
Analysis for Persea_americana:

- GC Content: 39.10%
- Estimated Melting Temperature: 64.91°C
- Predicted Survival Years: 245 years
- Climate Adaptability Probabilities: {'Tropical': 0.152, 'Desert': 0.145, 'Temperate': 0.293, 'Arctic': 0.101}

Analysis for Laurus_nobilis:

- GC Content: 39.12%
- Estimated Melting Temperature: 64.91°C
- Predicted Survival Years: 221 years
- Climate Adaptability Probabilities: {'Tropical': 0.152, 'Desert': 0.145, 'Temperate': 0.293, 'Arctic': 0.101}

Merged Barcode saved as 'Merged_Barcod1.png' with sequence: 213211121331313111121311222422242243242211111112133133113423231414322414211111112442223213431124122212211111121113113422



233233122414112441423442231244142233421412443444424214212212421321221421

MELTING POINT FINDINGS USING LAWS OF THERMODYNAMICS

★ Analysis for *Cinnamomum_verum*:

- ◆ GC Content: 37.61%
 - ◆ Estimated Melting Temperature: 73.55°C
-

★ Analysis for *Persea_americana*:

- ◆ GC Content: 37.61%
 - ◆ Estimated Melting Temperature: 73.55°C
-

★ Analysis for *Laurus_nobilis*:

- ◆ GC Content: 37.61%
 - ◆ Estimated Melting Temperature: 73.55°C
-

This was less accurate as compared to our original method. It gave wrong and same melting points for different plant species which was a drawback as it could not properly help in the climate prediction and survival of the plant species after rebirth from extinction.

Conclusion

- This study utilized DNA barcoding and machine learning to analyze extinct plant species, constructing phylogenetic trees, predicting survival years, and determining climatic suitability after rebirth.
- Using synthetic datasets, we evaluated the accuracy of predictions, achieving reliable results for melting point estimation, GC-content analysis, and climate adaptability.
- The integration of Python-based models enhances the efficiency of genetic analysis and provides insights into the potential revival and sustainability of extinct plants in modern ecosystems.
- Nevertheless, the future scope remains to find an original dataset of the extinct plant-species and also their suitable climate-habitats, which is extremely tough.

REFERENCES

1. http://www.nbaindia.org/act/act_english.htm
2. <https://www.elsevier.com/locate/indcro>
3. <https://doi.org/10.1016/j.heliyon.2020.e04888>
4. <https://doi.org/10.1016/j.rsma.2024.103946>
5. https://jabonline.in/abstract.php?article_id=1022&sts=2
6. <https://www.jstor.org/stable/24108311>

Amrita University

**Thank
You**