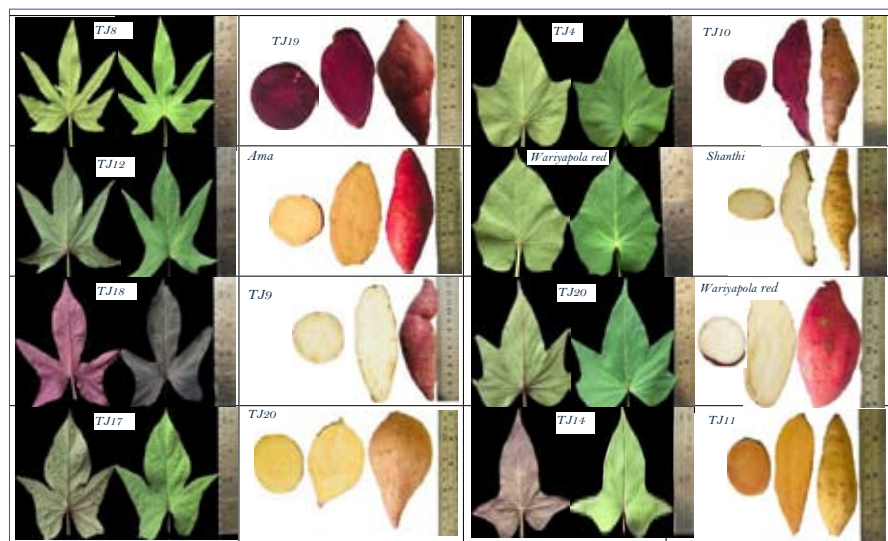


# Journal of the National Science Foundation of Sri Lanka





NATIONAL  
SCIENCE  
FOUNDATION

## JOURNAL OF THE NATIONAL SCIENCE FOUNDATION OF SRI LANKA

### Editorial Board

Ajit Abeysekera (Editor in Chief)  
J.K.D.S. Jayanetti  
L.P. Jayatissa  
P. Prasad M. Jayaweera  
Jagath Manatunge  
S.S.N. Perera  
Rohini de A. Seneviratne  
Saman Seneweera  
S.A.H.A. Suraweera  
P. Wijekoon  
M.J.S. Wijeyaratne

### Language Editor

R.D. Guneratne  
M.C.M. Iqbal

### Editorial Office

Nadeeja Wickramarachchi (Principal  
Scientific Officer)  
Upuli Ratnayake (Scientific Officer)  
Supun de Silva (Information Officer)  
Bhagya Dasanayaka (Scientific Officer)

Graphic Designing & Typesetting  
Kanchana Sewwandi

### Contact details

Editorial Office, National Science Foundation,  
47/5, Maitland Place, Colombo 07, Sri Lanka.

E-mail : [jnsf@nsf.gov.lk](mailto:jnsf@nsf.gov.lk)  
Phone : +94-11- 2696771  
JNSF online submission portal :  
<https://jnsfsl.sljol.info/about/submissions>  
JNSF home page :  
<http://www.nsf.gov.lk/index.php/nsfscience>  
magazine

**Publication :** Published quarterly (March, June, September and December) by the National Science Foundation of Sri Lanka.

**Manuscripts:** Research Articles, Research Communications, Reviews and Correspondences in all fields of Science and Technology can be submitted for consideration for publication. A guide to the preparation of manuscripts is provided in each issue. The guidelines may also be obtained by visiting the NSF website or JNSF online submission portal.

**Disclaimer:** No responsibility is assumed by the National Science Foundation of Sri Lanka for statements and opinions expressed by contributors to this Journal.

**Publication :** A publication fee of US\$ 250 will be levied for each manuscript in two stages except, when the corresponding author is affiliated with a Sri Lankan institution.

- A processing fee of US\$ 20 will be levied for each manuscript at peer-review stage.
- Remaining US\$ 230 will be charged for accepted articles at the time of publication.

**Copyright :** © National Science Foundation of Sri Lanka

Articles in the Journal of the National Science Foundation of Sri Lanka are Open Access articles published under the Creative Commons CC-BY-ND License (<http://creativecommons.org/licenses/by/4.0/>). This license permits the use, distribution and reproduction, commercial and non-commercial, provided that the original work is properly cited and has not been changed anyway.

**Indexing :** The JNSF is indexed in Science Citation Index Expanded, Journal Citation Reports/Science Edition, BIOSIS Previews, Zoological Record, Biological Abstracts, Chemical Abstracts, Scopus, DOAJ, TEEAL, Ulrich's, AGRICOLA and EBSCOhost, CAB Abstracts, SafetyLit, Journal TOCs, EBSCO Applied Science & Technology Source Ultimate

# JOURNAL OF THE NATIONAL SCIENCE FOUNDATION OF SRI LANKA

Volume 53 Number 1 March 2025

## CONTENTS

### EDITORIAL

- 1 **Embracing AI in Scholarly Publishing – Enhancing Integrity and Expanding Access**  
*Roshan Ragel*

### RESEARCH ARTICLES

- 3 **BRAF (V600E) gene mutation in patients with thyroid cancer in Sri Lanka: A pilot study**  
*N Ekanayaka, S Rathnayake, P Ratnayake, K Perera and P Udagama*
- 13 **Networking model based on fuzzy graphs of almost primary fuzzy ideals of near-rings**  
*A Muhammad, A Taouti, WA Khan and M Ashraf*
- 21 **Physiochemical characterization of critically endangered *Coscinium fenestratum* (Gaertn.) Colebr. seed fat for potential niche applications in cosmetics and nutritional supplements**  
*KAH Thathsara and SDM Chinthaka*
- 33 **Coincidence of inductive photoperiod and early vegetative phase leads to early flowering and increased yield in selected Sri Lankan traditional rice varieties**  
*S Weerakoon, B Nanayakkara, C Abayasekara, H Abeyundara and I Liyanage*
- 45 **Recreational water quality in the river Mahaweli, Sri Lanka: Enumeration and antibiotic sensitivity of *Escherichia coli* and sociological aspects**  
*S Weerakoon, B Nanayakkara, C Abayasekara, H Abeyundara and I Liyanage*
- 57 **Evaluation of genotypic variability and yield performance of selected indigenous and improved *Ipomoea batatas* L. genotypes under poly sack cultivation**  
*MFZ Safwa, TN Samarasinghe, AN Wijewardhana, HAPA Shyamalee, and AL Ranawake*
- 73 **Machine learning methods to classify engineering students' ability to perform recreational archery activities using biomechanical data**  
*NRH Basri, MS Mohhtar and WNLW Mahadi*
- 85 **Molecular phylogeography of *Echis carinatus* revisited: Insights from the Sri Lankan population of saw-scaled viper (Serpentes: Viperidae: *Echis*)**  
*A Murugananthan, A Gnanathan, T Kumanan, KP Amarasinghe and S Pirasath*



**Cover:** Leaf characteristics and corresponding qualitative storage root characteristics of studied *I. batatas* genotypes  
See *J.Natn.Sci.Foundation Sri Lanka* 2025 **53**(1): 57 - 72

## EDITORIAL

# Embracing AI in scholarly publishing : Enhancing integrity and expanding access

Integrating artificial intelligence (AI) into scholarly publishing heralds a transformative era beyond mere efficiency. As AI technologies become increasingly embedded in academic research, they offer not only the potential to enhance research processes but also to bolster the integrity and accessibility of scholarly communication.

AI and generative AI (GenAI) technologies are reshaping the academic landscape by providing tools that can assist with data analysis, literature reviews, and even the drafting of complex documents. Publishers and related service providers like COPE, ICMJE, and Elsevier have outlined guidelines that underscore AI's role as a supportive tool rather than a standalone author. These guidelines advocate for transparency in AI use and ensuring that AI's application in research is clear and justified, enhancing rather than obscuring human authors' accountability.

Globally, AI is being used to empower researchers across disciplines by enhancing how they communicate, explore, and analyze knowledge. Tools like *Grammarly* help non-native English-speaking researchers express themselves more clearly and confidently. These technologies are not about getting published per se, but about removing linguistic barriers that can obscure the clarity of thought or novelty of ideas. By enabling researchers to communicate their work more effectively, such tools promote richer academic discourse and more inclusive participation in global scholarly conversations. In fields like genomics, medicine, and climate science, AI is accelerating discoveries by processing large-scale datasets with speed and precision. Beyond efficiency, these tools promote a more data-driven approach by enhancing feature engineering, automating data interpretation, and enabling more exhaustive data searches. By reducing cognitive biases, such as researchers seeking self-

affirming patterns, GenAI helps surface insights that might otherwise be overlooked. The result is not just faster science, but deeper, more reproducible research that pushes the boundaries of what we can understand and achieve.

To fully harness the potential of AI, it is essential to cultivate a deep understanding of AI capabilities and limitations within the academic community. Educational initiatives should be launched to increase AI literacy among researchers, editors, and reviewers. This knowledge will empower them to use AI tools effectively and ethically, ensuring that these technologies are applied in ways that truly enhance the research process.

Equitable access to AI tools is crucial for leveling the playing field in academic research. Institutions should strive to provide widespread access to AI resources, enabling researchers from diverse backgrounds and disciplines to leverage these powerful tools. This democratization will foster a more inclusive scholarly environment where innovative ideas can flourish regardless of geographical or financial constraints.

As AI becomes more prevalent in academic settings, it is imperative to establish and evolve ethical guidelines that not only govern its use but also shape how we present research. Future scholarly publishing may no longer follow a one-size-fits-all format instead, it could adapt to promote transparency around the use of AI tools. This includes disclosing where and how AI assisted in research, from writing to data analysis, allowing readers to better assess the provenance and reliability of findings. Guidelines must also continue addressing data privacy, bias mitigation, and reproducibility, ensuring that AI contributes meaningfully to the values of openness, rigor, and responsible knowledge dissemination.

AI in scholarly publishing represents a frontier of immense possibility but it is an opportunity only if we act responsibly. Its true potential will be realized only if we commit to enhance academic integrity, increase AI literacy across the research community, and democratize

access to powerful AI tools. When guided by these principles, AI can serve as a trusted ally in the pursuit of academic excellence, helping us preserve and even elevate the core values of scholarly communication in a rapidly evolving digital world.

**Roshan Ragel**

## RESEARCH ARTICLE

### Molecular Genetics

# BRAF (V600E) gene mutation in patients with thyroid cancer in Sri Lanka: A pilot study

N Ekanayaka<sup>1</sup>, S Rathnayake<sup>1</sup>, P Ratnayake<sup>1</sup>, K Perera<sup>2</sup> and P Udagama<sup>1\*</sup>

<sup>1</sup> Department of Zoology and Environment Sciences, Faculty of Science, University of Colombo 03, Sri Lanka.

<sup>2</sup> National Cancer Institute of Sri Lanka, Apeksha Hospital, Maharagama, Sri Lanka.

Submitted: 02 March 2024; Revised: 16 December 2024; Accepted: 23 December 2024


**Abstract:** Thyroid cancer (TC) is the most common endocrine malignancy worldwide. Occasionally, inadequate treatment of malignancies and overtreatment of benign conditions result due to inconclusive diagnosis, while poor clinical outcomes occur due to cancer recurrence or resistance to existing therapies. Elucidation of the molecular genetic basis of TC potentiates its diagnosis and treatment. BRAF(V600E) is a commonly occurring mutation in TC. This retrospective cross-sectional study was undertaken to determine the prevalence of this mutation in a cohort of Sri Lankan TC patients. DNA was extracted from formalin-fixed paraffin-embedded (FFPE) thyroid tissue samples (malignant subtypes, n = 50 (papillary, follicular, medullary, anaplastic, and oncocyctic thyroid carcinomas); benign type, follicular adenoma, n = 26) using the phenol-chloroform method. Mutant allele specific PCR (sensitive to 1-5% mutant allele detection in a wild-type background) amplified BRAF wild-type, and mutant alleles. Genotypes were established by visualizing PCR amplicons (125 bp) of each allele type on agarose gels. Based on the allele specific PCR used for molecular diagnosis, most of TC patients in the study population (n = 68) were of the heterozygous genotype. One individual with papillary TC was homozygous for the mutant allele, and the remaining seven were homozygous for the wild-type allele. As this mutation was not restricted to malignant subtypes in comparison to the benign subtype (p > 0.05), its applicability as a diagnostic marker remains contentious. These preliminary findings underpin the potential for optimizing TC management through molecular diagnosis using available targeted therapy.

**Keywords:** BRAF(V600E) gene mutation, malignant types, molecular marker, targeted therapy, thyroid cancer.

## INTRODUCTION

Thyroid cancer (TC) is the most common endocrine malignancy, with a continuously increasing disease burden, with an exponential rise in cases both globally and locally. TC is described under four main pathological types; papillary thyroid carcinoma (PTC), follicular thyroid carcinoma (FTC), medullary thyroid carcinoma (MTC), and anaplastic thyroid carcinoma (ATC) (Ge *et al.*, 2020). Other rare types include oncocyctic carcinoma (OC, previously known as Hürthle cell carcinoma), and insular cancer (Liska & Galbavy, 2005). The most prominent subtype (80%) is PTC (Ito *et al.*, 2014).

TC has a higher global mortality rate than other endocrine malignancies (Kobawala *et al.*, 2016). Currently, TC represents approximately 2.1% of all newly diagnosed tumours worldwide (Crispo *et al.*, 2019), with an unprecedented rise in the cumulative number of living cases of TC (Xing, 2019). This disease affects women threefold more frequently than men (Warakomski *et al.*, 2018). Since the latter part of the last century, TC has increased at a rapid rate, especially among women in developed countries (Jayarajah *et al.*, 2018). In countries such as the United States (Jung *et al.*, 2014), Canada (Liu *et al.*, 2001) and Australia (Grodski *et al.*, 2008), the incidence of TC has more than doubled over the last few decades; developing nations in the Asian region recorded

\* Corresponding author (preethi@zoology.cmb.ac.lk;  <https://orcid.org/0000-0001-5431-0715>)



This article is published under the Creative Commons CC-BY-ND License (<http://creativecommons.org/licenses/by-nd/4.0/>). This license permits use, distribution and reproduction, commercial and non-commercial, provided that the original work is properly cited and is not changed in any way.

a greater incidence of TC, compared to the West (Kilfoy *et al.*, 2009). China has witnessed a four-fold increase in TC over 25 years (1983 - 2007) (Wang & Wang, 2015). The increasing incidence rate has evoked much interest and concern towards the management of TC.

The National Cancer Control Programme of the Ministry of Health, Sri Lanka maintains the most comprehensive database on information regarding TC in Sri Lanka. Accordingly, the local incidence of TC has increased significantly, as in other parts of the world (Jayarajah *et al.*, 2018). In 2014, the thyroid gland became the fifth leading cancer site among Sri Lankans (*Cancer Incidence Data Sri Lanka, 2014*). Currently, it is the third most common cancer among Sri Lankan women, having accounted for 8.7% of the new female cancer cases in 2020 (International Agency for Research on Cancer, 2020).

Although TC elicits excellent prognosis, conventional diagnostic tests occasionally lead to inconclusive results, that may either cause inadequate treatment of aggressive, high-risk thyroid malignancies or overtreatment of low-risk, benign conditions (Xing, 2019). In addition, despite large costs and resources incurred on prevailing therapeutic methods, some patients manifest poor clinical outcomes due to resistance to existing treatment options, and cancer recurrence. These complications contribute to the continuously increasing disease burden and associated cost of treatment, highlighting the need for novel diagnostics and treatment modalities.

Molecular targeted therapies for TC are based on the concept of precision medicine. A major challenge faced by clinicians is balancing the therapeutic approach towards patients with different types of TC (Gil *et al.*, 2020), which can be resolved using molecular markers for efficient TC diagnosis, and molecular targeted therapy.

Most of the genetic alterations implicated in TC lead to the aberrant activation of the RAS–RAF–MEK–MAP kinase cell signaling pathway (Frasca *et al.*, 2008). Four types of somatic genetic alterations are extensively studied due to their potential in the diagnosis and prognosis of follicular cell-derived thyroid carcinomas (TCs). These are BRAF mutations, RAS mutations, RET/PTC, and PAX8/PPAR $\gamma$  rearrangements (Bozec *et al.*, 2013). In addition, TERT promoter mutations, TP53 mutations, and NTRK fusions have also been reported in TC (Prasad *et al.*, 2016; Volante *et al.*, 2021). RAS mutations occur in about 21% of PTC and 57% of FTC (Howell *et al.*,

2013). RET/PTC rearrangement is the second most common genetic alteration in PTC, accounting for a prevalence of approximately 20% of PTC cases (Bozec *et al.*, 2013). PAX8/PPAR $\gamma$  rearrangements occur in approximately 30 – 35% of FTC cases (Raman & Koenig, 2014). In familial MTC, germ-line mutations in the RET gene have been reported while the inactivation of the p53 tumor-suppressor gene occurs in ATC (Xing, 2005). The prevalence of TERT promoter mutation is lower in PTC (~ 12%), but higher in poorly differentiated or undifferentiated types of TC such as ATC, accounting for about 50% in ATC cases (Soares *et al.*, 2021).

BRAF (B-type Raf kinase protein) gene mutations are known to play a significant role in the manifestation and progression of thyroid tumours. BRAF is an integral part of the RAS/MAP kinase pathway (Xing, 2005). The most common BRAF mutation observed in TC, especially the PTC subtype, is the BRAF (V600E) mutation (Santonastaso *et al.*, 2019) which causes a valine (V) to glutamate (E) substitution at the 600th amino acid residue of the protein (Bozec *et al.*, 2013). Upon mutation, the V600E substitution disrupts the hydrophobic interactions and retains the protein in a catalytically active conformation (Moulana *et al.*, 2018), causing elevated kinase activity and tumorigenesis (Lassalle *et al.*, 2010). It is considered as an “early driver” mutation in TC (Volante *et al.*, 2021). The reported prevalence of BRAF V600E mutation in PTC ranges from 40 – 80%, while it is less prevalent (1.4%) (Kebebew *et al.*, 2007) or absent (Tennakoon *et al.*, 2017) in the FTC subtype, which is mostly associated with RAS mutations. In poorly differentiated TC subtypes such as ATC, BRAF V600E prevalence ranges from 10% - 50% (Volante *et al.*, 2021). Its presence has also been reported in oncocytic carcinoma (Hürthle cell carcinoma), but within a small fraction of cells within the tumour area (McFadden & Sadow, 2021).

BRAF mutational analysis has been established as part of the diagnostic protocol for TC in Italy (Santonastaso *et al.*, 2019), while revised guidelines of the American Thyroid Association recommend the use of molecular markers including BRAF during cases of indeterminate fine needle aspiration (FNA) test results (Haugen *et al.*, 2016).

This pilot study aims to assess the prevalence of the BRAF(V600E) mutation in a cohort of thyroid cancer patients in Sri Lanka, highlighting the need to investigate its prognostic impact for effective molecular diagnosis, and management in the local context.



## MATERIALS AND METHODS

### Study population

The study was based on patients reported to the National hospital, Kandy, the second largest tertiary care state medical institution in Sri Lanka. Formalin-fixed paraffin-embedded (FFPE) cancerous thyroid tissue samples of patients who had been initially diagnosed with thyroid cancer by histopathological examination by an experienced histopathologist, were included in the study. Those with co-morbidities and indistinct diagnosis were excluded. Prior to the acquisition of samples, inclusion of the tumour area in the FFPE thyroid tissue was confirmed by the same histopathologist who confirmed the initial diagnosis.

Thus, FFPE samples of patients with thyroid cancer sub-types of PTC, FTC, MTC and ATC were included in the study. Samples of patients diagnosed with follicular adenoma (FA) were selected as controls due to the reported absence of BRAF V600E mutation (0 %) in FFPE samples of this benign tumour type (Sapio *et al.*, 2006; Yoo *et al.*, 2016). Accordingly, FFPE samples were retrospectively selected from medical records, reported within the year 2020.

As advocated by WHO guidelines in 1991, the sample size calculation was carried out using the following formula (Charan and Biswas, 2013):

$$\text{Sample size} = \frac{Z_{1-\alpha/2}^2 p (1 - p)}{d^2}$$

where,

$n$  = Desired sample size population <10,000

$Z$  = Standard normal deviate – set at 1.96 at 95% confidence level

$p$  = Proportion of subjects (considering the 5-year prevalence proportion) – 0.02492%

$d$  = absolute precision or sampling error tolerated = 5%

A sample size of 37 was obtained when the formula was applied based on the island-wide 5-year prevalence data of thyroid cancer patients.

### PCR analysis of samples for BRAF(V600E) mutation

DNA was extracted using the phenol-chloroform method, from 10 µm thick microtome sections of FFPE

thyroid tissue samples (Sambrook & Russel, 1982). The quality of the extracted DNA was determined by agarose gel electrophoresis. Two separate allele-specific PCRs (targeting the wild type and the mutant alleles) were performed on each DNA sample, on Mutant allele specific PCR amplification (MASA) to detect the presence of the BRAF(V600E) gene mutation (Sapio *et al.*, 2006). The wild type allele of the BRAF gene was amplified using a forward primer specific for the wild type allele (5'-GTGATTTTGGTCTAGCTACAGT-3'), with the common reverse primer (5'-GGCCAAAATTTAATCAGTGGA-3') (Davies *et al.*, 2002). The mutant allele of the BRAF gene, with the transversion mutation, was amplified using the forward primer specific for the mutant allele (5'-GTGATTTTGGTCTAGCTACAGA-3') and the common reverse primer.

PCR amplicons were visualized on 2% agarose gels (Santonastaso *et al.*, 2019). A band corresponding to 125 bp indicated the presence of each allele. A DNA sample confirmed by sequencing to contain the mutant allele was used as the positive control. The genotype of the BRAF gene of each study subject was thus established.

### Clinical and socio-demographic information of study subjects

Clinical information such as the histological TC subtype diagnosis, age at incidence, and gender were obtained from the clinical records of the patient. In addition, information on the diet (daily dietary intake patterns based on the types of food) and habits (smoking, alcohol consumption, exercise) associated with the day-to-day lifestyle of the study population were collected using an interviewer-administered questionnaire. For this purpose, the post-surgery patients whose FFPE samples were selected for the study were interviewed on their subsequent visits to the hospital clinic for routine checkups.

### Statistical analysis

SPSS version 26.0 (IBM, USA) software for Microsoft Windows was used for the statistical analyses of data. The Chi-square test was used to compare categorical variables. The Mann-Whitney U test and Kruskal-Wallis test were used to assess the significance of the difference of a variable among groups, where appropriate. The significance level was set at  $p < 0.05$ .

## Ethical clearance

Ethical approval for the study was granted by the Ethics Review Committee of the Institute of Biology, Sri Lanka (Application No: ERC IOBSL 224 10 2020).

## RESULTS AND DISCUSSION

### Prevalence of TC types, demographic, and clinical data of the study cohort

The cohort of malignant TC samples ( $n = 50$ ) of the five TC sub-types (PTC, FTC, MTC, ATC, OC) were representative of the prevalence of each sub-type in the country (Jayarajah *et al.*, 2018). A comparison between the percentages of samples collected for each malignant sub-type, against its prevalence is presented in Table 1.

The study group comprised a total of 76 individuals (67 females and 9 males); the selected samples represented all malignant types of TC present in Sri Lanka (a total of 50;  $n = 33$  of PTC, 11 of FTC, 02 of ATC, 03 of OC, and 01 MTC), while follicular adenoma (FA) samples ( $n = 26$ ) represented a benign thyroid tumour condition. Of the 33 PTC samples, 15 were papillary microcarcinomas.

Demographic details of the study subjects are presented in Table 2. An overall female preponderance was evident, both with the malignant and the benign tumor conditions. Yet, the proportion of females and males with malignant tumours did not significantly differ from the benign tumor condition ( $p > 0.05$ ). Though not statistically significant, the mean age of males at the onset of thyroid tumour was higher than that of females ( $p > 0.05$ ).

**Table 1:** Percentage distribution of each malignant thyroid cancer sub-type within the study cohort and the respective island-wide prevalence

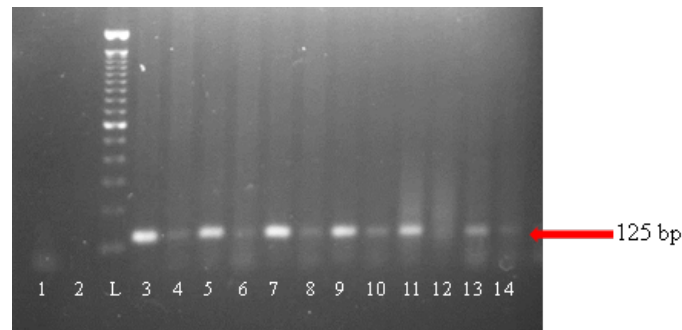
	Subtype of Thyroid Cancer				
	PTC	FTC	MTC	ATC	OC
Prevalence within the study cohort	66%	14.47%	1.31%	2.63%	3.94%
Local prevalence <sup>a</sup>	69%	18.4%	1.8%	3.71%	7.1%

<sup>a</sup> (Jayarajah *et al.*, 2018); PTC- Papillary thyroid cancer; FTC - Follicular thyroid cancer; MTC – Medullary thyroid cancer; ATC – Anaplastic thyroid cancer; OC- oncocytic carcinoma.

**Table 2:** Summary of demographic and clinical data of study subjects

	Diagnosis					
	PTC	FTC	MTC	ATC	OC	FA
Number of patients	42	12	01	02	03	45
Number of females	35	11	1	2	3	36
Number of males	7	1	-	-	-	9
Number of FFPE tissue samples obtained	33	11	01	02	03	26
Mean age (years) ( $\pm$ SD)	45.8 ( $\pm 13.5$ )	52.7 ( $\pm 16.1$ )	73	66 ( $\pm 4.00$ )	52.3 ( $\pm 11.5$ )	46.4 ( $\pm 13.0$ )
Mean age of females (years) ( $\pm$ SD)	43.6 ( $\pm 12.4$ )	49.7 ( $\pm 14.5$ )	73	66 ( $\pm 4.00$ )	52.3 ( $\pm 11.5$ )	44.8 ( $\pm 12.7$ )
Mean age of males (years) ( $\pm$ SD)	49.6 ( $\pm 9.4$ )	73	-	-	-	52.9 ( $\pm 11.6$ )

Values are presented as mean  $\pm$  SD. PTC – Papillary thyroid cancer; FTC - Follicular thyroid cancer; MTC – Medullary thyroid cancer; ATC – Anaplastic thyroid cancer; OC- Oncocytic carcinoma; FA – Follicular adenoma



**Figure 1:** Visualization of BRAF wild-type and mutant allele-specific PCR products of 6 DNA samples. Lanes 1 and 2 - Reagent controls, for the wild type (WT) and the mutant allele, respectively; L – 100 bp ladder; 3, 5, 7, 9, 11, 13 – PCR products of WT alleles and 4, 6, 8, 10, 12, 14 - Mutant alleles of the identical 6 samples.

**Table 3:** BRAF genotype distribution of study subjects

Type of thyroid tumour	Homozygous for wild-type allele	Homozygous for mutant allele	Heterozygous
PTC (n = 33)	04	01	28
FTC (n = 11)	-	-	11
MTC (n = 01)	-	-	01
ATC (n = 02)	-	-	02
OC (n = 03)	-	-	03
FA (n = 26)	03	-	23

PTC – Papillary thyroid cancer; FTC – Follicular thyroid cancer; MTC – Medullary thyroid cancer; ATC – Anaplastic thyroid cancer; OC – oncocytic carcinoma; FA – Follicular adenoma.

### BRAF genotype of the study subjects

The molecular analysis (Figure 1) established that a majority of the study group were heterozygous (n = 68) for the BRAF gene, while one individual diagnosed with PTC, exclusively carried the mutant allele. The rest of the study participants harbored only the wild type allele (n = 4 of PTC; n = 3 of FA). The BRAF genotype of the study participants are presented in Table 3.

Positivity for the BRAF(V600E) mutation was not significant in malignant subtypes, when compared with the benign subtype ( $p > 0.05$ ). The frequency of the BRAF(V600E) mutation among the thyroid tumor subtypes were 87.88% of PTC, 100% of FTC/ MTC/ ATC/ OC, and 88.46% of FA.

As the genetic constitution of a population varies from region to region, and Sri Lanka has a higher degree

of endemism than non-island nations, it was essential to determine the prevalence of the BRAF(V600E) mutation in a cohort of Sri Lankan TC patients. In the current study, of the individuals diagnosed with the conventional PTC (n = 18), only 4 lacked the mutant allele. However, all papillary microcarcinomas (n = 15), a subcategory of PTC, were positive for the mutation. The overall BRAF positivity of PTC subtype (~88%) accounted for a higher proportion than previously reported (Xing, 2005; Silver *et al.*, 2021).

In addition to PTC, positivity for the BRAF V600E mutation was also observed in all FFPE samples of other TC subtypes (FTC, ATC, MTC, and OC) and in most FA samples. Existing studies which report the prevalence of BRAF V600E mutation include study cohorts from Chinese (Xing, 2005; Cong *et al.*, 2024) or European populations (Sapio *et al.*, 2006).

A significant difference was not observed when BRAF positivity in malignant TC subtypes was compared with that of the benign category, FA ( $p > 0.05$ ). While most studies that checked for the presence of BRAF V600E in FA had reported a 0% frequency (Sapio *et al.*, 2006; Yoo *et al.*, 2016), conversely, in a study based on BRAF V600E detection in FNA biopsies, Chen *et al.* (2018) reported that of 36 samples positive for BRAF V600E mutation, 5 were confirmed benign, indicating a prevalence of approximately 14% (Chen *et al.*, 2018). Therefore, although exceedingly rare, BRAF V600E mutation may be present in the benign tumour type FA. It is important to note that the FA samples were not reassessed by histopathology to exclude the missed possibility of the samples carrying the encapsulated follicular variant of PTC. In TC pathogenesis, there is a possibility for the encapsulated variant of PTC to arise in FA. This remains a challenge in FA management and TC diagnosis (Singh *et al.*, 2018). The FA patients, whose FFPE samples harboured the mutant allele, could have a predisposition to malignancy, which could lead to the follicular variant of PTC (Tennakoon *et al.*, 2017; Singh *et al.*, 2018). Longitudinal studies of FA patients are necessary to confirm this possibility.

Considering the genotypes, a single PTC patient sample was homozygous for the mutant allele, while all other BRAF-positive samples were heterozygous, carrying only one mutant allele. Based on the available literature, the presence of one mutant allele is sufficient to induce phenotypic alterations compared to the homozygous wild-type genotype (Sapio *et al.*, 2006). Studies based on PTC have indicated that the percentage of BRAF V600E positive alleles in a tumour area, which mostly consists of a mixture of tumour cells carrying both wild-type and mutant BRAF, is positively correlated with the tumour burden (Guerra *et al.*, 2012; Cheng *et al.*, 2014).

This study was based on the mutant-allele-specific PCR method (Sapio *et al.*, 2006), which has a high sensitivity of detection of 1-5% BRAF V600E mutation in a wild-type background (Yang *et al.*, 2017), comparable to other available, more sensitive, testing methods such as real-time PCR (1-2% mutation detection) with hybridizing probes, and amplification refractory mutation system-PCR (0.5% mutation detection). Other molecular methods such as Sanger sequencing (15 – 20%), high-resolution melting analysis (5-10%), *etc.* have relatively lower sensitivities (Sapio *et al.*, 2006; Huang *et al.*, 2013; Pisareva *et al.*, 2014). Therefore, varying sensitivity of the methods used to detect the mutation may yield incongruent results.

## Risk factors of TC

Risk factors assessed in relation to thyroid tumour onset were family history, gender, age, dietary intake, habits, and body mass index (BMI) of which none significantly correlated with thyroid tumour onset ( $p > 0.05$ ). Most study subjects were between 31–50 years old, at the diagnosis of thyroid tumour. Obesity was prominently observed in the current study group. A significant increment in the mean age of BRAF(V600E) positive individuals, compared to BRAF(V600E) negative individuals, was observed in the PTC subtype ( $p \leq 0.05$ ).

The onset of thyroid tumours in females at an earlier age is speculated to be a result of the sex hormones and pregnancy-related hormones of females at this age, which are clear distinctions from the male physiology (Rahbari *et al.*, 2010). Overall, most patients were aged between 31 and 50 years. When considering the BMI status of study subjects, most individuals belonged to the obese category. The outcome of this pilot study warrants to be confirmed via an in-depth investigation using a larger cohort of TC patients.

The observed female-to-male ratio (6.5:1) was comparable with the age-standardized global incidence rate of TC between the two genders (International Agency for Research on Cancer, 2020). The mean age of onset of TC in males was slightly higher than that of females. This observation was in agreement with the trends observed in previous studies conducted using the data of the Surveillance, Epidemiology and End Results (SEER) Program of the US (Rahbari *et al.*, 2010).

The high disease burden of TC has significant implications for a nation's healthcare system and resource allocation. Surgical facilities and radioactive iodine, the available therapeutic modalities for TC, are in short supply in many developing countries, including Sri Lanka (Jayarajah *et al.*, 2018). Hence, application of molecular markers in TC diagnosis and treatment could benefit the free healthcare system and therefore, the patients of any developing nation.

Oncogenic BRAF gene mutations are significant in TC-based research due to their association with the onset and progression of the disease (Xing, 2005; Wang *et al.*, 2018). Of these, the most commonly reported and extensively studied mutation concerning TC is the BRAF(V600E) mutation (Xing, 2005; Silver *et al.*, 2021). The prevalence and potential of this gene mutation to serve as a molecular marker for TC have

been assessed in several countries. Subsequently, this has led to FDA-approved genetic screening tests for BRAF(V600E) mutation before treatment; drugs which target this mutation and its aberrant products are currently under clinical trials (Crispo *et al.*, 2019). In Sri Lanka, although the clinical significance of molecular markers of TC, including the BRAF(V600E) mutation, has been acknowledged in few recent publications (Tennakoon *et al.*, 2017; Moulana *et al.*, 2018), experimental studies have not been conducted to assess the prevalence of this mutation. Genetic screening is not yet routinely performed on TC patients, locally.

Based on this study, although the applicability of BRAF (V600E) as a prognostic or diagnostic marker for TC remains ambiguous, its potential to serve as a therapeutic marker remains significant. There are small-molecule drugs, recently approved by the FDA, designed to specifically target this mutation's aberrant protein product, such as "vemurafenib" and "dabrafenib" (Crispo *et al.*, 2019). These may have a positive impact on the subgroup of TC patients harboring the mutation. In Sri Lanka, these drugs are yet to be approved for clinical use. Preclinical trials performed elsewhere have indicated that these drugs are capable of restoring the radioactive iodine (RAI) uptake in BRAF (V600E) mutant cells which are resistant to RAI therapy (Crispo *et al.*, 2019). Currently, the FDA has approved the use of the targeted BRAF inhibitor, dabrafenib, along with the MEK inhibitor trametinib, for patients with any type of advanced solid tumours carrying the BRAF V600E mutation. This combination of targeted drugs led to highly promising results in a Phase II clinical trial based on patients (n = 16) with ATC carrying the BRAF V600E mutation, reporting an overall response rate of 69% in patients (Zhang *et al.*, 2023). The study concluded that screening for the mutation should be performed in ATC patients due to the potential of targeted therapy to improve the clinical outcomes of the patients (Subbiah *et al.*, 2018). Hence, screening for the BRAF mutation holds the promise of facilitating more informed therapeutic decisions, especially for patients who do not respond to RAI therapy (Crispo *et al.*, 2019).

Analysis of additional mutations is increasingly utilized for TC prognosis (Xu, 2023; Chindris *et al.*, 2024). The presence of BRAFV600E alongside mutations in other genes such as PIK3CA, the TERT promoter, or TP53 is linked to poorer clinical outcomes. Emerging data indicate that the combination of specific genetic alterations with BRAF mutations could serve as better prognostic tools and predictors of TC aggressiveness (Ahmadi & Landa, 2024).

This pilot study investigated the prevalence of the BRAF V600E mutation among TC patients in Sri Lanka, motivated by its established prognostic and therapeutic significance. A notable limitation of this pilot study is the absence of mutation burden quantification, which may provide a better understanding of the prevalence of BRAF V600E mutation within the tumour area of each TC subtype and the benign tumour, FA. Detailed analysis of PTC subtypes and longitudinal studies focusing on follicular adenoma (FA) patients harboring the BRAF mutation to assess potential predisposition to malignancy in Sri Lankan patients are potential areas of future study. Additionally, the use of fine-needle aspiration (FNA) biopsy samples, alongside formalin-fixed paraffin-embedded (FFPE) tissues, could offer a complementary approach to assess the robustness of diagnostic methodologies of thyroid cancer.

## CONCLUSION

BRAF(V600E) mutation is prevalent in a high proportion of Sri Lankan patients (91%) with thyroid cancer subtypes, and is also mirrored in the benign thyroid tumour, follicular adenoma (FA). However, as the frequency of BRAF(V600E) mutation was not distinct in TC when compared with the benign tumour, its applicability as a diagnostic marker is contentious. Yet, if present, it may serve as an effective therapeutic marker targeted in TC therapy. This can facilitate alternative therapeutic options, especially for TC patients who are resistant to RAI therapy, and for those who manifest cancer recurrence.

## Acknowledgment

Dr. Palitha Ratnayake, Consultant Histopathologist, and other Medical staff of the Histopathology Laboratory, National Hospital, Kandy for providing samples, and the University of Colombo, Sri Lanka for the funding, are gratefully acknowledged.

## REFERENCES

- Ahmadi, S., & Landa, I. (2024). The prognostic power of gene mutations in thyroid cancer. *Endocrine Connections*, 13(2), Article ID: 230297. <https://doi.org/10.1530/EC-23-0297>
- Bozec, A., Ilie, M., Lassalle, S., Hofman, V., Benaim, G., Long, E., Santini, J., & Hofman, P. (2013). Usefulness of Ancillary Methods for Diagnosis, Prognosis and Targeted Therapy in Thyroid Pathology. *Current Medicinal Chemistry*, 20(5), 639–654. <https://doi.org/10.2174/092986713804999376>
- Cancer Incidence and Mortality Data Sri Lanka*. (2014).

- National Cancer Control Programme, Ministry of Health, Sri Lanka. [https://www.nccp.health.gov.lk/storage/post/pdfs/Cancer%20Registry%20-2014%20%20\(March\).pdf](https://www.nccp.health.gov.lk/storage/post/pdfs/Cancer%20Registry%20-2014%20%20(March).pdf)
- Charan, J., & Biswas, T. (2013). How to calculate sample size for different study designs in medical research? *Indian Journal of Psychological Medicine*, 35(2), 121–126. <https://doi.org/10.4103/0253-7176.116232>
- Chen, X., Zhou, Q., Wang, F., Zhang, F., Du, H., Zhang, Q., Wu, W., & Gong, X. (2018). Value of BRAF V600E in high-risk thyroid nodules with benign cytology results. *American Journal of Neuroradiology*, 39(12), 2360–2365. <https://doi.org/10.3174/ajnr.A5898>
- Cheng, S.-P., Hsu, Y.-C., Liu, C.-L., Liu, T.-P., Chien, M.-N., Wang, T.-Y., & Lee, J.-J. (2014). Significance of Allelic Percentage of BRAF c.1799T > G; A (V600E) Mutation in Papillary Thyroid Carcinoma. *Annals of Surgical Oncology*, 21(S4), 619–626. <https://doi.org/10.1245/s10434-014-3723-5>
- Chindris, A. S., Rivera, M., Ma, Y., Nair, A., Liu, Y., Wang, X., Necela, B. M., Kachergus, J. M., Casler, J. D., Brett, C., Mejia, A. M. R., Bernet, V. J., Iii, J. A. C., Knutson, K. L., Thompson, E. A., & Smallridge, R. C. (2024). BRAFV600E / pTERT double mutated papillary thyroid cancers exhibit immune gene suppression. *Frontiers in Endocrinology* 1–14. <https://doi.org/10.3389/fendo.2024.1440722>
- Cong, R., Ouyang, H., Zhou, D., Li, X., & Xia, F. (2024). BRAF V600E mutation in thyroid carcinoma: a large-scale study in Han Chinese population. *World Journal of Surgical Oncology*, 22(1), 259. <https://doi.org/10.1186/s12957-024-03539-7>
- Crispo, F., Notarangelo, T., Pietrafesa, M., Lettini, G., Storto, G., Sgambato, A., Maddalena, F., & Landriscina, M. (2019). Braf inhibitors in thyroid cancer: Clinical impact, mechanisms of resistance and future perspectives. *Cancers*, 11(9). MDPI AG. <https://doi.org/10.3390/cancers11091388>
- Davies, H., Bignell, G. R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M. J., Bottomley, W., Davis, N., Dicks, E., Ewing, R., Floyd, Y., Gray, K., Hall, S., Hawes, R., Hughes, J., Kosmidou, V., ... Futreal, P. A. (2002). Mutations of the BRAF gene in human cancer. *Nature*, 417(6892), 949–954. <https://doi.org/10.1038/nature00766>
- Frasca, F., Nucera, C., Pellegriti, G., Gangemi, P., Attard, M., Stella, M., Loda, M., Vella, V., Giordano, C., Trimarchi, F., Mazzone, E., Belfiore, A., & Vigneri, R. (2008). BRAF(V600E) mutation and the biology of papillary thyroid cancer. *Endocrine-Related Cancer*, 15(1), 191–205. <https://doi.org/10.1677/ERC-07-0212>
- Ge, J., Wang, J., Wang, H., Jiang, X., Liao, Q., Gong, Q., Mo, Y., Li, X., Li, G., Xiong, W., Zhao, J., & Zeng, Z. (2020). The BRAF V600E mutation is a predictor of the effect of radioiodine therapy in papillary thyroid cancer. *Journal of Cancer*, 11(4), 932–939. <https://doi.org/10.7150/jca.33105>
- Gil, M. S. R., Pozas, J., Molina-Cerrillo, J., Gómez, J., Pian, H., Pozas, M., Carrato, A., Grande, E., & Alonso-Gordoa, T. (2020). Current and future role of tyrosine kinases inhibition in thyroid cancer: From biology to therapy. *International Journal of Molecular Sciences*, 21(14), 1–29. <https://doi.org/10.3390/ijms21144951>
- Grodski, S., Brown, T., Sidhu, S., Gill, A., Robinson, B., Learoyd, D., Sywak, M., Reeve, T., & Delbridge, L. (2008). Increasing incidence of thyroid cancer is due to increased pathologic detection. *Surgery*, 144(6), 1038–1043. <https://doi.org/10.1016/j.surg.2008.08.023>
- Guerra, A., Fugazzola, L., Marotta, V., Cirillo, M., Rossi, S., Cirello, V., Forno, I., Moccia, T., Budillon, A., & Vitale, M. (2012). A high percentage of BRAFV600E alleles in papillary thyroid carcinoma predicts a poorer outcome. *Journal of Clinical Endocrinology and Metabolism*, 97(7), 2333–2340. <https://doi.org/10.1210/jc.2011-3106>
- Haugen, B. R., Alexander, E. K., Bible, K. C., Doherty, G. M., Mandel, S. J., Nikiforov, Y. E., Pacini, F., Randolph, G. W., Sawka, A. M., Schlumberger, M., Schuff, K. G., Sherman, S. I., Sosa, J. A., Steward, D. L., Tuttle, R. M., & Wartofsky, L. (2016). 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*, 26(1), 1–133. <https://doi.org/10.1089/thy.2015.0020>
- Howell, G. M., Hodak, S. P., & Yip, L. (2013). RAS Mutations in Thyroid Cancer. *The Oncologist*, 18(8), 926–932. <https://doi.org/10.1634/theoncologist.2013-0072>
- Huang, T., Zhuge, J., & Zhang, W. W. (2013). Sensitive detection of BRAF V600E mutation by Amplification Refractory Mutation System (ARMS)-PCR. *Biomarker Research* <https://doi.org/10.1186/2050-7771-1-3>
- International Agency for Research on Cancer. (2020). *Globocan 2020*.
- Ito, Y., Yoshida, H., Kihara, M., Kobayashi, K., Miya, A., & Miyauchi, A. (2014). BRAFV600E mutation analysis in papillary thyroid carcinoma: Is it useful for all patients? *World Journal of Surgery*, 38(3), 679–687. <https://doi.org/10.1007/s00268-013-2223-2>
- Jayarajah, U., Fernando, A., Prabashani, S., Fernando, E. A., & Seneviratne, S. A. (2018). Incidence and histological patterns of thyroid cancer in Sri Lanka 2001-2010: An analysis of national cancer registry data. *BMC Cancer*, 18(1), 1–7. <https://doi.org/10.1186/s12885-018-4083-5>
- Jung, C. K., Little, M. P., Lubin, J. H., Brenner, A. V., Wells, S. A., Jr., Sigurdson, A. J., & Nikiforov, Y. E. (2014). The Increase in Thyroid Cancer Incidence During the Last Four Decades Is Accompanied by a High Frequency of BRAF Mutations and a Sharp Increase in RAS Mutations. *The Journal of Clinical Endocrinology and Metabolism*, 99(2), E276. <https://doi.org/10.1210/JC.2013-2503>
- Kebebew, E., Weng, J., Bauer, J., Ranvier, G., Clark, O. H., Duh, Q. Y., Shih, D., Bastian, B., & Griffin, A. (2007). The prevalence and prognostic value of BRAF mutation in thyroid cancer. *Annals of Surgery*, 246(3), 466–470. <https://doi.org/10.1097/SLA.0b013e318148563d>
- Kilfoy, B. A., Zheng, T., Holford, T. R., Han, X., Ward, M. H., Sjoedin, A., Zhang, Y., Bai, Y., Zhu, C., Guo, G. L., Rothman, N., & Zhang, Y. (2009). International patterns and trends in thyroid cancer incidence, 1973-2002. *Cancer Causes and Control*, 20(5), 525–531. <https://doi.org/10.1007/s10552-009-0000-0>

- 008-9260-4
- Kobawala, T. P., Trivedi, T. I., Gajjar, K. K., Patel, D. H., Patel, G. H., & Ghosh, N. R. (2016). Significance of Interleukin-6 in Papillary Thyroid Carcinoma. *Journal of Thyroid Research*, 2016. <https://doi.org/10.1155/2016/6178921>
- Lassalle, S., Hofman, V., Ilie, M., Butori, C., Bozec, A., Santini, J., Vielh, P., & Hofman, P. (2010). Clinical Impact of the Detection of BRAF Mutations in Thyroid Pathology: Potential Usefulness as Diagnostic, Prognostic and Theragnostic Applications. *Current Medicinal Chemistry*, 17(17), 1839–1850. <https://doi.org/10.2174/092986710791111189>
- Liska, J., & Galbavy, S. (2005). Thyroid tumors : Histological classification and genetic factors involved in the development of thyroid cancer. *Endocrine Regulations*, 39(October), 73–83. <https://pubmed.ncbi.nlm.nih.gov/16468229/>
- Liu, S., Semenciw, R., Ugnat, A. M., & Mao, Y. (2001). Increasing thyroid cancer incidence in Canada, 1970-1996: Time trends and age-period-cohort effects. *British Journal of Cancer*, 85(9), 1335–1339. <https://doi.org/10.1054/bjoc.2001.2061>
- McFadden, D. G., & Sadow, P. M. (2021). Genetics, Diagnosis, and Management of Hürthle Cell Thyroid Neoplasms. *Frontiers in Endocrinology*, 12(June), 1–9. <https://doi.org/10.3389/fendo.2021.696386>
- Moulana, F. I., Priyani, A. A. H., de Silva, M. V. C., & Dassanayake, R. S. (2018). BRAF-Oncogene-Induced Senescence and the Role of Thyroid-Stimulating Hormone Signaling in the Progression of Papillary Thyroid Carcinoma. *Hormones and Cancer*, 9(1). <https://doi.org/10.1007/s12672-017-0315-4>
- Pisareva, E., Gutkina, N., Kovalenko, S., Kuehnepfel, S., Hartmann, A., Heinzerling, L., Schneider-Stock, R., Lyubchenko, L., & Shamanin, V. A. (2014). Sensitive allele-specific real-time PCR test for mutations in BRAF codon V600 in skin melanoma. *Melanoma Research*, 24(4), 322–331. <https://doi.org/10.1097/CMR.0000000000000090>
- Prasad, M. L., Vyas, M., Horne, M. J., Virk, R. K., Morotti, R., Liu, Z., Tallini, G., Nikiforova, M. N., Christison-Lagay, E. R., Udelsman, R., Dinauer, C. A., & Nikiforov, Y. E. (2016). NTRK fusion oncogenes in pediatric papillary thyroid carcinoma in northeast United States. *Cancer*, 122(7), 1097–1107. <https://doi.org/10.1002/cncr.29887>
- Rahbari, R., Zhang, L., & Kebebew, E. (2010). Thyroid cancer gender disparity. In *Future Oncology* (Vol. 6, Issue 11, pp. 1771–1779). NIH Public Access. <https://doi.org/10.2217/fon.10.127>
- Raman, P., & Koenig, R. J. (2014). Pax-8-PPAR- $\gamma$  3 fusion protein in thyroid carcinoma. *Nature Reviews Endocrinology*, 10(10), 616–623. <https://doi.org/10.1038/nrendo.2014.115>
- Santonastaso, C., Di Paolo, M., Tommaselli, A., Morelli, L., Pugliese, S., Troisi, A., Muto, T., & Di Martino, S. (2019). A simple and high sensitive method for detection of BRAF 1799T>A (V600E) mutation in the thyroid fine needle aspirate. *World Cancer Research Journal*, 6, 1–6. <http://frodo.wi.mit.edu/primer3/>
- Sapio, M. R., Posca, D., Troncone, G., Pettinato, G., Palombini, L., Rossi, G., Fenzi, G., & Vitale, M. (2006). Detection of BRAF mutation in thyroid papillary carcinomas by mutant allele-specific PCR amplification ( MASA ). 341–348. <https://doi.org/10.1530/eje.1.02072>
- Silver, J. A., Bogatchenko, M., Pusztaszeri, M., Forest, V. I., Hier, M. P., Yang, J. W., Tamilia, M., & Payne, R. J. (2021). BRAF V600E mutation is associated with aggressive features in papillary thyroid carcinomas  $\leq 1.5$  cm. *Journal of Otolaryngology - Head and Neck Surgery*, 50(1), 1–8. <https://doi.org/10.1186/s40463-021-00543-9>
- Singh, K., Pujani, M., Chauhan, V., Agarwal, C., & Dhingra, S. (2018). Encapsulated Follicular Variant of Papillary Thyroid Carcinoma Arising in a Follicular Adenoma: a Diagnostic Dilemma. *Indian Journal of Surgical Oncology*, 9(3), 414–417. <https://doi.org/10.1007/s13193-018-0801-3>
- Soares, P., Póvoa, A. A., Melo, M., Vinagre, J., Máximo, V., Eloy, C., Cameselle-Teijeiro, J. M., & Sobrinho-Simões, M. (2021). Molecular Pathology of Non-familial Follicular Epithelial-Derived Thyroid Cancer in Adults: From RAS/BRAF-like Tumor Designations to Molecular Risk Stratification. *Endocrine Pathology*, 32(1), 44–62. <https://doi.org/10.1007/s12022-021-09666-1>
- Subbiah, V., Cabanillas, M. E., Kreitman, R. J., Wainberg, Z. A., Cho, J. Y., Keam, B., Schellens, J. H. M., Soria, J. C., Wen, P. Y., Zielinski, C., Urbanowitz, G., Mookerjee, B., Wang, D., & Rangwala, F. (2018). Dabrafenib and trametinib treatment in patients with locally advanced or metastatic BRAF V600-mutant anaplastic thyroid cancer. *Journal of Clinical Oncology*, 36(1), 7–13. <https://doi.org/10.1200/JCO.2017.73.6785>
- Tennakoon, T. M. P. B., Rushdhi, M., Ranasinghe, A. D. C. U., & Dassanayake, R. S. (2017). Values of molecular markers in the differential diagnosis of thyroid abnormalities. *Journal of Cancer Research and Clinical Oncology*, 143(6), 913–931. <https://doi.org/10.1007/s00432-016-2319-9>
- Volante, M., Lam, A. K., Papotti, M., & Tallini, G. (2021). Molecular Pathology of Poorly Differentiated and Anaplastic Thyroid Cancer: What Do Pathologists Need to Know? *Endocrine Pathology*, 32(1), 63–76. <https://doi.org/10.1007/s12022-021-09665-2>
- Wang, F., Zhao, S., Shen, X., Zhu, G., Liu, R., Viola, D., Elisei, R., Puxeddu, E., Fugazzola, L., Colombo, C., Jarzab, B., Czarniecka, A., Lam, A. K., Mian, C., Vianello, F., Yip, L., Riesco-Eizaguirre, G., Santisteban, P., O'Neill, C. J., ... Xing, M. (2018). BRAF V600E confers male sex disease-specific mortality risk in patients with papillary thyroid cancer. *Journal of Clinical Oncology*, 36(27), 2787–2795. <https://doi.org/10.1200/JCO.2018.78.5097>
- Wang, Y., & Wang, W. (2015). Increasing incidence of thyroid cancer in Shanghai, China, 1983-2007. *Asia-Pacific Journal of Public Health*, 27(2), NP223–NP229. <https://doi.org/10.1177/1010539512436874>
- Warakomski, J., Romuk, E., Jarzab, B., Krajewska, J., & Siemińska, L. (2018). Concentrations of selected Adipokines, interleukin-6, and Vitamin D in patients with papillary thyroid carcinoma in respect to thyroid cancer stages. *International Journal of Endocrinology*, 2018. <https://doi.org/10.1155/2018/4921803>

- Xing, M. (2005). BRAF mutation in thyroid cancer. *Endocrine-Related Cancer*, 12(2), 245–262. <https://doi.org/10.1677/erc.1.0978>
- Xing, M. (2019). Entering an Era of Precision Management of Thyroid Cancer. *Endocrinology and Metabolism Clinics of North America*, 48(1), xvii–xviii. <https://doi.org/10.1016/j.ecl.2018.12.001>
- Xu, B. (2023). Molecular alterations of follicular cell-derived thyroid neoplasms. *Diagnostic Histopathology*, 29(11), 487–494. <https://doi.org/10.1016/j.mpdhp.2023.07.007>
- Yang, Z., Zhao, N., Chen, D., Wei, K., Su, N., Huang, J. F., Xu, H. Q., Duan, G. J., Fu, W. L., & Huang, Q. (2017). Improved detection of BRAF V600E using allele-specific PCR coupled with external and internal controllers. *Scientific Reports*, 7(1), 1–12. <https://doi.org/10.1038/s41598-017-14140-2>
- Yoo, S. K., Lee, S., Kim, S. J., Jee, H. G., Kim, B. A., Cho, H., Song, Y. S., Cho, S. W., Won, J. K., Shin, J. Y., Park, D. J., Kim, J. Il, Lee, K. E., Park, Y. J., & Seo, J. S. (2016). Comprehensive Analysis of the Transcriptional and Mutational Landscape of Follicular and Papillary Thyroid Cancers. *PLoS Genetics*, 12(8), 1–23. <https://doi.org/10.1371/journal.pgen.1006239>
- Zhang, L., Feng, Q., Wang, J., Tan, Z., Li, Q., & Ge, M. (2023). Molecular basis and targeted therapy in thyroid cancer: Progress and opportunities. *Biochimica et Biophysica Acta - Reviews on Cancer*, 1878(4), 188928. <https://doi.org/10.1016/j.bbcan.2023.188928>



## RESEARCH ARTICLE

### Fuzzy Mathematics

# Networking model based on fuzzy graphs of almost primary fuzzy ideals of near-rings

A Muhammad<sup>1</sup>, A Taouti<sup>2</sup>, WA Khan<sup>3</sup> and M Ashraf<sup>1</sup>

<sup>1</sup> Department of Mathematics, University of Wah, Wah Cantt, Pakistan.

<sup>2</sup> ETS-Maths and NS Engineering Division, HCT, University City P.O. Box 7947, Sharjah, United Arab Emirates.

<sup>3</sup> Division of Science and Technology, Department of Mathematics, University of Education Lahore, Attock Campus, Pakistan.

Submitted: 06 March 2024; Revised: 01 September 2024; Accepted: 09 January 2025

**Abstract:** Sometimes traditional mathematical methods do not work properly to find the solutions of the problems with uncertainties in different field of sciences. The reason is that there are lot of complexities and ambiguities which occur in real-world problems. Researchers from all across the world have created new theories such as fuzzy set theory, rough set theory, *etc.*, to overcome such circumstances. Over the past few decades numerous fuzzy algebraic structures and substructures have been studied, like fuzzy rings, fuzzy near-rings, fuzzy ideals of near-rings, and so on. In this study, we present the notion of almost primary fuzzy ideals (almost primary  $f$ -ideals) of near-rings and provide their application through the fuzzy graphs associated with them. Firstly, we provide numerous characterizations of almost primary  $f$ -ideals in near-rings. We observe that every almost primary  $f$ -ideal is not a primary fuzzy ideal (PFI) in near-rings, and show when a primary fuzzy ideal is an almost primary  $f$ -ideals in near-rings. Some inter-relationships between a classical almost primary ideal (API) and almost primary  $f$ -ideals of near-rings are also established. Finally, we introduce the fuzzy graph structure related to APFIs of near-rings and present a model of general networks.

**Keywords:** Almost primary fuzzy ideal in near-rings, fuzzy graphs, fuzzy ideals of near-rings, fuzzy prime ideal of near-rings.

## INTRODUCTION

Zadeh (1965) introduced the notion of fuzzy sets (FSs). In FSs, each element was allocated a value from the interval  $[0, 1]$ . Due to the flexible nature of FSs,

many generalizations of FSs have been explored in the literature. The very first extension of the FSs was also presented by Zadeh (1975) and termed as the interval-valued FSs. Afterwards, different extensions of FSs such as intuitionistic FSs (Atanassov & Atanassov, 1999), bipolar FSs (Zhang, 1998), Pythagorean FSs (Razaq & Alhamzi, 2023), picture FSs (Cuong, 2014), *etc* have been added in the literature.

FSs were utilized in different fields especially where uncertainties were involved. FSs have resolved many issues in decision-making procedures involving vague information, enabling more flexible and sensitive assessments in contexts such as multi-criteria decision-making. Fuzzy logic controllers, particularly in engineering, are useful for controlling complicated systems like robots and automated temperature control because they approximate human thinking. Due to their ability to handle pixel value discrepancies, FSs are especially useful in image processing, where they aid in tasks like edge recognition and noise reduction. On the other hands, fuzzy graphs (FGs) were first introduced by Rosenfeld (1975). FGs have fuzzy associations between nodes which extend the idea of classical graphs. Features of FGs become helpful in solving problems with uncertainties in different fields, such as networking, medicine, engineering *etc*. They have also been used in clustering algorithms to enhance the way that overlapping data points and unclear group memberships are involved.

\* Corresponding author (sirwak2003@yahoo.com;  <https://orcid.org/0000-0002-7009-5493>)



This article is published under the Creative Commons CC-BY-ND License (<http://creativecommons.org/licenses/by-nd/4.0/>). This license permits use, distribution and reproduction, commercial and non-commercial, provided that the original work is properly cited and is not changed in anyway.

FSs and FGs become useful tools for controlling ambiguity and imprecision in a variety of areas from data mining to artificial intelligence (Bloch & Ralescu, 2023).

In the last few decades, many concepts of classical rings, near-rings, and near-semirings have been shifted towards FSs theory. In this regard, the concepts of fuzzy ideals (in short,  $f$ -ideals) in rings, near-rings, and near-semirings have been introduced in the literature. Liu (1982) introduced  $f$ -ideals in rings. If for every  $x, y \in R, J(x \cdot y) \geq J(x)$  (resp.,  $J(x \cdot y) \geq J(y)$ ), then we call it a  $f$ -left (resp. right) ideal of a ring  $R$ . The fuzzy prime ideal in a ring was first introduced by Mukherjee et al. (1987). The primary  $f$ -ideal in ring was introduced by Malik and Mordeson (1991). Various kinds of prime ideals of near-rings have been introduced by Birkenmeier et al. (1993), Elavarasan (2011), and Groenewald (1991). Birkenmeier et al. (1993) discussed the link between the prime ideal and type-1 prime ideal. Elavarasan (2011) introduced the almost prime ideal in near-rings. The concepts of  $v$ -primary ideals and their graphs have been discussed by Muhammad et al. (2023). Recently, Khan et al. (2018) introduced almost prime ideals in  $\Gamma$ -near-rings. The concepts of  $f$ -ideal and prime  $f$ -ideal of  $\mathcal{N}$  were introduced by Abou-Zaid (1991). Different types of prime  $f$ -ideals in near-rings were introduced by Kedukodi et al. (2009).

Recently, Razaq and Alhamzi (2023) examined many algebraic aspects of Pythagorean  $f$ -ideals in classical rings. Onar et al. (2023) extended the understanding of algebraic structures within G-rings by introducing the concept of non-symmetric 2-absorbing vague weakly complete G-ideals in commutative G-rings. Jebapresitha (2024) explored the hybrid structure of FSs within near-rings, effectively transformed  $f$ -ideals into hybrid ideals and fuzzy maximal ideals into hybrid maximal ideals. The transformation of  $f$ -ideals into hybrid ideals, such as fuzzy maximal ideals into hybrid maximal ideals, addressed the uncertainty problems more efficiently. Subsequently, Batool et al. (2023) discovered intuitionistic multi-fuzzy sub-near-rings and ideals within near-rings, exploring foundational operations and relationships within these structures. Adak et al. (2023) studied the concepts of semi-prime ideals within the context of picture FSs. They introduced the methods for constructing picture fuzzy regular and intra-regular ideals along with fundamental insights related to these constructions.

Numerous graph structures associated with algebraic structures such as rings, near-rings, ideals of rings and near-rings, etc., have been discussed in the literature.

Similarly, fuzzy graphs linked with fuzzy algebraic structures have also been explored. The graphs based on ring structure consist of ring elements as vertices of graphs, while the edges of the graphs represent some operations of the rings, including addition and multiplication. In contrast to conventional graphs, the vertices and edges of the FGs represent a membership value that measures the degree of membership and the strength of the link. FGs of rings become useful in many fields like coding theory, cryptography or network analysis. Algebraic structures usually interact with uncertain data because they enable the modelling of imprecise interactions inside the ring. This method allows for studying uncertainties that may occur in real-world applications, and allows for a more clear understanding of the interactions inside the ring by adding fuzziness to the graphs (Mordeson et al., 2023).

Jagadeesha et al. (2023), introduced equiprime fuzzy graphs associated with a near-ring in relation to the level ideal of a  $f$ -ideal. They discussed the theoretical properties of both the graphs and ideals, and demonstrated their key characteristics such as vertex cut and connectedness influenced by the properties of the  $f$ -ideal. Srinivas and Prasad (2024) introduced the notion of a c-prime  $f$ -graph associated with a near-ring. They focused on the relationship between the properties of a  $f$ -ideal and their corresponding fuzzy graphs. Additionally, a connection between near-ring homomorphism and graph homomorphism was also described by them. They also considered that the homomorphic image of a c-prime  $f$ -ideal remains a c-prime  $f$ -ideal.

## Research gap

From literature review, we come to know that fuzzy set theory in relation to rings and near-rings is playing a key role in solving the problems lying in different fields of sciences. Within this fuzzy framework, classical rings and ideals have been transformed into fuzzy rings and fuzzy ideals. In this regard, prime and primary  $f$ -ideals of near-rings have been defined parallel to rings. However, almost primary  $f$ -ideals of near-rings have not been discussed to date. To fill this gap, we initiate the idea of almost primary  $f$ -ideal in near-rings. In addition, we explore fuzzy graphs associated with these ideals and present their application in networking.

## Novelty

The novelty of this study can be described as follows.

- Firstly, we introduce the notion of almost primary  $f$ -ideals of near-rings.
- Numerous characterizations of almost primary  $f$ -ideals in near-rings are discussed.

- iii. We consider that every almost primary  $f$ -ideal is not a primary fuzzy ideal (PFI) in near-rings and find a condition when a primary fuzzy ideal is almost primary  $f$ -ideals in near-rings.
- iv. For near-rings, we establish some inter-relationships between the classical almost primary ideal (API) and almost primary  $f$ -ideals.
- v. Finally, we present the fuzzy graph structures related to APFIs of near-rings and their application in networking.

## METHODOLOGY

This section consists of basic concepts related to the existing algebraic structures like rings, near-rings, semi-near-rings, and their ideals and fuzzy ideals. These concepts will help to understand the presented work.

An algebraic structure  $(\mathcal{N}, +, \cdot)$  is called a (right) near-ring, if  $\mathcal{N}$  is a group (not necessarily Abelian) under “+”, semigroup under “ $\cdot$ ” and fulfils (right) distributive law. For the elementary notions of near-rings, we refer Pilz (2011). Right ideal of  $\mathcal{N}$  is a subset  $\mathcal{J}$  of  $\mathcal{N}$ , where (i)  $(\mathcal{J}, +)$  is normal subgroup of  $(\mathcal{N}, +)$ , (ii) For all  $\mathcal{J}\mathcal{N} \subseteq \mathcal{J}$ . Similarly,  $\mathcal{J}$  is a left ideal, if (i)  $(\mathcal{J}, +) \trianglelefteq (\mathcal{N}, +)$ , (ii)  $a_1 \cdot (a_2 + i) - a_1 \cdot a_2 \in \mathcal{J}$ , for every  $a_1, a_2 \in \mathcal{N}$  and  $i \in \mathcal{J}$ . An ideal  $\mathcal{J}$  of  $\mathcal{N}$  is called prime, if for every  $\mathcal{J}_1, \mathcal{J}_2 \subset \mathcal{N}$ ,  $\mathcal{J}_1\mathcal{J}_2 \subseteq \mathcal{J} \Rightarrow \mathcal{J}_1 \subseteq \mathcal{J}$  or  $\mathcal{J}_2 \subseteq \mathcal{J}$ . Various kinds of prime ideals of near-rings have been established in (Birkenmeier *et al.*, 1993; Elavarasan, 2011; Groenewald, 1991).

Now we present some useful terminologies which would be helpful in understanding the forthcoming literature.

**Definition 1.** (Groenewald, 1991)  $\mathcal{J}$  is called a 0-prime ideal of  $\mathcal{N}$ , if  $\mathcal{J}_1\mathcal{J}_2 \subseteq \mathcal{J} \Rightarrow \mathcal{J}_1 \subseteq \mathcal{J}$  or  $\mathcal{J}_2 \subseteq \mathcal{J}$ , where  $\mathcal{J}_1, \mathcal{J}_2$  are ideals of near-ring.

**Definition 2.** (Groenewald, 1991) An ideal  $\mathcal{J}$  is called a prime ideal, if  $\mathcal{J}_1\mathcal{J}_2 \subseteq \mathcal{J} \Rightarrow \mathcal{J}_1 \subseteq \mathcal{J}$  or  $\mathcal{J}_2 \subseteq \mathcal{J}$ , where  $\mathcal{J}_1$  &  $\mathcal{J}_2$  are ideals of near-ring.

**Definition 3.** (Groenewald, 1991)  $\mathcal{J}$  is said to be 3-prime ideal of  $\mathcal{N}$ , if  $a\mathcal{N}b \subseteq \mathcal{J}$ , then  $a \in \mathcal{J}$  or  $b \in \mathcal{J}$ , for  $a, b \in \mathcal{N}$ .

**Definition 4.** (Elavarasan, 2011)  $\mathcal{J}$  is called an almost prime ideal if for  $\mathcal{J}_1\mathcal{J}_2 \subseteq \mathcal{J}$  and  $\mathcal{J}_1\mathcal{J}_2 \not\subseteq \mathcal{J}^2 \Rightarrow \mathcal{J}_1 \subseteq \mathcal{J}$  or  $\mathcal{J}_2 \subseteq \mathcal{J}$  where  $\mathcal{J}_1$  &  $\mathcal{J}_2$  are ideals of  $\mathcal{N}$ .

**Theorem 1.** (Elavarasan, 2011) Let  $\mathcal{N}$  be a near-ring and  $\mathcal{J}$  be an almost prime ideal of  $\mathcal{N}$ . Then  $\mathcal{J}^2 = \mathcal{J}$ , if  $\mathcal{J}$  is a prime ideal.

**Corollary 1.** (Elavarasan, 2011) Let  $\mathcal{J}$  be a proper ideal of  $\mathcal{N}$  and if  $\mathcal{J}$  is an almost prime with  $(\mathcal{J}^2: \mathcal{J}) \subseteq \mathcal{J}$ , then  $\mathcal{J}$  is a prime ideal.

**Definition 5.** (Zadeh, 1965) A fuzzy set  $T$  on a non-empty set  $U$  is described as  $= \{(u, r(u)): r(u) \in [0, 1]\}$ .

**Definition 6.** (Rosenfeld, 1975) A fuzzy graph is a pair  $G = (C, D)$  where  $C = \{\rho_c\}$  and  $D = \{\rho_d\}$  such that  $\rho_c: V \rightarrow [0, 1]$  and  $\rho_d: V \times V \rightarrow [0, 1]$ . Then,  $\rho_d(x, w) \leq \rho_c(x) \wedge \rho_c(w)$ .

**Definition 7.** (Mukherjee *et al.*, 1987) A  $f$ -ideal  $\mathcal{P}$  of a  $R$  is a prime fuzzy ideal (prime  $f$ -ideal), if  $\mathcal{J}_1 \circ \mathcal{J}_2 \subseteq \mathcal{P}$  implies  $\mathcal{J}_1 \subseteq \mathcal{P}$  or  $\mathcal{J}_2 \subseteq \mathcal{P}$ , where  $\mathcal{J}_1, \mathcal{J}_2 \in R$ .

**Definition 8.** (Malik & Mordeson, 1991) An ideal  $\mathcal{P}$  is called a primary  $f$ -ideal of a ring, if the values of  $\mathcal{P}$  is not a constant and  $\mathcal{P}_1 \circ \mathcal{P}_2 \subseteq \mathcal{P}$  implies  $\mathcal{P}_1 \subseteq \mathcal{P}$  or  $\mathcal{P}_2 \subseteq \sqrt{\mathcal{P}}$ , where  $\mathcal{P}_1, \mathcal{P}_2$  are  $f$ -ideals.

**Definition 9.** (Abou-Zaid, 1991) An  $f$ -ideal  $\mathcal{J}$  is called a right  $f$ -ideal of  $\mathcal{N}$ , if  $\mathcal{J}$  is a normal  $f$ -subgroup under addition and  $\mathcal{J}(n_1 \cdot n_2) \geq \mathcal{J}(n_1)$ , for all  $n_1, n_2 \in \mathcal{N}$ . Similarly,  $\mathcal{J}$  is called a left  $f$ -ideal of  $\mathcal{N}$ , if  $\mathcal{J}$  is normal  $f$ -subgroup under addition and  $\mathcal{J}(a \cdot (b + i) - a \cdot b) \geq \mathcal{J}(i)$ , for every  $a, b, i \in \mathcal{N}$ .

**Definition 10.** (Kedukodi *et al.* 2009) An ideal  $\mathcal{P}$  of  $\mathcal{N}$  is said to be a  $c$ -prime  $f$ -ideal, if for every  $a, b \in \mathcal{N}$ ,  $\max\{\mathcal{P}(a), \mathcal{P}(b)\} \geq \mathcal{P}(a \cdot b)$ . Similarly,  $\mathcal{P}$  is called a 3-prime  $f$ -ideal, if  $\max\{\mathcal{P}(a), \mathcal{P}(b)\} \geq \mathcal{P}(a \cdot n \cdot b)$ , for every  $a, b, n \in \mathcal{N}$ .

**Definition 11.** (Mukherjee & Sen, 1987) A  $f$ -ideal  $\mathcal{P}$  of  $R$  is called an primary  $f$ -ideal if for  $a, b \in R$  such that  $\mathcal{P}(ab) \geq \mathcal{P}(a)$  implies  $\mathcal{P}(b^n) \geq \mathcal{P}(ab)$ .

**Theorem 2.** (Mukherjee & Sen, 1987) Let  $\mathcal{P}$  be an ideal of  $R$ . Then its characteristic function is a fuzzy primary ideal of  $R$  if and only if  $\mathcal{P}$  is a primary ideal.

**Theorem 3.** (Mukherjee & Sen, 1987) Suppose  $R$  is a ring, then every  $f$ -prime ideal in  $R$  is a  $f$ -almost prime ideal in  $R$ .

## RESULTS AND DISCUSSIONS

In this section, we introduce the notion of almost primary  $f$ -ideals in near-rings and explore some of the generalizations of the almost primary  $f$ -ideals of near-rings. In addition, we also provide several characterizations of these ideals.

**Definition 12.** An  $f$ -ideal  $\mathcal{P}$  of  $\mathcal{N}$  is said to be

Tables set. 1

+	0	1	2	3	4	5	6	7
0	0	1	2	3	4	5	6	7
1	2	3	0	1	7	6	4	5
2	2	3	0	1	5	4	7	6
3	3	0	1	2	6	7	5	4
4	4	7	5	6	2	0	1	3
5	5	6	4	7	0	2	3	1
6	6	4	7	5	1	3	0	2
7	7	5	6	4	3	1	2	0

almost primary  $f$ -ideal, if  $\mathcal{P}(a \cdot b) > \mathcal{P}(1_{\mathcal{N}})$  and  $(\mathcal{P} \circ \mathcal{P})(a \cdot b) = \mathcal{P}(1_{\mathcal{N}})$ , we have  $\mathcal{P}(a \cdot b) \geq \mathcal{P}(a^n)$  or  $\mathcal{P}(a \cdot b) \geq \mathcal{P}(b^n)$ , for all  $a, b, c \in \mathcal{N}$  and some  $n \in \mathbb{Z}^+$ , where

$$(\mathcal{P} \circ \mathcal{P})(c) = \begin{cases} \sup(\min\{\mathcal{P}(a), \mathcal{P}(b)\}) & \text{for } a \cdot b = c \\ 0, & a \cdot b \neq c \end{cases}$$

**Example 1.** Let us consider a near-ring defined in tables set 1.

·	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7
2	0	2	0	2	2	2	0	0
3	0	3	2	1	5	4	6	7
4	0	4	2	5	4	5	6	7
5	0	6	2	4	5	4	6	7
6	0	6	0	6	0	0	0	0
7	0	7	0	7	2	2	0	0

Consider a mapping  $\mathcal{P}: \mathcal{N} \rightarrow [0,1]$  such that  $\mathcal{P}(0) = \mathcal{P}(1) = \mathcal{P}(3) = \mathcal{P}(5) = \mathcal{P}(1_{\mathcal{N}}) = 0, \mathcal{P}(2) = \mathcal{P}(4) = \mathcal{P}(6) = \mathcal{P}(7) = \alpha$ , where  $\alpha \in (0,1]$ . Here,  $\mathcal{P}$  is an almost primary  $f$ -ideal. To verify the definition we chose  $3, 5 \in \mathcal{N}$  such that  $\mathcal{P}(3 \cdot 6) = \mathcal{P}(6) = \alpha > \mathcal{P}(1_{\mathcal{N}}) = 0$  and  $(\mathcal{P} \circ \mathcal{P})(6) = \sup_{6=3 \cdot 6}(\min\{\mathcal{P}(3), \mathcal{P}(6)\}) = \sup_{6=3 \cdot 6}(\min\{0, \alpha\}) = 0 = \mathcal{P}(1_{\mathcal{N}})$ , then  $\Rightarrow \mathcal{P}(3 \cdot 6) = \alpha \geq \mathcal{P}(3) = 0$  or  $\mathcal{P}(3 \cdot 6) \geq \mathcal{P}(6^2) = \mathcal{P}(0) = 0$ . Similarly,  $5, 7 \in \mathcal{N}$  such that  $\mathcal{P}(5 \cdot 7) = \mathcal{P}(7) = \alpha > \mathcal{P}(1_{\mathcal{N}}) = 0$   $\mathcal{P}(5 \cdot 7) = \mathcal{P}(7) = \alpha > \mathcal{P}(1_{\mathcal{N}}) = 0$  and  $(\mathcal{P} \circ \mathcal{P})(7) = \sup_{7=5 \cdot 7}(\min\{\mathcal{P}(5), \mathcal{P}(7)\}) = \sup_{7=5 \cdot 7}(\min\{0, \alpha\}) = 0 = \mathcal{P}(1_{\mathcal{N}})$ ,  $\Rightarrow \mathcal{P}(5 \cdot 7) = \alpha \geq \mathcal{P}(5) = 0$  or  $\mathcal{P}(5 \cdot 7) \geq \mathcal{P}(7^2) = \mathcal{P}(0) = 0$ . According to the definition of primary  $f$ -ideal, we can easily verify that  $\mathcal{P}$  is not a primary  $f$ -ideal. For this, let  $2, 4 \in \mathcal{N}$  such that  $\mathcal{P}(2 \cdot 4) = \mathcal{P}(2) = \alpha \leq \mathcal{P}(2) = 0$  but  $\mathcal{P}(2 \cdot 4) = \mathcal{P}(2) = \alpha > \mathcal{P}(4^2) = \alpha$ , which contradicts the definition of primary  $f$ -ideal of  $\mathcal{N}$ .

**Remarks.** Every almost primary  $f$ -ideal is not a primary  $f$ -ideal of near-rings.

**Proposition 1.** Let  $\mathcal{N}$  be a near-ring. Then every primary  $f$ -ideal of  $\mathcal{N}$  is an almost primary  $f$ -ideal of  $\mathcal{N}$ .

*Proof.* Let  $\mathcal{P}$  be a primary  $f$ -ideal of  $\mathcal{N}$  and  $a, b \in \mathcal{N}$  such that  $\mathcal{P}(a \cdot b) > \mathcal{P}(1_{\mathcal{N}})$  and  $(\mathcal{P} \circ \mathcal{P})(a \cdot b) = \mathcal{P}(1_{\mathcal{N}})$ . As  $\mathcal{P}$  is primary  $f$ -ideal then  $\mathcal{P}(a \cdot b) \leq \mathcal{P}(a)$  and  $\mathcal{P}(a \cdot b) > \mathcal{P}(b^n)$  for some  $n \in \mathbb{Z}^+$ , so without loss of generality, we have  $\mathcal{P}(b^n) = \min\{\mathcal{P}(a), \mathcal{P}(b)\} = \mathcal{P}(1_{\mathcal{N}})$ , then  $(\mathcal{P})(a \cdot b) > \mathcal{P}(b^n) = \mathcal{P}(1_{\mathcal{N}})$  since  $(\mathcal{P} \circ \mathcal{P})(a \cdot b) = \mathcal{P}(1_{\mathcal{N}})$ , then  $\mathcal{P}(a \cdot b) \geq \mathcal{P}(b^n)$ . This suggests that  $\mathcal{P}$  is an almost primary  $f$ -ideal of near-ring.

However, the converse of the above proposition need not be true, in general. In this context, we provide an example as evidence.

**Example 2.** Let  $\mathcal{N} = \{0,1,2,3,4,5,6,7\}$  be a right near-ring described in tables set 2.



Tables set 2

+	0	1	2	3	4	5	6	7
0	0	1	2	3	4	5	6	7
1	1	2	3	0	7	6	4	5
2	2	3	0	1	5	4	7	6
3	3	0	1	2	6	7	5	4
4	4	7	5	6	2	0	1	3
5	5	6	4	7	0	2	3	1
6	6	4	7	5	1	3	0	2
7	7	5	6	4	3	1	2	0

·	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7
2	0	2	0	2	2	2	0	0
3	0	3	2	1	5	4	6	7
4	0	4	2	5	4	5	6	7
5	0	6	2	4	5	4	6	7
6	0	6	0	6	0	0	0	0
7	0	7	0	7	2	2	0	0

As,  $\mathcal{N} = \{0,1,2,3,4,5,6,7\}$  and  $\mathcal{I} = \{0,2\}$ . Set mapping  $\mathcal{P}: \mathcal{N} \rightarrow [0,1]$  is given as

$$\mathcal{P}(x) = \begin{cases} 1, & \text{if } x \in \mathcal{I}^2 \\ \alpha, & \text{if } x \in \mathcal{I} - \mathcal{I}^2, x \in \mathcal{N} \text{ and } \alpha \in (0,1] \\ 0, & \text{else} \end{cases}$$

Now, we must demonstrate that  $\mathcal{P}$  is an almost primary  $f$ -ideal of  $\mathcal{N}$ . Consider the following cases.

**Case I.** Suppose  $x, y \in \mathcal{N}$ . If  $x \in \mathcal{I}$  and  $y \notin \mathcal{I}$  then  $x - y \notin \mathcal{I}$ . Therefore  $\mathcal{P}(x - y) = \mathcal{P}(1_{\mathcal{N}}) = 0 = \mathcal{P}(y) = \min\{\mathcal{P}(x), \mathcal{P}(y)\}$  and

$$\mathcal{P}(x \cdot y) = \begin{cases} 1, & \text{if } x \cdot y \in \mathcal{I}^2 \\ \alpha, & \text{if } x \cdot y \in \mathcal{I} - \mathcal{I}^2, \text{ where } x \in \mathcal{N} \text{ and } \alpha \in (0, 1]. \end{cases}$$

If  $\mathcal{P}(x \cdot y) = 1$ , then  $\mathcal{P}(x \cdot y) \geq \mathcal{P}(x^n)$  or  $\mathcal{P}(y^n)$ , where  $n \in \mathbb{Z}^+$  and  $\mathcal{P}(x^n) = \mathcal{P}(y^n) = \min\{\mathcal{P}(x), \mathcal{P}(y)\}$  and if  $\mathcal{P}(x \cdot y) = \alpha$ , then  $x \cdot y \in \mathcal{I} - \mathcal{I}^2$ , so  $\mathcal{P}(x^n) = \alpha$ , then  $\min\{\mathcal{P}(x), \mathcal{P}(y)\} = \alpha$ . As a result,  $\mathcal{P}(x \cdot y) \geq \mathcal{P}(x^n)$  or  $\mathcal{P}(y^n)$ .

**Case II.** Assume,  $x \in \mathcal{I}$  and  $y \in \mathcal{I} - (\mathcal{I} \circ \mathcal{I})$ , then  $\min\{\mathcal{P}(x), \mathcal{P}(y)\} = \alpha$  &  $x - y \in \mathcal{I}$ . Therefore,  $\mathcal{P}(x - y) = \alpha = \mathcal{P}(y^n) = \min\{\mathcal{P}(x), \mathcal{P}(y)\}$  and  $x \cdot y \in \mathcal{I}^2$  implies that  $\mathcal{P}(x \cdot y) = 1 \geq \min\{\mathcal{P}(x), \mathcal{P}(y)\} = \mathcal{P}(x^n)$  or  $\mathcal{P}(y^n)$ .

**Case III.** If  $x \in \mathcal{I}$  and  $y \in \mathcal{I}$ , then  $\min\{\mathcal{P}(x), \mathcal{P}(y)\} = 0$  &  $\mathcal{P}(x - y) \geq 0$ . Further to this,  $\mathcal{P}(x - y) \geq \min\{\mathcal{P}(x), \mathcal{P}(y)\}$ . Since  $\mathcal{I}$  is a primary ideal in  $\mathcal{N}$ , at that point  $x^n \notin \mathcal{I}$  and  $y^n \in \mathcal{I}$  for some  $\mathbb{Z}^+$ . So,  $x \cdot y \notin \mathcal{I}$ , this implies  $\mathcal{P}(x \cdot y) = \min\{\mathcal{P}(x), \mathcal{P}(y)\} = 0$ .

Hence,  $\mathcal{P}$  is  $f$ -ideal of  $\mathcal{N}$ . Now, we have to show that  $\mathcal{P}$  is an almost primary  $f$ -ideal of  $\mathcal{N}$ .

Suppose,  $x, y \in \mathcal{N} | \mathcal{P}(x \cdot y) = 0$  and  $(\mathcal{P} \circ \mathcal{P})(x \cdot y) = 0$ , where  $(\mathcal{P} \circ \mathcal{P})(x \cdot y) = 0$  such  $x$  and  $y$  existed, so  $\min\{\mathcal{P}(x), \mathcal{P}(y)\} = 0$ . So consider that  $\min\{\mathcal{P}(x), \mathcal{P}(y)\} = \mathcal{P}(x)$ , then  $\mathcal{P}(x) = 0$ , implies  $x \in \mathcal{I}$ . But  $\mathcal{P}(x \cdot y) = 0$  leads that  $x \cdot y \in \mathcal{I}$ . Therefore,  $y^n \in \mathcal{I}$  (since  $\mathcal{I}$  is a primary ideal in  $\mathcal{N}$ ). Assume  $y \in \mathcal{I}^2$  then there exists  $a, b \in \mathcal{I}$  with  $y = a \cdot b$  and  $(\mathcal{P} \circ \mathcal{P})(x \cdot y) = 0$  which contradicts the selection of  $x$  and  $y$ , so  $y \in \mathcal{I} - \mathcal{I}^2$  and  $x \cdot y \in \mathcal{I} - \mathcal{I}^2$ , implies that  $\mathcal{P}(x \cdot y) \geq \mathcal{P}(y^n)$ .  $\mathcal{P}$  is hence an almost primary  $f$ -ideal in  $\mathcal{N}$ .  $\mathcal{P}$  still contains three values, therefore  $\mathcal{P}$  is not a primary  $f$ -ideal of  $\mathcal{N}$ .

**Proposition 2.** Let  $\mathcal{N}$  be a near-ring and if  $\mathcal{P}$  is a  $f$ -ideal of  $\mathcal{N}$  satisfying the condition that for  $a, b \in \mathcal{N}$  and  $s, t \in [0, 1]$  with  $a_s \cdot b_t \in \mathcal{P}$  and  $a_s \cdot b_t \notin \mathcal{P} \circ \mathcal{P}$ ,  $a_s \cdot b_t \notin \mathcal{P} \circ \mathcal{P}$ , where  $a_s \in \mathcal{P}$  or  $b_t^n \in \mathcal{P}$ , for some  $n \in \mathbb{Z}^+$ , then  $\mathcal{P}$  is an almost primary  $f$ -ideal of near-ring.

*Proof.* Assume  $a \cdot b \in \mathcal{N}$  such that  $a \cdot b = 0$ ,  $(\mathcal{P} \circ \mathcal{P})(a \cdot b) = 0$  and  $s, t \in (0, 1]$  with  $s = t = \mathcal{P}(a \cdot b)$ , then  $a_s b_t \in \mathcal{P}$ . Further to this,  $a_s b_t > 0$ , then  $b_t \notin \mathcal{P} \circ \mathcal{P}$  so either  $a_s \in \mathcal{P}$  or  $b_t^n \in \mathcal{P}$  as  $(\mathcal{P} \circ \mathcal{P})(a \cdot b) = 0$  implies that  $\max\{(\mathcal{P}(a), \mathcal{P}(b))\} = \mathcal{P}(a)$ , then  $a_s(a) = s < \mathcal{P}(a)$ , hence  $a_s \in \mathcal{P}$  or  $b_t^n \in \mathcal{P}$  implies that  $b_t^n(b) > \mathcal{P}(b^n)$ . Therefore,  $t \geq \mathcal{P}(b^n) \Rightarrow \mathcal{P}(a \cdot b) > \mathcal{P}(b^n)$ , for some  $n \in \mathbb{Z}^+$ . Hence  $\mathcal{P}$  is an almost primary  $f$ -ideal of  $\mathcal{N}$ .

**Theorem 4.** Let  $\mathcal{P}$  be an ideal of  $\mathcal{N}$ . Then its characteristic function  $\lambda_{\mathcal{P}}$  is an almost primary  $f$ -ideal of  $\mathcal{N}$  if and only if  $\mathcal{P}$  is an almost primary ideal of  $\mathcal{N}$ .

*Proof.* Let  $\mathcal{P}$  be an almost primary ideal of  $\mathcal{N}$ . Let  $a, b \in \mathcal{N} \mid \mathcal{P}(a \cdot b) > \mathcal{P}(1_{\mathcal{N}})$  and  $\mathcal{P} \circ \mathcal{P}(ab) = \mathcal{P}(1_{\mathcal{N}})$ . As we know that  $\lambda_{\mathcal{P}}(a)$  is either 0 or 1, so for  $\lambda_{\mathcal{P}}(a \cdot b) > \lambda_{\mathcal{P}}(a)$  that  $\lambda_{\mathcal{P}}(a) = 0$  implies  $\lambda_{\mathcal{P}}(a \cdot b) = 1$ . Since for  $a \cdot b \in \mathcal{P}$  and  $a \notin \mathcal{P}$ , according to almost primary ideal  $b^n \in \mathcal{P}$  for some positive integer  $n$ . Hence  $\lambda_{\mathcal{P}}(b^n) = 1 = \lambda_{\mathcal{P}}(a \cdot b)$ , so  $\lambda_{\mathcal{P}}$  is an almost primary  $f$ -ideal of  $\mathcal{N}$ . Conversely suppose that  $\lambda_{\mathcal{P}}$  is an almost primary  $f$ -ideal of  $\mathcal{N}$ , then for  $a, b \in \mathcal{N} \mid \mathcal{P}(a \cdot b) > \mathcal{P}(1_{\mathcal{N}})$  and  $\mathcal{P} \circ \mathcal{P}(a \cdot b) = \mathcal{P}(1_{\mathcal{N}})$  implies  $\mathcal{P}(a \cdot b) > \mathcal{P}(a)$  or  $\mathcal{P}(a \cdot b) > \mathcal{P}(b^n)$ . To prove  $\mathcal{P}$  is an almost primary ideal, let  $a, b \in \mathcal{P}$  and  $a \notin \mathcal{P}$  then  $\lambda_{\mathcal{P}}(a \cdot b) = 1$  and  $\lambda_{\mathcal{P}}(a) = 0$ . From definition of almost primary  $f$ -ideal there is a positive integer  $n$  such that  $\lambda_{\mathcal{P}}(a \cdot b) > \mathcal{P}(b^n)$ . Hence  $\lambda_{\mathcal{P}}(b^n) = 1$  implies that  $b^n \in \mathcal{P}$ . This prove that  $\mathcal{P}$  is an almost primary ideal of  $\mathcal{N}$ .

**Proposition 3.** Let  $\mathcal{N}$  and  $\mathcal{M}$  be two near-rings and  $g: \mathcal{N} \rightarrow \mathcal{M}$  be a homomorphism. If  $h: \mathcal{M} \rightarrow [0, 1]$  is an almost primary  $f$ -ideal, then  $g^{-1}(h)$  is an almost primary  $f$ -ideal of  $\mathcal{N}$ .

*Proof.* We start with proving that  $g^{-1}(h)$  is a  $f$ -ideal of  $\mathcal{N}$ . Assume  $a, b \in \mathcal{N}$ , then  $g^{-1}(h)(a \cdot b) = h(g(a \cdot b)) = h(g(a) \cdot g(b)) \geq \min\{h(g(a)), h(g(b))\} = \min\{g^{-1}(h)(a), g^{-1}(h)(b)\}$

and

$$g^{-1}(h(a \cdot b)) = h(g(a \cdot b)) = h(g(a), g(b)) \geq \max\{h(g(a), h(g(b))\} = \max\{g^{-1}(h)(a), g^{-1}(h)(b)\}.$$

Now, we suppose that  $h$  is an almost primary  $f$ -ideal of  $\mathcal{N}$ . Suppose  $a, b \in \mathcal{N}$  such that  $g^{-1}(h(a \cdot b)) > g^{-1}(h)(1_{\mathcal{N}})$  and  $(g^{-1}(h) \cdot g^{-1}(h))(a \cdot b) = h(1_{\mathcal{N}})$ , so  $g^{-1}(h(a \cdot b)) = \max\{\min\{g^{-1}(a), g^{-1}(b)\}\} = g^{-1}(h(1_{\mathcal{N}}))$ , then  $(h \circ h)(g(a \cdot b)) = (h \circ h)(g(a)g(b)) = h(1)$ . But, since  $g^{-1}(h(a \cdot b)) > g^{-1}(h)(1_{\mathcal{N}})$ , then  $h(g(a) \cdot g(b)) = \max\{h(g(a)), h(g(b))\} = \max\{g^{-1}(h)(a), g^{-1}(h)(b)\}$ .

### Fuzzy graph of almost primary fuzzy ideal of near-ring

In this section, we explore some fuzzy graphs associated with almost primary  $f$ -ideals of near-rings and provide their application in networking. We give a clue to model any network through our provided application. Moreover, we also discuss the properties of fuzzy graphs of almost primary  $f$ -ideal in near-rings.

**Definition 13.** A fuzzy graph of almost primary  $f$ -ideal  $\mathcal{P}: \mathcal{N} \rightarrow [0, 1]$  is denoted by  $G = (V, \mu, \rho)$ , where  $V$  is the

set of vertices,  $\mu: \mathcal{N} \rightarrow [0, 1]$  is the fuzzy vertex set of  $G$  and  $\rho: \mathcal{N} \cdot \mathcal{N} \rightarrow [0, 1]$  is the fuzzy edge set of  $G$ . In a fuzzy graph of almost primary ideal there is an edge between two vertices  $a$  and  $b$  if  $\rho(a, b) = \rho(a \cdot b) \leq \min\{\mu(a), \mu(b)\}$  and  $\rho(a, b) = \rho(b \cdot a) \leq \min\{\mu(b), \mu(a)\}$ , where  $a, b \in \mathcal{N}$ .

**Example 3.** Consider that  $\mathcal{P}$  is an almost primary  $f$ -ideal of  $\mathcal{N}$  in example 2. To construct the fuzzy graph of  $\mathcal{P}$ , we set the values for fuzzy vertices as  $\mu(1) = \rho(1) = 0.1, \mu(2) = \rho(2) = 0.2, \mu(3) = \rho(3) = 0.3, \mu(4) = \rho(4) = 0.4, \mu(5) = \rho(5) = 0.5, \mu(6) = \rho(6) = 0$  and  $\mu(7) = \rho(7) = 0$ . We can easily find that  $\rho(3, 4) = \rho(3 \cdot 4) = \rho(5) = 0.5 \not\leq \min\{\mu(3), \mu(4)\} = 0.4$ , so there is no fuzzy edge between fuzzy vertices 0.3 and 0.4. Similarly, a fuzzy edge does not exist between fuzzy vertices 0.5, 0.1 and 0.5, 0.2. This fuzzy graph of an almost primary  $f$ -ideal is depicted in Fig. 1.

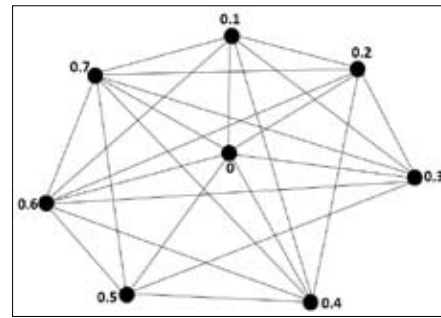


Figure 1: Fuzzy graph of almost primary  $f$ -ideal

There is no loop in Fig. 1, so a fuzzy graph of an almost primary  $f$ -ideal is a simple graph. Also, the graph is not a star graph because some fuzzy vertices are not connected. Subsequently, it is not a regular as well as complete graph because all fuzzy vertices are not having the same degree. The fact which prevents a fuzzy graph from becoming regular, complete, and a star graph is the operation of a near-ring denoted by “ $\cdot$ ”. Because in a near-ring structure for all  $a, b \in \mathcal{N}$ , it is not necessary that  $a \cdot b = b \cdot a$ . Furthermore, a fuzzy graph of almost primary  $f$ -ideal is not a strong fuzzy graph.

### Important Note:

Fuzzy graphs are used in networking models to handle uncertainty and vagueness in network connections. In such models, edges between nodes are assigned a membership value (ranging from 0 to 1) instead of a binary presence or absence. This allows for more flexible and realistic modelling of networks where connections may not be entirely certain or strong, such as in social

networks, communication networks, or reliability analysis. Fuzzy graphs help in optimizing routes, improving fault tolerance and enhancing the overall efficiency of network systems.

Refer to the above example 3, we can fix any issue related to social networking, computer networking, interconnected networks, etc., by exploiting the above given fuzzy structure in near-rings.

## CONCLUSION

Recently, many classical ideals in different types of rings have been shifted in the paradigm of fuzzy sets. We have observed that the notion of almost primary  $f$ -ideal has been defined in rings but not extended for the near-rings. To fill this gap, we have initiated the notion of almost primary  $f$ -ideal in near-rings. In this article, we have also provided several characterizations of almost primary  $f$ -ideal of a near-ring. We have provided several examples and counter-examples in our study. Furthermore, we have explored the relations of almost primary  $f$ -ideal with primary  $f$ -ideal of a near-ring and concluded that every primary  $f$ -ideal is an almost primary  $f$ -ideal, but the converse is not true, in general. We have also discussed the homomorphism of two near-rings  $\mathcal{N}$  and  $\mathcal{M}$  denoted by  $g: \mathcal{N} \rightarrow \mathcal{M}$  along with almost primary  $f$ -ideal  $h: \mathcal{M} \rightarrow [0, 1]$ . Moreover, we have also identified the fuzzy graphs associated with almost primary  $f$ -ideal of near-rings. We have also provided some properties of these newly established graphs. We discussed how the operation of near-ring affected fuzzy graph of almost primary  $f$ -ideal of near-ring. No doubt, almost primary  $f$ -ideal and their relations with classical ideals will help us to understand the linkage of classical ideal theory and fuzzy ideal theory. Finally, fuzzy graph of almost primary ideal will also help us to understand the theoretical fuzzy structure and modelling of different networks. One could extend this work towards other generalized structures like semi-near-rings, hemi-rings etc.

## REFERENCES

- Abou-Zaid S. (1991). On fuzzy subnear-rings and ideals. *Fuzzy sets and systems*, 44(1), 139-146. [https://doi.org/10.1016/0165-0114\(91\)90039-s](https://doi.org/10.1016/0165-0114(91)90039-s)
- Adak, A. K., Gunjan, M. K., & Agarwal, N. K. (2023). Picture fuzzy semi-prime ideals. *Journal of fuzzy extension and applications*, 4(2), 115-124. <https://doi.org/10.22105/jfea.2023.395704.1261>
- Atanassov, K. T., (1999). *Intuitionistic fuzzy sets* (pp. 1-137). Physica-Verlag HD. [https://doi.org/10.1007/978-3-7908-1870-3\\_1](https://doi.org/10.1007/978-3-7908-1870-3_1)
- Batool, N., Hussain, S., Kausar, N., Munir, M., Li, R. Y. M., & Khan, S. (2023). Decision making under incomplete data: intuitionistic multi fuzzy ideals of near-ring approach. *Decision Making: Applications in Management and Engineering*, 6(1), 564-582. <https://doi.org/10.31181/dmame04012023b>
- Birkenmeier, G., Heatherly, H., & Lee, E. (1993). Prime ideals in near-rings. *Results in Mathematics*, 24(1), 27-48. <https://doi.org/10.1007/bf03322315>
- Bloch, I., & Ralescu, A. (2023). *Fuzzy Sets Methods in Image Processing and Understanding*. Springer Cham. <https://doi.org/10.1007/978-3-031-19425-2>
- Cuong, B. C. (2014). Picture fuzzy sets. *Journal of Computer Science and Cybernetics*, 30(4), 409-409. <https://doi.org/10.15625/1813-9663/30/4/5032>
- Elavarasan, B. (2011). Generalizations of prime ideals in near-rings. *International Journal of Open Problems computer Science and Mathematics*, 4(4), 47-53. <https://www.i-csrs.org/Volumes/ijopcm/vol.4/vol.4.4.6.S.11.pdf>
- Groenewald, N. J. (1991). Different prime ideals in near-rings. *Communications in Algebra*, 19(10), 2667-2675. <https://doi.org/10.1080/00927879108824287>
- Jagadeesha, B., Kedukodi, B. S., & Kuncham, S. P. (2023). Equiprime fuzzy graph of a near-ring with respect to a level ideal. *Matematički Vesnik*, 75(4), 247-264. <https://doi.org/10.57016/mv-zuxv3844>
- Jebapresitha, B. (2024). Hybrid structure of maximal ideals in near rings. *Complex & Intelligent Systems*, 1-14. <https://doi.org/10.1007/s40747-024-01486-0>
- Khan, W. A., Muhammad, A., Taouti, A., & Maki, J. (2018). Almost prime ideal in gamma near ring. *European Journal of Pure and Applied Mathematics*, 11(2), 449-456. <https://doi.org/10.29020/nybg.ejpam.v11i2.3193>
- Kedukodi, B. S., Kuncham, S. P., & Bhavanari, S. (2009). Equiprime, 3-prime and c-prime fuzzy ideals of nearrings. *Soft Computing*, 13(10), 933-944. <https://doi.org/10.1007/s00500-008-0369-x>
- Liu, W. J. (1982). Fuzzy invariant subgroups and fuzzy ideals. *Fuzzy Sets and Systems*, 8, 133-139. [https://doi.org/10.1016/0165-0114\(82\)90003-3](https://doi.org/10.1016/0165-0114(82)90003-3)
- Malik, D. S., & Mordeson, J. N. (1991). Fuzzy maximal, radical and primary ideals of a ring. *Information sciences*, 53(3), 237-250. [https://doi.org/10.1016/0020-0255\(91\)90038-v](https://doi.org/10.1016/0020-0255(91)90038-v)
- Mordeson, J. N., Mathew, S., & Gayathri, G. (2023). *Fuzzy Graph Theory: Applications to Global Problems* (Vol. 424). Springer Nature. [https://doi.org/10.1007/978-3-031-23108-7\\_5](https://doi.org/10.1007/978-3-031-23108-7_5)
- Muhammad, A., Ashraf, M., & Khan, W. A. (2023). Different types of primary ideals and their graphs in near-rings. *Songklanakarinn Journal of Science and Technology*, 45(2), 173-181. <https://sjst.psu.ac.th/journal/45-2/2.pdf>
- Mukherjee, T. K., & Sen, M. K. (1987). On fuzzy ideals of a ring I. *Fuzzy sets and systems*, 21(1), 99-104. [https://doi.org/10.1016/0165-0114\(87\)90155-2](https://doi.org/10.1016/0165-0114(87)90155-2)
- Onar, S., Hila, K., Etemad, S., Akgül, A., De la Sen, M., & Rezapour, S. (2023). 2-Absorbing vague weakly complete  $\Gamma$ -ideals in  $\Gamma$ -rings. *Symmetry*, 15(3), 740. <https://doi.org/10.3390/sym15030740>

- Pilz, G. (2011). *Near-rings: the theory and its applications*, Elsevier. [https://doi.org/10.1016/s0304-0208\(08\)x7135-x](https://doi.org/10.1016/s0304-0208(08)x7135-x)
- Razaq, A., & Alhamzi, G. (2023). On Pythagorean fuzzy ideals of a classical ring. *AIMS Math*, 8(2), 4280-4303. <https://doi.org/10.3934/math.2023213>
- Rosenfeld, A. (1975). Fuzzy graphs. In *Fuzzy sets and their applications to cognitive and decision processes* (pp. 77-95). Academic press. <https://doi.org/10.1016/B978-0-12-775260-0.50008-6>
- Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, 8(3), 338-353. [https://doi.org/10.1016/s0019-9958\(65\)90241-x](https://doi.org/10.1016/s0019-9958(65)90241-x)
- Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning-I. *Information sciences*, 8(3), 199-249. [https://doi.org/10.1016/0020-0255\(75\)90046-8](https://doi.org/10.1016/0020-0255(75)90046-8)
- Zhang, W. R. (1998, May). Bipolar fuzzy sets. In *1998 IEEE international conference on fuzzy systems proceedings. IEEE world congress on computational intelligence (Cat. No. 98CH36228)* (Vol. 1, pp. 835-840). IEEE. <https://doi.org/10.1109/fuzzy.1998.687599>

2010 Mathematics Subject Classification. 16Y30, 16Y99, 03G25.



## RESEARCH ARTICLE

### Oleochemistry

# Physiochemical characterization of critically endangered *Coscinium fenestratum* (Gaertn.) Colebr. seed fat for potential niche applications in cosmetics and nutritional supplements

KAH Thathsara and SDM Chinthaka

Department of Chemistry, Faculty of Applied Sciences, University of Sri Jayawardenepura, Nugegoda, Sri Lanka.

Submitted: 16 January 2024; Revised: 18 December 2024; Accepted: 29 January 2025

**Abstract:** *Coscinium fenestratum* (Sinhala *Veniwalgatta*), a medicinal plant native to Sri Lanka and India, is widely utilized in traditional medical practices such as Ayurveda, Unani and Siddha. The global cosmetic and nutritional industries are increasingly shifting towards plant-based feedstock, leading to the exploration of novel sources of fats and oils. However, the seed and fat of *Coscinium fenestratum* remain underutilized with no prior studies conducted on them. Therefore, this study aims to characterize the seed oil of *C. fenestratum* by determining its fatty acid (FA) composition as their methyl esters, chemical constituents in unsaponifiable matter, and other physiochemical properties. The oil was extracted using the Soxhlet extraction method. The ash and moisture contents of the seeds, acid value (AV), iodine value (IV), smoke point and thermal stability of the oil were determined. Fatty acid methyl esters and unsaponifiable matter were analyzed using gas chromatography-mass spectrometry. Results showed an oil yield of  $46.94 \pm 0.01\%$  with a moisture content of  $4.10 \pm 0.03\%$ , ash content of  $2.30 \pm 0.28\%$ , AV of  $2.34 \pm 0.10$  mg KOH/g, IV of  $56.33 \pm 0.32$  g I<sub>2</sub>/100 g, smoke point of  $202.9 \pm 3.9$  °C, decomposition temperature of  $416.88 \pm 1.74$  °C, and unsaponifiable matter yield of  $0.57 \pm 0.01\%$ . The dominant FAs, stearic (C18:0) and oleic (C18:1), contributed to  $50.54 \pm 0.88\%$  and  $39.21 \pm 0.86\%$  of the composition, respectively. The main constituents of the unsaponifiable matter were stigmasterol (1.91 mg/g), gamma-sitosterol (1.29 mg/g), campesterol (0.40 mg/g), fucosterol (0.37 mg/g), and squalene (0.13 mg/g). In conclusion, the findings of this study suggest that *C. fenestratum* seed fat has promising potential to be valorized in both cosmetic and nutritional supplement industries.

**Keywords:** *Coscinium fenestratum*, cosmetics and nutritional supplements, fatty acids, seed fat, unsaponifiable matter.

## INTRODUCTION

The global beauty market has consistently grown at an approximate annual rate of 4.5% over the last two decades. Manufacturers are strategically incorporating plant-based feedstocks such as fats and oils into cosmetic formulations in alignment with consumer demand for natural and sustainable products and their health and dermatological properties (Łopaciuk & Łoboda, 2013). Fish oils have long dominated the nutritional supplement industry due to the high content of essential fatty acids (FAs), eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), which are renowned for their protective properties against various diseases, particularly cardiovascular conditions (Lenihan-Geels *et al.*, 2013). However, concerns about the declining population of fish due to overfishing and ecological impact have promoted the replacement of fish oils with novel plant oils. This has created a new global trend to search for underexplored and underutilized plant resources to incorporate into the cosmetic and nutritional industries (Nasopoulou & Zabetakis, 2012).

Lipids are the main constituent in plant oils, present as triglycerides and play vital roles in various biological

\* Corresponding author (sdmchin@sjp.ac.lk;  <https://orcid.org/0000-0002-2798-5810>)



This article is published under the Creative Commons CC-BY-ND License (<http://creativecommons.org/licenses/by-nd/4.0/>). This license permits use, distribution and reproduction, commercial and non-commercial, provided that the original work is properly cited and is not changed in anyway.

functions. The human body can synthesize saturated and monounsaturated FAs but cannot produce polyunsaturated fatty acids (PUFAs). Assessing the saturated/unsaturated FA ratio in seed oils provides insights into their potential as essential FAs and suitability for incorporation into food ingredients (Yao & Xu, 2021). The FA composition of plant oils is highly variable and influences their suitability for different applications. Long-chain saturated fatty acids (SFAs) such as myristic (C14:0), palmitic (C16:0), and stearic (C18:0) acids offer skin softening, smoothing, and protective properties. Plant oils with higher SFA content are favourable for cosmetics but may not be suitable as nutritional supplements because of their adverse health effects (Johnson, 1978; Rabasco Alvarez & González Rodríguez, 2000). Understanding the saturated/unsaturated FA ratio is crucial for the suitability of plant oils in various applications. A higher proportion of unsaturated fatty acids (UFAs) can lead to deterioration of oil quality due to natural oxidation reactions associated with them (Kostik *et al.*, 2013).

Unsaponifiable matter in fats and oils is another class of chemical constituents that are found in minor levels compared to FAs. Despite their low levels, they contribute significantly to cosmetic and health benefits. Unsaponifiable matter in plant oils mainly contains, but is not limited to, substances such as phytosterols, carotenoids, terpenes, triterpene alcohols, tocopherols and fatty alcohols (Fontanel, 2013). These constituents, whether working synergistically or individually, offer health benefits such as antioxidant, anticancer, anti-inflammatory and dermatological properties and they are also proven to contribute to the prevention of cardiovascular diseases (Caligiani *et al.*, 2010). Among the constituents in unsaponifiable matter, squalene is the most commercially explored and utilized compound due to its various cosmetic and health benefits. Despite the diverse bioactivities of squalene, its potential applications in cosmetics are limited due to the lack of natural resources other than fish oils (Kim & Karadeniz, 2012). Therefore, most squalene and its derivatives, which are currently used in cosmetic and nutritional supplement industries, originate from crude oil via synthetic routes. Furthermore, the inclusion of phytosterols in dietary supplements will offer significant health benefits such as direct inhibition of cholesterol absorption, etc. (Ogbe *et al.*, 2015). However, some studies have shown that elevated concentrations of certain phytosterols in plasma can lead to risk factors (Glueck *et al.*, 1991).

Plant extracts, including seed oils, are increasingly used in cosmetics due to their biodegradability and lower toxicity compared to synthetic or semi synthetic

substances. Currently, various seed oils such as almond oil, apricot kernel oil, castor oil, sunflower oil, jojoba oil, and plant fats such as shea-butter, kokum butter and mango kernel butter are extensively used in the cosmetic industry. Soybean and rapeseed oils, rich in PUFAs, are considered promising alternatives to fish oils (Nasopoulou & Zabetakis, 2012). Plant oils such as olive, flax seed, moringa seed, grape seed and perilla seed are currently used in dietary supplements as alternatives to fish oils due to their FAs and unsaponifiable matter composition for their health-enhancing properties (Caligiani *et al.*, 2010). The chemical composition of unsaponifiable fractions highlights the superiority of olive oil, jojoba oil, avocado oil, shea butter and kokum butter in commercial applications.

The global cosmetic and nutritional supplement industry is increasingly turning to plant seed oils as renewable feedstock causing a substantial need to explore novel sources of fats and oils. The increasing global population and the escalating demand for plant oils in both the food and industrial sectors pose significant challenges. The main challenge is the extremely high cost of some seed oils due to their rarity and short supply. This opens new avenues to explore alternative sources of fats and oils, focusing on non-conventional and underutilized seeds to address the shortage of currently used plant oils.

Sri Lanka is a tropical country with very high biodiversity and has a wide variety of oil-rich seeds that are widespread across all geographic regions of the country. However, many of these oil seeds are poorly characterized or not chemically characterized at all. *Coscinium fenestratum* (Gaertn.) Colebr. is among the unexplored oil seed species in Sri Lanka. It is a critically endangered plant belonging to the family Menispermaceae due to its over-exploitation for its proven medicinal properties of root, bark and stem. It is native to Sri Lanka and the Western Ghats of India, known locally as “*Veniwalgatta*”. The stem of *C. fenestratum* is extensively used in various ayurvedic, siddha and indigenous medicine formulations (Kothalawala *et al.*, 2020). Surprisingly, there is no report of the utilization of seeds in these medicinal practices. Therefore, the chemical composition of the seeds of *C. fenestratum* remains unexplored even though the stem has been extensively studied for its medicinal properties and chemical composition (Rai *et al.*, 2013). A wide array of commercial cosmetics and nutritional supplements such as soap, bath oils, face wash, body lotions and fairness creams and medical concoctions that include *C. fenestratum* stem, root or bark extracts have emerged over the last few decades (Warakagoda & Subasinghe,

2014). However, seeds or seed oils have been ignored in these formulations due to the lack of knowledge of their chemical composition.

Therefore, the selection of *C. fenestratum* seeds in this study is driven by their medicinal value compared to other parts of the plant as well as the lack of prior investigative data for seeds. This study focused on determining the physiochemical properties, FA composition and unsaponifiable matter composition of *C. fenestratum* seeds with the ultimate goal of identifying potential applications in the cosmetic and nutritional supplement industries.

## MATERIALS AND METHODS

### Extraction of *C. fenestratum* seed oil and determination of oil content

Seeds were collected from the Yagirala forest reserve (6.365506, 80.172835) in the Kalutara District in September based on random sampling. A quantity of 20-30 seeds was obtained from well-ripened, dried fruits of *C. fenestratum* and only high-quality, pest and fungi-free seeds selected for further analysis. After removing the seed coat, kernels were ground and sieved to obtain a fine powder. This powder was oven-dried at 60 °C for 2-3 hours. Approximately 25 g of ground seed kernels were used for oil extraction. Soxhlet extraction was conducted using 150.0 cm<sup>3</sup> of HPLC-grade hexane for six hours. Anhydrous sodium sulphate was added after refluxing to remove traces of moisture, and the solvent was then removed using a rotary evaporator. The crude oil percentage (w/w) was calculated based on the dried ground seed kernel powder (Keneni *et al.*, 2021). Seed oil extraction was duplicated as necessary to ensure accuracy.

### Determination of moisture content in seed kernel

Moisture content was determined by the modified AOAC 934.06 method. The ground seed kernel sample was dried at 103 ± 2 °C for 2 h in an oven and repeated till a constant weight was achieved. The moisture content was calculated as a percentage of the initial seed sample weight (Zang *et al.*, 2017). Moisture content was determined using the following formula (1):

$$\text{Percentage of Moisture Content (\%)} = \frac{x - y}{x} \times 100$$

x- Weight of the sample used for moisture determination  
y- Weight of dried sample of constant weight

### Determination of ash content in seed kernel

The modified AOAC 942.05 method was followed using a muffle furnace. A known weight of the dried ground seed kernel was ignited at 550 °C for 5 h. The weight of the resulting ash was measured at room temperature and ash content was calculated as a percentage of the initial seed sample (Yadav *et al.*, 2022) using the following formula (2):

$$\text{Ash Content (\%)} = \frac{W_3 - W_1}{W_2 - W_1} \times 100$$

W1 - Weight of the empty crucible

W2 - Weight of the crucible and the sample before ashing

W3 - Weight of the crucible and ash

### Determination of acid value

The Acid Value (AV) was determined using a modified A.O.C.S Official method, Cd 3a-63, 2006. A weight of 0.2 g of *C. fenestratum* seed fat was dissolved in a freshly prepared, pre-neutralized methanol-diethyl ether 1:1 (v:v) ratio. This was neutralized with a 0.05 M KOH solution using phenolphthalein as the indicator (Abdullahi *et al.*, 2021). The AV was calculated using the following formula (3).

Acid Value =

$$\frac{\text{Volume of KOH solution} \times 56.1 \times \text{KOH concentration}}{\text{Weight of Oil}}$$

### Determination of iodine value

The modified AOAC 920.159 method was followed using Wij's solution to determine the Iodine Value (IV) of the seed fat. The sample was titrated against 0.1 M Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub> solution, and the IV was calculated using the following formula (4) (Pardeshi, 2020):

Iodine Value =

$$\frac{(V_{\text{blank}} - V_{\text{sample}}) \times \text{Na}_2\text{S}_2\text{O}_3 \text{ concentration} \times 12.69}{\text{Weight of Oil}}$$

### Determination of smoke point

The smoke point was determined by heating 0.5 mL of the seed fat on a hot plate and noting the temperature at which continuous smoke started to appear. The temperature was measured using an IR thermometer (Falade *et al.*, 2008).

### Determination of fatty acid composition by gas chromatography – mass spectrometry (GC-MS).

To identify and quantify the FAs present in the extraction, the seed fat was converted to Fatty Acid Methyl Esters (FAME). Approximately 0.05 g of oil was mixed with 5.00 cm<sup>3</sup> of iso-octane in a stoppered test tube. To this solution, 200 µL of 2 M methanolic KOH was added, stoppered, and vigorously shaken for 30 seconds until cloudiness appeared. Finally, approximately 1 g of sodium dihydrogen orthophosphate was added and shaken until the cloudiness disappeared. The upper layer of the freshly prepared sample was subjected to GC-MS analysis after drying with anhydrous sodium sulphate and filtration. From the filtered sample, 1.0 µL was injected at a split ratio of 1:20 to an Agilent GC model-7890A and MS model-5975C with a non-polar ultra-inert capillary column containing 5% phenyl methyl siloxane (19091S-433UI HP-5MS with an internal diameter of 250 µm, column length of 30.0 m, and film thickness of 0.25 µm). The temperature program included an initial temperature of 100.0 °C held for 3.0 minutes, increased to 240.0 °C at a rate of 3.0 °C/min and then held for 7.0 minutes. Samples were prepared in four replicates and the average was considered for quantification and comparison.

Compound identification was performed by comparing the total ionic current spectrum of each peak of *C. fenestratum* seed fat extract with that from the NIST mass spectral library and the retention time window of each matching FAMES of the mass spectrum of the Supelco 37 component FAME reference mix standard (Sigma Aldrich). The quantification was carried out using the internal standard method by considering the response factors with heptadecanoic acid methyl ester (Sigma Aldrich) as the internal standard.

### Extraction of unsaponifiable matter

A modified AOAC 933.08 procedure was used to extract unsaponifiable matter from *C. fenestratum* seed fat. Around 5 g of fat was refluxed with 50.0 cm<sup>3</sup> of 1 M methanolic KOH. The unsaponifiable matter was extracted using liquid-liquid extraction with HPLC-grade hexane. The solvent was removed by rotary evaporation and acetone was added to eliminate volatile components. The weight of the obtained unsaponifiable matter was measured to calculate the yield (w/w) as a percentage of the initial fat weight (Giacometti, 2001; Janporn *et al.*, 2015).

### Identification and quantification of unsaponifiable matter using gas chromatography-mass spectrometry (GC-MS)

The white unsaponifiable matter was dissolved in HPLC-grade hexane, dried with anhydrous sodium sulphate, and analysed using GC-MS. An Agilent GC model-7890A and MS model-5975C were used with the same column as for FAME analysis.

The temperature program started at 70.0 °C was kept for 1.5 minutes, then increased to 280.0 °C at a rate of 10.0 °C/min and held for 38.0 minutes at 280.0 °C. Compound identification followed the same method for FAME identification and quantification was done using an external calibration method with reference standards from Sigma Aldrich.

### Thermal analysis

Thermal analysis was conducted using the Discovery SDT 650 model, equipped with simultaneous Thermogravimetric Analysis (TGA) and Differential Scanning Calorimetry (DSC) capabilities (Trios software). The analysis involved heating the *C. fenestratum* seed fat sample from room temperature to 600.0 °C at a rate of 10.0 °C/min under a nitrogen atmosphere. The resulting mass changes were plotted using Origin software to generate TGA and Differential Thermogravimetry (DTG) curves (Solís-Fuentes *et al.*, 2010).

### Determination of thermal stability of peanut oils as the reference

The seed kernels of fully ripened peanuts, procured from the local market, were extracted by the Soxhlet method outlined in seed oil extraction section (Keneni *et al.*, 2021). The thermal stability of the extracted peanut oil was then analyzed according to the procedure described in above (Solís-Fuentes *et al.*, 2010).

## RESULTS AND DISCUSSION

Figure 1 shows the seed morphology and seed fat of *C. fenestratum*. The seed kernel of *C. fenestratum* has a distinct yellow color and the seed kernel is solid white at room temperature (Figure 1). Determination of the percentage of crude oil content is significant to evaluate the economic feasibility and commercial potential of *C. fenestratum* seeds as an oil-bearing source. Compared to shea butter (47.5% yield) extracted from shea nuts,

which is extensively used in the food, pharmaceutical, and cosmetic industries, *C. fenestratum* seeds had a calculated oil yield of  $46.94 \pm 0.01\%$  (Table 1) through solvent extraction (Abdul-mumeen *et al.*, 2019). This suggests that *C. fenestratum* seeds can be considered as high-oil-yielding seeds and could be a potentially advantageous and cost-effective resource for large-scale industrial applications.

To assess the quality of the extracted oils, we determined the moisture content of the seed kernel. While previous studies reported a moisture content of 28% for fresh seeds of *C. fenestratum*, it is important to note that the seeds collected for our study were well-ripened, dried and fallen from the plant (Ramasubbu *et al.*, 2012). Consequently, the moisture content determined in this study (Table 1) was significantly lower at  $4.10 \pm 0.03\%$ .



**Figure 1:** Seeds (A), kernels (B) and seed fat of *C. fenestratum* (C)

**Table 1:** Physiochemical properties of *C. fenestratum* seed kernel.

Physiochemical property	Average $\pm$ SD
Yield of crude Oil (%)	$46.94 \pm 0.01$
Moisture content (%)	$4.10 \pm 0.03$
Ash content (%)	$2.30 \pm 0.28$

**Table 2:** Physiochemical properties of the seed fat of *C. fenestratum* seed.

Physiochemical property	Average $\pm$ SD
Acid value (mg KOH/g)	$2.34 \pm 0.10$
Iodine value (g I <sub>2</sub> /100g)	$56.33 \pm 0.32$
Yield of unsaponifiable matter (%)	$0.57 \pm 0.01$
Smoke point (°C)	$202.9 \pm 3.9$
Decomposition temperature (°C)	$416.88 \pm 1.74$

The low moisture content in the kernel is a positive indicator, suggesting extended storage stability and a lower risk of spoilage, which is consistent with previous research findings (Onyeike & Acheru, 2002).

The ash content of *C. fenestratum* seed kernel was determined to be  $2.30 \pm 0.28\%$  (Table 1), which is close to the value obtained for groundnut seed kernel in a previous study and notably higher than the values for coconut (0.43%), dika nut (0.63%), melon (1.50%), and palm kernel (1.50%) (Onyeike & Acheru, 2002). This finding provides us an initial insight into the mineral content and nutritional value of *C. fenestratum*, since the ash content indicates the total minerals present in the seed kernel. Essentially, it represents the inorganic residue remaining after the removal of water and organic matter. This suggests a significant mineral richness in *C. fenestratum* seed kernels, suggesting potential nutritional benefits and making it a promising candidate for dietary inclusion (Zang *et al.*, 2017). However, further analysis of metal ion content is required for a comprehensive assessment of the benefits associated with mineral content. This study is currently underway in our group.

The AV of the *C. fenestratum* seed fat was  $2.34 \pm 0.10$  mg KOH/g (Table 2). This value reflects the free fatty acid content of the oil and serves as an important indicator of the rancidity of fats and oils. A significantly low value suggests its suitability for potential edible purposes of the oil. The AV of *C. fenestratum* seed fat was lower compared to the values for sesame, soybean, sunflower and rapeseed oils. This positions the oil as a stable and durable alternative for various applications with further treatments. The IV of *C. fenestratum* seed fat was  $56.33 \pm 0.32$  g I<sub>2</sub>/100g (Table 2), which corresponds to the degree of unsaturation of the FA profile. Oils with higher IVs typically contain more UFAs and are more susceptible to oxidative rancidity. In contrast, the relatively low IV of *C. fenestratum* seed fat indicates a higher proportion of SFAs, suggesting greater resistance to oxidative rancidity and enhanced oxidative stability (Kyari, 2008).

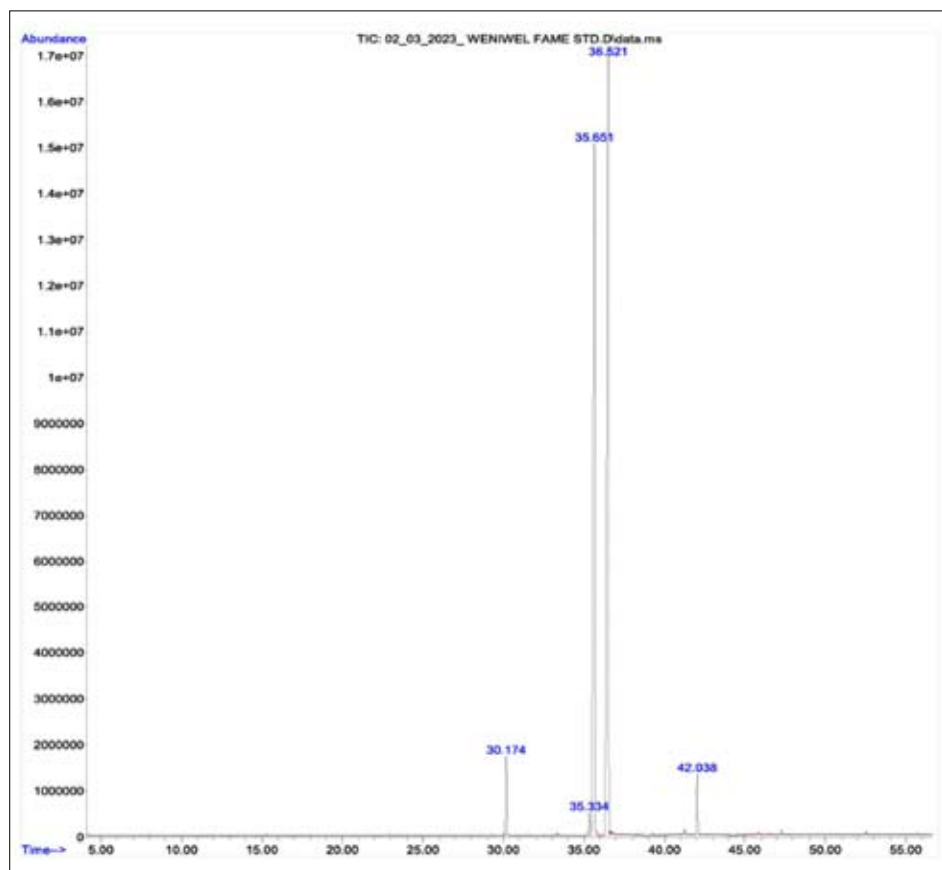
Having a low IV of  $56.33 \pm 0.32$  g I<sub>2</sub>/100g (Table 2) indicates a high amount of SFAs. A previous study suggested that oils with higher IVs are susceptible to photocatalytic reactions (Kyari, 2008).

The smoke point is an important parameter when assessing the stability and suitability of fats and oils for high-temperature applications. It is determined by the threshold temperature at which a continuous wisp of smoke is produced upon heating. The higher the smoke point the higher the thermal stability of the oil.

According to Canadian government specifications, oils that are subjected to high temperature applications, especially frying, require a smoke point exceeding 200 °C (Przybylski *et al.*, 2020). The value in this study for *C. fenestratum* seed fat is  $202.9 \pm 3.9$  °C (Table 2). Therefore, it can be considered as a moderately thermally stable oil.

#### Analysis of fatty the acid profile of *C. fenestratum* seed fat

The percentage composition of each identified FA is given in Table 3, and the total ionic current spectrum of the FAMES of *C. fenestratum* seed fat is shown in Figure 2.



**Figure 2:** Total ion chromatogram of the FAME present in *C. fenestratum* seed fat obtained from the GC-MS analysis in this study. The retention times of relevant FA are; 30.174 min (C16:0), 35.334 min (C18:2), 35.651 min (C18:1), 36.521 min (C18:0), 42.038 min (C20:0)

The gas chromatography-mass spectrometry (GC-MS) method was used to identify and quantify the FAs as their methyl esters of *C. fenestratum* seed fat. Typically, when the level of SFAs significantly exceeds that of UFAs, most plant triglycerides remain solid at room temperature. However, exceptions exist, as seen in palm oil and coconut oil.

According to the FA profile in Table 3, the total composition of FAs is  $97.01 \pm 1.25\%$ . The most abundant are the SFAs with a total percentage of  $56.65 \pm 0.91\%$ . The specific amounts of SFAs are palmitic (C16:0) 3.14

$\pm 0.18\%$ , stearic (C18:0)  $50.54 \pm 0.88\%$  and arachidic (C20:0)  $2.62 \pm 0.13\%$ . It is also noteworthy to observe that very long-chain fatty acids including C20 to C24 are present in minor levels ( $3.24 \pm 0.14\%$ ). These very long-chain fatty acids are associated with various cosmetic properties. The total percentage of UFAs is  $40.36 \pm 0.86\%$ . The most abundant UFAs are the oleic acid (C18:1)  $39.21 \pm 0.86\%$  with linoleic (C18:2)  $0.85 \pm 0.07\%$  and gondoic (C20:1)  $0.30 \pm 0.02\%$  acids present at minor levels. The calculated SFA/UFA ratio is 1.40, which supports the appearance of a solid physical state.

**Table 3:** Fatty acid composition of *C. fenestratum* seed fat (n=4)

Fatty Acid		%
Saturated Fatty Acids		
C14:0	Myristic acid	0.03 ± 0.01
C16:0	Palmitic acid	3.14 ± 0.18
C18:0	Stearic acid	50.54 ± 0.88
C20:0	Arachidic acid	2.62 ± 0.13
C22:0	Behenic acid	0.20 ± 0.04
C24:0	Lignoceric acid	0.12 ± 0.04
Unsaturated Fatty Acids		
C18:1	Oleic acid	39.21 ± 0.86
C18:2	Linoleic acid	0.85 ± 0.07
C20:1	Gondoic acid	0.30 ± 0.02
Total FAs		97.01 ± 1.25
Total Unsaponifiable Matter		0.57 ± 0.01
Unidentified FAs		2.42 ± 1.25

SFAs such as stearic and palmitic and UFAs such as oleic acids are noted for their ability to enhance skin permeation, while linoleic acid is recognized for providing moisturizing and skin healing effects (Vermaak *et al.*, 2011). Having a higher SFA composition offers advantages in resisting oxidative rancidity (Jayadas & Nair, 2006). Another notable observation in the FA composition is the simplicity of the composition. Only

two FAs, steric and oleic, contribute to  $89.75 \pm 1.23\%$  of the total FA content. This simple composition makes the *C. fenestratum* seed fat a potential candidate for a wide variety of cosmetic and dietary supplement applications without concerns for other FAs that have potential health risks such as cholesterol elevation effects. This also stamps *C. fenestratum* seed fat as a stearic acid and oleic acid-rich clean plant source for other industrial applications in the lubrication industry.

#### Comparison of FA composition of *C. fenestratum* seed fat with commonly used seed oils for possible use in cosmetics and dietary supplements

This study focuses on identifying the potential of *C. fenestratum* seed fat to be incorporated into cosmetics and nutritional supplements. To achieve this objective the FA composition of *C. fenestratum* seed fat was compared with plant oils and fats that are already established in these industries. The primary objective of this comparison is to assess the possibility of substituting *C. fenestratum* seed fat with expensive fats and oils currently utilized in the industry. Our comparison reveals distinct similarities in the FA profile of *C. fenestratum* with that of shea butter, sal fat, mango kernel fat, and kokum butter (Table 4). All these seed fats are already used in various cosmetic and dietary supplement formulations and are high in oleic and stearic acid, which are in demand in the cosmetic industry.

**Table 4:** Comparison of FA profile of *C. fenestratum* seed fat with selected commercially available fats (Rabasco Alvarez & González Rodríguez, 2000; Firestone, 2013)

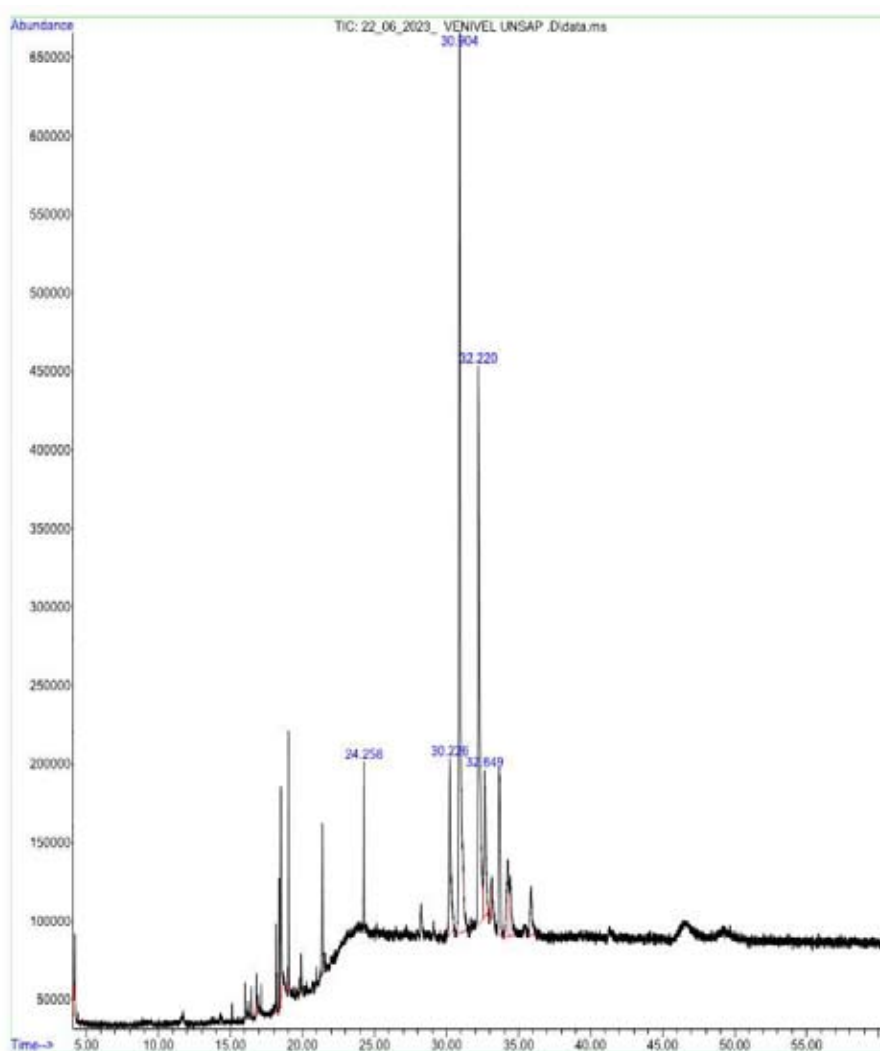
Species	SFA (%)			UFA (%)	
	C16:0	C18:0	C20:0	C18:1	C18:2
	Palmitic	Stearic	Arachidic	Oleic	Linoleic
Shea butter	0.5-8.5	35.1-47.4	0.1-2	43.5-50	4-5
Sal fat	5.3-23	33-57	1-8	31-52	0.3-5
Mango kernel fat	3-18	26-57	1.6-6	38-50	1-13
Kokum butter	1.4-5	49-60.4	-	39-49	1-2
<i>C. fenestratum</i> fat	3.0 -3.3	49.7 -51.4	2.5 – 2.8	38.4– 40.1	0.8 - 0.9

Mango kernel fat, a newly introduced source in the cosmetic industry, has shown promise in producing stable emulsions (Shukla & Kragballe, 1998). Cocoa butter, known for its use in night creams, lip balms, and

bath oils, and shea butter, widely employed in skin care applications in massage creams, makeup, and baby-care products, possess unique characteristics due to their high content of unsaponifiable matter. This unsaponifiable

matter is rich in terpenic alcohols and phytosterols, imparting soothing properties and sun protection abilities to cosmetic products. As a result, shea butter is extensively incorporated into various cosmetic products, such as hair care, baby care, sun care, and body lotions, in suitable amounts (Shukla & Kragballe, 1998). The significant presence of oleic acid in *C. fenestratum* seed fat makes it suitable for use as excipients in cosmetic formulations and as penetration enhancers in pharmaceutical formulations. Kokum butter is a highly stable exotic butter, widely recognized for its applications in skin and hair care products, acne treatments, skin tonics, and as an enhancer for cocoa butter in the global market (Shukla &

Kragballe, 1998). While sal butter is locally employed in cooking, soap production, and the formulation of skin and hair care products, it is often combined with mango butter to enhance emollient and oxidative properties (Shukla & Kragballe, 1998). Furthermore, some studies have shown that UFAs exhibit an effect on reducing blood levels of cholesterol, LDL and HDL, while all SFAs, except stearic acid (C18:0), are believed to have an opposing effect on this physiological aspect (Williams, 2000). Based on this FA profile comparison, *C. fenestratum* seed fat has a promising potential to substitute or to be used as a mixture with all the above seed fats.



**Figure 3:** Constituents of unsaponifiable matter present in *C. fenestratum* seed fat. They are; retention time (min) = 24.258 (Squalene), 30.226 (Campesterol), 30.904 (Stigmasterol), 32.220 (gamma-Sitosterol) and 32.649 (Fucosterol)



### Analysis of unsaponifiable fraction of *C. fenestratum* seed fat

*C. fenestratum* seed fat has an unsaponifiable matter content of  $0.57 \pm 0.01\%$  (Table 2) when calculated as a percentage of total lipids. Compared to the unsaponifiable contents of cocoa (0.5%), kokum (0.5%), mango (0.8%), sal (0.9%) and shea butter (6.0%), *C. fenestratum* seed fat showed a close similarity to all except shea butter (Shukla & Kragballe, 1998). Shea butter is a unique cosmetic ingredient due to its very high percentage of unsaponifiable matter making it difficult to find a suitable substitute.

Figure 3 shows the total ionic current spectrum of the unsaponifiable fraction of *C. fenestratum* seed fat and Table 5 shows the weight of each unsaponifiable matter constituent identified per 1 g of the *C. fenestratum* seed fat. The most abundant constituent is stigmasterol (1.91 mg/g) followed by gamma-sitosterol (1.29 mg/g) and campesterol (0.40 mg/g). Other minor constituents present are fucosterol (0.37 mg/g) and squalene (0.13 mg/g).

**Table 5:** The major constituents of the unsaponifiable matter of *C. fenestratum* seed fat.

Compound	Concentration (mg/g)
Stigmasterol	1.91
gamma-Sitosterol	1.29
Campesterol	0.40
Fucosterol	0.37
Squalene	0.13

Phytosterols, which are steroidal compounds similar in structure and function to cholesterol, have demonstrated cholesterol-lowering effects in human and animal plasma mitigating the risk of coronary heart diseases. Besides cardiovascular benefits, phytosterols also possess anti-inflammatory properties, lowering the effects of total cholesterol and low-density lipoprotein levels (Ogbe *et al.*, 2015).  $\beta$ -sitosterol, campesterol, and stigmasterol are prevalent phytosterols. Among these,  $\beta$ -sitosterol is noted for its various health benefits. It is reported to have cholesterol-lowering, anti-inflammatory, antibacterial and antifungal properties. Moreover,  $\beta$ -sitosterol may help to inhibit the development of cancer and has antioxidant

and UV stabilizing properties. Therefore, the presence of stigmasterol (1.91 mg/g), gamma-sitosterol (1.29 mg/g) and campesterol (0.40 mg/g) in *C. fenestratum* seed fat (Table 5) enhances the potential applicability of this fat in both cosmetics and nutritional supplements (Phuong *et al.*, 2018).

A relatively low level of squalene (0.13 mg/g) is found in *C. fenestratum* seed fat (Table 5). Squalene is arguably one of the most popular ingredients in cosmetic and dietary supplement formulations. It is mainly sourced from animal fats such as whales, sharks, salmon, etc. Studies have highlighted potential benefits of squalene including, but not limited to, reducing skin damage from UV radiation, lowering LDL and cholesterol levels, and antitumor and anticancer effects against various cancers. Incorporating squalene into the human diet is recommended without adverse health risks (Lozano-Grande *et al.*, 2018). Despite the diverse bioactivities of squalene and its potential applications in cosmetics, pharmaceuticals and nutrition, human intake remains low due to limited natural sources and extraction methods.

Fucosterol is also found as a minor constituent in *C. fenestratum* seed fat. A study done on the constituents present in the unsaponifiable matter of 10 Indian seed oils highlighted the presence of fucosterol in minor quantities (Saeed *et al.*, 1991). Due to the presence of fucosterol, the incorporation of *C. fenestratum* seed fat could provide therapeutic properties associated with fucosterol, including anti-cancer, anti-diabetic, reduction of oxidative stress and cholesterol-lowering effects (Maeda, 2013).

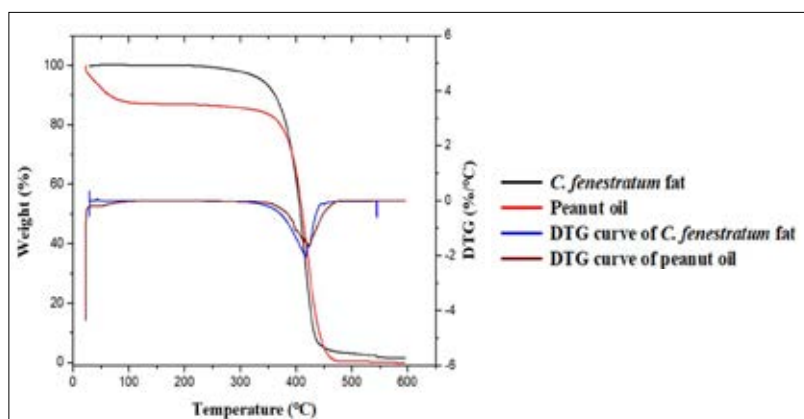
### Thermal stability *C. fenestratum* seed fat

Thermogravimetric Analysis (TGA) and Derivative Thermogravimetry (DTG) curves for *C. fenestratum* seed fat and peanut oil, illustrate the mass loss at increasing temperatures in an inert nitrogen atmosphere (Figure 4). The peak in the DTG curve signifies the temperature at which the maximum fat mass loss occurs in a nitrogen atmosphere, serving as an indicator of the thermal stability of fat. A heating rate of  $10\text{ }^{\circ}\text{C}/\text{min}$  was chosen in this study for its proven effectiveness in providing comprehensive insights into the thermal behaviour of oils, as established in previous studies (Jayadas & Nair, 2006).

The onset temperature in TGA analysis refers to the temperature at which the TGA curve first deviates from the baseline before the occurrence of the thermal event. The onset temperature, determined at a 2% loss of

mass accounts for potential weight loss due to moisture and volatile compounds (Jayadas & Nair, 2006). The comparison in Table 6 reveals that the onset temperature of the thermal decomposition of *C. fenestratum* seed fat is higher than that of sesame and coconut oils but lower than for sunflower oil (Jayadas & Nair, 2006). This analysis provides valuable information about the relative thermal stability of *C. fenestratum* seed fat compared to other commonly used oils. The final decomposition temperature of *C. fenestratum* seed fat obtained from the DTG curve was  $416.88 \pm 1.74^\circ\text{C}$  (Table 1).

This temperature signifies the point at which the fat undergoes complete degradation, and understanding this parameter is pivotal for assessing the thermal stability and potential applications of *C. fenestratum* seed fat. The decomposition temperature is compared with that of peanut oil ( $422.22 \pm 1.73^\circ\text{C}$ ), which is considered as one of the most thermally stable oils. The close decomposition temperature to the peanut oil confirms the high thermal stability of *C. fenestratum* seed fat, a crucial property for industrial applications of plant oil.



**Figure 4:** TGA and DTG curves of *C. fenestratum* seed fat and peanut oil (for comparison) obtained from thermogravimetric analysis

**Table 6:** Onset temperatures of thermal degradation obtained by *C. fenestratum* seed fat compared with commercially available plant oils. (Jayadas & Nair, 2006)

Name of the oil	Onset temperature ( $^\circ\text{C}$ )
Sunflower oil	345
Sesame oil	282
Coconut oil	257
Peanut oil	128
<i>C. fenestratum</i> fat	293

## CONCLUSION

This study reveals that mature seeds of *C. fenestratum* are a new source of high oil/fat bearing seeds, a property demanded in industrial applications of oil seeds. The ripened seeds yielded substantial oil content of  $46.94 \pm 0.01\%$ , along with favorable attributes such as low

moisture content ( $4.10 \pm 0.03\%$ ), AV ( $2.34 \pm 0.10$  mg KOH/g), IV ( $56.33 \pm 0.32$  g I<sub>2</sub>/100g), and a promising FA profile comparable to shea, sal, mango and kokum butter. The most abundant SFA is stearic acid, indicating potential applications in cosmetics for its emollient properties while the most abundant UFA is oleic acid, which is considered as a component of heart-healthy fat. The simple FA profile also signifies its potential use in other industrial applications as well. *C. fenestratum* seed fat also shows thermal stability closer to that of peanut oil. The chemical constituents in unsaponifiable matter indicate enhanced cosmetic properties such as moisturizing, anti-inflammatory and antioxidant effects. From a nutritional standpoint, the presence of phytosterols and squalene indicates potential applications in dietary supplements. While these preliminary findings are promising, further confirmation studies are required to confirm its full potential in cosmetics and dietary supplements. Future investigations could include comprehensive tests such as shelf-life, toxicity assessments, microbial analysis, and phase behaviour studies etc. These additional steps

will provide a thorough understanding of the potential industrial, cosmetic, and nutritional applications of *C. fenestratum* seed fat. However, *C. fenestratum* is considered a rare and critically endangered vine plant that is native to Sri Lanka and small parts of South Asia. Despite its critically endangered status, the plant is extensively used in a large number of commercial herbal medicine products. The plant is almost entirely sourced for commercial products from natural habitats due to the extreme difficulty of cultivation. However, this plant produces large number of seeds during its annual fruiting season, which are currently not used in any commercial applications. Therefore, we believe that the finding of this study would not pose a new threat to the critically endangered *C. fenestratum* plant and would provide more insight to its already known medicinal properties.

### Acknowledgements

We would like to acknowledge the financial support provided by university research grant ASP/01/RE SCI/2021/29, University of Sri Jayewardenepura. We also wish to acknowledge Mr. Koshala Gunawardana, forest manager of the University of Sri Jayewardenepura, for his support for seed collection and taxonomical identification of *C. fenestratum* and the Central instrument facility of the Faculty of Applied sciences, University of Sri Jayewardenepura for their support in chemical and thermal analysis.

### REFERENCES

- Abdul-mumeen I., Beauty D. & Adam A. (2019). Shea butter extraction technologies : Current status and future perspective. *African Journal of Biochemistry Research* 13(2), 9–22. DOI: 10.5897/AJBR2018.1007
- Abdullahi B.M., Garba A., Salihu A. & Saliu M.A. (2021). Effect of degumming on physicochemical properties of fatty acid ethyl esters obtained from *Acacia nilotica* seed oil. *Bioresource Technology Reports* 14, 1-8. DOI: 10.1016/j.biteb.2021.100678
- Caligiani A., Bonzanini F., Palla G., Cirilini M. & Bruni R. (2010). Characterization of a potential nutraceutical ingredient: Pomegranate (*Punica granatum* L.) seed oil unsaponifiable fraction. *Plant Foods for Human Nutrition* 65(3), 277–283. DOI: 10.1007/s11130-010-0173-5
- Falade O.S., Adekunle A.S., Aderogba M.A., Atanda S.O., Harwood C. & Adewusi S.R. (2008). Physicochemical properties, total phenol and tocopherol of some *Acacia* seed oils. *Journal of the Science of Food and Agriculture* 88, 263–268. DOI: 10.1002/jsfa.3082
- Firestone D. (ed.) (2013) *Physical and Chemical Characteristics of Oils, Fats, and Waxes*, 3<sup>rd</sup> edition. AOCS press.
- Fontanel D. (2013). *Unsaponifiable Matter in Plant Seed Oils*, pp. 365. Springer International Publishing, New York, USA.
- Giacometti J. (2001). Determination of aliphatic alcohols, squalene,  $\alpha$ -tocopherol and sterols in olive oils: Direct method involving gas chromatography of the unsaponifiable fraction following silylation. *Analyst* 126(4), 472–475. DOI: 10.1039/b007090o
- Glueck C., Speirs J., Tracy T., Streicher P., Illig E. & Vande-grift J. (1991). Relationships of serum plant sterols (phytosterols) and cholesterol in 595 hypercholesterolemic subjects, and familial aggregation of phytosterols, cholesterol, and premature coronary heart disease in hypercholesterolemic probands and their first-degree relatives. *Metabolism* 38, 136–140. DOI: 10.1016/0026-0495(91)90013-M
- Janporn S., Ho, C.T., Chavasit V., Pan M.H., Chittakorn S., Ruttarattanamongkol K. & Weerawatanakorn M. (2015). Physicochemical properties of *Terminalia catappa* seed oil as a novel dietary lipid source. *Journal of Food and Drug Analysis* 23(2), 201–209. DOI: 10.1016/j.jfda.2014.06.007
- Jayadas N.H. & Nair K.P. (2006). Coconut oil as base oil for industrial lubricants-evaluation and modification of thermal, oxidative and low temperature properties. *Tribology International* 39(9), 873–878. DOI: 10.1016/j.triboint.2005.06.006
- Johnson D.H. (1978). The use of fatty acid derivatives in cosmetics and toiletries. *Journal of the American Oil Chemists' Society* 55(4), 438–443. DOI: 10.1007/BF02911909
- Keneni Y.G., Bahiru L.A. & Marchetti J.M. (2021). Effects of different extraction solvents on oil extracted from *Jatropha* seeds and the potential of seed residues as a heat provider. *Bioenergy Research* 14, 1207–1222. DOI: 10.1007/s12155-020-10217-5
- Kim S.K. & Karadeniz F. (2012). Biological Importance and Applications of Squalene and Squalane. In *Advances in Food and Nutrition Research, Marine Medicinal Foods* (ed. S.K. Kim), pp. 223-233. Elsevier Academic Press, Amsterdam, Netherlands. DOI: 10.1016/B978-0-12-416003-3.00014-7. 65:223-33
- Kostik V., Memeti S. & Bauer B. (2013). Fatty acid composition of edible oils and fats. *Journal of Hygienic Engineering and Design* 4, 112–116.
- Kothalawala S.D., Edward D., Harasgama J.C., Ranaweera L., Weerasena O.V.D.S.J., Niloofa R., Ratnasooriya W.D., Premakumara G.A.S. & Handunnetti S.M. (2020). Immunomodulatory activity of a traditional Sri Lankan concoction of *Coriandrum sativum* L. and *Coscinium fenestratum* G. *Evidence-based Complementary and Alternative Medicine* 2020. DOI: 10.1155/2020/9715060
- Kyari M.Z. (2008). Extraction and characterization of seed oils. *International Agrophysics* 22, 139–142.
- Lenihan-Geels G., Bishop K.S. & Ferguson L.R. (2013). Alternative sources of omega-3 fats: Can we find a sustainable substitute for fish?. *Nutrients* 5(4), 1301–1315. DOI: 10.3390/nu5041301
- Lopaciuk A. & Łoboda M. (2013). Global beauty industry trends in the 21st century. In: Active citizenship by knowledge management & innovation. *Proceedings of*

- the Management, Knowledge and Learning International Conference 2013* (eds. V. Dermol, N. T. Širca & G. Dakovic), 19-21 June. Zadar, Croatia, pp. 1079–1087.
- Lozano-Grande M.A., Gorinstein S., Espitia-Rangel E., Dávila-Ortiz G. & Martínez-Ayala A.L. (2018). Plant sources, extraction methods, and uses of squalene. *International Journal of Agronomy* 2018, 1-13. DOI: 10.1155/2018/1829160
- Maeda H. (2013). Anti-obesity and anti-diabetic activities of algae. In: *Functional ingredients from algae for foods and nutraceuticals* (ed. H. Domínguez), pp. 453-472. Woodhead Publishing Series in Food Science, Technology and Nutrition. DOI: 10.1533/9780857098689.2.453
- Nasopoulou C. & Zabetakis I. (2012). Benefits of fish oil replacement by plant originated oils in compounded fish feeds. A review. *LWT- Food Science and Technology* 47(2), 217–224. DOI: 10.1016/j.lwt.2012.01.018
- Ogbe R.J., Ochalefu D.O., Mafulul S.G. & Olaniru O.B. (2015). A review on dietary phytosterols: their occurrence, metabolism and health benefits. *Asian Journal of Plant Science and Research* 5(4), 10–21. Available at: www.pelagiaresearchlibrary.com.
- Onyeike E.N. & Acheru G.N. (2002). Chemical composition of selected Nigerian oil seeds and physicochemical properties of the oil extracts. *Food Chemistry* 77(4), 431–437. DOI: 10.1016/S0308-8146(01)00377-6
- Pardeshi D.S. (2020). Evaluation of analytical results obtained from standard AOAC method and accelerated wijs method for the determination of iodine value of different brands of mustard edible oils. *International Journal of Innovative Science and Research Technology* 5(8), 1323–1327. DOI: 10.38124/ijisrt20aug679
- Phuong D.L., Thuy N.T., Long P.Q., Quan P.M., Thuy T.T.T., Minh P.T.H., Kuo P. & Thang T.D. (2018). Fatty acid, tocopherol, sterol compositions and antioxidant activity of three Garcinia seed oils. *Records of Natural Products* 12(4): 323–331. DOI: 10.25135/rnp.32.17.09.051054
- Przybylski R., Mag T., Eskin N.A.M. & McDonald B.E. (2005). Canola oil. *Bailey's industrial oil and fat products* 2, 61-122. DOI: 10.32388/mihx3w
- Rabasco Alvarez A.M. & González Rodríguez M.L. (2000). Lipids in pharmaceutical and cosmetic preparations. *Grasas y Aceites* 51(1–2), 74–96. DOI:10.3989/gya.2000.v51.i1-2.409
- Ramasubbu R., Prabha A.C. & Kumuthakalavalli R. (2012). Seed biology of *Coscinium fenestratum* (Gaertn.) Colebr.- A critically endangered medicinal plant of Western Ghats. *Journal of Medicinal Plants Research* 6(6), 1094–1096. DOI: 10.5897/jmpr11.558
- Rai R.V., Rajesh P.S. & Kim H.M. (2013). Medicinal use of *Coscinium fenestratum* (Gaertn.) Colebr.: An short review. *Oriental Pharmacy and Experimental Medicine* 13,1–9. DOI: 10.1007/s13596-012-0094-y
- Saeed M.T., Agarwal R., Khan M.W.Y., Ahmad F., Osman S.M., Akihisa T., Suzuki K. & Matsumoto T. (1991). Unsaponifiable lipid constituents of ten Indian seed oils 68(3), 193–197.
- Shukla V.K.S. & Kragballe K. (1998). Exotic butters as cosmetic lipids. *INFORM - International News on Fats, Oils and Related Materials* 9, 512–516.
- Solís-Fuentes J.A., Camey-Ortiz G., del Rosario Hernández-Medel M., Pérez-Mendoza F. & Durán-de-Bazúa C. (2010). Composition, phase behavior and thermal stability of natural edible fat from rambutan (*Nephelium lappaceum* L.) seed. *Bioresource Technology* 101(2), 799–803. DOI: 10.1016/j.biortech.2009.08.031
- Vermaak I., Kamatou G.P.P. Komane-Mofokeng B., Viljoen A.M. & Beckett K. (2011). African seed oils of commercial importance - Cosmetic applications. *South African Journal of Botany* 77(4), 920–933. DOI: 10.1016/j.sajb.2011.07.003
- Warakagoda P. & Subasinghe S. (2014). In vitro Seed Germination of *Coscinium fenestratum* (Gaertn.) Colebr. *Annual Research & Review in Biology* 4(23), 3549–3565. DOI: 10.9734/arrb/2014/10558
- Williams C.M. (2000). Dietary fatty acids and human health. *Animal Research* 49(3), 165-180. DOI: 10.1051/animres:2000116
- Yadav G.G., Murthy H.N. & Dewir Y.H. (2022). Nutritional composition and in vitro antioxidant activities of seed kernel and seed oil of *Balanites roxburghii*: An underutilized species. *Horticulturae* 8(9). DOI: 10.3390/horticulturae8090798
- Yao Y. & Xu B. (2021). New insights into chemical compositions and health promoting effects of edible oils from new resources. *Food Chemistry* 364(June), 1-11. DOI: 10.1016/j.foodchem.2021.130363
- Zang C.U., Jock A.A., Garba I.H. & Chindo I.Y. (2017). Physicochemical and phytochemical characterization of seed kernel oil from desert date (*Balanites Aegyptica*). *Journal of Chemical Engineering and Bioanalytical Chemistry* 2(1), 49–61. DOI: 10.25177/jcebc.2.1.1

## RESEARCH ARTICLE

### Genetics and Plant Breeding

# Coincidence of inductive photoperiod and early vegetative phase leads to early flowering and increased yield in selected Sri Lankan traditional rice varieties

WHDU Pushpakumari<sup>1</sup>, NVT Jayaprada<sup>2</sup>, LALW Jayasekara<sup>3</sup>, G Senanayake<sup>4</sup>, DMJB Senanayake<sup>5</sup> and S Geekiyanage<sup>4\*</sup>

<sup>1</sup> Board of Study in Agriculture, Faculty of Graduate Studies, University of Ruhuna, Matara, Sri Lanka.

<sup>2</sup> Department of Agricultural Technology, Faculty of Technology, University of Colombo, Pitipana, Sri Lanka.

<sup>3</sup> Department of Mathematics, Faculty of Science, University of Ruhuna, Matara, Sri Lanka.

<sup>4</sup> Department of Agricultural Biology, Faculty of Agriculture, University of Ruhuna, Mapalana, Sri Lanka.

<sup>5</sup> Rice Research and Development Institute, Bathalagoda, Sri Lanka.

Submitted: 15 August 2024; Revised: 13 December 2024; Accepted: 06 January 2025


**Abstract:** Despite the nutritional and environmental benefits, Sri Lankan traditional rice (SLTR) is under-cultivated due to the photoperiod sensitivity in most varieties. This experiment was carried out to determine the effect of photoperiod on the days to flowering (DF) and number of spikelets per first panicle (SP) of selected SLTR varieties *Masuran* (5530), *Heras* (6412) and *Hathe pas dawase wee* (4237), by exposing to different photoperiods and developing a gene network model. Three week old plants were exposed to time controlled sunlight for three photoperiods of 11 hour and 45 minutes representing short-day photoperiod (SD), 12 hours, and 12 hours and 05 minutes in the photoperiod chamber, and to the prevailing natural photoperiod, which ranged from 12 hours and 14 minutes to 12 hours and 02 minutes, and 12 hours and 15 minutes to 12 hours and 03 minutes, in two locations in a completely randomized design with three replicates. DF and SP were recorded. The ANOVA revealed that DF and SP of each accession were significantly affected by photoperiod ( $p < 0.05$ ). The SD resulted in significantly reduced DF in three accessions. The increased DF reduced the SP in accessions 5530 ( $r = -0.51$ ) and 6412 ( $r = -0.53$ ) according to the correlation analysis ( $p < 0.05$ ). The photoperiod effect on SP was explained through quadratic regression models, which indicated an optimum photoperiod for the highest SP ( $p < 0.05$ ). Gene network model analysis suggested distinct interactions among pleiotropic genetic factors in response to photoperiod, and tested rice accessions. *Hd1* and *Ghd7* interactions mediated late-flowering under long-day conditions, while the absence of the *Ghd7* pathway mitigated

the delayed flowering response. *DTH8*-mediated suppression of *Ehd1* and *Hd3a* under long day conditions led to delayed flowering and a simultaneous increase in SP. Alternatively, the *Ehd1-RFT1* pathway facilitated early flowering under short-day conditions in the absence of *Hd1*. Plant responses and suggested gene interactions provide valuable insights for manipulating DF and yield in breeding SLTR.

**Keywords:** Days to flowering, gene network model, increased yield, photoperiod, Sri Lankan traditional rice.

## INTRODUCTION

Rice grown in Asia is a major component of food security in today's context (Zhao *et al.*, 2020). However, the rice sector will face significant challenges in meeting the future demand, driven by the projected global population growth from 8.9 to 10.6 billion by 2050 (United Nations, 2019). Rice production is mainly dependent on the rice varieties, environmental factors, and inputs. Response to the environmental clues of photoperiod and temperature are common among most of the rice varieties, which are critical for the transition from vegetative phase to initiation of flowering. The temperature perception pathway and photoperiod pathway converge for the initiation (Vicentini *et al.*, 2023). The effect of photoperiod and temperature

\* Corresponding author (sudarshanee@agbio.ruh.ac.lk;  <https://orcid.org/0000-0002-3771-2680>)



This article is published under the Creative Commons CC-BY-ND License (<http://creativecommons.org/licenses/by-nd/4.0/>). This license permits use, distribution and reproduction, commercial and non-commercial, provided that the original work is properly cited and is not changed in anyway.

on the early vegetative phase of a Sri Lankan traditional rice core collection has been reported in previous studies. The average number of days to reach the fifth leaf stage in the core collection was significantly lower under a high temperature regime of  $36.9 \pm 0.4$  °C compared to the low temperature regime of  $34.0 \pm 1.0$  °C, irrespective of the photoperiod (Rathnathunga & Geekiyanage, 2017a; Rathnathunga et al., 2019). According to Vergara and Chang (1985), rice is a facultative short-day plant that flowers early under short day exposure, while some rice accessions can flower under long-day conditions as well. During the photoperiod sensitive phase, the rice plant must receive the photoperiod required for flowering initiation: the critical photoperiod is dependent on the variety and determines the time of transition from the vegetative phase to flowering initiation. When a rice plant does not meet the critical photoperiod, it remains in the vegetative phase. The shortest time to initiate flowering is achieved when plants are exposed to the optimum photoperiod. The requirement for shorter day lengths to promote faster flowering is typically fulfilled with day lengths shorter than 13.5 hours (Itoh et al., 2010). Extremely photoperiod sensitive varieties of Sri Lankan traditional rice remain in the vegetative phase if they are not exposed to the critical photoperiod in the latter part of the *Maha* season under natural field conditions (Rathnathunga et al., 2016 a). The optimum photoperiod is an important phenomenon in the ecological adaptation of rice varieties for climate change resilience (Vergara & Chang, 1985).

The genetic basis of the photoperiod effect on days to flowering (DF) of rice has been intensively studied. Specific pathways are assigned to the flowering time quantitative trait loci, some of which behave in accordance with Mendelian laws (Hori et al., 2016). Identification of a key flowering time gene in rice named *Heading date 1* (*Hd1*) of the CCT Family transcription factor was a major breakthrough in flowering time studies (Yano et al., 2000). *Hd1* is activated by *OsGI* and *Hd1* promotes the florigen *Hd3a* under short-day conditions (Kojima et al., 2002; Hamaya et al., 2003). The rice flowering locus *T 1* (*RFT1*) encodes a mobile flowering signal that induces floral transition under short-day environments (Komiya et al., 2008). In addition, *Early heading date 1* (*Ehd1*), *B* type response regulator is rice-specific and reported to be a mediator of the key flowering time pathway network under both short-day and long-day conditions (Vincentini et al., 2023). Strong repressors of *Ehd1* under long-days include *grain number*, *plant height and heading date 7* (*Ghd7*, also known as *Hd4* and *Ghd8*) (Xue et al., 2008) and *DTH8* (also known as *Hd5* or *Ghd8*) (Wang et al., 2021), which affect DF and yield related agronomic traits. Interactions of non-functional *Hd1*

with *Ghd7* and *DTH8* are established to explain the time of flowering (Sun et al., 2022). Under long-days, *Hd1* represses flowering (Yano et al., 2000). Flowering repression of *Hd1* under long-days is through repressing *Ehd1*, which indicates the importance of non-functional *Hd1* and *Ehd1* in artificial selection of rice for wider environments (Wei et al., 2015). In addition, the presence of unique non-Circadian clock genes of rice, which act as master switches of transition from the vegetative phase to the reproductive phase, such as *Rice Indeterminate 1* (*RIDI1*), also play an important role in floral transition (Wu et al., 2008). The effect of photoperiod on flower opening time in rice was recently studied. The variability in flower opening time is reported to be significantly lower in photoperiod sensitive cultivars compared to photoperiod-insensitive cultivars for flowering initiation (Deb, 2024).

Sri Lankan traditional rice germplasm is a diverse collection of accessions for variation in flowering time, yield and several agronomic characters (Rathnathunga et al., 2014; Padukkage et al., 2015 ; Pushpakumari et al., 2016; Rathnathunga et al., 2016a). Early maturing rice accessions are grouped as short-day sensitive and long-day sensitive for flowering initiation (Padukkage et al., 2017). Diverse genes are reported to be responding to short-day and long-day conditions in early flowering sub-populations of *indica* and *japonica* rice (Han et al., 2016). In a study using temperate *japonica* rice grown in Northeast China through molecular markers and flowering time records, *DTH2* and *Hd18* were identified as the major flowering promoters. *Hd2*, *Hd4*, and *Hd5* were identified as the major flowering repressors. Furthermore, *Hd6* and *Hd16* were identified as minor flowering repressors, and *Hd17* was minor flowering promoter (Li et al., 2018). CRISPR/Cas9-mediated multiplex genome editing for manipulating flowering time in rice was reported (He et al., 2024). Flowering time genes of *Hd2*, *Ghd7*, and *DTH8* were edited in rice, and the mutant lines of the above genes flowered earlier without a significant yield loss. The successful development of homozygous mutant lines with early flowering and no yield reduction provides a foundation for breeding rice varieties with shorter growth periods and greater adaptability for rice breeding applications.

Several attempts have been made to determine the effect of the number of DF on grain yield. Early flowering during an inductive short-day season increased the grain yield of *Ma wee* accessions under field conditions, whereas delayed flowering of these accessions in a non-inductive season reduced the yield (Pushpakumari et al., 2017). On the other hand, several studies reported a

significant positive correlation between DF of different rice accessions and their grain yield (Pushpakumari & Geekiyanage, 2014; Pushpakumari *et al.*, 2016; Rathnathunga & Geekiyanage, 2017b; Jayasinghe *et al.*, 2023). Furthermore, in several other studies, variation in DF among accessions of the same rice variety within a single growing season was positively correlated with vegetative growth while negatively correlated with grain yield in the varieties of *Kalu heenati* (Pushpakumari & Geekiyanage, 2015), *Sudu wee* (Rathnathunga & Geekiyanage, 2016), *Hondarawala* (Rathnathunga *et al.*, 2016b) and the varieties of *Suduru samba* and *Pachchaperumal* (Rathnathunga & Geekiyanage, 2017b). Additionally, a positive correlation was found between plant height and yield in selected *Ma wee* accessions (Jayasinghe *et al.*, 2023).

These research findings highlight the necessity of investigating the genetic basis of the response to inductive photoperiod exposure during the photoperiod-sensitive phase of rice to manipulate yield. Based on the above reports, we hypothesize that there is genetic variation in Sri Lankan traditional rice varieties in response to inductive photoperiod exposure during the photoperiod-sensitive phase, which could be used to manipulate yield.

Therefore, the present study was conducted to determine the effect of photoperiod, through exposure to time controlled-natural sunlight during a defined period of vegetative phase on DF and yield under Sri Lankan conditions. Additionally, the study aimed to identify the gene network associated with photoperiod sensitivity and yield manipulation, in order to utilize the generated knowledge for future breeding purposes.

## MATERIALS AND METHODS

### Plant material

Three traditional rice accessions (accession 5530 of variety *Masuran*, accession 6412 of variety *Heras*, and accession 4237 of variety *Hathe pas dawase wee*) were selected based on the seasonal variation of flowering time during the *Yala* (in 2019) and *Maha* (in 2018/2019) seasons (unpublished data).

### Experimental location, design and duration

The study was carried out at two locations in Ibbagamuwa, Kurunegala District, Sri Lanka in the agro-ecological zone IL1b (7.3500° N, 80.4333° E) from August 2021 to February 2022, and Vantharumoolai of the Batticaloa

District, Sri Lanka, in agroecological zone DL2b (7.7178° N, 81.6894° E) from August 2022 to February 2023. The pots were filled with a mixture of mud and sandy loam soil in the Ibbagamuwa experimental site, and with sandy soil in the Vantharumoolai experimental site. Pots with germinating seeds were exposed to sunlight. The exposure of plants to sunlight was restricted by a black cover for three photoperiod regimes of 11 h and 45 min, 12 h, and 12 h and 05 min from the 19<sup>th</sup> and 21<sup>st</sup> day of seed establishment at Ibbagamuwa and Vantharumoolai respectively, for a period of 30 ds. The choice of photoperiods—11 h and 45 min, 12 h, and 12 h and 5 min—was based on the natural variations in photoperiods under Sri Lankan conditions for plant exposure for one month during the early vegetative phase. The 11 h and 45 min photoperiod corresponds to the naturally prevailing short-day condition for about one month in the *Maha* season, which is supposed to facilitate traditional rice flowering behaviour under short day conditions in Sri Lanka. The chosen photoperiods of 12 h, and 12 h and 5 min are realistic representations of the day-neutral and long-day conditions that prevail during this period, ensuring that the experiment closely reflects the natural environmental conditions for the rice plants during the early vegetative phase. Such subtle differences were considered necessary to identify sensitivities in flowering regulation genes at the basic vegetative stage. After the treatment, the plants were exposed to photoperiods available naturally without any restriction. The control plants were exposed to photoperiods available naturally without any restriction for light exposure throughout the period. The experiment was laid out with three replicates in a completely randomized design.

### Plant data collection and analysis

The DF and number of spikelets per first panicle (SP) of each accession under 4 photoperiods were recorded according to the descriptors for rice characters (Rathnathunga *et al.*, 2014; Padukkage *et al.*, 2015; Pushpakumari *et al.*, 2016). and mean values were calculated. The DF and SP data were separately analyzed through two way-ANOVA for significant effects of three imposed photoperiods, three rice accessions and their interaction. This was followed by Tukey's honestly significant difference (HSD) post-hoc test to identify significant differences in the DF and SP of the different treatments. Secondly, an analysis of simple main effects was conducted to explore the interaction effects further and to understand the impact of either variable (photoperiod or accession) at each level on the other variable. Following this, Fisher's least significant difference (LSD) test was employed to make pairwise comparisons.



A correlation analysis was carried out for the relationship between the DF and SP, while a regression analysis was carried out to determine the effect of photoperiod on SP for each accession at each location. Both methods of analysis were performed using IBM SPSS Statistics software (version 23).

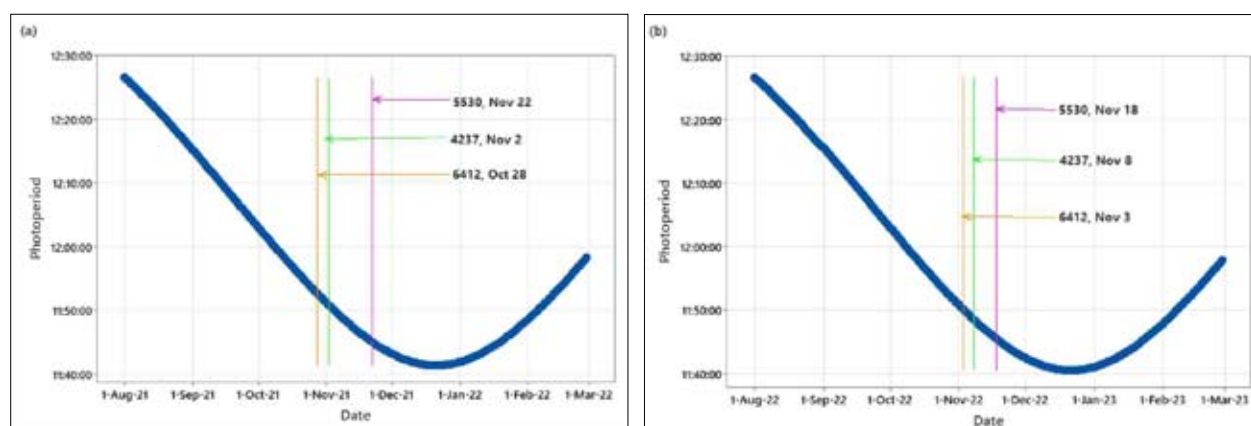
### Gene network model generation

Information on genes explaining the correlation between flowering and SP was collected from KEGG (<https://www.genome.jp/kegg/pathway.html>) and STRING (<https://string-db.org/>) databases and prior literature. The network table files were created out of these details. Network data were imported into the Cytoscape interface. Interaction parameters were defined by specifying columns of data containing the source interaction, target interaction, and interaction type. Network models were generated using Cytoscape 3.8. In the generated network, nodes (or vertices) represent genes, while edges represent relationships of interactions. The generated network was analyzed to decipher gene interactions.

## RESULTS AND DISCUSSION

### Natural photoperiod variation at two experimental sites

The natural photoperiod was experienced by rice plants from the first to the 18<sup>th</sup> and 21<sup>st</sup> day of seed germination, ranging over time from 12 hours and 21 minutes to 12 hours and 14 minutes, and 12 hours and 23 minutes to 12 hours and 15 minutes, in Ibbagamuwa and Vantharumoolai during the periods from 15.08.2021 to 02.09.2021 and 10.08.2022 to 31.08.2022, respectively. During the 30-day period of photoperiod treatments, the natural photoperiod variation was from 12 hours and 14 minutes to 12 hours and 02 minutes, and 12 hours and 15 minutes to 12 hours and 03 minutes in Ibbagamuwa and Vantharumoolai, respectively. After the 30-day treatment period, all plants were again exposed to sunlight without restriction. Plants experienced a reducing natural photoperiod during the experimental period, which ranged at the end of treatment at 12 hours and 02 minutes and 12 hours and 03 minutes, to 11 hours and 44 minutes and 11 hours and 45 minutes, on the 100<sup>th</sup> day in Ibbagamuwa and Vantharumoolai, respectively (Figure 1).



**Figure 1:** Natural variation of photoperiod during the experiment and flowering time of three accessions in Ibbagamuwa, Kurunegala (a) and Vantharumoolai, Batticaloa (b).

### Effect of photoperiod on days to flowering

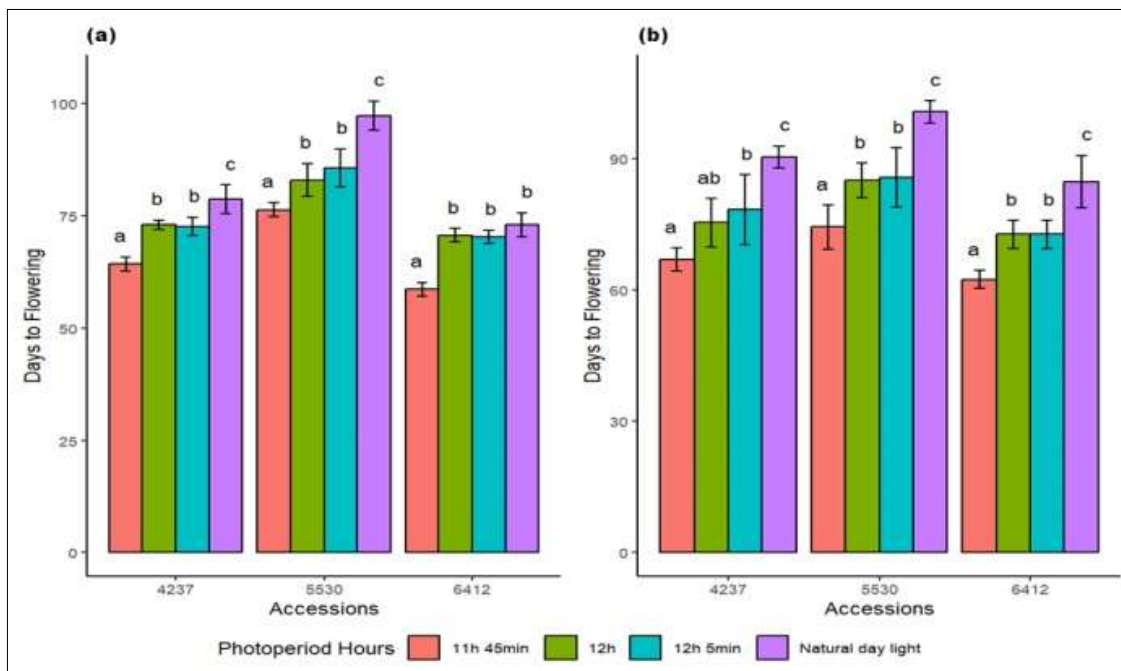
Exposure to photoperiod during the early vegetative phase affected the number of DF of three accessions significantly at each location. Three accessions responded at different levels to photoperiod. There was a significant interaction between rice accession and photoperiod for DF (Table 1). Simple main effect analysis revealed that the photoperiod of 11 hours and 45 minutes significantly

produced the least number of DF in the 3 accessions of 5530 ( $76.33 \pm 0.51$  and  $74.33 \pm 1.67$ ), 6412 ( $58.67 \pm 0.50$  and  $62.33 \pm 0.69$ ), and 4237 ( $64.33 \pm 0.61$  and  $67 \pm 0.88$ ) at Ibbagamuwa and Vantharumoolai, respectively (Supplementary Table S1 and Figure 2). Photoperiods of 12 hours, and 12 hours and 5 minutes during the treatment period, had influenced the DF positively, which is evident through regression analysis. Furthermore, exposure to the natural photoperiod which



was a long day condition for the 18<sup>th</sup> day and 21<sup>st</sup> day of seed germination in Ibbagamuwa and Vanthuramoolai,

respectively, increased the DF significantly (Figure 1 and Figure 2).



**Figure 2:** Effect of photoperiod on days to flowering of the three selected accessions at two locations in Sri Lanka. (a) Ibbagamuwa, (b) Vanthuramoolai. Significantly different values of each accession at each location are given in different letters.

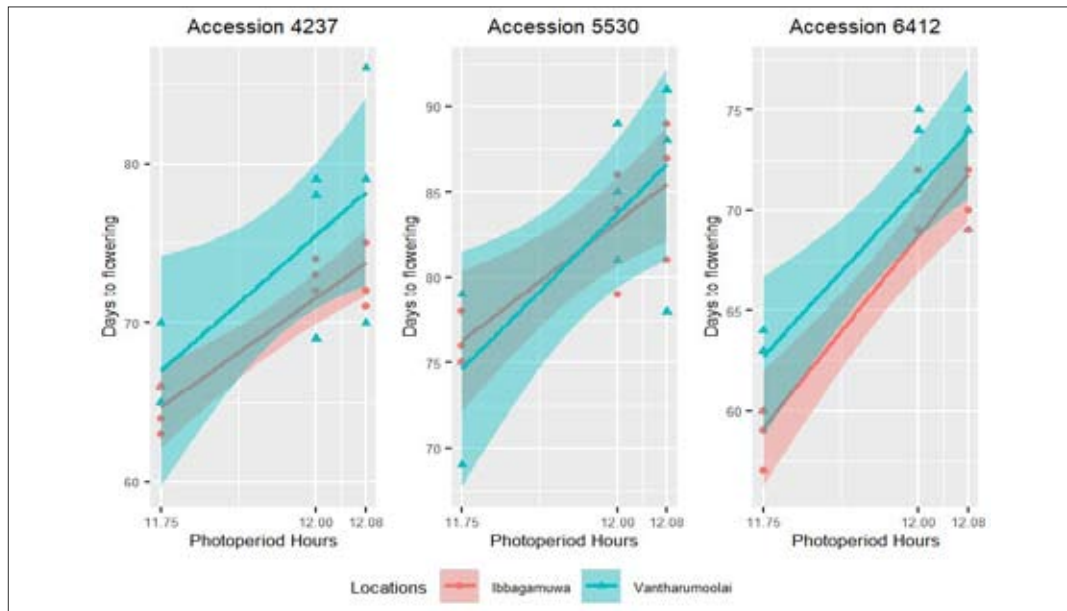
**Table 1:** Analysis of Variance for the effect of photoperiod and rice accession on days to flowering and number of spikelets per first panicle

Variables	F value					
	Photoperiod	Ibbagamuwa Accession	Photoperiod x Accession	Photoperiod	Vantharumoolai Accession	Photoperiod x Accession
<b>DF</b>	66.541*	159.791*	3.483*	39.829*	25.033*	0.177
<b>SP</b>	901.639*	166.374*	108.500*	727.498*	1049.494*	130.762*

\* F value is significant at  $p < 0.05$

There was a variation in DF among three accessions under similar conditions of each photoperiod, suggesting their genetic diversity in responding to similar growing conditions. The linear regression relationships for the two locations under the tested photoperiods did not overlap in accessions 4237 and 6412, while those of 5530 overlapped at two locations under the tested photoperiod conditions (Figure 3).

The photoperiod response curves have been used by Chandraratna (1955) to determine the optimum photoperiod at the minimum number of DF. During this experiment, the photoperiod of 11 hours and 45 minutes was tested as it is close to the lowest photoperiod available under natural conditions in Sri Lanka. According to the regression analysis, short day conditions are favourable for early flowering (Figure 3). Therefore, the optimum



**Figure 3:** Effect of photoperiod on days to flowering of the three accessions based on the regression analysis (Accession 4237;  $DF = -103.5 + 14.66 \text{ Photoperiod}$ ,  $R\text{-Sq} = 89.3\%$  (Ibbagamuwa),  $DF = -131.7 + 17.34 \text{ Photoperiod}$ ,  $R\text{-Sq} = 48.8\%$  (Vantharumoolai), Accession 5530;  $DF = -85.60 + 14.14 \text{ Photoperiod}$ ,  $R\text{-Sq} = 65.1\%$  (Ibbagamuwa),  $DF = -144.3 + 19.10 \text{ Photoperiod}$ ,  $R\text{-Sq} = 58.0\%$  (Vantharumoolai), Accession 6412;  $DF = -175.2 + 20.43 \text{ Photoperiod}$ ,  $R\text{-Sq} = 94.3\%$  (Ibbagamuwa),  $DF = -142.2 + 17.87 \text{ Photoperiod}$ ,  $R\text{-Sq} = 80.6\%$  (Vantharumoolai).

inductive photoperiod of all 3 accessions for early flowering should be 11 hours and 45 minutes, which resulted in the lowest DF under tested conditions or photoperiods less than that. Accession 4237 was tested in a previous experiment by Padukkage *et al.* (2017) for DF under 8 hours, 12 hours and 14 hours of light throughout the plant life span. Accession 4237 flowered early under 8 hours of light over day-neutral conditions and long-day conditions of 14 hours of light. However, eight hours of sunlight exposure is not naturally available under tropical conditions in Sri Lanka. Early flowering at 11 hours and 45 minutes for a period of 1 month in the early vegetative phase, which is only 15 minutes less than day-neutral conditions, creates an interesting question to reveal the genetic factors that are sensitive to such a marginal photoperiod difference under tropical conditions.

### Effect of photoperiod on number of spikelets per the first panicle

Photoperiod and rice accession were significant factors affecting the SP at both locations according to the analysis of variance (Table 1). Accessions 5530 and 6412 produced the highest number of spikelets per first panicle under the photoperiod of 11 hours and 45 minutes at

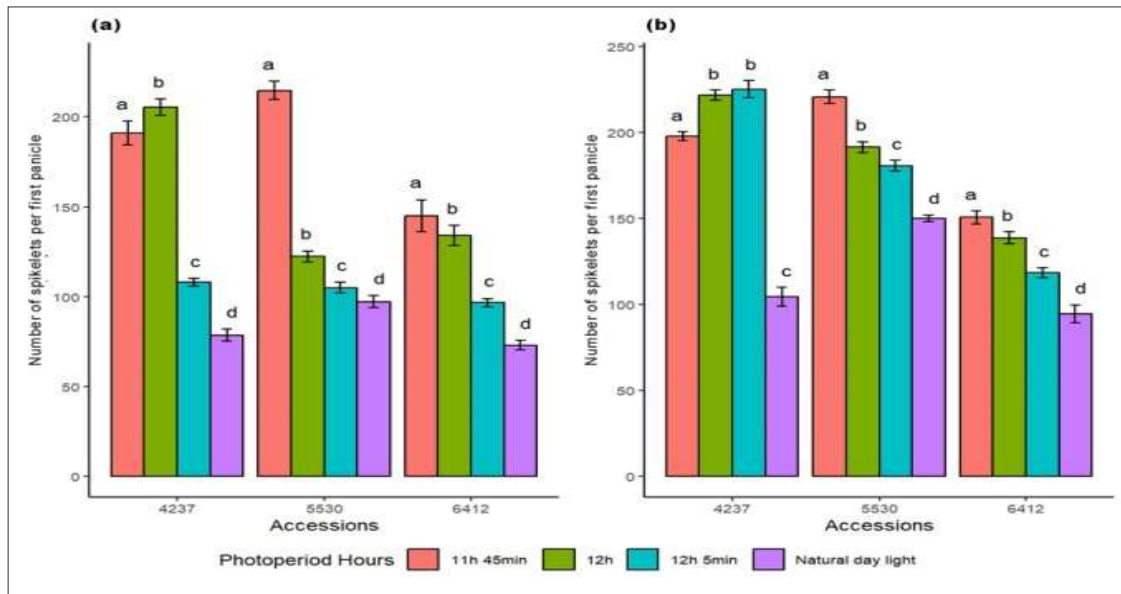
Ibbagamuwa ( $214.6 \pm 0.9$ ,  $145 \pm 3$ ) and Vantharumoolai ( $220 \pm 4$ ,  $150.6 \pm 1.4$ ) respectively. The SPs of accession 4235 under 12 hours and 12 hours and 05 minutes were not lower than that under 11 hours and 45 minutes, while the natural photoperiod had the significant effect of lowering the number of spikelets according to the single main effect analysis (Supplementary Table S2).

### Relationship between days to flowering and number of spikelets per first panicle

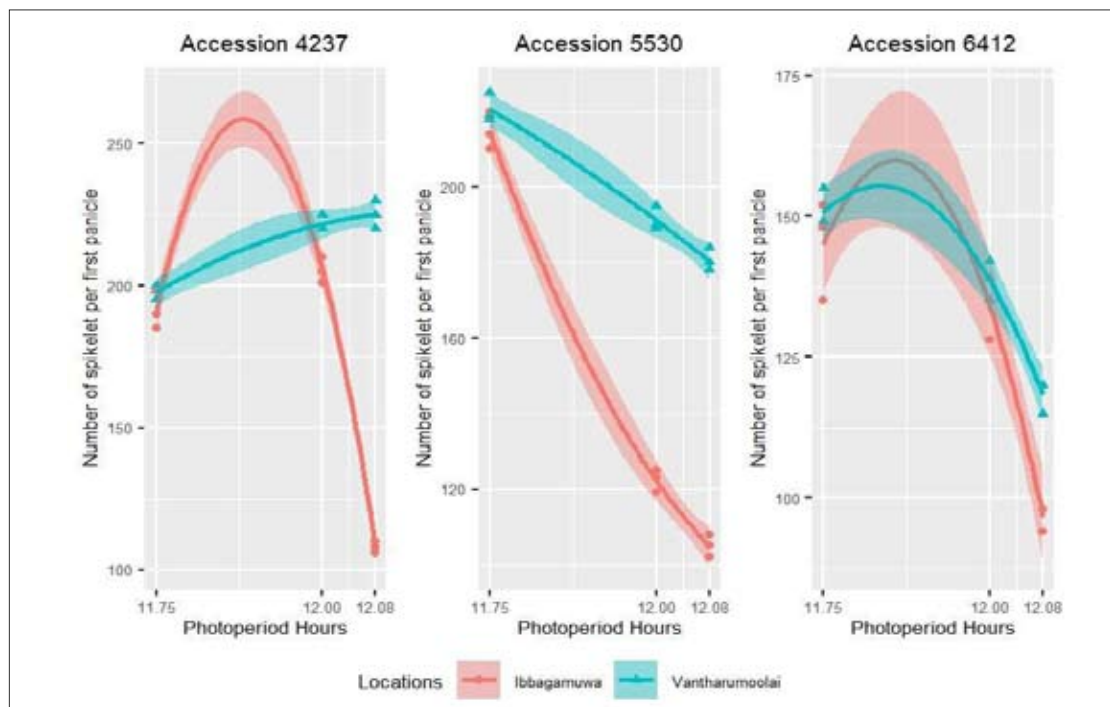
The correlation analysis revealed that the number of DF is negatively correlated with SP in accessions 5530 and 6412 with a correlation coefficient of 0.51 and 0.53 respectively at the 0.05% level of significance (Table 2).

**Table 2:** Correlation between days to flowering and the number of spikelets per first panicle

Accession	Correlation coefficient (r)	p value
4237	0.1878	0.4555
5530	-0.5168	0.0280
6412	-0.5341	0.0224



**Figure 4:** Effect of photoperiod on number of spikelets per first panicle of the three selected accessions at two locations in Sri Lanka (a) Ibbagamuwa, (b) Vantharumoolai. Significantly different values of each accession at each location are given in different letters.



**Figure 5:** Relationship between photoperiod exposure and number of spikelets per first panicle at two locations. (Ibbagamuwa: Accession 6412:  $SP = -38806 + 6820 \text{ Photoperiod} - 298.0 \text{ Photoperiod}^2$  ( $R^2 = 99.6\%$ ); Accession 5530:  $SP = -38806 + 6820 \text{ Photoperiod} - 298.0 \text{ Photoperiod}^2$  ( $R^2 = 99.6\%$ ); Accession 4237:  $SP = -451862 + 77127 \text{ Photoperiod} - 3288 \text{ Photoperiod}^2$  ( $R^2 = 99.2\%$ ); Vantharumoolai: Accession 6412:  $SP = -87730 + 15019 \text{ Photoperiod} - 641.4 \text{ Photoperiod}^2$  ( $R^2 = 95.8\%$ ), Accession 5530:  $SP = -35809 + 6200 \text{ Photoperiod} - 266.7 \text{ Photoperiod}^2$  ( $R^2 = 97.4\%$ ), Accession 4237:  $SP = 4972 - 856 \text{ Photoperiod} + 38.4 \text{ Photoperiod}^2$  ( $R^2 = 94.4\%$ )).

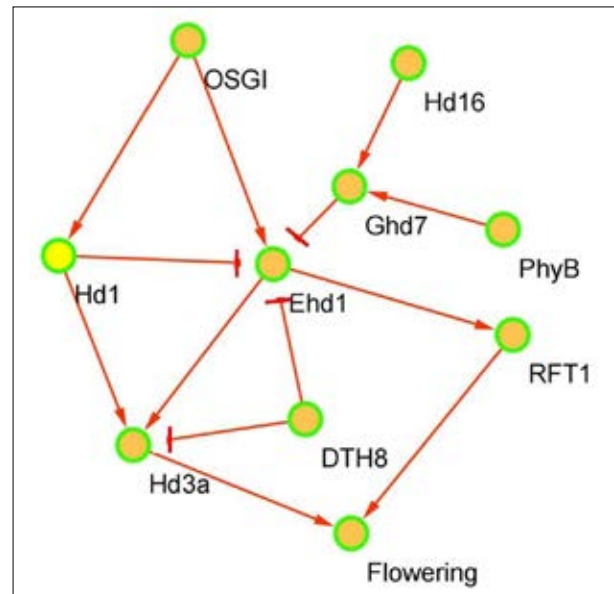
Long-aged rice varieties are considered to be high yielding over short-aged rice varieties. Previous studies on the relationship between DF of different traditional rice accessions and SP under similar field conditions had reported positive correlations between DF and SP (Rathnathunga *et al.*, 2016a; Rathnathunga & Geekiyanage, 2016; Jayasinghe *et al.*, 2023). According to a photoperiod chamber experiment, yield was reduced in the rice varieties *Sulai*, *Kohu ma wee*, and *Devaraddili* under long-day conditions (Geekiyanage *et al.*, 2012). In contrast, timing of planting affects DF and SP in selected *Ma wee* accessions, where the extended DF of the same accession due to seasonal effects reduced the yield (Pushpakumari *et al.*, 2017).

Chandrarathna (1955) considered the minimum photoperiod for the minimum DF as the optimum photoperiod without considering the effect of DF on the yield. During this study, we developed the regression models of SP and photoperiod for each accession to determine the optimum photoperiod for the highest SP (Figure 5). The optimum photoperiod for the highest SP can be deduced through the regression curve. Accessions 5530 and 6412 produce the maximum SP at the photoperiod of 11 hours and 45 minutes at both locations. According to the postulated SP for the range of tested photoperiods, it is suggested that the photoperiod under 11 hours and 45 minutes giving the lowest DF is not the optimum photoperiod for SP for accession 4237, reconfirming the non-correlation of DF and SP in accession 4237 (Table 2).

### Gene network model analysis for identification of key flowering time genes involved in vegetative growth and yield

Different gene models were generated to predict the possible gene interactions in the three tested rice accessions for the DF and yield under three photoperiods (Figure 6). The model explains the behaviour of accession 5530 as follows: under short-day conditions, *OSG1* promotes the *Hd1*, which promotes the *Ehd1* and *Hd3a* expression for flowering initiation (Endo-Higashi & Izawa, 2011). In addition to the effect of *Hd1* on flowering, there are reports on its pleiotropic effects on plant height and yield traits of SP and number of grains per panicle (NGP), in the genetic background of the *indica* rice variety ZS97 (Zhang *et al.*, 2012; Yan *et al.*, 2011).

Therefore, the significant increase in SP under the short-day photoperiod of 11 hours and 45 minutes in accession 5530 could be due to the direct influence of *Hd1* as reported by Zhang *et al.* (2012).



**Figure 6:** Proposed generalized gene network models for rice accessions 4237, 5530, and 6412 on flowering initiation under short-day and long-day photoperiods.

In the gene pathway model for accession 5530 under long-day condition, the presence of functional *Ghd7* is suggested for the interaction with *Hd1* to delay flowering (Figure 6). A night break of even 10 minutes results in delayed flowering through adversely affecting *Hd3a* expression (Ishikawa *et al.*, 2005). Therefore, the model indicates the unavailability of *Hd3a* under 12 hours, and 12 hours and 05 minutes in accession 5530. According to the model, *DTH8* is not present in the accession 5530. As a result, the delayed number of DF under long-day conditions negatively influenced the SP. The red-light photoreceptor, *Phytochrome B* and *Hd16* induce *Ghd7*, leading to delayed flowering under long-day conditions. It is postulated that *Phytochrome B* may affect the *Hd1* function adversely at the post-transcription level (Zhou *et al.*, 2021).

The accession 6412 exhibited the shortest DF ( $59.0 \pm 1.8$  days) among the three tested accessions under short-day conditions. Accordingly, accession 6412 is not supposed to have functional *Ghd7*, expression of which delays flowering under short-day conditions also (Zhang *et al.*, 2019). Two *Hd1*-induced pathways were proposed for accession 6412 with the presence of *Hd3a*, *Hd1*, *Ehd1* and *RFT1* (Figure 6). Among the previously derived pathways, those involving *Ghd7* were presumed to be absent in this accession, allowing *Hd1* to initiate

early flowering in 6412 under short-day conditions. *Hd16* was reported to shorten flowering time by 20 days under long days and to extend DF under short days in the non functional *Ghd7* background (Hori *et al.*, 2013; Kwon *et al.*, 2014). *Hd1* promoted early heading under two pathways, where one promoted *Ehd1* and then *Hd3a*, as observed in the 5530 accession. In the second pathway, *Ehd1* expression under short-day conditions promotes *RFT1* expression (Zhu *et al.*, 2017). Simultaneously, *Hd1* induces higher SP under short-day conditions as well.

A genetic model with two pathways was proposed for 4237 with *DTH8*, *Ehd1*, and *RFT1* (Figure 6). In the absence of *Hd1*, *Ehd1* may have induced the expression of *RFT1*, leading to early flowering in accession 4237 under short-day conditions. The presence of *DTH8* in the model explains the increase of SP under long-day conditions of 12 hours and 5 minutes with delayed flowering; *DTH8* down-regulates the transcriptions of *Ehd1* and *Hd3a* under long-day conditions (Wei *et al.*, 2010). Our results can be supported by the previous work reporting the findings that *DTH8* plays an important role in the signal network of photoperiodic flowering as a suppressor, as well as in the regulation of yield (Wei *et al.*, 2016).

The *Hd1* floral repressor activity basically requires *Ghd7* in long-day conditions in functional *DTH8* background (Nemoto *et al.*, 2016). The heading repression in long-day conditions by *Hd1* also depends on *DTH8* in functional *Ghd7* background (Du *et al.*, 2017). Post-transcriptional interactions of *Hd1-Ghd7* negatively influences flowering initiation by binding to the *Ehd1* promoter and suppressing the expression of *Ehd1* under long-day conditions (Zhou *et al.*, 2021).

The comprehensive mechanism by which these genes interactively control the various extents of photoperiod sensitivity in the same genetic background needs to be further addressed. Our results shed light on gene network manipulations in the future for breeding Sri Lankan traditional rice to contribute to food security under climate change.

## CONCLUSION

The short-day photoperiod of 11 hours and 45 minutes, which is close to the prevailing shortest photoperiod under tropical Sri Lankan conditions during the vegetative phase of accessions 4237, 5530 and 6412 is the optimum photoperiod for significant early flowering.

Early flowering under the photoperiod of 11 hours and 45 minutes is correlated with the highest number of spikelets per first panicle ( $214.6 \pm 0.9$ ,  $145 \pm 3$  at Ibbagamuwa) and ( $220 \pm 4$ ,  $150.6 \pm 1.4$  at Vantharumoolai) in accessions 5530 and 6412 respectively ( $p < 0.05$ ). The number of spikelets per first panicle is in a regression relationship with photoperiod exposure. According to the regression analysis, the optimum photoperiod of 11 hours and 45 minutes produces the highest SP in accessions 5530 and 6412 under tested conditions. Gene network model analysis revealed the pleiotropic genetic factors leading to higher yield and early flowering under inductive photoperiod. The gene network model suggests that specific genes, *Hd1*, *Ehd1*, and *Ghd7*, are involved in the photoperiodic response of the three rice accessions. In accession 5530, the significant increase in SP under the short-day photoperiod of 11 hours and 45 minutes aligns with the role of *Hd1* in promoting flowering and yield traits under these conditions. The varied responses of the three accessions to photoperiods, including differences in days to flowering and number of spikelets per first panicle, are strongly influenced by the functional presence or absence of *Ghd7* and *DTH8* under short-day and long-day conditions.

## Acknowledgements

The authors acknowledge the Plant Genetic Resources Centre (PGRC), Gannoruwa, Sri Lanka for provision of the traditional rice seeds.

## REFERENCES

- Chandraratna, M. F. (1955). Genetics of photoperiod sensitivity in rice. *Journal of Genetics*, 53, 215–223. <https://doi.org/10.1007/BF02993976>
- Deb, D. (2024). Is the time of anthesis in rice (*Oryza sativa*) influenced by photoperiod? *Biologia Futura*, 75, 453–458. <https://doi.org/10.1007/s42977-024-00223-5>
- Du, A., Tian, W., Wei, M., Yan, W., He, H., Zhou, D., Huang, X., Li, S., & Ouyang, X. (2017). The DTH8-Hd1 module mediates day-length-dependent regulation of rice flowering. *Molecular Plant*, 10, 948–96. <https://doi.org/10.1016/j.molp.2017.05.006>
- Endo-Higashi, N. & Izawa, T. (2011) Flowering time genes heading date 1 and early heading date 1 together control panicle development in rice. *Plant and Cell Physiology*, 52(6), 1083–1094. <https://doi.org/10.1093/pcp/pcr059>.
- Geekiyange, S., Madurangi, S. A. P., & Rathnathunga, E.U.U. (2012). Effect of photoperiod on flowering time and attributed traits of selected Sri Lankan rice varieties. *Proceedings of the 10th Academic Sessions*, University of

- Ruhuna. 70.
- Han, Z., Zhang, B., Zhao, H., Ayaad, M., & Xing, Y. (2016). Genome-Wide Association Studies Reveal that Diverse Heading Date Genes Respond to Short and Long Day Lengths between Indica and Japonica Rice. *Frontiers in Plant Science*, 7,1270. <https://doi.org/10.3389/fpls.2016.01270>
- Hayama, R., Yokoi S., Tamaki S., Yano M., & Shimamoto K. (2003). Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature*, 422(6933), 719-72. <https://doi.org/10.1038/nature01549>
- He, D., Zhou, R., Huang, C., Li, Y., Peng, Z., Li, D., Duan, W., Huang, N., Cao, L., Cheng, S., Sun, L., Zhan, X., & Wang, S. (2024). Improvement of flowering stage in Japonica rice variety Jiahe212 by using CRISPR/Cas9 system. *Plants*, 13(15), 2166. <https://doi.org/10.3390/plants13152166>
- Hori, K., Ogiso-Tanaka E., Matsubara K., Yamanouchi U., Ebana K., & Yano M. (2013). Hd16, a gene for casein kinase I, is involved in the control of rice flowering time by modulating the day-length response. *The Plant Journal*, 76(1), 36–46. <https://doi.org/10.1111/tpj.12268>
- Hori, K., Matsubara, K., & Yano, M. (2016). Genetic control of flowering time in rice: integration of Mendelian genetics and genomics. *Theoretical and Applied Genetics*, 129(12), 2241–2252. <https://doi.org/10.1007/s00122-016-2773-4>.
- Ishikawa, R., Tamaki, S., Yokoi, S., Inagaki, N., Shinomura, T., Takano, M., & Shimamoto, K., 2005. Suppression of the floral activator Hd3a is the principal cause of the night break effect in rice. *The Plant Cell*, 17, 3326–3336. <https://doi.org/10.1105/tpc.105.037028>
- Itoh, H., Nonoue, Y., Yano, M., & Izawa, T. (2010). A pair of floral regulators sets critical day length for Hd3a florigen expression in rice. *Nature Genetics*, 42(7), 635-638. <https://doi.org/10.1038/ng.606>
- Jayasinghe, H.M.A.S.B., Senanayake, D.M.J.B., Rathnathunga, E.U.U., & Geekiyanage, S. (2023). Correlation of plant height and crop age with yield potential in Sri Lankan traditional rice accessions of *Ma wee*: a comparative study in Kamburupitiya with selected new improved rice varieties. *Tropical Agricultural Research and Extension*, 26(4), 264-272. <https://doi.org/10.4038/tare.v26i4.5666>.
- Kojima, S., Takahashi, Y., Kobayashi, Y., Monna, L., Sasaki, T., Arak, T., & Yano, M. (2002). Hd3a, a rice ortholog of the Arabidopsis FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. *Plant & Cell Physiology*, 43(10), 1096–1105. <https://doi.org/10.1093/pcp/pcf156>
- Komiya, R., Ikegami, A., Tamaki, S., Yokoi, S., & Shimamoto, K. (2008). Hd3a and RFT1 are essential for flowering in rice. *Development*, 135(4), 767–774. <https://doi.org/10.1242/dev.008631>
- Kwon, C. T., Yoo, S. C., Koo, B. H., Cho, S. H., Park, J. W., Zhang, Z., Li, J., Li, Z., & Paek, N. C. (2014). Natural variation in Early flowering1 contributes to early flowering in japonica rice under long days. *Plant, Cell & Environment*, 37(1), 101–112. <https://doi.org/10.1111/pce.12134>
- Li, X., Sun, Y., Tian, X., Ren, Y., Tang, J., Wang, Z., Cheng, Y., & Bu, Q. (2018). Comprehensive identification of major flowering time genes and their combinations, which determined rice distribution in Northeast China. *Plant Growth Regulation*, 84(3), 593–602. <https://doi.org/10.1007/s10725-017-0364-2>
- Nemoto, Y., Nonoue, Y., Yano, M., & Izawa, T. (2016). Hd1a CONSTANS ortholog in rice, functions as an Ehd1 repressor through interaction with monocot-specific CCT-domain protein Ghd7. *The Plant Journal*, 86(3), 221–233. <https://doi.org/10.1111/tpj.13168>
- Padukkage, D., Rathnathunga, E.U.U., Pushpakumari, W.H.D.U., Geekiyanage, S., Dissanayake, N., Senanayake, S. G. J. N., & Benthota, A. S. S. (2015). *Handbook on Varietal Diversity of Sri Lankan Traditional rice (Oryza sativa) Volume II*. National Research Council of Sri Lanka. ISBN: 978-955-0263-04-2
- Padukkage, D., Senanayake, G. & Geekiyanage, S. (2017). Photoperiod sensitivity of very early maturing Sri Lankan rice for flowering time and plant architecture. *Open Agriculture*, 2(1), 580–588. <https://doi.org/10.1515/opag-2017-0061>.
- Pushpakumari, W.H.D.U., Padukkage, D., Rathnathunga, E.U.U., Geekiyanage, S., Dissanayake, N., Senanayake, S.G.J.N., & Benthota, A.S.S. (2016). *Handbook on Varietal Diversity of Sri Lankan Traditional rice (Oryza sativa) Volume III*. National Research Council of Sri Lanka. ISBN: 978-955-0263-059.
- Pushpakumari, W.H.D.U., & Geekiyanage, S. (2014). Yield potential of Medicinally Important Sri Lankan Traditional Rice. *Sri Lanka Journal of Indigenous Medicine*, 4(2), 309–312.
- Pushpakumari, W.H.D.U., Senanayake, G., & Geekiyanage, S. (2017). Photoperiod driven days to flowering variation affect vegetative growth and yield in Sri Lankan traditional rice (*Oryza sativa* L.) ‘*Ma wee*’, *Pakistan Journal of Botany*, 49(3), 945–954.
- Pushpakumari, W.H.D.U., & Geekiyanage, S. (2015). Evaluation of Yield Variation in Sri Lankan Traditional Rice (*Oryza sativa* . L) *Kalu Heenati. Proceedings of KDU International Research Symposium*: 24–26. <http://ir.kdu.ac.lk/handle/345/997>.
- Rathnathunga, E.U.U, Padukkage, D, Pushpakumari, W.H.D.U, Geekiyanage, S, Dissanayake, N, Senanayake, S.G.J.N., & Benthota, A.S.S. (2014). *Handbook on Varietal Diversity of Sri Lankan Traditional rice (Oryza sativa) Volume I*. National Research Council of Sri Lanka. ISBN: 978-955-0263-01-1.
- Rathnathunga, E.U.U. & Geekiyanage, S. (2016). Within variety flowering time variation leads to yield variation in Sri Lankan traditional rice “*Sudu wee*”. *Ceylon Journal of Science*, 45(2), <https://doi.org/10.4038/cjs.v45i2.7386>
- Rathnathunga, E.U.U., Senanayake, G., Dissanayake, N., Seneweera, S., & Geekiyanage, S. (2016a). Development of a mini core collection from Sri Lankan traditional rice for flowering time variation. *Australian Journal of Crop Science*, 10, 1357-1367. <https://doi.org/10.21475/ajcs.2016.10.09.p7865>.
- Rathnathunga, E.U.U., Senanayake, S.G.J.N., Dissanayake,



- N., Seneweera, S., & Geekiyanage, S. (2016b). Vegetative Growth and Yield Associated Flowering Time Variation in Sri Lankan Rice “Hondarawala”. *Journal of Agricultural Sciences*, 11(1), :42-52. <https://doi.org/10.4038/jas.v11i1.8079>
- Rathnathunga, E.U.U., & Geekiyanage, S. (2017a). Responses of Sri Lankan traditional rice to photoperiod at early vegetative stage. *Pakistan Journal of Botany*, 49(1), 63–66. <https://doi.org/10.4038/jnsfsr.v47i1.8928>
- Rathnathunga, E.U.U. & Geekiyanage, S. (2017b). Morphological diversity of Sri Lankan traditional rice varieties “Pachchaperumal” and “Sudurumba”. *Open Agriculture*, 2(1), 552–560. <https://doi.org/10.1515/opag-2017-0058>.
- Rathnathunga, E.U.U., Senanayake, G., Seneweera, S., & Geekiyanage S. (2019). Varietal diversity of Sri Lankan traditional rice based on sensitivity to temperature and photoperiod at vegetative stage. *Journal of the National Science Foundation of Sri Lanka*, 47(1), 51–68. <https://doi.org/10.4038/jnsfsr.v47i1.8928>
- Sun, C., He, C., Zhong, C., Liu, S., Liu, H., Luo, X., Li, J., Zhang, Y., Guo, Y., Yang, B., Wang, P., & Deng, X. (2022). Bifunctional regulators of photoperiodic flowering in short day plant rice. *Frontiers in Plant Science*, 13, 1044790. <https://doi.org/10.3389/fpls.2022.1044790>
- United Nations (2019) *World Population Prospects 2019: Data booklet*. Department of Economic and Social Affairs Population.
- Vergara, B.S., & Chang, T. (1985). *The flowering response of the rice plant to photoperiod - A Review of the Literature*. Los Banos, Laguna, Philippines: The International Rice Research Institute.
- Vicentini, G., Biancucci, M., Mineri, L., Chirivì, D., Giaume, F., Miao, Y., Kyojuka, J., Brambilla, V., Betti, C., & Fornara, F. (2023). Environmental control of rice flowering time. *Plant Communications*, 4(5), 100610. <https://doi.org/10.1016/j.xplc.2023.100610>
- Wang, X., Zhou, P., Huang, R., Zhang, J., & Ouyang, X. (2021). A daylength recognition model of photoperiodic flowering. *Frontiers in Plant Science*, 12, 778515 <https://doi.org/10.3389/fpls.2021.778515>
- Wei, X., Xu, J., Guo, H., Jiang, L., Chen, S., Yu, C., Zhou, Z., Hu, P., Zhai, H., & Wan, J. (2010). DTH8 suppresses flowering in rice, influencing plant height and yield potential simultaneously. *Plant Physiology*, 153(4), 1747–1758. <https://doi.org/10.1104/pp.110.156943>
- Wei, F.-J., Tsai, Y.C., Wu, H.-P., Huang, L.-T., Chen, Y.-C., Chen, Y.-F., Wu, C.-C., Tseng, Y.-T., & Hsing, Y.C., (2015). Both *Hd1* and *Ehd1* are important for artificial selection of flowering time in cultivated rice. *Plant Science*, 238, 1-10. <https://doi.org/10.1016/j.plantsci.2015.09.005>
- Wu, C., You, C., Li, C., Long, T., Chen, G., Byrne, M.E., & Zhang, Q. (2008). RID1, encoding a Cys2/His2-type zinc finger transcription factor, acts as a master switch from vegetative to floral development in rice. *Proceedings of the National Academy of Sciences*, 105(35), 12915-12920. <https://doi.org/10.1073/pnas.0806019105>
- Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L., Zhou, H., Yu, S., Xu, C., Li, X., & Zhang, Q., (2008). Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nature Genetics*, 40, 761–767. <https://doi.org/10.1038/ng.143>
- Yan, W.H., Wang, P., Chen, H.X., Zhou, H.J., Li, Q.P., Wang, C.R., Ding, Z.H., Zhang, Y.S., Yu, S.B., Xing, Y.Z., & Zhang, Q.F. (2011). A major QTL, *Ghd8*, plays pleiotropic roles in regulating grain productivity, plant height, and heading date in rice. *Molecular Plant*, 4(2), 319–330. <https://doi.org/10.1093/mp/ssp070>
- Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., Baba, T., Yamamoto, K., Umehara, Y., Nagamura, Y., et al., (2000). *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*. *The Plant Cell*, 12(12), 2473–2484. <https://doi.org/10.1105/tpc.12.12.2473>.
- Zhang, Z.H., Wang, K., Guo, L., Zhu, Y.J., Fan, Y.Y., Cheng, S.H., & Zhuang, J.Y. (2012). Pleiotropism of the Photoperiod-Insensitive Allele of *Hd1* on Heading Date, Plant Height and Yield Traits in Rice. *PloS One*, 7(12), 52538. <https://doi.org/10.1371/journal.pone.0052538>
- Zhang, Z.H., Zhu, Y.J., Wang, S.L., Fan, Y.Y., & Zhuang, J. (2019). Importance of the interaction between heading date genes *Hd1* and *Ghd7* for controlling yield traits in rice. *International Journal of Molecular Sciences*, 20(3), 516. <https://doi.org/10.3390/ijms20030516>
- Zhao, M., Lin, Y., & Cheng, H., (2020). Improving nutritional quality of rice for human health. *Theoretical and Applied Genetics*, 133(5), 1397–1413. <https://doi.org/10.1007/s00122-019-03530>.
- Zhu, Y.J., Fan, Y.Y., Wang, K., Huang, D.R., Liu, W.Z., Ying, J.Z., & Zhuang, J.Y. (2017). Rice Flowering Locus T1 plays an important role in heading date influencing yield traits in rice. *Scientific Reports*, 7(1), 4918. <https://doi.org/10.1038/s41598-017-05302-3>





## RESEARCH ARTICLE

### Microbiology

# Recreational water quality in the river Mahaweli, Sri Lanka: Enumeration and antibiotic sensitivity of *Escherichia coli* and sociological aspects

S Weerakoon<sup>1</sup>, B Nanayakkara<sup>1\*</sup>, C Abayasekara<sup>1</sup>, H Abeysundara<sup>2</sup> and I Liyanage<sup>1</sup>

<sup>1</sup> Department of Botany, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka.

<sup>2</sup> Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka.

Submitted: 23 November 2023; Revised: 03 September 2024; Accepted: 16 January 2025

**Abstract:** Although recreational water quality affects health, in Sri Lanka, recreational water quality is hardly assessed. Faecal contamination of water poses health risks as faeces are a source of pathogens. The current study aimed to investigate *Escherichia coli* in recreational water by enumerating and assessing antibiotic sensitivity, and to conduct a sociological survey. Water samples ( $n = 72$ ) were collected from six bathing locations along the river Mahaweli during rainy and dry periods, and subjected to membrane filtration. Typical blue colonies on m-FC agar were enumerated. An antibiotic sensitivity test was performed on *E. coli* isolates using seven antibiotics commonly used in Sri Lanka. *Escherichia coli* counts in 95.0% of water samples exceeded the permissible limit (235 CFU/100 mL). Mean *E. coli* counts were significantly higher ( $p < 0.05$ ) during the rainy period compared to the dry period. In both periods, mean counts were significantly higher ( $p < 0.05$ ) in more populated locations than less populated locations. Antibiotic-resistant *E. coli* were present in varying degrees in all locations. Isolates were resistant to ciprofloxacin (56.5%), cefotaxime (4.8%), amoxicillin-clavulanate (4.2%), amikacin (3.0%), and ticarcillin-clavulanate (0.6%). None were resistant to imipenem or meropenem. The percentages of isolates resistant to amikacin, amoxicillin-clavulanate, and ciprofloxacin were significantly higher in less populated locations compared to more populated locations ( $p < 0.05$ ), while the percentage for cefotaxime was higher in more populated locations. According to the sociological survey ( $n = 60$ ), inadequate sewage facilities and urban runoff are likely reasons of pollution. Elevated *E. coli* counts and the presence of antibiotic-resistant *E. coli* render the water unsuitable for recreational purposes.

**Keywords:** Antibiotic resistance, bathing water, faecal contamination, faecal indicator organisms, river water.

## INTRODUCTION

In terms of water quality, recreational water includes water of rivers, lakes, reservoirs etc., that is used for bathing, swimming, fishing, boating, whitewater rafting, and surfing (US EPA, 2022). The use of recreational water has significant benefits for health and well-being, including relaxation, rest, cultural and religious practices, exercise, and aesthetic pleasure (Crouse *et al.*, 2018; White *et al.*, 2020). However, pathogen-contaminated recreational water environments can pose a significant risk to the health of water users. Routes of exposure to infectious agents in water depend on the degree of water contact. The probability of ingestion of water is greater for whole-body contact activities including swimming, surfing, diving etc. Microbial pathogens are the most common public health hazard in recreational water (WHO, 2021). Faecal contamination of water may introduce pathogenic microorganisms that are mainly derived from human sewage or excreta from other warm-blooded animals (WHO, 2021).

The sources of faecal contamination include point sources and non-point or diffuse sources. Point sources

\* Corresponding author (buddhie@sci.pdn.ac.lk;  <https://orcid.org/0000-0002-7110-4216>)



This article is published under the Creative Commons CC-BY-ND License (<http://creativecommons.org/licenses/by-nd/4.0/>). This license permits use, distribution and reproduction, commercial and non-commercial, provided that the original work is properly cited and is not changed in anyway.

are contaminants that enter a waterway from a single, identifiable source. Non-point sources refer to diffuse contamination that does not originate from a single discrete source (Soni, 2019). Point sources include effluents from wastewater treatment plants, discharges of raw wastewater, discharges of industrial effluents, direct contamination from recreational water users, discharges of urban stormwater, and combined sewer overflow which contains a mixture of urban runoff water and wastewater (Brauwere *et al.*, 2014; WHO, 2021). Urban runoff is a key diffuse source that contaminates river water (Müller *et al.*, 2020).

Recreational water use can lead to a number of health problems, including gastrointestinal illness, and also non-enteric diseases caused by respiratory, ear, eye, and skin infections (Fleisher *et al.*, 1993; Kay *et al.*, 1994; Fewtrell & Kay, 2015). The waterborne disease outbreaks associated with enteric illness are attributed to microbial agents including bacteria (*Escherichia coli*, *Shigella* spp., *Vibrio cholerae*, *Salmonella* spp., *Yersinia enterocolitica* and *Campylobacter jejuni*), viruses (Enterovirus, Rotavirus and Adenovirus), and protozoa (*Giardia lamblia*, *Cryptosporidium parvum*, and *Entamoeba histolytica*) (Fewtrell & Kay, 2015; Neil *et al.*, 2018; Lugo *et al.*, 2021). Worldwide, waterborne gastrointestinal illnesses account for an estimated 16.4 million cases annually. The number of enteric disease outbreaks associated with recreational water that is caused by pathogens transmitted by ingesting, or by having contact with contaminated water, has considerably increased (Neil *et al.*, 2018). Therefore, it is necessary to assess the microbiological quality of recreational water to ensure that no health risk due to pathogenic microorganisms is associated with its use. Since it is impossible to measure the abundance of all possible pathogenic microorganisms, the general practice is to quantify the abundance of only one or a few faecal indicator bacteria (FIB), *i.e.*, organisms that are selected to be indicative of faecal pollution and can therefore serve as general indicators of microbiological water quality (Brauwere *et al.*, 2014). Total coliforms, faecal coliforms, Enterococci, and *E. coli* are used as faecal indicators (Ishii & Sadowsky, 2008; Odonkor & Ampofo, 2013).

An ideal indicator bacterium should be present in the intestinal tracts of warm-blooded animals. Besides, indicators should be present in faeces, sewage, and sewage-polluted waters whenever the pathogens are also present, and absent in waters without faecal pollution. The levels of indicators should have some direct relationship to the density of pathogens in the environment and be proportional to the extent of the pollution. Furthermore,

an indicator organism should be non-pathogenic and present in greater numbers than pathogens. The survival rate in water must be at least similar to that of pathogens and it should be unable to grow and multiply in the environment. Further, an indicator bacterium should be detected and quantified by rapid, easy, and inexpensive methods (Borrego & Figueras, 1997; Havelaar *et al.*, 2001; Ishii & Sadowsky, 2008; Lugo *et al.*, 2021). *Escherichia coli* meets most of these criteria and is widely used as an indicator organism for monitoring faecal contamination of recreational freshwater ecosystems (Odonkor & Ampofo, 2013; Rossi *et al.*, 2020). For recreational activities, *E. coli* levels should not exceed 235 CFU/100 mL of water for full-body contact (US EPA, 2012).

Antibiotics are compounds that are intended to treat bacterial infections (Calhoun *et al.*, 2022). Antibiotics can be classified in several ways based on their mode of action, molecular structures and spectrum of activity, *i.e.*,  $\beta$ -lactams, quinolones, sulphonamides, macrolides, and tetracyclines (Kümmerer, 2009; Etebu & Ariekpar, 2016).

Antibiotic resistance is the ability of bacteria to resist the effects of an antibiotic to which they were previously sensitive. Infections with bacteria that express antibiotic resistance genes can be difficult or even impossible to treat, posing a serious threat to public health (Amarasiri *et al.*, 2020). High mortality rates are associated with multidrug-resistant bacterial infections (Wu *et al.*, 2021). After using antibiotics, a large proportion of the antibiotics that are absorbed are excreted into the environment by faeces or urine in non-metabolized forms or coupled with polar molecules. Therefore, antibiotics and their degradation products are considered emergent contaminants of water (Oberlé *et al.*, 2012), and may lead to the development of resistance in target and non-target organisms in the environment (Kümmerer, 2009). The leftover antibiotics are disposed of and become part of urban waste. These also can end up in river water through urban runoff or leachate from solid waste dump sites (Wimalasuriya *et al.*, 2011).

In veterinary clinical practice, antibiotics are widely used to treat a variety of bacterial infections. However, the use and improper use of antibiotics in animal husbandry, and agriculture has led to the emergence of several drug-resistant bacterial communities (Economou & Gousia, 2015; Ranasinghe *et al.*, 2022). The runoff from livestock fields into the river may lead to the spread of antibiotic-resistant bacteria in the river water.

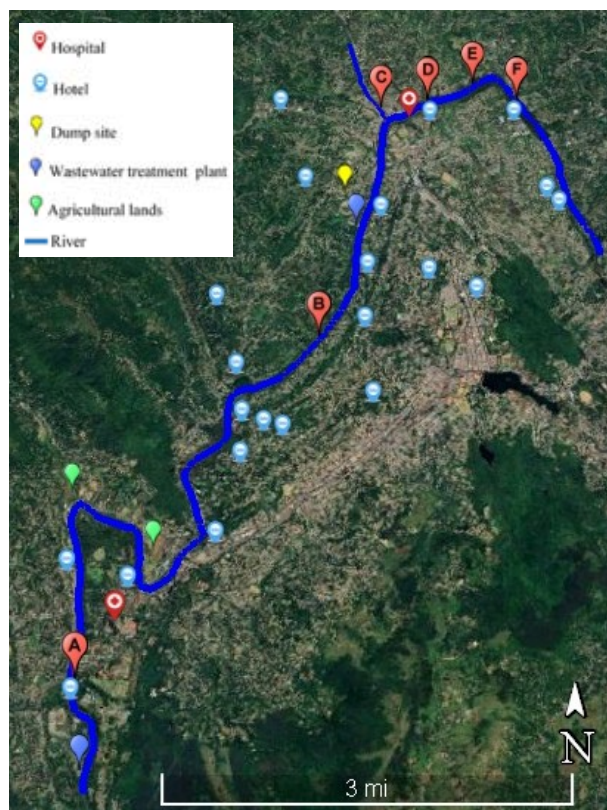
Regardless of the multi-faceted health hazards, unlike developed countries, in Sri Lanka, people are generally

ignorant about faecal contamination of recreational water. The current study aimed to enumerate *E. coli* in water at selected bathing locations of the river Mahaweli and assess the sensitivity of isolates to selected antibiotics that are of common clinical use in Sri Lanka. As a part of this study, a sociological survey was conducted to gather information on water quality and implications of contamination at the selected bathing locations.

## MATERIALS AND METHODS

### Study area and sample collection

Water samples were collected from six bathing locations in the river Mahaweli from Peradeniya to Katugastota, Sri Lanka, during rainy and dry periods (Figure 1). Each sample was 200 mL and was collected from the surface of the river, at each bathing location. Six replicate water samples were collected per location per period. In total, 72 water samples were collected.



**Figure 1:** Selected bathing locations along the river Mahaweli between Peradeniya and Katugastota, and possible sources of water contamination. A - Devil's Bridge, B - Dodamwala, C - Ranawana Road, D - Madawala Road 1, E - Madawala Road 2, F - Madawala Road 3.

The six bathing locations were classified according to proximity to a town area, as defined by a population density of  $\geq 2000$ , which was considered as the level of urbanization. Accordingly, locations A and B were in less-populated areas whereas locations C, D, E and F were in more-populated areas.

### Isolation and enumeration of *Escherichia coli*

Water samples were analyzed using the membrane filtration apparatus (Rocker, Taiwan). The samples were diluted where necessary, to ensure the colonies were within the countable range. A 100 mL aliquot of each water sample was separately filtered through a sterile, 0.45  $\mu\text{m}$  pore-sized, gridded membrane filter paper (Aisimo Corporation) using a vacuum filtration system. After each filtration, the membrane filter paper was placed on m-FC agar (HIMEDIA) and incubated for 24 - 48 h at 44.5 °C. An elevated temperature of 44.5 °C was used to inhibit the growth of other fecal coliforms and promote the growth specifically of *E. coli*. After incubation, the typical blue colonies grown on the plate were counted and reported as colony forming units (CFU). The colonies that had the typical blue colour indicative of *E. coli* were taken as *E. coli* and were confirmed by using eosin methylene blue (EMB) agar (HIMEDIA) with which *E. coli* produces dark colonies with a metallic green sheen, MacConkey agar (HIMEDIA) with which *E. coli*, being lactose-positive, produces pink colonies, and Simmons citrate agar (HIMEDIA) where the colour of the medium remains green as *E. coli* is citrate-negative.

*Escherichia coli* colonies had the characteristic metallic green sheen on EMB agar and colonies on MacConkey agar appeared pink. When inoculated with *E. coli*, the green color of Simmons citrate agar remained green, producing no color change. The *E. coli* colonies were used to prepare stock cultures, which were preserved at -80 °C, to be used for antibiotic sensitivity testing.

### Antibiotic sensitivity test

#### Quality control test

Quality control tests were performed according to the guidelines specified by the Clinical and Laboratory Standards Institute (CLSI, 2020). Disks impregnated with seven antibiotics which are clinically used in Sri Lanka were employed; amikacin (30  $\mu\text{g}$ ), amoxicillin-clavulanate (20/10  $\mu\text{g}$ ), cefotaxime (30  $\mu\text{g}$ ), ciprofloxacin (5  $\mu\text{g}$ ), imipenem (10  $\mu\text{g}$ ), meropenem (10  $\mu\text{g}$ ), and ticarcillin-clavulanate (75/10  $\mu\text{g}$ ). A five-day quality control test was performed continuously, for two reference

strains of *E. coli* (ATCC 35218 and ATCC 25922), to ensure that the disks function well. *Escherichia coli* ATCC 35218 was used for quality control of amoxicillin-clavulanate (20/10 µg) and ticarcillin-clavulanate (85 µg). *Escherichia coli* ATCC 25922 was used for quality control of amikacin (30 µg), cefotaxime (30 µg), ciprofloxacin (5 µg), imipenem (10 µg), and meropenem (10 µg). The diameters of zones of inhibition obtained were checked against the standard quality control ranges (CLSI, 2020).

#### **Antibiotic sensitivity test for *Escherichia coli* isolates obtained from water**

Isolated, pure colonies of *E. coli* obtained by sub-culturing from the previously prepared freezer cultures were transferred separately to sterile normal saline (0.85% NaCl). The inoculum density was standardized by making the turbidity of the suspensions similar to that of 0.5 McFarland standard. The inocula thus prepared were used to inoculate Mueller Hinton agar (MHA) (HIMEDIA) plates.

Antibiotic disks were equidistantly placed on MHA plates. Two replicates were prepared for each *E. coli* isolate. The plates were incubated at 35 °C for 18 h and the diameter of the zone of inhibition was measured for each antibiotic. Two perpendicular measurements were taken for each zone of inhibition, and the values were averaged. Among the 72 water samples (n = 36 in dry period, n = 36 for rainy period), one *E. coli* isolate per sample was subjected to antibiotic testing. Accordingly, 72 *E. coli* isolates were subjected to antibiotic sensitivity testing, in duplicate and antibiotic susceptibility / resistance was ascertained in comparison with the zone diameter breakpoints given under the CLSI guidelines (CLSI, 2020).

#### **Sociological survey**

A questionnaire comprising multiple choice questions and closed questions (yes/no) was designed to gather information on the water quality of the six selected bathing locations. The questionnaire was prepared in Sinhala, Tamil, and English, to cater to the proficiency of the respondents. A Google form-based pre-test was done. The questionnaire was distributed among the target group (n = 60 in total; n = 10 per location), among the general public near the selected bathing locations.

#### **Statistical analysis**

Minitab version 21.2 was used for analyzing the data. The *E. coli* counts obtained for the bathing water samples collected in the dry and rainy periods were analyzed using parametric tests. A previous set of data collected during the year 2020-2021 was analyzed to explore any changes in the trends of *E. coli* counts with time. Differences at  $p \leq 0.05$  (95% significance level) were considered statistically significant, whereas differences at  $p > 0.05$  were considered statistically not significant. The one-way analysis of variance (one-way ANOVA) test was used in determining differences in mean counts between selected locations. The two-way analysis of variance (two-way ANOVA) test was used to study the interaction effect in mean counts due to the location and the period. The Mann-Whitney test was used to determine the differences in median counts between the dry and rainy periods in selected locations, while the two sample t-test was used in determining the differences in mean counts between more populated and less populated locations.

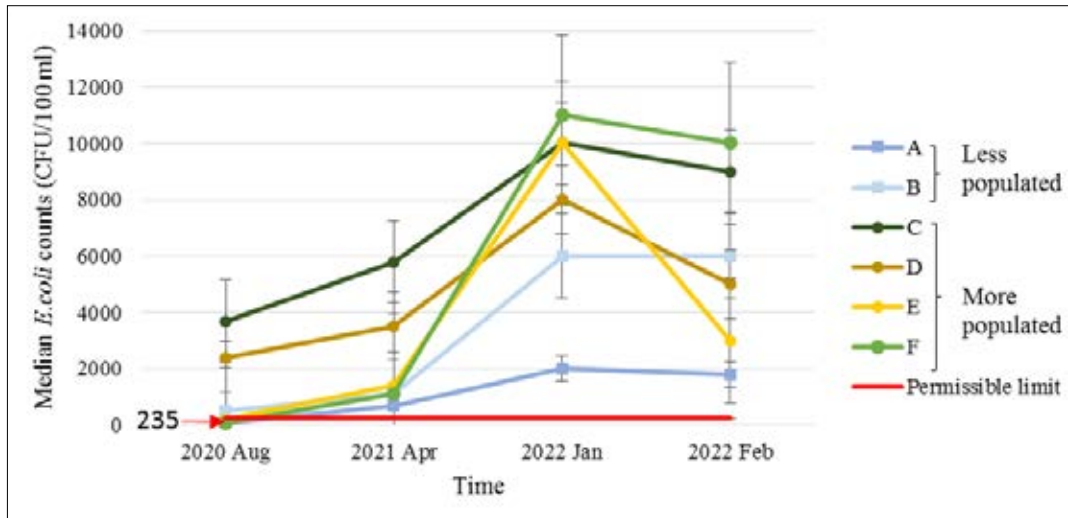
The two-proportion test was used in determining the differences in the percentage of resistant *E. coli* isolates for each antibiotic between the rainy and dry periods and also to determine the differences in the percentage of resistant *E. coli* isolates for each antibiotic between the more populated and less populated locations.

## **RESULTS AND DISCUSSION**

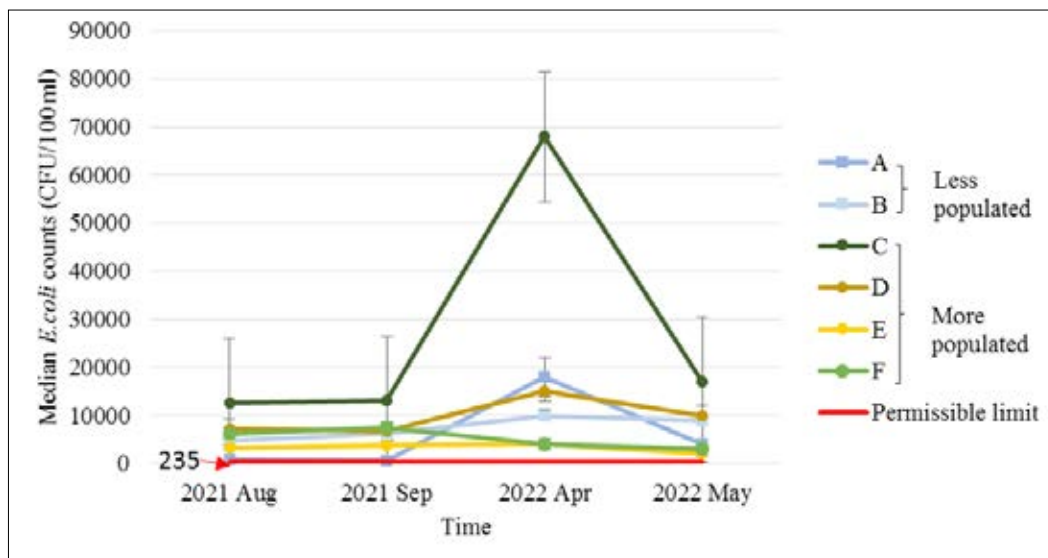
#### **Enumeration of *Escherichia coli***

The *E. coli* counts in 95.0% of the bathing water samples exceeded the permissible limit of 235 CFU/100 mL (US EPA, 2012) accepted for recreational water. Counts of *E. coli* were higher in the rainy period compared to those in the dry period in locations A, B, C, D, and E, while counts were higher during the dry period in location F. An overall increase in *E. coli* counts was observed in both the dry and rainy periods of the year 2022, compared to 2020-2021, in selected locations (Figures 2 and 3). It could be due to increased effluents from wastewater treatment plants, discharges of industrial effluents and urban runoff compared to the past. Due to the COVID-19 pandemic in Sri Lanka, people were under a lockdown at home during the period from 2020 – 2021, reducing the amount of urban waste production and resultant contamination of water, compared to 2022, when community activities were restored.





**Figure 2:** *Escherichia coli* counts in the dry period from the year 2020 - 2022.

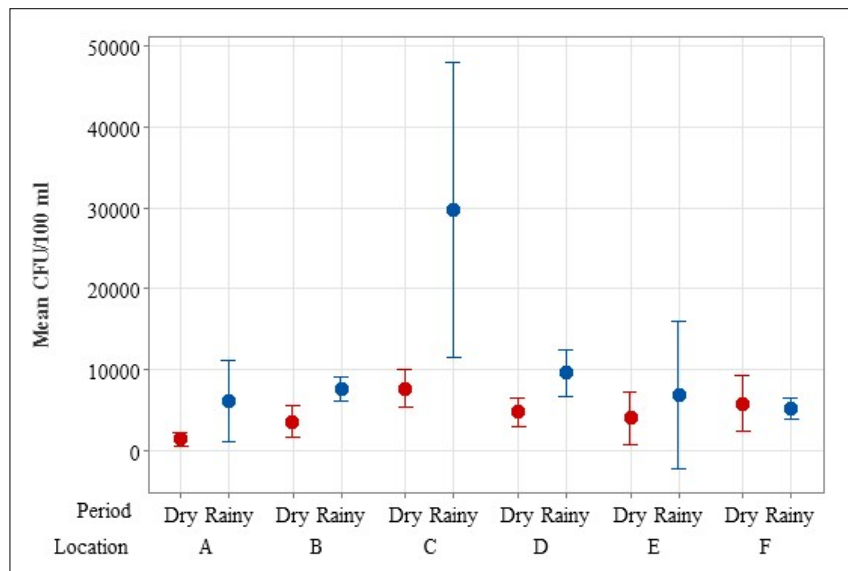


**Figure 3:** *Escherichia coli* counts in the rainy period from the year 2021 - 2022.

Two of the more populated locations (C and D) had higher counts during the rainy period. It can be speculated that urban runoff may be high during rain. The lower counts in locations E and F in the rainy period may be due to water being diluted by increased volume and flow rate of water during rain. Water tends to stagnate during the dry period and this might be accountable for the higher counts observed in the dry period for location F. Location B is of particular interest as it is located close to a municipal

solid waste dump site. The *E. coli* counts observed in this location could be attributed to the leachate from the dump site.

According to the mean data (Figure 4), location C showed significantly higher values ( $p = 0.001$ ) for the rainy period compared to all other locations, while having the highest value for the dry period as well.



**Figure 4:** *Escherichia coli* counts in dry and rainy periods in different locations.

Location C lies in close proximity to a hospital. According to the survey, hospital waste is a possible source of river water contamination in this location. Further, it was revealed that toilet pits are directly drained into the river, at or near location C. These may be the reasons for elevated *E. coli* counts in location C. Mean *E. coli* counts in both more populated and less populated locations showed significantly higher values (more populated:  $p = 0.012$ ; less populated:  $p = 0.002$ ) for the rainy period compared to the dry period. In turn, when locations are compared, in both the rainy and dry periods, mean counts were significantly higher in more populated locations than less populated locations (rainy period:  $p = 0.047$ ; dry period:  $p < 0.001$ ). All in all, there was a significant difference in mean counts due to location ( $p = 0.001$ ) and there was a significant difference in mean counts due to period ( $p = 0.001$ ). There was a significant interaction effect between the location and sampling period ( $p = 0.002$ ) (Figure 4).

Both point and non-point sources are greatly variable in space (*i.e.*, depending on the area considered) and in time (depending on hydrological and meteorological conditions, or accidental events). In rural areas, the sources of human origin are often negligible compared to animal sources (cattle, birds, and wildlife). Some animals may defecate directly in the water (birds or land animals at drinking points). In densely populated areas,

the faecal contaminants enter the natural water through both point and non-point sources such as discharges of raw wastewater and sewage, discharges of industrial effluents, direct contamination from recreational water users and urban runoff (Servais *et al.*, 2007; Ouattara *et al.*, 2011; Brauwere *et al.*, 2014).

According to the results, the microbiological quality of river water was subject to seasonal variations. During the rainy period, pollutants can enter river water through runoff. The rain washes the particles which have been deposited on nearby surfaces into the water source through leaching processes. This may depend on the type of soil and physical forms of pollution. In dry periods the water table becomes very low and the evaporation rate increases. Potgieter *et al.* (2006) report that this affects the oxygen content which decreases the multiplication of the bacteria in water, and that low temperatures could reduce the amount of available oxygen and disturb the bacterial metabolic processes. Sunlight can be an important factor responsible for the death or loss of culturability of microorganisms of faecal origin in aquatic systems (Brauwere *et al.*, 2014). As demonstrated by several studies, temperature has a significant effect on the decay rates of faecal coliforms in fresh, saline, and estuarine waters (Barcina *et al.*, 1986; Flint, 1987; Evison, 1988; Craig *et al.*, 2004). Thus, the level of contamination of water will depend on the seasonal variation.



## Antibiotic sensitivity testing

### Quality control test

After a five-day quality control test, the inhibition zone diameters produced by all seven antibiotics tested [amoxicillin-clavulanate (20/10 µg, augmentin), ticarcillin-clavulanate (75/10 µg, timentin), amikacin (30 µg), cefotaxime (30 µg), ciprofloxacin (5 µg), imipenem (10 µg), and meropenem (10 µg)], were found to be within the quality control range (CLSI, 2020).

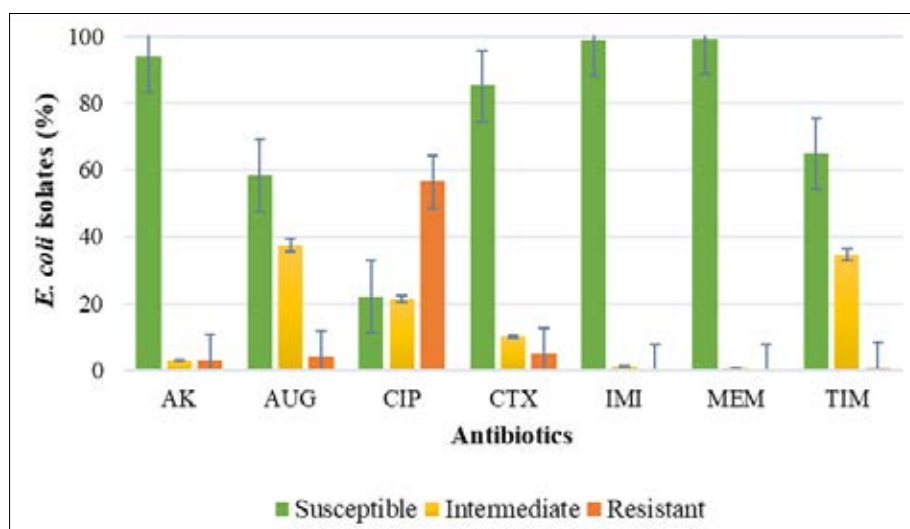
### Antibiotic sensitivity testing for *Escherichia coli* isolates

Antibiotic-resistant *E. coli* isolates were present in all six sampling locations, during both rainy and dry periods. *Escherichia coli* isolates were resistant to ciprofloxacin (56.5%), cefotaxime (4.8%), augmentin (4.2%), amikacin (3.0%), and timentin (0.6%). None of the isolates was resistant to imipenem or meropenem (0%) (Figure 5). ciprofloxacin is the most frequently prescribed empirical antimicrobial for the treatment of urinary tract infections in Sri Lanka (Wijekoon *et al.*, 2014). Further, the consumption of beta-lactam antibacterials (augmentin, timentin, cefotaxime) and quinolones (ciprofloxacin) is high in Sri Lanka (Sri Ranganathan *et al.*, 2021).

Improper use of antibiotics for humans and in animal husbandry has led to the emergence of antibiotic-resistant bacteria (Serwecińska, 2020; Ranasinghe *et al.*, 2022). This could be the reason behind the high incidence of antibiotic-resistant *E. coli* in the water samples.

While certain isolates showed intermediate sensitivity to selected antibiotics, a considerable number of isolates were susceptible to the antibiotics (Figure 5). The proportions of isolates that were susceptible to each antibiotic were as follows; meropenem 99.4%, imipenem 98.8%, amikacin 94.0%, cefotaxime 85.1%, ticarcillin-clavulanate (timentin) 64.9%, amoxicillin-clavulanate (augmentin) 58.3%, and ciprofloxacin 22.0%. Thus, it can be argued that the selected antibiotics still exhibit potency against *E. coli*, though this varies with the isolate.

Multi-drug resistance (MDR) is defined as non-susceptibility to at least one agent in three or more antimicrobial categories (Ibrahim *et al.*, 2012). In a previous study carried out by Liyanage *et al.*, (2021) at the same locations, 1.0% of the *E. coli* isolates was multi-drug resistant (year 2021). No MDR isolates were reported in the samples collected in 2022 (current study).



**Figure 5:** Percentages of *E. coli* isolates that were resistant, intermediate and susceptible to the antibiotics from the year 2020 to 2022. AK – Amikacin, AUG – Augmentin (Amoxicillin-clavulanate), CIP – Ciprofloxacin, CTX – Cefotaxime, IMI – Imipenem, MEM – Meropenem, TIM – Timentin (Ticarcillin-clavulanate). Breakpoint values were taken from CLSI (2020).

The data was analysed to infer differences between more populated versus less populated locations and between rainy and dry periods. The percentage of isolates that were resistant to amikacin ( $p = 0.007$ ), augmentin ( $p = 0.023$ ) and ciprofloxacin ( $p = 0.026$ ) were significantly higher in less populated locations compared to the more populated locations. No significant differences were observed in the percentages of isolates resistant to imipenem, meropenem, and ticarcillin-clavulanate (Timentin) between more populated and less populated locations ( $p > 0.05$ ).

Antibiotics, antibiotic residues (with functional groups), antibiotic-resistant isolates and a myriad other bacteria can be present in hospital waste (Mackul'ak *et al.*, 2021). The runoff of hospital waste into the river may lead to the spread of antibiotic-resistant bacteria in the river water. Prolonged exposure to pharmaceutically active compounds in water can lead to changes in the bacterial genomes leading to antibiotic resistance. The resistant genes can horizontally transfer to susceptible strains, contributing to the development and spread of resistance (Mackul'ak *et al.*, 2021). Also, due to urban runoff and residues from industrial processes, a significant amount of antibiotic-resistant faecal bacteria and antibiotic residues can reach the river. These could be the reasons behind the occurrence of antibiotic-resistant bacteria in more populated locations. Antibiotic-resistant bacteria were also present in less populated locations. A less populated sampling site (B) along the river is located close to the municipal solid waste dump site at Gohagoda. The daily collection of solid waste in the Kandy city is disposed of to this dump site. Due to migration of leachate, the river water can be contaminated with antibiotic-containing wastes including pharmaceuticals (Wimalasuriya *et al.*, 2011; Athukorala *et al.*, 2021). This could be a reason for the high percentage of antibiotic-resistant isolates in location B, particularly in the rainy period, when more leachate is produced. Lack of sanitary facilities, excessive use and exposure to antimicrobial agents, and detergents may be further reasons for the buildup of resistant strains as reported earlier (Huijbers *et al.*, 2015; Leonard *et al.*, 2015). Unsupervised self-medication, inappropriate prescriptions, improper use (misuse) of antibiotics, poor hygiene, and inadequacy of duration of use are some reasons behind the emergence and spread of antibiotic resistance in Sri Lanka (Abhayasinghe *et al.*, 2021; Sri Ranganathan *et al.*, 2021).

It is likely that the water in most of the selected bathing locations in the river Mahaweli is highly contaminated with faecal matter, and it indicates the possible presence of pathogens in the water. Thus, the

results of the current study emphasize the need for proper monitoring of recreational water quality in Sri Lanka. Depending on the observations made through monitoring, regulatory actions including closure of water bodies for public access, should be taken, especially if the level of contamination is high. Such regulatory action is taken in developed countries, where recreational water quality is much heeded (Power *et al.*, 2005; Janezic *et al.*, 2013). Action needs to be taken to control the disposal of waste at these locations with the support of the government, in conjunction with increased public awareness. Awareness programs would definitely serve to ensure public health and safety.

### Sociological survey

As revealed by the survey, compared to the past, the use of river water for daily living activities has decreased, with increased access to pipe-borne water at these locations. Yet, some people continue to use the water of the river Mahaweli at the selected locations for daily living activities, since they have no other sources of water. All respondents had observed a notable difference in the clarity of the water between the dry and rainy periods. The use of river water in the rainy period is lower than that in the dry period. Other than recreational activities including bathing, some people use the river water for their daily living activities including washing clothes, gardening, and animal bathing. Inadequate municipal sewage facilities, industrial effluent discharge, and urban runoff are likely reasons for the contamination of river water at the selected bathing locations. Urban runoff appears to have a significant impact on water quality, especially at locations C and D (more populated), while it does play a role in contamination of water at the other locations as well. In addition to the above reasons, hospital waste is a possible source of water contamination in location C.

### CONCLUSION

The water quality in the sampling locations undergoes seasonal variations. In 95% of the bathing water samples, the *E. coli* counts exceeded the permissible limit of 235 CFU/100 mL of water, and the counts vary in the rainy and dry periods. Antibiotic-resistant *E. coli* isolates were present in the water at all selected bathing locations. Compared to the year 2021, the *E. coli* counts observed in the year 2022 were higher, at all selected bathing locations. Compared to the year 2021, some resistant isolates were observed in 2022, for previously effective antibiotics. Elevated *E. coli* counts and the presence of antibiotic-resistant *E. coli* render the water at these

locations unsuitable for recreational purposes. The sociological survey revealed that improper management and disposal of waste lead to contamination of water at these locations. This emphasizes the need for proper monitoring and action being taken by authorities to ensure public health and safety.

## Acknowledgements

We acknowledge the financial assistance from University of Peradeniya (URG/2019/29/S).

## REFERENCES

- Abhayasinghe, R., Ekanayake, N., Priyasad, I., Amarajeewa, O., Abeyrathna, S., Althaf, K.R., Tennegedara, A.L., & Liyanapathirana, V.C. (2021). Understanding on antibiotics and antibiotic resistance among a group of internet users in Sri Lanka and development of a simple online educational tool. *Sri Lankan Journal of Infectious Diseases*, 11(2):83-93. DOI: <http://dx.doi.org/10.4038/sljid.v11i2.8412>
- Amarasiri, M., Sano, D., & Suzuki, S. (2020). Understanding human health risks caused by antibiotic resistant bacteria (ARB) and antibiotic resistance genes (ARG) in water environments: Current knowledge and questions to be answered. *Critical Reviews in Environmental Science and Technology*, 50(19), 2016-2059. DOI: 10.1080/10643389.2019.1692611
- Athukorala, P.A.T.V., Madawala, H.M.S.P., Nanayakkara, B.S., & Abeyesundara, H.T.K. (2021). Antibiotic resistance: A comparative study using polluted and unpolluted soil. Proceedings of the Annual Research Congress of the Postgraduate Institute of Science (RESCON) 2021. p. 38.
- Barcina, I., Arana, I., Iriberrí, J., & Egea, L. (1986). Factors affecting the survival of *E. coli* in a river. *Hydrobiologia*, 141(3), 249-253. <https://doi.org/10.1007/BF00014218>
- Borrego, J.J., & Figueras, M.J. (1997). Microbiological quality of natural waters. *Microbiologia-Madrid*, 13, 413-426.
- Brauwere, A., Ouattara, N.K., & Servais, P. (2014). Modeling fecal indicator bacteria concentrations in natural surface waters: a review. *Critical Reviews in Environmental Science and Technology*, 44(21), 2380-2453. DOI: 10.1080/10643389.2013.829978
- Calhoun, C., Wermuth, H. R., & Hall, G. A. (2022). Antibiotics. <https://www.ncbi.nlm.nih.gov/books/NBK535443/>
- CLSI – Clinical and Laboratory Standards Institute (2020). M100 – Performance standards for antimicrobial susceptibility testing. Clinical and Laboratory Standards Institute, Wayne, USA.
- Craig, D.L., Fallowfield, H.J., & Cromar, N.J. (2004). Use of microcosms to determine persistence of *Escherichia coli* in recreational coastal water and sediment and validation with *in situ* measurements. *Journal of Applied Microbiology*, 96(5), 922-930. DOI:10.1111/j.1365-2672.2004.02243.x
- Crouse, D.L., Balram, A., Hystad, P., Pinault, L., van den Bosch, M., Chen, H., Rainham, D., Thomson, E.M., Close, C.H., van Donkelaar, A., & Martin, R.V. (2018). Associations between living near water and risk of mortality among urban Canadians. *Environmental Health Perspectives*, 126(7), 077008. DOI: <https://doi.org/10.1289/EHP3397>
- Economou, V., & Gousia, P. (2015). Agriculture and food animals as a source of antimicrobial-resistant bacteria. *Infection and Drug Resistance*, pp.49-61. DOI: 10.2147/IDR.S55778
- Etebu, E., & Arikekpar, I. (2016). Antibiotics: Classification and mechanisms of action with emphasis on molecular perspectives. *International Journal of Applied Microbiology and Biotechnology Research*, 4(2016), 90-101.
- Evison, L.M. (1988). Comparative studies on the survival of indicator organisms and pathogens in fresh and sea water. *Water Science and Technology*, 20(11-12), 309-315. DOI: <https://doi.org/10.2166/wst.1988.0300>
- Fewtrell, L., & Kay, D. (2015). Recreational water and infection: a review of recent findings. *Current Environmental Health Reports*, 2(1), 85-94. DOI: 10.1007/s40572-014-0036-6
- Fleisher, J.M., Jones, F., Kay, D., Stanwell-Smith, R., Wyer, M.D., & Morano, R. (1993). Water and non-water-related risk factors for gastroenteritis among bathers exposed to sewage-contaminated marine waters. *International Journal of Epidemiology*, 22(4), 698-708. DOI: 10.1093/ije/22.4.698
- Flint, K.P. (1987). The long-term survival of *Escherichia coli* in river water. *Journal of Applied Bacteriology*, 63(3), 261-270. DOI: <https://doi.org/10.1111/j.1365-2672.1987.tb04945.x>
- Havelaar, A., Blumenthal, U.J., Strauss, M., Kay, D., & Bartram, J. (2001). Guidelines: the current position. Water Quality: Guidelines, Standards and Health (Fewtrell L., Bartram J., eds). WHO Water Series. London: IWA Publishing, 17-42.
- Huijbers, P.M., Blaak, H., de Jong, M.C., Graat, E.A., Vandenbroucke-Grauls, C.M., & de Roda Husman, A.M. (2015). Role of the environment in the transmission of antimicrobial resistance to humans: a review. *Environmental Science and Technology*, 49(20), 11993-12004. DOI: 10.1021/acs.est.5b02566
- Ibrahim, M.E., Bilal, N.E., & Hamid, M.E. (2012). Increased multi-drug resistant *Escherichia coli* from hospitals in Khartoum state, Sudan. *African Health Sciences*, 12(3), 368-375. DOI: 10.4314/ahs.v12i3.19
- Ishii, S., & Sadowsky, M.J. (2008). *Escherichia coli* in the environment: implications for water quality and human health. *Microbes and Environments*, 23(2), 101-108. DOI:10.1264/jsme2.23.101
- Janežic, K. J., Ferry, B., Hendricks, E. W., Janiga, B. A., Johnson, T., Murphy, S., & Daniel, S. L. (2013). Phenotypic and genotypic characterization of *Escherichia coli* isolated from untreated surface waters. *The Open Microbiology Journal*, 7(1). DOI: 10.2174/1874285801307010009
- Kay, D., Jones, F., Wyer, M.D., Fleisher, J.M., Salmon, R.L., Godfree, A.F., Zelenauch-Jacquotte, A., & Shore, R. (1994). Predicting likelihood of gastroenteritis from sea bathing: results from randomized exposure. *The Lancet*, 344(8927), 905-909. DOI: 10.1016/S0140-6736(94)92267-5
- Kümmerer, K. (2009). Antibiotics in the aquatic environment – a review: Part I. *Chemosphere*, 75(4), 417-434. DOI:

- 10.1016/j.chemosphere.2008.11.086
- Leonard, A.F., Zhang, L., Balfour, A.J., Garside, R., & Gaze, W.H. (2015). Human recreational exposure to antibiotic resistant bacteria in coastal bathing waters. *Environment International*, 82, 92-100. DOI: 10.1016/j.envint.2015.02.013
- Liyanage, L.R.I.S., Nanayakkara, B.S., Abayasekara, C.L., Galketiyaheewage, S.U., & Abeysundara, H.T.K. (2021). *Escherichia coli* in recreational water in selected sites of the river Mahaweli between Peradeniya and Katugastota. *Sri Lankan Journal of Infectious Diseases*, 11, p.7. DOI: 10.4038/sljid.v11i0.8443
- Lugo, J. L., Lugo, E. R., & Puente, M. D. L. (2021). A systematic review of microorganisms as indicators of recreational water quality in natural and drinking water systems. *Journal of Water and Health*, 19(1), 20-28.
- Mackulák, T., Cverenkárová, K., Vojs Staňová, A., Fehér, M., Tamáš, M., Škulcová, A. B., & Bírošová, L. (2021). Hospital wastewater—source of specific micropollutants, antibiotic-resistant microorganisms, viruses, and their elimination. *Antibiotics*, 10(9), 1070. DOI: <https://doi.org/10.3390%2Fantibiotics10091070>
- Müller, A., Österlund, H., Marsalek, J., & Viklander, M. (2020). The pollution conveyed by urban runoff: A review of sources. *Science of the Total Environment*, 709, 136125. DOI: <https://doi.org/10.1016/j.scitotenv.2019.136125>
- Neil, K.P., Yoder, J.S., Hall, A.J., & Bowen, A. (2018). Enteric diseases transmitted through food, water, and zoonotic exposures. In *Principles and Practice of Pediatric Infectious Diseases*. 397-409. DOI:10.1016/B978-0-323-40181-4.00059-1
- Oberlé, K., Capdeville, M.J., Berthe, T., Budzinski, H., & Petit, F. (2012). Evidence for a complex relationship between antibiotics and antibiotic-resistant *Escherichia coli*: from medical center patients to a receiving environment. *Environmental Science and Technology*, 46(3), 1859-1868. DOI: 10.1021/es203399h
- Odonkor, S.T., & Ampofo, J.K. (2013). *Escherichia coli* as an indicator of bacteriological quality of water: an overview. *Microbiology Research*, 4(1), 2. DOI:10.4081/mr.2013.e2
- Ouattara, N.K., Passerat, J., & Servais, P. (2011). Faecal contamination of water and sediment in the rivers of the Scheldt drainage network. *Environmental Monitoring and Assessment*, 183(1), 243-257. DOI: 10.1007/s10661-011-1918-9
- Potgieter, N., Mudau, L.S., & Maluleke, F.R.S. (2006). Microbiological quality of groundwater sources used by rural communities in Limpopo Province, South Africa. *Water Science and Technology*, 54(11-12), 371-377. DOI: <https://doi.org/10.2166/wst.2006.890>
- Power, M.L., Littlefield-Wyer, J., Gordon, D.M., Veal, D.A., & Slade, M.B. (2005) Phenotypic and genotypic characterization of encapsulated *Escherichia coli* isolated from blooms in two Australian lakes. *Environmental Microbiology*, 7(5): 631-640. DOI: 10.1111/j.1462-2920.2005.00729.x
- Ranasinghe, R. A. S. S., Satharasinghe, D. A., Anwarana, P. S., Parakatawella, P. M. S. D. K., Jayasooriya, L. J. P. A. P., Ranasinghe, R. M. S. B. K., Rajapakse, R. P. V. J., Huat, J. T. Y., Rukayadi, Y., Nakaguchi, Y., & Nishibuchi, M., (2022). Prevalence and antimicrobial resistance of *Escherichia coli* in chicken meat and edible poultry organs collected from retail shops and supermarkets of North Western Province in Sri Lanka. *Journal of Food Quality*, 2022, 962698. <https://doi.org/10.1155/2022/8962698>
- Rossi, A., Wolde, B. T., Lee, L. H., & Wu, M. (2020). Prediction of recreational water safety using *Escherichia coli* as an indicator: case study of the Passaic and Pompton rivers, New Jersey. *Science of the Total Environment*, 714, 136814. DOI: 10.1016/j.scitotenv.2020.136814
- Servais, P., Billen, G., Goncalves, A., & Garcia-Armisen, T. (2007). Modelling microbiological water quality in the Seine river drainage network: past, present and future situations. *Hydrology and Earth System Sciences*, 11(5), 1581-1592. DOI: <https://doi.org/10.5194/hess-11-1581-2007>
- Serwecińska, L. (2020). Antimicrobials and antibiotic-resistant bacteria: a risk to the environment and to public health. *Water*, 12(12), p.3313. DOI: <https://doi.org/10.3390/w12123313>
- Soni, H. B. (2019). 2. Categories, causes and control of water pollution-a review. *Life Sciences Leaflets*, 107, 4.
- Sri Ranganathan, S., Wanigatunge, C., Senadheera, G.P.S.G., & Beneragama, B.V.S.H. (2021). A national survey of antibacterial consumption in Sri Lanka. *PloS One*, 16(9), 0257424. DOI: <https://doi.org/10.1371/journal.pone.0257424>
- United States Environmental Protection Agency (US EPA) (2012). Recreational Water Quality Criteria documents. <https://www.epa.gov/sites/default/files/2015-10/documents/rwqc2012.pdf>
- United States Environmental Protection Agency (US EPA) (2022). Recreational Waters. <https://www.epa.gov/report-environment/recreational-waters>
- Weerakoon, W.M.S.H., Nanayakkara, B.S., Abayasekara, C.L., Abeysundara, H.T.K., & Liyanage, L.R.I.S., (2022). Enumeration and Antibiotic Sensitivity of *Escherichia coli* and a Sociological Survey on Water Quality in the river Mahaweli, Sri Lanka [Abstract]. Annual Research Congress of the Postgraduate Institute of Science, University of Peradeniya, Sri Lanka. p.149.
- White, M.P., Elliott, L.R., Gascon, M., Roberts, B., & Fleming, L.E. (2020). Blue space, health and well-being: A narrative overview and synthesis of potential benefits. *Environmental Research*, 191, 110169. DOI: <https://doi.org/10.1016/j.envres.2020.110169>
- Wijekoon, C.N., Dassanayake, K.M.M.P., & Pathmeswaran, A. (2014). Antimicrobial susceptibility patterns and empirical prescribing practices in adult in-patients with urinary tract infection in a tertiary care hospital in Sri Lanka: Is there a need for changing clinical practices? *Sri Lankan Journal of Infectious Diseases*. 4(1):9-21. DOI: <http://dx.doi.org/10.4038/sljid.v4i1.6229>
- Wimalasuriya, K.M.D.D.C., Chandrathilake, T.H.R.C., Liyanage, B.C., & Gunatilake, J. (2011). *In-situ* water quality and economical leachate treatment system for

- Gohagoda dumping site, Sri Lanka.
- World Health Organization (2021). WHO guidelines on recreational water quality: volume 1: coastal and fresh waters. World Health Organization, Geneva.
- Wu, D., Ding, Y., Yao, K., Gao, W., & Wang, Y. (2021). Antimicrobial resistance analysis of clinical *Escherichia coli* isolates in neonatal ward. *Frontiers in Pediatrics*, 9, 670470. DOI: 10.3389/fped.2021.670470



## RESEARCH ARTICLE

### Plant Genetics

# Evaluation of genotypic variability and yield performance of selected indigenous and improved *Ipomoea batatas* L. genotypes under poly sack cultivation

MFZ Safwa<sup>1</sup>, TN Samarasinghe<sup>1</sup>, AN Wijewardhana<sup>1</sup>, HAPA Shyamalee<sup>1</sup>, AL Ranawake<sup>2\*</sup>

<sup>1</sup> Agriculture Research Station, Thelijawila, Sri Lanka.

<sup>2</sup> Department of Agricultural Biology, Faculty of Agriculture, University of Ruhuna, Mapalana, Sri Lanka.

Submitted: 05 May 2024; Revised: 16 January 2025; Accepted: 21 January 2025

**Abstract:** The genotypic diversity of *Ipomoea batatas* L. is important to enhance resilience against diseases, adaptability to diverse environments, and to provide a broader nutritional profile to address food security challenges. Poly sack cultivation is often considered superior to field cultivation because it offers better control over environmental factors. This study examined the diversity among seventeen indigenous *I. batatas* accessions compared to five improved varieties through comparative yield analysis under poly sack cultivation. Thirty-three traits were recorded following the standards set by the International Potato Center. Principal component (PC) analysis, cluster analysis and correlation analysis were performed using IBM SPSS 25 statistical software. The four PCs that exceeded Eigenvalue one, explained 76.55% of the total variance. Storage root characteristics and vine diameter contributed to PC1, while leaf, petiole, internodal and vine characteristics contributed to PC2. Four distinct clusters were identified at cluster distance ten in the hierarchical agglomerative clustering using Ward linkage. Cluster II comprised the thickest storage roots. Accessions TJ19, TJ11, TJ1, and TJ20 were the best-performing under poly sack cultivation yielding 2.93 kg, 2.43 kg, 2.42 kg and 2.27 kg, respectively. These genotypes show potential for promoting urban cultivation of *I. batatas* in poly sack cultivation. Yield variation of *I. batatas* genotypes under different cultivation techniques must be considered when integrating them into farming systems since the best-yielding genotypes in field cultivation did not necessarily achieve the highest yields in poly sack cultivation. The existing flesh colour, shape of storage root, yield variation, and canopy characteristics can be utilized for breeding purposes in the future. Future studies should focus on controlling vegetative growth to optimize yield through

pruning or trailing, as many vegetative growth parameters were inversely correlated with yield.

**Keywords:** Agro-morphological characteristics, field and poly sack cultivation, *Ipomoea batatas*, PCA, yield determinants of *Ipomoea batatas*.

## INTRODUCTION

Sweet potatoes (*Ipomoea batatas*) are believed to have originated in Central and South America, with evidence of their cultivation dating back over 5,000 years. The Incas in Peru and the Aztecs in Mexico, were among the first to cultivate and consume sweet potatoes (Nedunchezhiyan & Ray, 2010). Christopher Columbus is credited with introducing sweet potatoes to Europe (Kingsbury, 1992). According to De Silva (2020), the Batatas genus is locally called 'Sweet potato'. *I. batatas* is commonly perceived as indigenous to Sri Lanka due to its longstanding cultivation as an important traditional food (Premathilake, 1999).

The *I. batatas* genome is typically hexaploid with a chromosome count of  $2n = 6x = 90$ ; some studies have documented the existence of tetraploid variants with  $2n = 4x = 60$  (Roullier *et al.*, 2013). Globally, 8 million hectares harvest 89 million tons of *I. batatas* (United Nations, 2020).

\* Corresponding author ([lankaranawake@hotmail.com](mailto:lankaranawake@hotmail.com); <https://orcid.org/0000-0003-0517-9911>)



This article is published under the Creative Commons CC-BY-ND License (<http://creativecommons.org/licenses/by-nd/4.0/>). This license permits use, distribution and reproduction, commercial and non-commercial, provided that the original work is properly cited and is not changed in anyway.



The Department of Agriculture in Sri Lanka, has introduced several improved *I. batatas* varieties, but none of the varieties have been developed for biotic or abiotic tolerances.

*I. batatas* has an advantage over other vegetables because it has a shorter growth period, and inclement weather rarely results in complete crop loss (Bashaasha et al., 1995). Processing the storage roots of *I. batatas* enhances accessibility and reduces post-harvest waste (Arutselvan et al., 2023).

*I. batatas* is a fundamental source of carbohydrates, energy, and essential phytonutrients (Mohanraj and Sivasankar, 2014). It is expected to play an important role in addressing malnutrition (Naskar et al., 2008). Due to their unique nutritional value and processing methods, the National Aeronautics and Space Administration (NASA) has selected *I. batatas* as a potential food item for astronauts on space missions (Arutselvan et al., 2023). In addition to its storage roots, *I. batatas* vines and foliage are essential sources of vitamins, proteins and starch (Chakrabarti et al., 2014).

*I. batatas* is grown in South Africa when dry conditions are not suitable for cereals (Motsa et al., 2015). *I. batatas* can adapt to diverse climatic and soil conditions, which has positioned it as a resilient and essential component of agricultural systems (Solomon, 1999; Mwanga et al., 2021). Hence, *I. batatas* is considered as a food security crop providing a high yield in terms of edible energy per unit area over a short period of time.

Cultivation practices for *I. batatas* vary depending on local agro-climatic conditions. It typically thrives in well-drained sandy or loamy soils with pH levels of 5.5-6.5 and moderate rainfall. It is propagated primarily through vine cuttings or tubers and has a relatively short growth cycle of 3-5 months. Its yield is influenced by factors such as variety, soil fertility, and management practices, with average global yields ranging from 15-25 tons per hectare (Department of Agriculture, 2021).

Genetic diversity in *I. batatas* is remarkable, with significant accession diversity observed in landraces and wild relatives, which serve as valuable genetic resources for breeding programs (Abed et al., 2023; Mahaman Mourtala et al., 2023). The genetic diversity of *I. batatas* is more significant than that of potato, which is increasingly utilized by plant breeders searching for potential crop improvement (Woolfe, 1992). *I. batatas* accessions can share identical physical traits but differ genetically; conversely, genetically identical clones can

exhibit varying physical characteristics (Lebot, 2009). Some accessions of *I. batatas* germplasm in Sri Lanka have been previously evaluated (Premathilake, 1999; Vilochani et al., 2011). These diverse accessions can be utilized to enhance traits such as yield, resistance to pests and diseases, drought tolerance, and nutritional quality.

Storage root enlargement (sink) and leaf canopy photosynthetic efficiency (source) are both important for crop productivity and are primarily influenced by environmental factors (Ghosh et al., 1988; Palaniswami & Peter, 2008). However, genotypic variation and ecological conditions determine the source potential and the sink capacity (Ravi & Indira, 1998).

In Sri Lanka, *I. batatas* can be cultivated throughout the year under different growing conditions (rice fallow land, open highland, river valley basins, homestead, partial shade) in all the agroecological zones (De Silva and Premathilake, 1995). However, the extent of shortening the vegetative phase in *I. batatas* varies based on genetic factors and the local environment (Lamaro, 2020). Different *I. batatas* varieties respond differently to poly sack cultivation (Ndwamato et al., 2022).

Poly sack cultivation has been shown to be optimal for urban gardens (Pearson et al., 2010) and is particularly well-suited for infertile soils as well as drought-prone or flood-prone regions (Tixier & De Bon, 2006). Generally, poly sack cultivation of vegetables ensures four meals per week for a family (Pascal & Mwendu, 2009). Compared to traditional field methods, poly sack cultivation requires less space and labour (Essilfie et al., 2016), watering (Dubbeling & Massonneau, 2014), and weeding is more convenient (Badgely et al., 2007). Utilizing poly sack cultivation has the potential to expedite the vegetative phase of *I. batatas*, as the controlled environment provided by poly sacks can create optimal growing conditions leading to faster growth and development of the plants. This acceleration could reduce the time taken for the plants to mature, providing an advantage over weevil and other soil-borne pests (Johnson & Gurr, 2016).

While *I. batatas* has been discussed in various contexts, there is a noticeable gap in the existing literature on comparative yield analysis of different indigenous *I. batatas* accessions in Sri Lanka.

This study aims to understand the genetic diversity of *I. batatas* accessions and their yield potential to introduce planting materials and simultaneously conserve the genetic diversity of *I. batatas*. Further,

adopting poly sack cultivation excludes many barriers faced in the field. Additionally, a broader range of individuals, including rural and urban, young and adult, can contribute to sustainable agricultural practices. This study evaluated the yield performance of twenty-two indigenous and improved *I. batatas* genotypes under poly sack cultivation, with implications for urban agricultural practices.

The primary objectives of this study were to assess the diversity among indigenous accessions of *I. batatas* compared to improved varieties, identify the traits influencing yield, and evaluate the yield performance of each genotype under poly sack cultivation.

## MATERIALS AND METHODS

The indigenous accessions (denoted by *TJ* numbers) and improved *I. batatas* varieties (denoted by names) were used in this study. These accessions were *TJ1*, *TJ2*, *TJ4*, *TJ5*, *TJ7*, *TJ8*, *TJ9*, *TJ10*, *TJ11*, *TJ12*, *TJ14*, *TJ15*, *TJ16*, *TJ17*, *TJ18*, *TJ19*, *TJ20*, and the improved varieties *Ama*, *HORDI white*, *Ranabima*, *Shanthi*, and *Wariyapola-red*. Indigenous accessions were collected from different locations in Southern Sri Lanka namely *Tangalle*, *Kataragama*, *Tissamaharama*, *Sooriyawewa*, *Ambalantota*, *Weligama*, *Devinuwara (Dondra)*, *Akuressa*, *Kamburugamuwa*, *Pitabeddara*, *Hikkaduwa*, *Ahangama*, *Karapitiya*, *Ambalangoda*, and *Yakkalamulla*. The experiment comparing field and poly sack cultivation techniques was conducted at the Agriculture Research Station in Thelijawila, Matara. Sri Lanka (6°01'04.5"N, 80°30'15.5"E).

### Crop establishment in poly sacks and the field

The cuttings were collected from the field-established germplasm collection. In the field, 0.3 m elevated ridges were prepared using the same soil mixtures that were used for the poly sacks, and the cuttings were planted with a spacing of 60 x 20 cm. Poly sacks with dimensions of 55 x 35 cm were used for the experiment. The sacks were filled with a potting mixture of topsoil, cow dung, and partially burnt paddy husk in a 1:1:1 ratio. The total weight of the potting mixture in each sack was 10 kg. Each cutting measuring approximately 25 cm in length was planted in the poly sacks (Supplementary Figure 1). Poly sacks were maintained under the same environmental conditions as the field experiment. The experiment was conducted from August to December in 2022 and 2023. Data were collected at the maturity stage, identified by yellowing of the lower leaves and latex exudation from damaged stems.

Irrigation was practiced daily during the first week. The plants were established with the onset of the southwest monsoon, and weed management was done manually. Fertilization was carried out by following the recommendations provided by the Department of Agriculture, Sri Lanka: basal dressing was applied two weeks after planting (urea 60 kg/ha, TSP 120 kg/ha, and MOP 120 kg/ha), and top dressing was applied six weeks after planting (urea 60 kg/ha, and MOP 60 kg/ha). Fertilizer was applied based on the calculated per-plant requirement, considering the polybag and field area.

### Experimental design and data collection

The experiment was conducted in poly sacks using a randomized complete block design with three replicates, each replicate consisting of ten poly sacks. The experiment was repeated in two seasons. All qualitative and quantitative data were collected according to the descriptors for *I. batatas* (Huaman, 1999), developed by the joint contribution of the International Potato Center (CIP), the Asian Vegetable Research and Development Center (AVRDC), and the International Board for Plant Genetic Resources (IBPGR).

### Parameters evaluated for characterization of *I. batatas*

The qualitative canopy characteristics included the general outline of the leaf lamina, lobe prominence, central lobe shape, petiole pigmentation, abaxial vein pigmentation, and foliage colour. Qualitative storage root characteristics included shape variations of the root, skin colour, and predominant and secondary flesh colour. The vegetative characteristics studied were mature leaf length (cm), internodal length (cm), number of leaves per plant, number of branches per plant, vine length (cm), yield per plant (kg), cortex thickness (mm), length of storage roots (cm), and width of storage roots (cm). Quantitative data were analyzed using SPSS software (SPSS Inc., 2012).

### Data analysis

Analysis of variance (ANOVA) was performed to assess the overall significance of differences among genotypes for various traits. Duncan's multiple range test (DMRT) was applied as a post-hoc test to identify significant differences in yield among genotypes. Principal component analysis (PCA) was used to reduce dimensionality and identify key traits contributing to variability among genotypes. Cluster analysis, using Ward's linkage method, grouped genotypes based on similarities in observed traits. Correlation analysis was conducted to explore relationships between yield and other agronomic traits.

## RESULTS AND DISCUSSION

### Diversity of indigenous *I. batatas* accessions

There was significant diversity among the examined genotypes of *I. batatas* in various qualitative and quantitative characteristics. The leaf outline shape showed distinct variation: triangular, cordate, hastate or lobed outline (Supplementary Figure 2).

Most genotypes, except *TJ9*, *TJ17*, and *TJ18*, had green mature leaves (Supplementary Figure 2). *TJ9*, *TJ17*, and *TJ18* showed yellow-green leaves, green leaves with purple veins on the upper surface, and both

leaf surfaces purple, respectively. Both mature and immature leaves were green in *TJ1*, *TJ8*, *TJ20*, *Shanthi*, and *HORDI-white*. Other accessions were green and purple. Interestingly in the group of accessions with purple pigments on leaves, *TJ18* mature leaves were purple, but immature leaves were green while *TJ4*, *TJ5*, *TJ7*, *TJ9-TJ15*, *TJ16*, *TJ19*, *TJ20*, *Ama*, *Ranabima*, and *Wariyapola-red*, showed various purple pigmentation patterns in the immature leaves (Table 1, Supplementary Figure 3).

*TJ1*, *TJ2*, *TJ4*, *TJ19*, *Ama*, *HORDI-white* and *Wariyapola-red* had green petioles while other accessions exhibited purple pigments. Among genotypes with green

**Table 1:** Qualitative leaf characteristics of *I. batatas* genotypes

Genotype	General outline	Lobe prominence	Central lobe shape	Petiole pigmentation	Abaxial vein pigmentation	Foliage color	
						Mature	Immature
<i>TJ1</i>	Cordate	NLL	TO	G	G	G	G
<i>TJ2</i>	Cordate	NLL	TO	G	G	G	GG
<i>TJ4</i>	Lobed	S	TR	G	G	G	GPE
<i>TJ5</i>	Lobed	S	TR	GPBE	PSMRB	G	GPE
<i>TJ7</i>	Cordate	NLL	TO	GPNL	PSMRB	G	GPE
<i>TJ8</i>	Lobed	VD	OB	GPNL	PSMRB	G	G
<i>TJ9</i>	Triangular	VS	TR	GPS	AVPP	YG	GG
<i>TJ10</i>	Cordate	NLL	SE	GPNL	G	G	PBS
<i>TJ11</i>	Lobed	M	L	GPNL	PSMRB	G	MP
<i>TJ12</i>	Lobed	D	L	GPNL	PSMRB	G	PBS
<i>TJ14</i>	Hastate	S	SE	GPSTP	PSSV	G	PBS
<i>TJ15</i>	Cordate	VS	SC	GPNL	G	G	GG
<i>TJ16</i>	Cordate	NLL	TO	GPNL	PSMRB	G	GPVUS
<i>TJ17</i>	Lobed	D	SE	MP	AVMP	GPVUS	G
<i>TJ18</i>	Lobed	D	L	MP	MP	PBS	G
<i>TJ19</i>	Cordate	NLL	TR	G	G	G	PBS
<i>TJ20</i>	Lobed	M	SE	GPBE	AVMP	G	G
<i>Ama</i>	Lobed	D	L	G	G	G	SP
<i>HORDI white</i>	Lobed	M	E	G	G	G	G
<i>Ranabima</i>	Triangular	VS	TR	GPS	G	G	GPNL
<i>Shanthi</i>	Cordate	NLL	TO	GPNL	GPNL	G	G
<i>Wariyapola-red</i>	Lobed	VS	TO	G	G	G	GPE

NLL: no lateral lobes, S: slight, VS: very slight, M: moderate, D: deep, VD: very deep, TO: toothed, TR: triangular, SE: semi-elliptic, L: lanceolate, SC: semi-circular, OB: oblanceolate, E: elliptic, G: green, PSMRB: a purple spot in main rib base, AVPP: all veins partially purple, PSSV: purple spots in several veins, AVMP: all veins mostly purple, MP: mostly purple, GPNL: green with purple near leaf, YG: yellow-green, GPVUS: green with purple veins on the upper surface, PBS: purple both surfaces, GG: greyish green, GPE: green with purple edge, MP: mostly purple, GPNL: green with purple near leaf base, SP: slightly purple, GPS: green with purple stripes, GPSTP: green with purple spot throughout petiole, GPBE: green with purple at both ends.

petioles, observed leaf margins were cordate in *TJ1*, *TJ2* and *TJ19* while *TJ4*, *Ama*, *HORDI-white* and *Wariyapola-red* were lobed (Supplementary Figure 2). In the group of purple pigmented petioles, *TJ5*, *TJ8*, *TJ11*, *TJ12*, *TJ17*, *TJ18* and *TJ20* leaf lamina were lobed whereas *TJ9* and *Ranabima* were triangular. Hastate leaf lamina was observed only in *TJ14* while *TJ7*, *TJ10*, *TJ15*, *TJ16* and *Shanthi* leaf lamina were cordate (Table 1, Supplementary Figure 2).

The leaf lobes showed different degrees of prominence: very slight, slight, moderate, deep, or very deep. The central lobes exhibited seven different shapes (Table 1). The number of lobes in the accessions exhibited a range of variation, transitioning from 3, 5 to 7 lobes. Rahajeng *et al.*, (2018) have also documented the presence of 3, 5, and 7 lobes in the *I. batatas* accessions they studied. The diversity in purple pigmentation in the leaves, veins, and petiole, along with the shape of leaf lamina and the features of the main lobe, can be used to

distinguish the studied *I. batatas* genotypes. *TJ17* and *TJ18* accessions possess ornamental value due to their vibrant canopy colors (Supplementary Figure 3).

The qualitative attributes of the storage roots of the *I. batatas* genotypes were assessed in terms of storage root shape, size variations, and skin and flesh colors (Table 2). Storage roots did not emerge in the *TJ17* and *TJ18* accessions.

The flesh of the storage root displayed colors, ranging from white, cream, pale orange, intermediate orange to vividly pigmented with anthocyanin (Supplementary Figure 4). Mau *et al.*, (2022) reported six accessions with anthocyanin pigmentation in the flesh and nine with anthocyanin pigmentation in the twigs. The tuber skin exhibited diverse colours: brownish orange, cream, dark purple, pink, purple-red, and white (Supplementary Figure 4). This diversity highlights the inherent breeding potential for various combinations of traits in offspring.

**Table 2:** Qualitative storage root characteristics of *I. batatas* genotypes

Genotype	Shape	Shape variation	Size variation	Skin color	Flesh color (predominant)	Flesh color (Secondary)	Dispersion of secondary flesh color
<i>TJ1</i>	Obovate	Slight	Slight	BO	White	Absent	A
<i>TJ2</i>	Long elliptic	Uniform	Slight	C	White	Cream	S
<i>TJ4</i>	Obovate	Slight	Slight	PR	Pale Yellow	Cream	BRC
<i>TJ5</i>	Round elliptic	Slight	Slight	PR	Cream	Absent	A
<i>TJ7</i>	Ovate	Moderate	Slight	BO	Pale orange	Cream	S
<i>TJ8</i>	Elliptic	Uniform	Slight	C	White	Cream	NRC
<i>TJ9</i>	Long elliptic	Slight	Moderate	PR	White	Absent	A
<i>TJ10</i>	Long curved	Slight	Moderate	DP	SPWA	Dark purple	S
<i>TJ11</i>	Round elliptic	Slight	Slight	C	I. Orange	Cream	S
<i>TJ12</i>	Long curved	Slight	Moderate	C	White	Cream	BRC
<i>TJ14</i>	Long irregular	Slight	Slight	DP	SPWA	White	NRF
<i>TJ15</i>	Long elliptic	Slight	Moderate	C	White	Cream	S
<i>TJ16</i>	Long irregular	Slight	Slight	DP	SPWA	White	S
<i>TJ19</i>	Elliptic	Slight	Slight	DP	SPWA	Dark purple	BRF
<i>TJ20</i>	Round	Moderate	Moderate	P	Cream	White	S
<i>Ama</i>	Long elliptic	Slight	Slight	PR	Pale orange	Cream	BRC
<i>HORDI-white</i>	Oblong	Slight	Moderate	W	White	Cream	NRC
<i>Ranabima</i>	Long elliptic	Slight	Moderate	PR	Cream	Pink	BRF
<i>Shanthi</i>	Elliptic	Slight	Slight	C	White	Absent	A
<i>Wariyapola-red</i>	Long curved	Slight	Moderate	PR	White	Absent	A

BO: brownish orange, C: cream, DP: dark purple, P: pink, PR: purple-red, W: white, SPWA: strongly pigmented with anthocyanins, I. Orange: intermediate orange, A: absent, S: scattered, BRC: broad ring in the cortex, NRF: narrow ring in the flesh, BRF: broad ring in the flesh, NRC: narrow ring in the cortex.

*I. batatas* accessions and improved varieties can be distinguished according to the following key based on qualitative leaf and tuber characteristics.

*Ipomoea batatas*

1. Canopy has green leaves with purple pigmentation
  2. Leaf margin is laterally lobed
    3. Petiole is pigmented with purple and green; some are moderately purple
      4. Leaf abaxial veins are mostly purple
        5. Mature leaves are purple pigmented ..... TJ18
        - 5'. Mature leaves are not purple pigmented; they're green
          6. Produce storage root..... TJ20
          - 6'. Do not produce storage root..... TJ17
      - 4'. Leaf abaxial veins are not mostly purple; exhibits a green mixed with purple colour
        7. Leaf margin is moderate to deeply lobed
          8. Storage root is carotene pigmented; orange flesh..... TJ12
          - 8'. Storage root is not carotene pigmented; white flesh..... TJ11
        - 7'. Leaf margin is not deeply lobed; slightly lobed
          9. Leaf veins have purple spots
            10. Storage root is strongly pigmented with anthocyanins; purple pigmented flesh..... TJ14
            - 10'. Storage root is not pigmented with anthocyanins; white or cream color flesh
              11. Mature leaf exhibits green with yellow pigmentation..... TJ9
              - 11'. Mature leaf has no yellow mixed green; only green colour..... TJ5
            - 9'. Leaf veins have no purple spots; green pigmented veins..... Ranabima
      - 3'. Petiole is green pigmented; no purple
        12. Storage root flesh is carotene pigmented..... Ama
        - 12'. Storage root flesh is not carotene pigmented; white or cream flesh
          13. Flesh exhibits a secondary colour: cream..... TJ4
          - 13'. Flesh has no secondary color distribution..... Wariyapola-red
    - 2'. Leaf margin is not laterally lobed; only central lobe is present
      14. Petiole is green with purple pigmented
        15. Purple spot in main rib base
          14. Storage root is strongly pigmented with anthocyanins; purple flesh..... TJ16
          - 14'. Storage root is not pigmented with anthocyanins; orange colour..... TJ7
        - 15'. No purple spots in main rib base
          15. storage root is anthocyanin pigmented; purple flesh..... TJ15
          - 15'. Storage root is not anthocyanin pigmented; white flesh..... TJ10
      - 14'. Petiole is green without purple pigmentation
        16. Tooth shaped central lobe..... TJ2
        - 16'. Triangular shaped central lobe..... TJ19
- 1'. Canopy has green foliage; no purple pigmentation
  15. Leaf margin is laterally lobed
    16. Deeply lobed; 7 lobes..... TJ8
    - 16'. Moderately lobed; 3 lobes ..... HORDI-white
  - 15'. Leaf margin is not laterally lobed; central lobe tooth shaped
    17. Storage root skin is carotene pigmented; brownish orange skin..... TJ1
    - 17'. Storage root skin is not carotene pigmented; white skin..... Shanthy

Quantitative canopy and storage root characteristics showed statistically significant differences among all the studied genotypes (Table 3). Variety *Ama* recorded the longest main vine length (140 cm) and internodal length (8.89 cm). *Wariyapola-red* had the shortest vine internodal length (2.16 cm) and the thickest root cortex (3.6 cm). In a previous study with sixty-two *I. batatas* accessions, vine length ranged from 83.08 to 283.80 cm, with an average of 178.65 cm (Rahajeng &

Rahayuningsih, 2017). The same study reported that the number of branches per plant ranged from 0.80 to 1.60. In this study, the improved variety '*HORDI-white*' and indigenous accession *TJ11* had the highest number of branches per plant (4 branches). The improved variety, *Shanthi*, had the lowest values for vine diameter, mature leaf length, petiole length, and storage root width (Table 3).

**Table 3:** DMRT grouping of all the evaluated morphological traits in *I. batatas* genotypes cultivated in poly sacks

Accession name	MVL (cm)	NLPP	NBPP	VIL (cm)	VD (mm)	MLL (cm)	PL (cm)	SRL (cm)	SRW (cm)	RCT (mm)	TY (kg)
<i>TJ1</i>	92.72 <sup>bc</sup>	101.06 <sup>ab</sup>	2.72 <sup>i</sup>	5.69 <sup>b</sup>	4.33 <sup>efgh</sup>	13.61 <sup>a</sup>	16.21 <sup>bc</sup>	10.03 <sup>cdef</sup>	4.45 <sup>cde</sup>	2.08 <sup>cd</sup>	2.41 <sup>b</sup>
<i>TJ2</i>	46.44 <sup>g</sup>	65.56 <sup>cdefg</sup>	2.94 <sup>ghi</sup>	2.40 <sup>i</sup>	4.27 <sup>fgh</sup>	9.25 <sup>hi</sup>	12.22 <sup>cdef</sup>	14.08 <sup>abc</sup>	4.63 <sup>bcd</sup>	2.08 <sup>cd</sup>	1.81 <sup>d</sup>
<i>TJ4</i>	65.61 <sup>cdefg</sup>	84.50 <sup>bc</sup>	3.33 <sup>cdefgh</sup>	5.46 <sup>bc</sup>	4.11 <sup>gh</sup>	9.91 <sup>ghi</sup>	11.04 <sup>def</sup>	9.30 <sup>ef</sup>	3.73 <sup>efg</sup>	2.65 <sup>bcd</sup>	0.16 <sup>k</sup>
<i>TJ5</i>	97.72 <sup>b</sup>	39.94 <sup>fgh</sup>	3.67 <sup>abcde</sup>	4.04 <sup>def</sup>	5.22 <sup>bc</sup>	13.43 <sup>a</sup>	15.88 <sup>bcd</sup>	12.47 <sup>bcd</sup>	4.87 <sup>bcd</sup>	1.68 <sup>d</sup>	0.40 <sup>j</sup>
<i>TJ7</i>	47.50 <sup>g</sup>	44.44 <sup>efgh</sup>	3.00 <sup>ghi</sup>	3.22 <sup>gh</sup>	4.56 <sup>defg</sup>	13.82 <sup>a</sup>	14.55 <sup>bcd</sup>	9.67 <sup>def</sup>	6.37 <sup>a</sup>	3.57 <sup>a</sup>	1.24 <sup>f</sup>
<i>TJ8</i>	65.06 <sup>cdefg</sup>	52.67 <sup>defgh</sup>	3.56 <sup>bcd</sup>	3.36 <sup>fgh</sup>	4.05 <sup>h</sup>	12.15 <sup>bcd</sup>	13.63 <sup>cdef</sup>	10.45 <sup>cdef</sup>	4.00 <sup>def</sup>	2.6 <sup>bcd</sup>	0.60 <sup>j</sup>
<i>TJ9</i>	41.67 <sup>g</sup>	72.00 <sup>cde</sup>	3.06 <sup>fghi</sup>	2.23 <sup>i</sup>	4.27 <sup>fgh</sup>	11.52 <sup>cde</sup>	12.07 <sup>cdef</sup>	12.53 <sup>bcd</sup>	3.08 <sup>g</sup>	1.85 <sup>cd</sup>	1.15 <sup>f</sup>
<i>TJ10</i>	57.06 <sup>efg</sup>	42.33 <sup>efgh</sup>	2.89 <sup>hi</sup>	2.19 <sup>i</sup>	4.61 <sup>def</sup>	10.04 <sup>fghi</sup>	12.89 <sup>cdef</sup>	15.83 <sup>ab</sup>	3.38 <sup>fg</sup>	1.60 <sup>e</sup>	1.41 <sup>e</sup>
<i>TJ11</i>	54.39 <sup>fg</sup>	114.50 <sup>a</sup>	4.11 <sup>a</sup>	3.51 <sup>efgh</sup>	3.50 <sup>h</sup>	10.82 <sup>efg</sup>	11.52 <sup>cdef</sup>	8.67 <sup>f</sup>	3.63 <sup>efg</sup>	1.63 <sup>d</sup>	2.42 <sup>b</sup>
<i>TJ12</i>	81.11 <sup>bcd</sup>	54.00 <sup>cdefgh</sup>	3.17 <sup>efghi</sup>	4.26 <sup>de</sup>	3.94 <sup>h</sup>	10.41 <sup>efgh</sup>	10.04 <sup>ef</sup>	13.22 <sup>abcde</sup>	3.40 <sup>fg</sup>	2.03 <sup>cd</sup>	0.84 <sup>h</sup>
<i>TJ14</i>	84.67 <sup>bcd</sup>	47.50 <sup>efgh</sup>	3.56 <sup>bcd</sup>	4.76 <sup>cd</sup>	4.61 <sup>def</sup>	12.58 <sup>abc</sup>	11.43 <sup>cdef</sup>	12.05 <sup>bcd</sup>	3.03 <sup>g</sup>	2.38 <sup>bcd</sup>	1.45 <sup>e</sup>
<i>TJ15</i>	40.17 <sup>g</sup>	68.56 <sup>cdef</sup>	4.06 <sup>ab</sup>	2.83 <sup>hi</sup>	5.27 <sup>ab</sup>	9.76 <sup>ghi</sup>	10.41 <sup>ef</sup>	16.67 <sup>a</sup>	4.92 <sup>bcd</sup>	2.40 <sup>bcd</sup>	1.02 <sup>g</sup>
<i>TJ16</i>	53.61 <sup>fg</sup>	60.94 <sup>cdefgh</sup>	3.78 <sup>abc</sup>	3.94 <sup>efg</sup>	4.78 <sup>de</sup>	13.46 <sup>a</sup>	15.66 <sup>bcd</sup>	12.32 <sup>bcd</sup>	4.08 <sup>def</sup>	1.75 <sup>cd</sup>	0.40 <sup>j</sup>
<i>TJ17</i>	48.39 <sup>g</sup>	31.11 <sup>h</sup>	2.89 <sup>hi</sup>	2.30 <sup>i</sup>	5.67 <sup>a</sup>	12.86 <sup>ab</sup>	21.89 <sup>a</sup>				0.00 <sup>l</sup>
<i>TJ18</i>	87.22 <sup>bcd</sup>	35.72 <sup>gh</sup>	3.78 <sup>abc</sup>	5.14 <sup>bc</sup>	4.17 <sup>fgh</sup>	13.41 <sup>ab</sup>	21.37 <sup>a</sup>				0.00 <sup>l</sup>
<i>TJ19</i>	78.33 <sup>bcd</sup>	34.06 <sup>gh</sup>	3.83 <sup>abc</sup>	3.76 <sup>efg</sup>	4.83 <sup>cd</sup>	12.52 <sup>abc</sup>	15.83 <sup>bcd</sup>	14.9 <sup>ab</sup>	5.08 <sup>bc</sup>	3.13 <sup>ab</sup>	2.93 <sup>a</sup>
<i>TJ20</i>	65.22 <sup>cdefg</sup>	37.67 <sup>fgh</sup>	3.72 <sup>abcde</sup>	3.38 <sup>fgh</sup>	4.56 <sup>defg</sup>	10.24 <sup>fgh</sup>	12.23 <sup>cdef</sup>	10.27 <sup>cdef</sup>	6.45 <sup>a</sup>	2.75 <sup>abc</sup>	2.27 <sup>c</sup>
<i>Ama</i>	140.00 <sup>a</sup>	37.83 <sup>fgh</sup>	3.22 <sup>defghi</sup>	8.89 <sup>a</sup>	3.94 <sup>h</sup>	10.93 <sup>efg</sup>	13.06 <sup>cdef</sup>	11.92 <sup>bcd</sup>	5.25 <sup>bc</sup>	2.75 <sup>abc</sup>	0.40 <sup>j</sup>
<i>HORDI-white</i>	48.22 <sup>g</sup>	79.78 <sup>bcd</sup>	4.11 <sup>a</sup>	3.27 <sup>gh</sup>	4.83 <sup>cd</sup>	12.85 <sup>ab</sup>	13.01 <sup>cdef</sup>	13.00 <sup>abcde</sup>	5.55 <sup>ab</sup>	2.5 <sup>bcd</sup>	0.84 <sup>h</sup>
<i>Ranabima</i>	138.11 <sup>a</sup>	55.44 <sup>cdefgh</sup>	3.83 <sup>abc</sup>	5.71 <sup>b</sup>	4.05 <sup>h</sup>	10.82 <sup>efg</sup>	18.60 <sup>ab</sup>	9.12 <sup>ef</sup>	2.75 <sup>gh</sup>	2.08 <sup>cd</sup>	0.45 <sup>j</sup>
<i>Shanthi</i>	91.89 <sup>bc</sup>	35.72 <sup>gh</sup>	3.44 <sup>cdefg</sup>	5.01 <sup>bc</sup>	2.00 <sup>i</sup>	8.92 <sup>i</sup>	8.72 <sup>f</sup>	13.55 <sup>abcd</sup>	2.10 <sup>h</sup>	1.78 <sup>cd</sup>	0.12 <sup>kl</sup>
<i>Wariyapola-red</i>	60.28 <sup>defg</sup>	63.94 <sup>cdefg</sup>	3.72 <sup>abcde</sup>	2.16 <sup>i</sup>	4.94 <sup>bcd</sup>	11.29 <sup>def</sup>	13.14 <sup>cdef</sup>	15.35 <sup>ab</sup>	4.6 <sup>bcd</sup>	3.6 <sup>a</sup>	0.37 <sup>j</sup>

MVL: main vine length, NLPP: number of leaves per plant, NBPP: number of branches per plant, VIL: vine internode length, VD: vine diameter, MLL: mature leaf length, PL: petiole length, SRL: storage root length, SRW: storage root width, RCT: root cortex thickness, TY: total yield

*TJ15* recorded the longest storage root length at 16.67 cm, while *TJ11* had the shortest at 8.67 cm. *TJ20* and *TJ7* demonstrated the maximum storage root width, measuring 6.45 cm and 6.37 cm respectively. The broadest vine diameter measuring 5.67 mm was observed in *TJ17*. Hence, *TJ17* and *TJ20* accessions have a

significant breeding potential for future research studies. Stem thickness has been identified as a favourable trait for extensive storage in *I. batatas* by Ndawmato *et al.*, (2022). Furthermore, Raemaekers (2001) confirmed that genetic and environmental factors control vegetative growth, including the vine diameter of *I. batatas*.

Principal component (PC) analysis resulted in four principal components with more than one Eigenvalue. PC1 was significantly influenced by vine diameter, storage root width, root cortex thickness, the best four storage root yield, and total yield (Table 4).

**Table 4:** Component matrix of *I. batatas* genotypes.

Quantitative characteristics	Component			
	1	2	3	4
Main vine length (cm)	-0.556	0.659	-0.388	0.097
Number of leave per plant	-0.081	-0.066	0.77	-0.38
Number of branches per plant	0.126	-0.116	0.378	0.051
Vine internode length (cm)	-0.467	0.65	-0.384	-0.164
Vine diameter (mm)	0.741	0.146	0.262	0.45
Mature leaf length (cm)	0.362	0.683	0.406	0.142
Petiole length (cm)	0.15	0.778	0.297	0.266
Storage root length (cm)	0.182	-0.586	-0.243	0.673
Storage root width (cm)	0.905	0.208	-0.141	-0.244
Root cortex thickness (mm)	0.663	0.206	-0.343	-0.029
Best four storage root yield (kg)	0.89	0.196	-0.238	-0.062
Total yield (kg)	0.702	-0.265	-0.221	-0.445

PC2 showed significant contributions from main vine length, petiole length, and mature leaf length, while PC3 was primarily constructed by the number of leaves per plant. The storage root length was the most prominent contributor to PC4. The four principal components described 76.5% of the total variance, while the principal components 1, 2, and 3 explained 32%, 21%, and 14%, respectively.

In a previous study, nineteen phenotypic traits elucidated 77.8% of the variance (Placide *et al.*, 2015). Rahajeng *et al.*, (2023) reported 76.3% of the variance described by seven principal components. Premathilake (1999) and Vilochani (2011) have reported the diversity in different *I. batatas* accessions in prior studies. In the present study Cluster I consisted of ten genotypes, showing similarities in mature leaf length, petiole length, and vine diameter. Cluster II and Cluster III consisted of three and two genotypes, respectively. The number of leaves per plant, vine diameter, and storage root width were common characteristics found in Cluster II, while

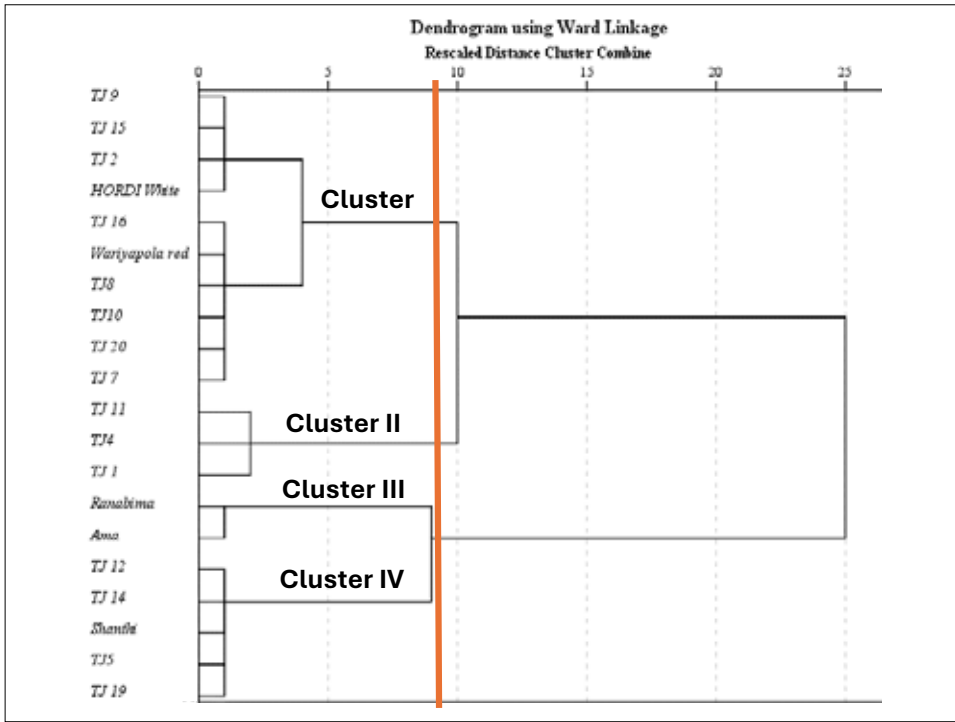
main vine length, number of branches per plant, mature leaf length, storage root length and root cortex thickness were common in Cluster III. Cluster IV consisted of genotypes with similarities in main vine length, number of branches per plant, vine internode length, and storage root length.

Character correlation is important in plant breeding (Kuswantoro, 2017) to determine the plant's architecture for optimal yield. However, none of the vegetative characteristics were significantly and positively correlated with storage root yield. The total yield was only correlated with storage root width (Table 5). Previous studies have also suggested that total yield was associated with root length and width (Lesatri *et al.*, 2021). However, in this study, total yield was not correlated with storage root length ( $r=-0.011$ ), but with storage root width ( $r=0.716$ ) (Table 5). Additionally, the storage root width was correlated with root cortex thickness ( $r=0.606$ ). The width and cortex thickness of storage roots are important for determining yield as they represent different aspects of the anatomy of *I. batatas*'s related to nutrient storage and physiological regulation (Ravi *et al.*, 2014). The cortex of sweet potatoes refer to the tissue surrounding the vascular system within the storage root playing a vital role in nutrient storage and transportation. A thicker cortex indicates greater storage capacity for nutrients and carbohydrates leading to higher yields and improved quality.

Total yield was significantly and negatively correlated with the main vine length ( $r=-0.495$ ). Excessive vegetative growth in *I. batatas* reduces tuber storage root yield due to resource competition. When the plant invests too much energy in producing and maintaining above-ground foliage, resources are diverted from tuber development resulting in smaller and fewer storage roots. Consistent with this, Sakaigaichi *et al.*, (2023) observed a substantial increase in yield in compact genotypes of *I. batatas*. However, high heritability estimates for the main vine length of *I. batatas*, suggest challenges in achieving compact genotypes through breeding programmes (Narayan *et al.*, 2022). Additionally, excessive foliage can cause shading limiting the sunlight reaching the lower parts of the plant and further inhibiting tuber growth.

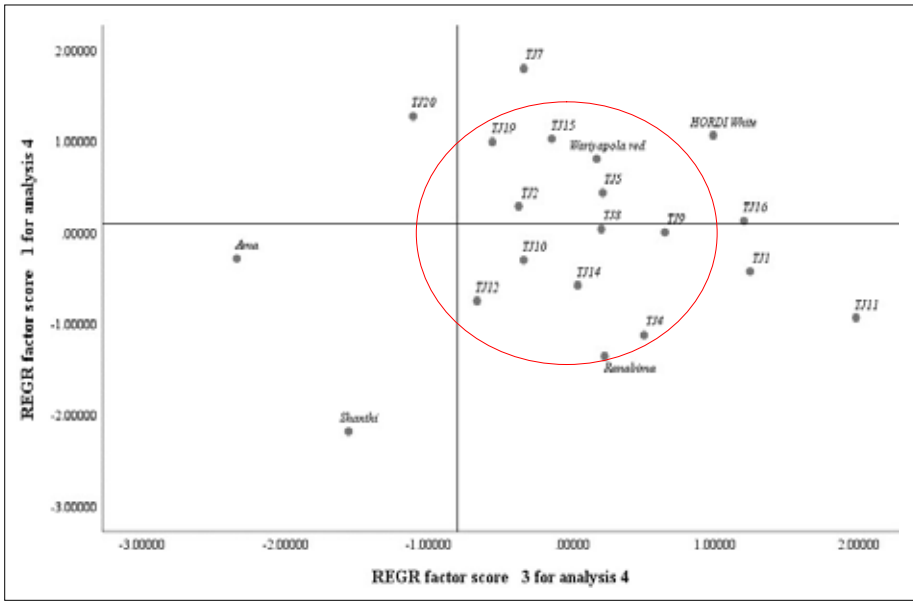
Main vine length showed a significant positive correlation with vine internodal length ( $r = 0.864$ ) and petiole length ( $r= 0.346$ ) of *I. batatas*. This aligns with findings by Abed *et al.*, (2023), who observed similar correlations with correlation coefficients of  $r = 0.83$  for internodal length and  $r = 0.67$  for petiole length. The number of leaves per plant did not positively correlate





**Figure 4:** Hierarchical agglomerative clustering among the accessions, a dendrogram of genetic distance made using PCA data employing the Ward criterion method.

The 2D scatter plot diagram showed moderate dispersion in all four quadrants. A tight cluster was visible, suggesting a strong relationship between some accessions. The improved varieties *Ama* and *Shanthi* had more dispersion, indicating a distant relationship. *TJ11*, *TJ19*, *TJ20*, and *TJ7* were also isolated from other genotypes. Among them, *TJ11*, *TJ19*, and *TJ20* were high-yielding accessions.



**Figure 5:** 2D scatter plot for *I. batatas* accessions.

with storage root characteristics ( $r=-0.146$ ). According to Widaryanto & Saitama (2017), the leaf area index of *I. batatas* was negatively correlated with the storage root yield. Proper pruning or management of vegetative growth can redirect resources to storage root production and enhance *I. batatas* yield. Panggabean *et al.*, (2014) have recommended shoot pruning to increase yield in

*I. batatas*. Future research should focus on improving storage root enlargement in *I. batatas* through shoot pruning by adjusting the number of leaves and the leaf area index (LAI). This can be accomplished by improving translocation efficiency to the storage roots, as indicated by positive changes in correlation values (Ndwamato *et al.*, 2022).

**Table 5:** Correlation matrix of *I. batatas* genotypes

Traits	MVL (cm)	NLPP	NBPP	VIL (cm)	VD (mm)	MLL (cm)	PL (cm)	SRL (cm)	SRW (cm)	RCT (mm)	BFSRW (kg)
NLPP	-.294										
NBPP	-.129	.155									
VIL	<b>.864</b>	-.132	-.159								
VD	-.311	-.003	.173	-.362							
MLL	-.004	.115	.017	.089	.469						
PL	<b>.346</b>	.015	-.008	.176	.410	.617					
SRL	-.314	-.282	.013	-.378	.290	-.352	-.360				
SRW	-.307	-.061	.033	-.148	<b>.572</b>	.345	.189	-.042			
RCT	-.114	-.250	.040	-.087	.314	.244	.105	.053	<b>.606</b>		
BFSRW	-.270	-.157	-.008	-.133	<b>.566</b>	.355	.144	.132	<b>.944</b>	<b>.684</b>	
TY	<b>-.495</b>	<b>-.146</b>	.052	-.346	.315	<b>-.135</b>	-.183	-.011	<b>.716</b>	.350	<b>.559</b>

MVL: main vine length, NLPP: number of leaves per plant, NBPP: number of branches per plant, VIL: vine internode length, VD: vine diameter, MLL: mature leaf length, PL: petiole length, SRL: storage root length, SRW: storage root width, RCT: root cortex thickness, BFSRY: best four storage root weight, TY: total yield

Duncan's multiple range test (DMRT) revealed significant yield differences across different genotypes in poly sack cultivation (Table 3).

Genotypes *TJ19*, *TJ11*, *TJ1*, and *TJ20* produced better yields under poly sack cultivation with 2.93 kg, 2.43 kg, 2.42 kg and 2.27 kg, respectively. The improved varieties *Ranabima*, *Shanthi*, *Wariyapola-red* and an indigenous accession *TJ1* yielded less than 1 kg per plant in poly sack cultivation; *TJ17* and *TJ18* recorded no yield under poly sack cultivation (Table 3), while *TJ16* was delayed in maturation in poly sacks. Maturation of *I. batatas* genotypes differs based on their genotype and environmental factors during cultivation (Bararyenya *et al.*, 2020). The production of storage root in *I. batatas* is influenced by sink-source; the sinking rate varies among varieties, resulting in yield variability. Some varieties prioritize shoot development, whereas others focus on enhancing root production (Ndwamato *et al.*, 2022). The yield variations among the genotypes occur due to genetic composition and environmental conditions (Ngailo *et al.*, 2013).

Generally, *I. batatas* has three growth phases based on dry matter partitioning. Shoot growth dominates during the first phase followed by the second phase, which involves consistent allocation between development of shoot and tubers (Ravi & Saravanan, 2012). High soil moisture extends the duration of total dry matter production, reducing the dry matter allocated for tubers and diverting it toward shoot growth (Satapathy *et al.*, 2005). The location where the study was carried out received prolonged rainfall that posed a potential challenge by saturating the soil. This excessive moisture limits the availability of oxygen in the root zone. Storage roots require sufficient  $O_2$  for their initiation and development, and excess soil moisture induces anaerobiosis, leading to decreased tuber production (Goswamy *et al.*, 1995). Surplus soil moisture results in a decrease in the number of tubers and size while causing an increase in the weight of fresh shoots. This is likely attributable to sink capacity reduction, obstructing photosynthate translocation to storage roots (Ravi & Saravanan, 2012). Additionally, excessive water can reduce proper aeration due to soil compression caused by the expansion of storage roots.

The observed variations in yield highlight the potential for optimizing agricultural practices to maximize overall crop yield.

## CONCLUSIONS

The findings of this research report the diversity of the studied *I. batatas* genotypes and their varying yield potential. The indigenous accession *TJ19* yielded the highest per plant yield (2.93 kg) followed by *TJ11* and *TJ1*, which recorded the second highest (2.42 kg), and *TJ20*, recorded the third highest (2.27 kg) per plant yield. This indicates that these accessions are the most suitable for poly sack cultivation. The four principal components exceeding Eigenvalue one collectively explained 76.55% of the overall variance. Storage root attributes and vine diameter influenced PC1. Four clusters were distinguished by utilizing Ward linkage in hierarchical agglomerative clustering at a cluster distance of ten. The most robust storage roots belonged to Cluster II. The 2D scatter plot illustrated moderate variation among the accessions. Vine diameter was the only vegetative factor that influenced the yield of *I. batatas* ( $r > 0.3$ ), while main vine length and number of leaves negatively affected the total yield ( $r > -0.3$ ). Therefore, understanding the determinants of *I. batatas* yield provides practical guidance for breeding programs. Accessions *TJ19*, *TJ11*, *TJ1* and *TJ20* were the best-yielding genotypes for poly sack cultivation. Further studies must be carried out to control the vegetative growth of them to maximize storage root yield. Additionally, more *I. batatas* accessions need to be collected and evaluated to exploit the advantage of *I. batatas* germplasm and yield performances must be evaluated under different agro-ecological zones.

## REFERENCES

- Abed, N. J., Oselebe, H. O., Chukwu, S. C., & Mourtala, I. Z. M. (2023). Assessment of the magnitude of genetic diversity among sweetpotato (*Ipomoea batatas* (L.) Lam) accessions collected from Nigeria and Niger based on agro-morphological and nutritional attributes. <https://doi.org/10.21203/rs.3.rs-3121316/v1>
- Arutselvan, R., Raja, K., Pati, K., Chauhan, V.B.S., & Nedunchezhiyan, M. (2023). Tuber Crops and their Potential in Food and Nutritional Security. *Pandemics and Innovative Food Systems*, 122-136. <http://dx.doi.org/10.1201/9781003191223-7>
- Badgley, C., Moghtader, J., Quintero, E., Zakem, E., Chappell, M.J., Aviles-Vazquez, K., Samulon, A., & Perfecto, I. (2007). Organic agriculture and the global food supply. *Renewable agriculture and food systems*, 22(2), 86-108. <https://doi.org/10.1017/S1742170507001640>
- Bashaasha, B., Mwanga, R.O.M., Ocitti p'Obwoya, C., & Ewell, P.T. (1995). Sweet potato in the farming and food systems of Uganda: A farm survey report. *International Potato Center (CIP), Nairobi, Kenya and National Agricultural Research Organization (NARO)*, Kampala, Uganda, 63.
- Bararyenya, A., Tukamuhabwa, P., Gibson, P., Grüneberg, W., Ssali, R., Low, J., Odong, T., Ochwo-Ssemakula, M., Talwana, H., Mwila, N., & Mwanga, R. (2020). Continuous Storage Root Formation and Bulking in Sweet potato. *Gates Open Research*, 9(3), 83. DOI: 10.12688/gatesopenres.12895.4.
- Chakrabarti, A., Kumari, R., Dey, A., & Bhatt, B.P. (2014). Sweet potato – an excellent source of livestock feed. *Krishi Sewa*. <https://www.krishisewa.com/articles/production-technology/391-sweet-potato.html>. Retrieved on September 27, 2023.
- De Silva, K. P. U., & Premathilake, A. (1995). Screening of sweet potato genotypes in rice fallow environments. Incorporation of users' criteria in variety development of sweet potato. *Selected Research Paper*, 1994/95( 2). <https://books.google.lk/books?id=yJ6SOTpGnw0C&printsec=frontcover>.
- De Silva, U. (2020). Desiya Ala Boga: *Sweet Potato*. Department of Agriculture Sri Lanka publication. 112-167. [https://drive.google.com/file/d/1L1AUUAVvUst8GYCGFgFYdrwYLRmzzDS\\_/view?usp=sharing](https://drive.google.com/file/d/1L1AUUAVvUst8GYCGFgFYdrwYLRmzzDS_/view?usp=sharing).
- Department of Agriculture. (2021). Sweet potato (*Ipomoea* L.). <https://doa.gov.lk/hordi-crop-sweet-potato/>
- Dubbeling, M., & Massonneau, E. (2014). Rooftop agriculture in a climate change perspective. *Urban Agriculture Magazine*, 27, 28-32.
- Essilfie, M., Dapaah, H., Tevor, J., & Darkwa, K. (2016). Number of nodes and part of vine cutting effect on the growth and yield of sweet potato (*Ipomoea batatas* (L.) in transitional zone of Ghana. *International Journal of Plant & Soil Science*, 9(5), 1-14. <http://dx.doi.org/10.9734/IJPSS/2016/22776>
- Goswami, S. B., Sen, H., & Jana, P. K. (1995). Tuberization and yield potential of sweet potato cultivars as influenced by water management practices. *Journal of Root Crops*, 21(2), 77-81. [http://dx.doi.org/10.1007/978-1-4020-9475-0\\_17](http://dx.doi.org/10.1007/978-1-4020-9475-0_17).
- Ghosh, S. P., Ramanujan, T, Jos, J. S., Moorthy, S. N., & Nair, R. G. (1988). Tuber Crops. *New Delhi: Oxford IBH Publishing*; 146. <https://doi.org/10.1371/annotation/936fe9b4-41cb-494d-87a3-a6d9a37c6c68>
- Huaman, Z. (1999). Sweet potato Germplasm Management (*Ipomoea batatas*). *Training manual*, 218. <http://www.sweetpotatoknowledge.org/wp-content/uploads/2016/04/Sweetpotato-Germplasm-Management-IPomoea-batatas-Training-Manual.pdf>.
- IBM Corp. Released. (2012). IBM SPSS Statistics for Windows, Version 25.0.
- Johnson, A. C., & Gurr, G. M. (2016). Invertebrate pests and diseases of sweet potato (*Ipomoea batatas*): a review and identification of research priorities for smallholder production. *Annals of Applied Biology*, 168(3), 291-320. <https://doi.org/10.1111/aab.12265>
- Kingsbury, J. M. (1992). Christopher Columbus as a botanist. *Arnoldia*, 52(2), 11-28.

- Kuswantoro, H. (2017). Genetic variability and heritability of acid-adaptive soybean promising lines. *Biodiversitas Journal of Biological Diversity*, 18(1). <https://doi.org/10.13057/biodiv/d180148>
- Lamaro, G. P. (2020). Acclimatization and performance evaluations of sweet potato (*Ipomoea batatas* L. Lam) on yield and yield components in different agro-ecologies of Northern Ethiopia. *Journal of Dryland Agriculture*, 6(5), 36-45. <https://doi.org/10.5897/JODA2019.0036>
- Lebot, V. (2009). Crop production science in horticulture 17, *Tropical root and tuber crops cassava, sweet potato, yams and aroids*. <http://books.google.com/books?id=rFwyrKRSMUMC>.
- Lestari, M.W. (2021). Yield determinants of some sweet potato (*Ipomoea batatas* L.) clones using principal component analysis. *Earth and Environmental Science*, 828(1). DOI 10.1088/1755-1315/828/1/012006
- Mahaman Mourtala, I. Z., Gouda, A. C., Baina, D. J., Maxwell, N. I. I., Adje, C. O., Baragé, M., & Happiness, O. O. (2025). Genetic diversity and population structure studies of West African sweetpotato [*Ipomoea batatas* (L.) ] collection using DArTseq. *PLoS one*, 20(1). <https://doi.org/10.1371/journal.pone.0312384>
- Mau, Y. S., Ndiwa, A. S., Arsa, I. G. A., Asa, G. V., Nana, A., Londingkene, J. A., Hosang, E. Y., & Kotta, N. R. E. (2022). Assessment of genetic diversity and characterization of distinctness, uniformity, and stability of newly bred sweet potato clones. *Biodiversitas Journal of Biological Diversity*, 23(11). <https://doi.org/10.13057/biodiv/d231146>
- Mohanraj, R., & Sivasankar, S. (2014). Sweet Potato (*Ipomoea batatas* [L.] Lam)-A valuable medicinal food: A review. *Journal of medicinal food*, 17(7), 733-741. <https://doi.org/10.1089/jmf.2013.2818>
- Motsa, N. M., Modi, A. T., & Mabhaudhi, T. (2015). Sweet potato (*Ipomoea batatas* L.) as a drought tolerant and food security crop. *South African Journal of Science*, 111(11-12), 1-8. <http://dx.doi.org/10.17159/sajs.2015/20140252>
- Mwanga, R.O., Swanckaert, J., da Silva Pereira, G., Andrade, M. I., Makunde, G., Grüneberg, W. J., Kreuze, J., David, M., De Boeck, B., Carey, E., & Ssali, R. T. (2021). Breeding progress for vitamin A, Iron and Zinc biofortification, drought tolerance, and sweet potato virus disease resistance in sweet potato. *Frontiers in Sustainable Food Systems*, 5. <https://doi.org/10.3389/fsufs.2021.616674>
- Narayan, A., Dileep, K., Alam, T., Singh, R. S., Giri, G. S., Prasad, R., & Singh, P. P. (2022). Genetic diversity in sweet potato (*Ipomoea batatas* L.). *The Pharma Innovation Journal*, 11(6), 2352-2355. <http://www.thepharmajournal.com/>
- Naskar, S. K., Mukherjee, A., Nedunchezhiyan, M., & Rao, K. R. (2008). Evaluation of sweet potato cultivars for quality traits. In: *New R and D Initiatives in Horticulture for Accelerated Growth and Prosperity, 3rd Indian Horticultural Congress*. <https://www.ijcmas.com/7-8-2018/Venkatraman%20Bansode,%20et%20al.pdf>
- Ndwamato, M., Yvonne, M., Given, S., & Pontsho, T. (2022). Effect of different types of sweet potato (*Ipomoea batatas*) cultivars on growth performance in woven polypropylene plastic bags. *Acta Agriculturae Scandinavica, Section B—Soil & Plant Science*, 72(1), 885-892. <https://doi.org/10.1080/09064710.2022.2117080>
- Nedunchezhiyan, M., & Ray, R. C. (2010). Sweet potato growth, development production and utilization: overview. Sweet potato: post-harvest aspects in food. *Nova Science Publishers Inc., New York*, 1-26. <http://www.novapublishers.com/>
- Ngailo, S., Shimelis, H., Sibiya, J., & Mtunda, K. (2013). Sweet potato breeding for resistance to sweet potato virus disease and improved yield: Progress and challenges. *African Journal of Agricultural Research*, 8(25), 3202-3215. DOI: 10.5897/AJAR12.1991
- Palaniswami, M. S., & Peter, K. V. (2008). Tuber & root crops. *New India Publishing*, 9.
- Panggabean, F. D. M., Mawarni, L., & Chairunnissa, T. (2014). Respons pertumbuhan dan produksi bengkuang (*Pachyrhizus erosus* (L.) Urban) terhadap waktu pemangkasan dan jarak tanam. *Jurnal Online Agroekotek*, 2(2), 702-71.
- Pascal, P., & Mwende, E. (2009). A garden in a sack: Experiences in Kibera, Nairobi. *Urban Agriculture Management*, 21, 38-40. <http://www.ruaf.org/>
- Pearson, L. J., Pearson, L., & Pearson, C. J. (2010). Sustainable urban agriculture: stock take and opportunities. *International journal of agricultural sustainability*, 8(1-2), 7-19. <http://dx.doi.org/10.3763/ijas.2009.0468>
- Placide, R., Shimelis, H., Laing, M., & Gahakwa, D. (2015). Application of principal component analysis to yield and yield related traits to identify sweet potato breeding parents. *Tropical Agriculture*, 92(1). <https://journals.sta.uwi.edu/ojs/index.php/ta/article/view/885>
- Premathilake, P. S. A. D. (1999). Sweet potato Germplasm Conservation and Breeding in the Department of Agriculture, Sri Lanka. *Michael Hermann, editors*, 2001, 6. <https://www.cabidigitallibrary.org/doi/pdf/10.5555/20073172905>
- Raemaekers, R. H. (2001). Crop Production in Tropical Africa. *Directorate General for International Cooperation (DGIC)*.
- Rahajeng, W., & Rahayuningsih, S.T. A. (2017). Evaluation of orange-fleshed sweet potato genotypes for yield and yield contributing parameters in two environments. *Nusantara Bioscience*, 9(3), 275-281. <https://doi.org/10.13057/nusbiosci/n090306>
- Rahajeng, W., Restuono, J., & Indriani, F. C. (2018). Assessment of Diversity in Sweet potato Accession using Quantitative Traits by Clusters Analysis Method. *Earth and Environmental Science*, 197(1). DOI 10.1088/1755-1315/197/1/012035
- Rahajeng, W., Restuono, J., & Indriani, F. C. (2023). The Relationship of Sweet Potato Germplasm Based on Morphological Characters. *Journal Biodiversity*, 8(1), 94-105. DOI: 10.15575/biodjati.v8i1.19331
- Ravi, V., & Indira, P. (1996). Investigation on the physiological factors limiting yield potential in sweet potato under drought stress. *Annual Report Trivandrum, India*. <https://scholar.google.com/scholar?oi=bibs&cluster=4847162939426448784&btnI=1&hl=en>

- Ravi, V., & Indira, P. (1998). Crop physiology of sweet potato. *Horticulture Reviews*, 23, 277-339.
- Ravi, V., Chakrabarti, S. K., Makesh Kumar, T., & Saravanan, R. (2014). Molecular regulation of storage root formation and development in sweet potato. *Horticultural Reviews*, 42, 157-208. <https://doi.org/10.1002/9781118916827.ch03>
- Ravi, V., & Saravanan, R. (2012). Crop physiology of sweet potato. *Fruit, Vegetable and Cereal Science and Biotechnology*, 6, 17-29.
- Roullier, C., Duputié, A., Wennekes, P., Benoit, L., Fernández Bringas, V. M., Rossel, G., Tay, D., McKey, D., & Lebot, V. (2013). Disentangling the origins of cultivated sweet potato (*Ipomoea batatas* (L.) Lam.). *PLoS One*, 8(5), e62707. <https://doi.org/10.1371/journal.pone.0062707>
- Satapathy, M. R., Sen, H., Chattopadhyay, A., & Mohapatra, B.K. (2005). Dry matter accumulation, growth rate and yield of sweet potato cultivars as influenced by levels of nitrogen and cutting management. *Journal of Root Crops*, 31, 129-132. <https://www.cabidigitallibrary.org/action/doSearch?do=Journal+of+Root+Crops>
- Sakaigaichi, T., Terajima, Y., Suematsu, K., Kamada, E., Kobayashi, A., & Kawata, Y. (2023). Analysis of sweet potato shoot traits diversity and its relationship with storage root yield under short-period cultivation. *Genetic Resources and Crop Evolution*, 1-15. <https://doi.org/10.1007/s10722-023-01633-5>
- Solomon, E. (1999). Yield performance of sweet potato cultivars at Werer under irrigation, *AgriTopia quarterly news letter of Ethiopian Agricultural Research Organization (EARO)*, Addis Ababa. 2-14.
- Tixier, P., & De Bon, H. (2006). Urban Horticulture. In: van Veenhuizen R (ed) Cities farming for the future. *Urban agriculture for sustainable cities*, 313-346. <https://worldveg.tind.io/record/5684/>
- United Nations. (2020) FAO statistics. <https://www.fao.org/faostat/en/#home>
- Vilochani, P. M., Mohotti, A. J., Ekanayake, E. M. D. S., & Wasala, S. K. (2011). Morphological Characterization of Sweet Potato [*Ipomoea batatas* (L.) Lam] Accessions. <http://dlib.pdn.ac.lk/handle/1/6531>
- Widaryanto, E., & Saitama, A. (2017). Research Article Analysis of Plant Growth of Ten Varieties of Sweet Potato (*Ipomoea batatas* L.) Cultivated in Rainy Season. *Asian Journal of Plant Science*, 16, 193-199. <https://scialert.net/qredirect.php?doi=ajps.2017.193.199&linkid=pdf>,
- Woolfe, J. A. (1992). Sweet potato: an untapped food resource. *Cambridge University Press*. [https://books.google.lk/books?id=\\_MWmIDzNMSYC&printsec=frontcover](https://books.google.lk/books?id=_MWmIDzNMSYC&printsec=frontcover)



## Supplementary Figures



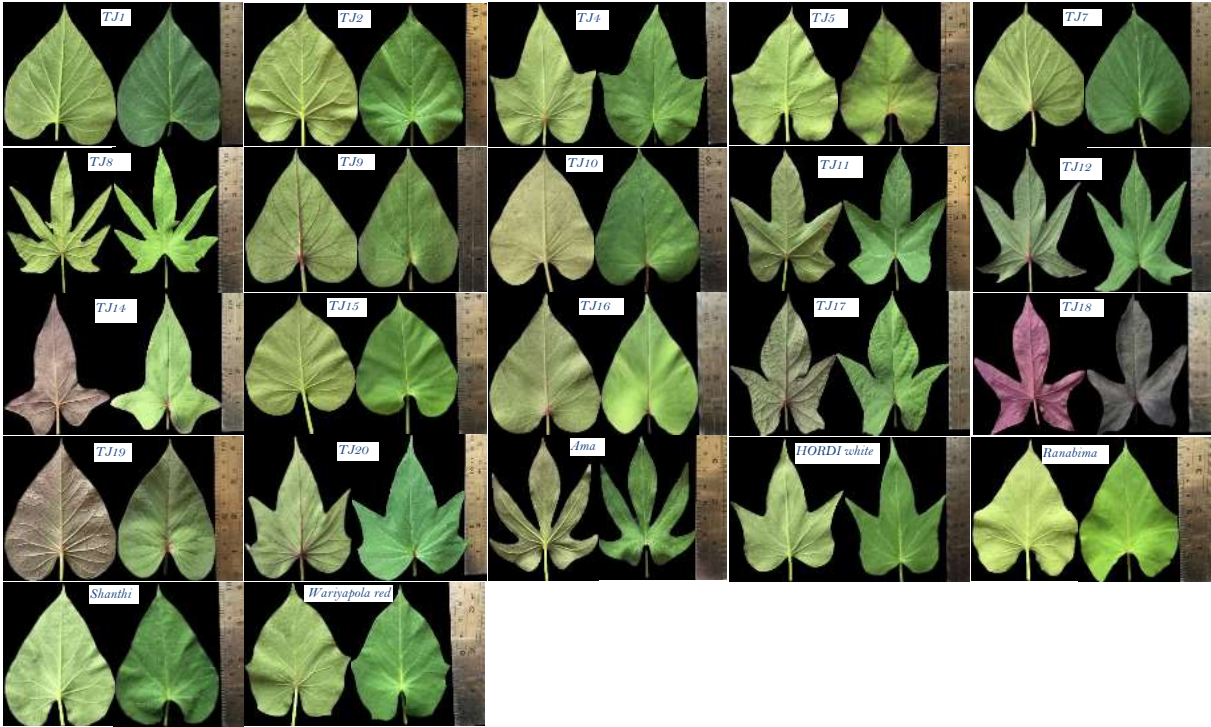
(1A)



(1B)

**Supplementary Figure 1:** (1A) Field established sweet potato genotypes.  
(1B) Poly sack cultivated sweet potato genotypes collected for the photograph.



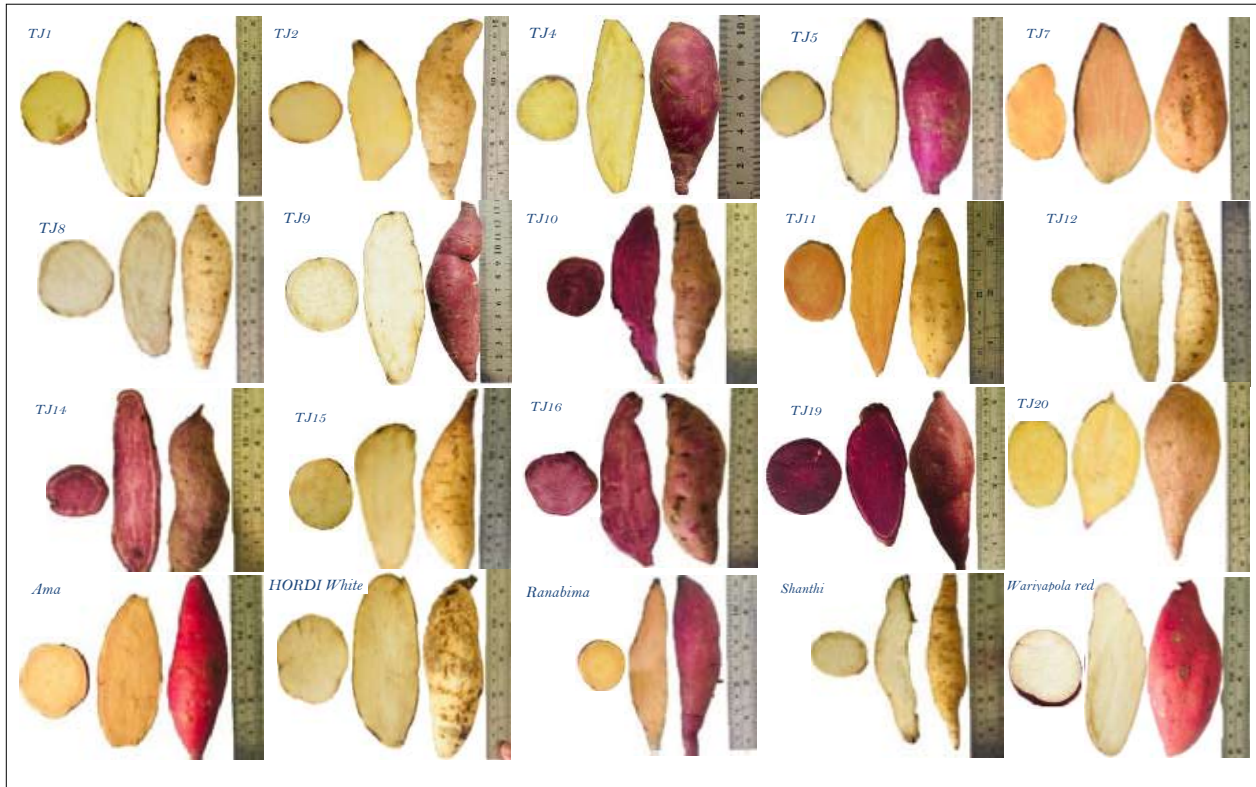


Supplementary Figure 2: Leaf characteristics of studied *I. batatas* genotypes



Supplementary Figure 3: Canopy appearance of studied *I. batatas* genotypes





**Supplementary Figure 4:** Qualitative storage root characteristics of *I. batatas* genotypes

## RESEARCH ARTICLE

### Machine Learning

# Machine learning methods to classify engineering students' ability to perform recreational archery activities using biomechanical data

NRH Basri<sup>1,2</sup>, MS Mohktar<sup>1,2\*</sup> and WNLW Mahadi<sup>3</sup>

<sup>1</sup> Centre for Innovation in Medical Engineering, Department of Biomedical Engineering, Faculty of Engineering, Universiti Malaya, 50603 Kuala Lumpur, Malaysia.

<sup>2</sup> Universiti Malaya Science, Technology, Engineering and Mathematics (STEM) Centre, Universiti Malaya, 50603 Kuala Lumpur, Malaysia.

<sup>3</sup> Department of Electrical Engineering, Faculty of Engineering, Universiti Malaya, 50603 Kuala Lumpur, Malaysia.

Submitted: 02 September 2024; Revised: 19 February 2025; Accepted: 25 February 2025

**Abstract:** Utilizing machine learning approaches to predict students' performance is a valuable tool for anticipating both low and high levels of achievement across different educational levels. In the case of engineering programmes, the students must exhibit proficient hands-on abilities that match their knowledge, to excel in their performance. When assessing the psychomotor abilities of engineering students during hands-on, practical, or laboratory tasks, the available assessment tools are limited. Utilizing both biomechanical data and observations of students' performance provides an alternate method for evaluating psychomotor skills. This study introduced a recreational archery game as a proof of concept to examine the impact of engineering students' biomechanical data-the upper-limb physical attributes, posture, and angle-on their performance as measured by the points they score in the archery activity. We recruited 104 engineering students without prior experience performing recreational archery and conducted the activity in laboratory settings. We developed a classification model combining fourteen predictors to categorize the students into two groups: those able to perform and those who are not. The model was run into three different machine learning classification algorithms including the decision tree, K-nearest neighbour, and support vector machine (SVM). Among all preset models, SVM shows the best accuracy, and then further optimization of the SVM model was conducted to achieve higher prediction model accuracy. The optimization of the SVM model increased the testing accuracy to 80 % with F1 scores of 0.75 and 0.83 in predicting the 'not able to perform' and 'able to perform' classes respectively. Thus, this study shows that a


combination of biomechanical data and machine learning is a potential tool for assessing engineering students' performance involving their psychomotor abilities.

**Keywords:** Activity performance, biomechanical analysis, machine learning, recreational archery.

## INTRODUCTION

Engineering students must graduate with commendable qualities to become exceptional engineers. In engineering education, three fundamental domains of Bloom's Taxonomy must be implemented. Bloom's Taxonomy delineates the domains of learning as the cultivation of cognitive, emotional, and psychomotor skills. The majority of students' cognitive abilities are cultivated through classroom instruction. The affective skills component, which pertains to the enhancement of emotions or attitudes, is cultivated through collaborative efforts. Students' psychomotor skills, often termed manual or physical skills, are typically cultivated in a laboratory environment (Baharom *et al.*, 2015a).

This study is a critical area of investigation as it focuses on the psychomotor abilities of engineering students to provide insights into how engineering students learn and perform hands-on tasks. This is an

\* Corresponding author (mas\_dayana@um.edu.my;  <https://orcid.org/0000-0001-7859-4396>)



This article is published under the Creative Commons CC-BY-ND License (<http://creativecommons.org/licenses/by-nd/4.0/>). This license permits use, distribution and reproduction, commercial and non-commercial, provided that the original work is properly cited and is not changed in anyway.

innovative area where engineering education intersects with cognitive science and human factors, leading to better educational strategies and technologies to enhance learning. The motivation of the study is to investigate the hands-on skill development of the students, bridging the gap between theory and practice, and addressing diverse learning styles. Investigating psychomotor skills taps into the potential of students who may learn better by doing rather than observing, allowing for more inclusive and effective teaching approaches.

Psychomotor abilities involve fine and gross motor skills, coordination, and precise control over tools and equipment. Investigating and enhancing these abilities helps students perform real-world tasks efficiently, such as operating machinery, prototyping, and performing technical repairs. The psychomotor domain consists of four main elements to test: physical (posture), motor, fitness, and play (sports skills). The psychomotor domain in learning refers to the set of skills that include the use and coordination of skeletal muscles, such as performing, manipulating, and creating. Referring to the Kemp model, four elements of design are an integral part of course development: students, objectives, methods, and evaluation. In our study, we focus on the element of subject content identification with task analysis related to goals and purpose (Baturay, 2008; Kusrini, 2018; Birgili, 2019).

Particularly because of the remote learning environment during the COVID-19 crisis, there has been a growing focus on students' psychomotor skills, primarily due to the absence of hands-on activities in practical courses during that period. Due to the practice-oriented components of the accreditation standards and to meet industry demands, the graduate engineers must have strong psychomotor skills in sync with their knowledge level. The evaluation and measurement of psychomotor skills is an important assessment for engineering courses, technical courses, practical courses, hands-on courses, and laboratory courses. Conventional assessment methods, such as reports, observation, and tests, are still being used to assess student's practical and hands-on skills. However, these methods heavily emphasize evaluating the student's cognitive skills (Weidner & Popp, 2007; Salim *et al.*, 2012; Baharom *et al.*, 2015b; Viscione *et al.*, 2017; Isa *et al.*, 2019). The assessment tools are limited, particularly when it comes to accessing technical and engineering students' psychomotor skills during hands-on, practical, or laboratory work. Many previous studies that worked on engineering students' psychomotor skills still relied on rubrics assessment. Assessing the motor performance during the hands-on,

biomechanical assessment of the student's posture and limb angles when performing the task is important to determine the effectiveness of learning to produce good outcomes.

This study identified skills-related activities and analyzed the coordination aspects of how engineering students use their senses and body parts to perform a motor activity or task smoothly and produce accurate outcomes. We designed a recreational archery activity to analyze the biomechanical parameters of the students, which are the upper limb physical posture and angles, thereby determining the performance level of the engineering students performing the task. Through the activity, the study aims to integrate an alternative to provide supporting quantitative data of effective learning on the engineering student's performance during recreational archery.

The posture and angle analysis during the hands-on recreational archery were performed by capturing the student's 2D motion using video recording and proceeding to analyses using a video annotation tool designed for sport analysis called Kinovea. The said software was chosen as it was an easily accessible tool that can be used free without subscription by educators, especially in engineering, to investigate more inclusive and effective learning methods. Kinovea is also a very simple and easily operated tool, and thus has been used in many fields before, especially sports science. The advantage of using Kinovea in this study when analysing the motor performance of the students is that it can be done in a class setting rather than conducting the activity in a laboratory with a fixed camera setting such as the Vicon system. This method is more convenient for educators as they can set up the camera settings according to their preferences and suitability to the activity they have chosen to observe (Nor Adnan *et al.*, 2018; Lau *et al.*, 2023).

Besides using a 2D motion analysis for motor performance, the study also applied the usage of educational data mining (EDM) in machine learning such as prediction models that have capabilities to inform and guide instructional practices, examine the efficiency and effectiveness of learning, provide meaningful feedback for teachers and learners, and formulate strategies for the evaluation of students, modifying learning environments to improve students' learning outcomes based on the early prediction of their performance. Through EDM, students' data will be analyzed to determine significant patterns that can improve students' knowledge, skills and overall institutional performance in general (Lee &

Kang, 2024; Pillitteri *et al.*, 2024). Lastly, this technique can provide early identification of future challenges or industry demand for special skills especially for engineering graduates (Ofori *et al.*, 2020; Albreiki *et al.*, 2021).

The decision tree (Rotgans & Schmidt) machine learning method operates by systematically partitioning a dataset into subsets based on a specific number of splits. At each stage of the decision-making process, the algorithm selects the most discriminative feature to split the dataset, aiming to maximize the homogeneity of the resulting subsets in terms of performance levels (Song & Ying, 2015; Charbuty & Abdulazeez, 2021; Panigrahi *et al.*, 2021). KNN classifiers provide a nuanced and intuitive framework for assessing subjects' performance levels by drawing on the concept of similarity among individuals. The K-nearest neighbour (KNN) algorithm operates on the premise that subjects with similar attributes are likely to exhibit comparable performance. KNN is particularly appealing in various research, as it has simplicity and flexibility (Johnson & Yadav, 2018; Ahmed *et al.*, 2020). This technique of classification can be applied to different types of distance metrics such as Euclidean, Cosine, Cubic, and Minkowski. Whereas support vector machine (SVM) is particularly popular for solving classification problems due to its ability to find a clear decision boundary between classes, especially in this study we want to classify the students into two different levels of performance, which are 'Not able to perform' and 'Able to perform' (Naicker *et al.*, 2020; Zulfiker *et al.*, 2020).

SVM can effectively handle large and complex datasets with many features, such as those collected from student performance and engagement tracking systems. With the kernel function, SVM can capture nonlinear relationships between input features and output performance levels. SVM models tend to generalize well to new, unseen data, which is important when predicting the future performance of students. In some cases, student datasets may be small, and SVM can still perform well when trained on limited samples (Bhavsar & Panchal, 2012; Cervantes *et al.*, 2020; Naicker *et al.*, 2020). SVM can be used to predict student performance based on historical data. In this study, a binary classification of whether a student is able or not able to perform can serve as the target labels, while features like demographic data such as gender, anthropometric measurements, and upper limb's joint angle were used as inputs.

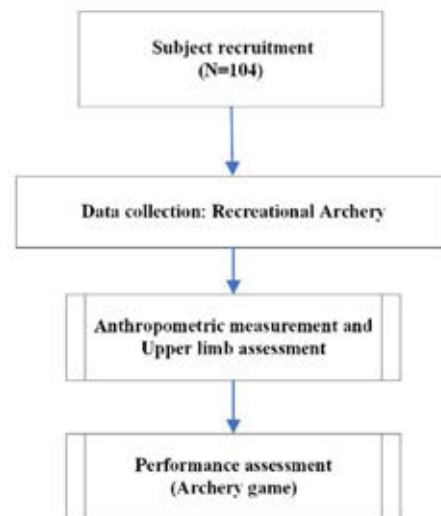
Thus, this study aims to provide classification

models on the engineering student's performance during recreational archery activity using 2D motion analysis and machine learning.

## MATERIALS AND METHODS

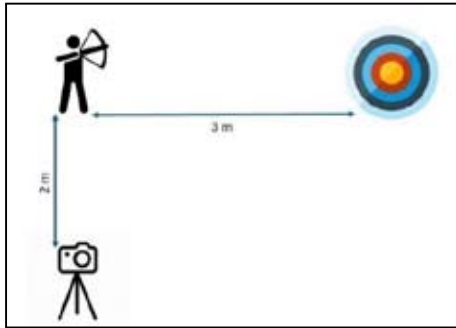
### Subject recruitment and study procedure

The study recruited a total of 104 subjects (N = 104; 42 males and 62 females). The criteria for recruiting subjects included their enrolment as engineering students, their age between 18 and 22 years, their ability to participate in physical activities in class and no prior experience in performing archery games. We collected written consent from each subject and conducted a briefing on the study procedure before the activity. There is no restriction on any of the subjects requesting to withdraw while the study is still ongoing.



**Figure 1:** Flow of the study data collection.

Figure 1 shows the flow of the study. The data collection was conducted with 1st and 2nd-year students from the biomedical engineering department. In this study, the archery game was selected as it can provide wide upper limb movement and immediate performance outcomes based on their scores on the archery target. Each subject was tasked with shooting nine arrows at a distance of 3 m. This task was recorded with a video camera. The setup of the camera is shown in Figure 2. Further analysis of the motion of the upper limb of the subjects during shooting was done in Kinovea (version 0.9.5, 2021).



**Figure 2:** Schematic diagram of archery activity settings in Human Centric Laboratory.

### Anthropometric measurements and upper limb assessment

Each subject underwent anthropometric measurements to determine their weight, height, body mass index (BMI), and arm muscle grade. Their upper limb length was also measured, such as shoulder to elbow, elbow to wrist, arm, hand, and trunk (shoulder to waist) length. The measurements were repeated three times, and the average was recorded on the anthropometric measurement form. The strength of the upper limbs of the students was also evaluated using the Oxford Scale (AKA Medical Research Council Manual Muscle Testing Scale). This method involves testing key muscles from the upper extremities against the examiner's resistance and grading the subject's strength on a 0 to 5 scale accordingly: 1) flicker of movement; 2) through full range actively with gravity counterbalanced; 3) through full range actively against gravity; 4) through full range actively against some resistance; and 5) through full range actively against strong resistance.

### Recreational archery performance

The activity setting was set where the archery target was put on the wall, and a 3-meter marker was determined to fix the distance between the subject and the target. Video cameras (Samsung Galaxy Tab S5e, 1080p HD Webcam) were positioned as in Figure 2, where the height was fixed perpendicular to the subject's centre and lateral to the subject's right-side position to capture their movement during the archery shooting. The right shoulder, elbow, wrist, and midsection of the participant was marked with tape prior to the commencement of the task as in Figure 3. This is important for later steps in analyzing the angles of the archer's wrist, elbow, shoulder, and trunk while performing archery. Throughout the task, the camera recorded footage of the archer.



**Figure 3:** Image view on the recording camera. Markers were placed on the shoulder, elbow, wrist and waist. Additional markers were placed during Kinovea analysis on the neck and knee. In the study, all participants were required to draw the bowstrings using their right arm.

The archery activity requires shooting in three separate sets of trials. During the task, each subject must draw a total of nine bows at a distance of three meters. The study used a bow and arrow playset (material: plastic, bow size: 64 cm x 21 cm x 3.5 cm, arrow length: 56 cm, target diameter: 20 cm, range distance: about 5 m). The target area has marks on it ranging from miss as '0' to '10' points. The subjects' scores for each attempt were recorded and calculated, and the maximum total score of the three sets was 90 points. The total marks for those three sets will be the outcome of the performance assessment. Table 1 lists the indicators for the subjects' performance levels.

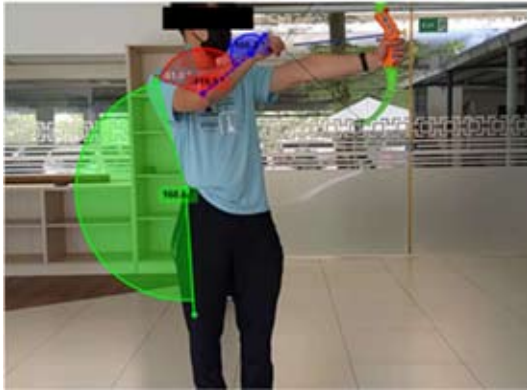
**Table 1:** Performance level of subjects based on their total archery score (Response).

Performance indicators	Total archery score
Able to perform	$\geq 10$
Not able to perform	0 – 9

In supporting the recreational archery performance scores, the upper limb angles were analyzed. For this analysis, the study uses Kinovea software, as it is easier to observe frame-by-frame motion and capture 2-dimensional kinematic angles at a specific frame. The software was feasible to use and a low-cost technology used often by researchers in biomechanics research



(Fernández-González *et al.*, 2020; Hisham *et al.*, 2017; Sharifnezhad *et al.*, 2021). The angles of the wrist (blue), elbow (red), shoulder (turquoise), and trunk (green) of the subject when they flexed or extended from each attempt were analyzed and recorded. Figure 4 shows four angles captured during the phase when the subjects were about to release the bow's arrow.



**Figure 4:** The image from Kinovea software where the video was imported and analysed to get the angles of the trunk, shoulder, elbow and wrist of their right hand when performing the task.

**Table 2:** Input predictors and response for the classification analysis.

Classification items	Details
Input predictor	Gender, weight, height, BMI, shoulder to elbow, elbow to wrist, hand, trunk and arm lengths, arm muscle grade, trunk, shoulder, elbow and wrist angles.
Response classes	2
Response class Names (Table 1)	Not able to perform (0), Able to perform (1)
Validation (Training and Testing)	10-fold cross-validation

### Student performance level classification model design

Machine learning classifiers have been adapted previously by researchers for classification purposes on students' performance (Araya *et al.*, 2022; Hashim *et al.*, 2020; Martínez-Carrascal *et al.*, 2020; Pallathadka *et al.*, 2023). In this step of the analysis, we will compare and

choose the best classifier that is most suitable to classify the students' performance levels using MATLAB 2024b software with the classification learner toolbox. The study chooses to compare three different classification algorithms which are DT (Rotgans & Schmidt, 2011), KNN and SVM.

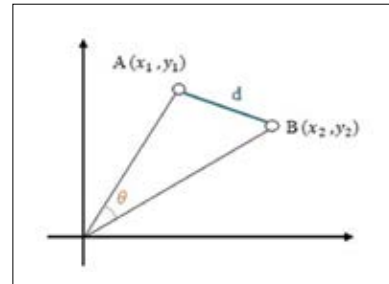
For DT, the split criterion was set using Gini's diversity index with the respective maximum number of splits. The formula of the Gini index is as in Equation 1, where the  $p_i$  is the probability of students being classified into a particular class (1- Not able to perform, 2-Able to perform).

$$Gini\ index = 1 - \sum_{i=1}^2 (p_i)^2 \quad (...01)$$

For KNN, we trained the subject's data with several models with different distance metrics, such as Euclidean (actual distance between two points) and cosine (angle between two points) as displayed in Figure 5.

$$Euclidean = \sqrt{\sum_{i=1}^n (A_i - B_i)^2} \quad ... (02)$$

$$Cosine\ angle = \cos \theta = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2 \sum_{i=1}^n B_i^2}} \quad ... (03)$$



**Figure 5:** Example of Euclidean distance and Cosine similarity angle of two points.

The SVM classifier models of various Kernel functions, such as linear, quadratic, cubic, and Gaussian, were compared in the study. The most accurate classifiers for the students' datasets use the kernel function defined in Equation 4. SVM created a linear decision boundary called a hyperplane that separates the training subject's data sets into their respective classes. One reasonable choice as the best hyperplane is the one that represents the largest separation or margin between the two classes.

$$K(x_i, x_j) = x_i \cdot x_j \quad \dots(04)$$

The quality of predictions depends heavily on the input features. The predictors classify the subjects' performance based on whether they fall into the 'perform' or 'not perform' classes. The validation is set to 10-fold cross-validation, to protect against the model overfitting. The model's performance is determined by its accuracy, confusion matrix, and F1 scores. The confusion matrix provides each model's performance with an actual and predicted output, called true positive (TP), true negative (TN), false positive (FP), and false negative (FN). This output is later used to calculate precision (the number of correct positive results divided by the number of positive results predicted by the classifier) and recall (the number of correct positive results divided by the number of all samples that should have been identified as positive). Lastly, the F1 scores that determine the harmonic mean between precision and recall is calculated using Equation 7.

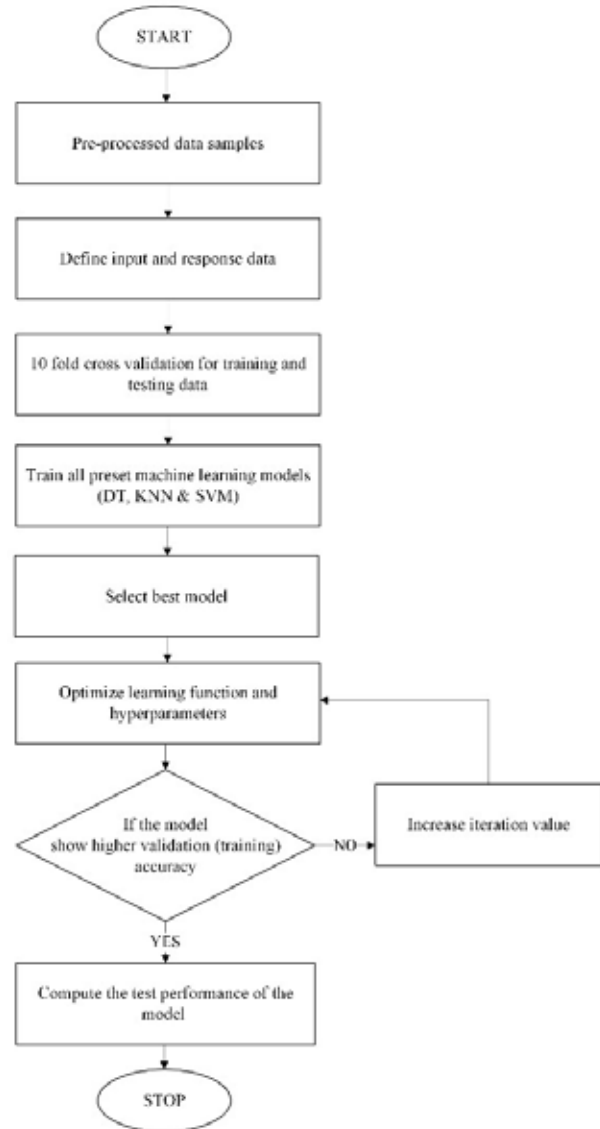
$$\text{Precision} = \frac{TP}{TP + FP} \quad \dots(05)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad \dots(06)$$

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad \dots(07)$$

To increase the model's accuracy, efficiency, and prediction ability, optimization of the classifiers was conducted, which involved fine-tuning key parameters and techniques. The optimization was run on the hyperparameters of the model and at each iteration, the model tried a different combination of hyperparameter values and plotted a graph with the minimum validation classification error observed up to that iteration. The hyperparameter of the SVM model was tuned in this study to optimize the performance of the model. The hyperparameters search range was set as follows; box constraint level: 0.001–1000, kernel scale: 0.001–1000, kernel function: linear, quadratic, cubic, and gaussian. To see whether a better result can be produced, it is advisable to run the optimization longer, therefore the iteration value was increased, in this study to 30.

**Ethical considerations:** The Universiti Malaya Research Ethics Committee has approved the study protocol under reference number UM.TNC2/UMREC\_3264.



**Figure 6:** Flowchart of the classifications using preset models and optimization of the models.

## RESULTS AND DISCUSSION

Recruited subjects consist of 40.4 % males and 61.5 % females. The demographic data of the subjects were tabulated in Table 3. The average height and weight of both male and female subjects slightly differ however their average BMI stay in the same range of healthy weight status between the value of 21 to 22. The average arm length of both genders also displays



a slight difference, not exceeding 10 cm. The average trunk length of the male subjects was recorded as higher than female subjects, due to the influence of the height of male subjects. However, the average arm grade muscle recorded for both male and female subjects falls into the same range between 4 to 5 grade which suggested that all subjects supposedly could conduct activity with their arm without any difficulties.

**Table 3:** Demographic data consists of no. of subjects, gender, and anthropometric measurements. Results are mean  $\pm$  SD (standard deviation).

Subjects' characteristics (N=104)	Mean $\pm$ SD	
Gender	Male, n = 42	Female, n = 62
Weight, kg	63.5 $\pm$ 10.4	54.38 $\pm$ 11.18
Height, cm	171 $\pm$ 6.24	159 $\pm$ 5.9
BMI, kg/m <sup>2</sup>	21.74 $\pm$ 3.19	21.54 $\pm$ 4.25
Shoulder to elbow length, cm	31.9 $\pm$ 3.8	28.5 $\pm$ 3.0
Elbow to wrist length, cm	26.4 $\pm$ 2.1	24.4 $\pm$ 1.7
Arm length, cm	58.3 $\pm$ 5.1	52.8 $\pm$ 4.1
Hand length, cm	18.5 $\pm$ 1.1	17.4 $\pm$ 0.9
Trunk length, cm	42.5 $\pm$ 4.6	36.8 $\pm$ 3.9
Arm muscle grade (0-5)	4.88 $\pm$ 0.22	4.67 $\pm$ 0.48
Trunk angle (degrees)	168 $\pm$ 5	168 $\pm$ 5.
Shoulder angle (degrees)	66 $\pm$ 28	91 $\pm$ 25
Elbow angle (degrees)	111 $\pm$ 30	81 $\pm$ 22
Wrist angle (degrees)	157 $\pm$ 11	160 $\pm$ 11
Total Archery Scores	16.02 $\pm$ 9.65	9.31 $\pm$ 8.49

### The performance level of students in recreational archery games

As stated in Table 1, the performance level of the subjects was determined by their total score in all archery sets.

Table 5 tabulates the number of students who achieved the score based on their performance level. The division of the subjects into two levels resulted in 52.88% of them achieving more than 10 total points, while 47.12% were unable to perform. The data aligns with the expectation that half of the students will be able to perform archery games for the first time.

**Table 4:** The subjects' performance is based on the total archery score of all three sets.

Performance level indicators	Total archery score	No. of subjects achieving the score	Percentages, %
Able to perform	$\geq 10$	55	52.88 %
Not able to perform	0 – 9	49	47.12 %

The analysis of the subjects' upper limb angles supported their performance in the archery activity. The angles of the trunk, shoulder, elbow, and wrist of the subjects' right hand were analyzed at the frame where they released the arrow. As the subjects quote that their posture and handling of the tools might be affecting their scores, Table 5 shows the average upper limb angles of different levels of performance. Table 5 shows that the posture of the body, indicated by referring to the trunk and upper limb angle, might suggest differences in their performance (Sarro *et al.*, 2021). The closer the trunk angle to 180 degrees (straight body posture), the higher the average total score achieved by the subjects. The shoulder must be angled within the range of 70 to 80 degrees when pulling the bowstring to achieve a favourable score on the target. Increasing the elbow angle and flexing the wrist inwards produced better total scores during the archery. A short-term coaching session may contribute positively to increasing their chances of getting better total scores thus improving their performance level (Beyaz *et al.*, 2024).

**Table 5:** Upper limb angles of different levels of performance

Subjects	Performance level	Angles (°)				Average total score
		Trunk	Shoulder	Elbow	Wrist	
n = 55	Able to Perform	169.08	75.41	103.09	157.27	19.50
n = 49	Not able to Perform	167.64	86.08	83.38	160.15	4.02

It is debatable whether gender affects the performance of the subjects as a previous study mentioned that gender does not influence determining one's archery skills (Destriani *et al.*, 2024). However, there was also a suggestion that females might perform poorly under pressure and there have been recent failures compared to males in archery games (Li & Zhao, 2024). In this study, the performance of male subjects was also better compared to female referring to their average total score achieved during the recreational archery activity. Therefore, we are considering gender as one of the predictors affecting their performance in our study. Besides that, several parameters such as arm length, trunk angle, and shoulder angle were determined by previous researchers to be affecting the scores of the archers. Previous studies by Noor and by Setiakarnawijaya *et al.* reported that, in addition to the arm muscle grade and strength, arm length significantly affected the accuracy of the archers' scores with a p-value of less than 0.05 (Setiakarnawijaya *et al.*, 2021; Noor, 2023). As both studies were investigating athletes, the athlete's anatomy is also very important; in their case, the arm length is very important for archery athletes. Their studies were focused on highlighting the importance of the physical aspect's contribution to determining archery achievement. Another parameter that appears

significant in previous studies was postural sway during shooting archery. This parameter was closely related to the position of the trunk during the phases of aiming and releasing the bow. A study by Zawi & Mohamed (2013) reported that the shooting score performance of skilled archers was significantly affected by the postural sway during the release phase at a p-value of less than 0.001. Reduction in postural sway meaning that the body trunk should be positioned towards the centre line suggesting an increasing trend of shooting scores. As archery mainly uses the upper limb to operate, a suitable shoulder position is important to minimize fluctuations in high-strength muscle performance. Several studies have shown that towards the releasing phase of archery, the position of the arm shoulder is the main key to achieving favourable results (Lin *et al.*, 2010; Kolayış & Ertan, 2016; Vendrame *et al.*, 2022).

### Students' performance on classification models

Classification with all fourteen measured parameters as input predictors was run in the preset DT, KNN and SVM models. The best of each classification algorithm was tabulated in Table 6.

**Table 6:** Comparison of the best models of each classification algorithm.

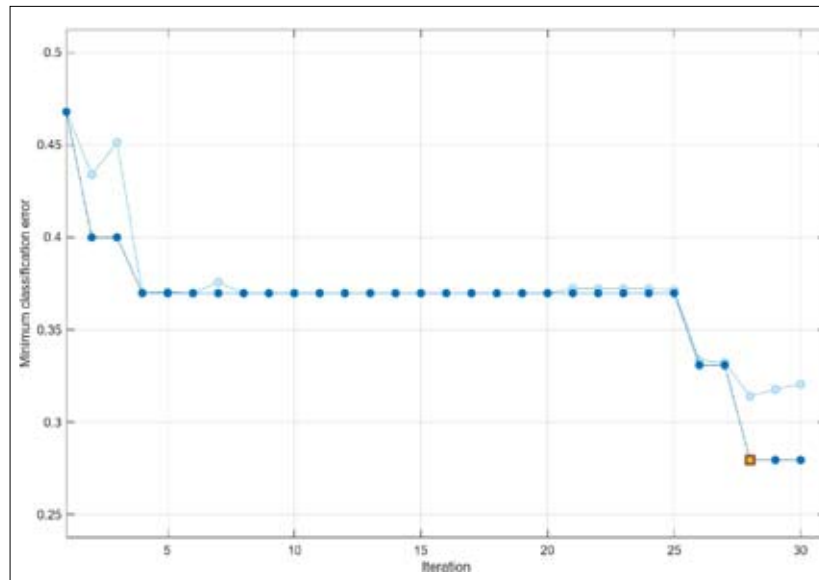
Models	Models performance			Model properties	
	Training accuracy (Validation) (%)	F1 score	Error rate (%)	Test accuracy (%)	Algorithm & hyperparameters
Model 1	50	0.50	50.0	70	<ul style="list-style-type: none"> <li>• Coarse DT</li> <li>• Maximum no of splits = 4</li> <li>• Split Criterion = Gini's diversity index</li> </ul>
Model 2	68.1	0.67	31.9	70	<ul style="list-style-type: none"> <li>• Medium KNN</li> <li>• No of neighbours = 10</li> <li>• Distance metric = Euclidean</li> <li>• Distance weight = Equal</li> </ul>
Model 3	69.2	0.69	30.9	70	<ul style="list-style-type: none"> <li>• Medium Gaussian SVM</li> <li>• Kernel function = Gaussian</li> <li>• Kernel scale = 3.7</li> <li>• Box constraint level = 1</li> </ul>

Out of all the algorithms compared, the SVM model produced the highest accuracy on validation and testing. To increase the performance of the SVM model, the optimization was run on the hyperparameters of the model and at each iteration, the model tried a different

combination of hyperparameter values and plotted a graph with the minimum validation classification error observed up to that iteration. To see whether a better result can be produced, it is advisable to run the optimization longer, therefore the iteration value was increased to 30.

From the SVM optimization, Model 4 gives a training (validation) accuracy of 72.3% and a test accuracy of 80%, where the best point hyperparameter was observed

in Figure 7 at the red square point with minimum classification error at 0.279, with kernel function: cubic and box constraint level at 0.00269.



**Figure 7:** Minimum classification error plot of Model 4 (Optimize SVM model).

**Table 7:** Per-class metrics of training (validation) and test sets of optimized SVM Model 4.

Per-class metrics	Model 4 (Optimize SVM)					
	Not able to perform class			Able to perform class		
	Precision	Recall	F1 score	Precision	Recall	F1 score
Training (Validation)	0.73	0.63	0.68	0.71	0.80	0.75
Test	1.00	0.60	0.75	0.71	1.00	0.83

As the training and testing accuracy of Model 4 increases after optimization, the performance in the ability to correctly classify the student's performance also increases. Table 7 tabulates the results of per-class metrics for Model 4. The F1 scores of Model 4 recorded the highest among all investigated models thus depicting the precision of the model in correctly classifying the students into their respective classes based on all fourteen predictors.

## CONCLUSION

The study's hypotheses lead to the conclusion that a student's biomechanical data, including physical and posture components like gender, arm muscle grade, arm length, shoulder angle, elbow angle, and trunk

angle, influences their ability to perform a physical task, specifically the recreational archery activity. Based on these parameters, the selected classifiers were able to classify the students into 'able to perform' and 'not able to perform' classes. In this study, after optimization, SVM classifiers displayed superior performance when applied to the student's data, with the highest accuracy and precise classification. Therefore, our study demonstrates that the integration of biomechanical data and machine learning has the potential to be a valuable tool for evaluating psychomotor abilities, particularly in the context of engineering students. Soon, to strengthen the assessment of engineering students' skills, we will further work on combining these parameters with the self-assessment of students' confidence levels before they perform the recreational archery task.

## Acknowledgement

The project is supported by Project Grant No: IF049-2021, GA025-2023, CORG004-2024

## REFERENCES

- Ahmed, M. R., Tahid, S. T. I., Mitu, N. A., Kundu, P., & Yeasmin, S. (2020). A comprehensive analysis on undergraduate student academic performance using feature selection techniques on classification algorithms. *11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*.
- Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student's performance prediction using machine learning techniques. *Education Sciences*, 11(9), 552.
- Araya, I., Beas, V., Stamulis, K., & Allende-Cid, H. (2022). Predicting student performance in computing courses based on programming behavior. *Computer Applications in Engineering Education*, 30(4), 1264-1276.
- Baharom, S., Khoiry, M. A., Hamid, R., Mutalib, A. A., & Hamzah, N. (2015a). Assessment of psychomotor domain in a problem-based concrete laboratory. *Journal of Engineering Science and Technology*, 10(1), 1-10.
- Baharom, S., Khoiry, M. A., Hamid, R., Mutalib, A. A., & Hamzah, N. (2015b). Assessment of psychomotor domain in a problem-based concrete laboratory. *Journal of Engineering Science and Technology*, 10(1), 1-10.
- Baturay, M. H. (2008). Characteristics of basic instructional design models. *Ekev Academic Review*, 12(34), 471-482.
- Beyaz, O., Eyraud, V., Demirhan, G., Akpınar, S., & Przybyla, A. (2024). Effects of short-term novice archery training on reaching movement performance and interlimb asymmetries. *Journal of Motor Behavior*, 56(1), 78-90.
- Bhavsar, H., & Panchal, M. H. (2012). A review on support vector machine for data classification. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 1(10), 185-189.
- Birgili, B. (2019). Comparative reflection on best known instructional design models: notes from the field. *Current Issues in Emerging eLearning*, 6(1), 5.
- Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189-215.
- Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(01), 20-28.
- Destriani, D., Yusfi, H., Destriana, D., Setyawan, H., García-Jiménez, J. V., Latino, F., Tafuri, F., Wijanarko, T., Kurniawan, A. W., & Anam, K. (2024). Results of beginner archery skills among adolescents based on gender review and shot distance. *Retos: Nuevas Tendencias en Educación Física, Deporte y Recreación*, (56), 887-894.
- Fernández-González, P., Koutsou, A., Cuesta-Gómez, A., Carratalá-Tejada, M., Miangolarra-Page, J. C., & Molina-Rueda, F. (2020). Reliability of kinovea® software and agreement with a three-dimensional motion system for gait analysis in healthy subjects. *Sensors*, 20(11), 3154.
- Hashim, A. S., Awadh, W. A., & Hamoud, A. K. (2020). Student performance prediction model based on supervised machine learning algorithms. *IOP conference series: materials science and engineering*.
- Hisham, N. A. H., Nazri, A. F. A., Madete, J., Herawati, L., & Mahmud, J. (2017). Measuring ankle angle and analysis of walking gait using kinovea.
- Isa, C. M. M., Joseph, E. O., Saman, H. M., Jan, J., Tahir, W., & Mukri, M. (2019). Attainment of program outcomes under psychomotor domain for civil engineering undergraduate students. *International Journal of Academic Research in Business and Social Sciences*, 9(13), 107-122.
- Johnson, J. M., & Yadav, A. (2018). Fault detection and classification technique for HVDC transmission lines using KNN. *Information and Communication Technology for Sustainable Development: Proceedings of ICT4SD 2016*.
- Kolayış, İ. E., & Ertan, H. (2016). Differences in activation patterns of shoulder girdle muscles in recurve archers. *Pamukkale Journal of Sport Sciences*, 7(1), 25-34.
- Kusrini, N. A. R. (2018). Comparative Theory on Three Instructional Designs: Dick and Carrey, Kemp, and Three Phases.
- Lau, J. S., Ghafar, R., Zulkifli, E. Z., Hashim, H. A., & Mat Sakim, H. A. (2023). Comparison of shooting time characteristics and shooting posture between high-and low-performance archers. *Annals of Applied Sport Science*, 11(2), 0-0.
- Lee, S., & Kang, M. (2024). A data-driven approach to predicting recreational activity participation using machine learning. *Research Quarterly for Exercise and Sport*, 1-13.
- Li, C., & Zhao, Y. (2024). Gender differences in skilled performance under failure stress.
- Lin, J.-J., Hung, C.-J., Yang, C.-C., Chen, H.-Y., Chou, F.-C., & Lu, T.-W. (2010). Activation and tremor of the shoulder muscles to the demands of an archery task. *Journal of sports sciences*, 28(4), 415-421.
- Martínez-Carrascal, J. A., Márquez Cebrián, D., Sancho-Vinuesa, T., & Valderrama, E. (2020). Impact of early activity on flipped classroom performance prediction: A case study for a first-year Engineering course. *Computer Applications in Engineering Education*, 28(3), 590-605.
- Naicker, N., Adeliyi, T., & Wing, J. (2020). Linear support vector machines for prediction of student performance in school-based education. *Mathematical Problems in Engineering*, 2020, 1-7.
- Noor, G. Z. (2023). An analysis of shooting accuracy towards archery athlete's arm length, arm strength, and body mass index (a study of Koni Bandung district, archery division). *Jurnal Kedokteran Diponegoro (Diponegoro Medical Journal)*, 12(1), 21-25.
- Nor Adnan, N. M., Ab Patar, M. N. A., Lee, H., Yamamoto, S.-I., Jong-Young, L., & Mahmud, J. (2018). Biomechanical analysis using Kinovea for sports application. *IOP conference series: materials science and engineering*.
- Ofori, F., Maina, E., & Gitonga, R. (2020). Using machine

- learning algorithms to predict students' performance and improve learning outcome: A literature based review. *Journal of Information and Technology*, 4(1), 33-55.
- Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J., & Phasinam, K. (2023). Classification and prediction of student performance data using various machine learning algorithms. *Materials today: Proceedings*, 80, 3782-3785.
- Panigrahi, R., Borah, S., Bhoi, A. K., Ijaz, M. F., Pramanik, M., Kumar, Y., & Jhaveri, R. H. (2021). A consolidated decision tree-based intrusion detection system for binary and multiclass imbalanced datasets. *Mathematics*, 9(7), 751.
- Pillitteri, G., Rossi, A., Bongiovanni, T., Puleo, G., Petrucci, M., Iaia, F. M., Sarmiento, H., Clemente, F. M., & Battaglia, G. (2024). Elite Soccer Players' Weekly workload assessment through a new training load and performance score. *Research Quarterly for Exercise and Sport*, 1-9.
- Rotgans, J. I., & Schmidt, H. G. (2011). Cognitive engagement in the problem-based learning classroom. *Advances in Health Sciences Education*, 16(4), 465-479.
- Salim, K. R., Puteh, M., & Daud, S. M. (2012). Assessing students' practical skills in basic electronic laboratory based on psychomotor domain model. *Procedia-Social and Behavioral Sciences*, 56, 546-555.
- Sarro, K. J., Viana, T. D. C., & De Barros, R. M. L. (2021). Relationship between bow stability and postural control in recurve archery. *European Journal of Sport Science*, 21(4), 515-520.
- Setiakarnawijaya, Y., Dlis, F., Tangkudung, J., & Asmawi, M. (2021). Correlation study between arm muscle endurance and arm length and accuracy of 30-meter arrow shots in a national round. *Journal of Physical Education and Sport*, 21, 2357-2363.
- Sharifnezhad, A., Raissi, G. R., Forogh, B., Soleymanzadeh, H., Mohammadpour, S., Daliran, M., & Bagherzadeh Cham, M. (2021). The validity and reliability of kinovea software in measuring thoracic kyphosis and lumbar lordosis. *Iranian Rehabilitation Journal*, 19(2), 129-136.
- Song, Y.-Y., & Ying, L. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*, 27(2), 130.
- Vendrame, E., Belluscio, V., Truppa, L., Rum, L., Lazich, A., Bergamini, E., & Mannini, A. (2022). Performance assessment in archery: a systematic review. *Sports Biomechanics*, 1-23.
- Viscione, I., D'Elia, F., Vastola, R., & Sibilio, M. (2017). Psychomotor Assessment in Teaching and Educational Research. *Athens Journal of Education*, 4(2), 169-177.
- Weidner, T. G., & Popp, J. K. (2007). Peer-assisted learning and orthopaedic evaluation psychomotor skills. *Journal of Athletic Training*, 42(1), 113.
- Zawi, K., & Mohamed, M. (2013). Postural sway distinguishes shooting accuracy among skilled recurve archers. *The Online Journal of Recreation and Sport*, 2(4), 21-28.
- Zulfiker, M. S., Kabir, N., Biswas, A. A., Chakraborty, P., & Rahman, M. M. (2020). Predicting students' performance of the private universities of Bangladesh using machine learning approaches. *International Journal of Advanced Computer Science and Applications*, 11(3), 672-679.



## RESEARCH ARTICLE

### Molecular Phylogeny

# Molecular phylogeography of *Echis carinatus* revisited: Insights from the Sri Lankan population of saw-scaled viper (Serpentes: Viperidae: *Echis*)

A Murugananthan<sup>1</sup>, A Gnanathasan<sup>2</sup>, T Kumanan<sup>3</sup>, KP Amarasinghe<sup>1</sup> and S Pirasath<sup>4</sup>

<sup>1</sup> Department of Parasitology, Faculty of Medicine University of Jaffna, Kokuvil, Sri Lanka.

<sup>2</sup> Department of Clinical Medicine Faculty of Medicine University of Colombo, Colombo, Sri Lanka.

<sup>3</sup> Department of Medicine, Faculty of Medicine University of Jaffna, Kokuvil, Sri Lanka.

<sup>4</sup> District General Hospital, Kilinochchi, Sri Lanka.

Submitted: 18 February 2024; Revised: 13 February 2025; Accepted: 27 February 2025


**Abstract:** The saw-scaled viper (*Echis carinatus*) is a member of the Viperidae family, known for its intricate taxonomic history. Initially, it was believed to be a unique Sri Lankan subspecies called *E. c. sinhaleyus*. The subsequent clinical evidence also suggests the presence of a distinct subspecies in Sri Lanka. However, its existence was later questioned, urging the need for genetic studies. Therefore, it was aimed to unravel the molecular phylogenetic affinities of the Sri Lankan population of *E. carinatus*. For the first time, the molecular phylogenetic affinities and phylogeography of the Sri Lankan population of *E. carinatus* were explored. This study was based on sequences of samples obtained from 12 locations across the Northern Province, using mitochondrial markers *Cytb*, *NADH4*, *16S*, and *12S*. Molecular phylogenetic analyses showed that the South Indian and Sri Lankan populations of *E. carinatus* form a reciprocally monophyletic clade, which is recovered as the sister group to the remaining groups of *E. carinatus*. Based on the divergence-time estimates, the divergence between Sri Lankan and South Indian population of *E. carinatus* is estimated to have occurred in the mid Pleistocene epoch. This study discloses that *E. carinatus* comprises two distinct subspecies namely, *E. c. carinatus* and *E. c. sochureki*. The Sri Lankan and South Indian populations belong to the subspecies *E. c. carinatus*, whereas *E. c. sochureki* comprises the remaining populations of *E. carinatus* distributed across northern, western, and eastern India (Rajasthan, Maharashtra, Odisha, and Goa), as well as in Pakistan, Sharjah, and Iraq.

**Keywords:** Biogeography, mitochondrial markers, molecular phylogeny, snakes, vipers.

## INTRODUCTION

Sri Lanka is a tropical island, bearing distinct groups of fauna with high degree of endemism (Meegaskumbura *et al.*, 2002; Bossuyt *et al.*, 2004). The terrestrial snake fauna of Sri Lanka comprises at least 89 species from 11 families (Botejue, 2020) of which a minimum of 49 (>50%) are endemic (Wickramasinghe *et al.*, 2019). Among them, only 5 species of snakes namely, Russell's viper (*Daboia russelii*) (Shaw & Nodder, 1797), Hump-nosed pit viper (*Hypnale hypnale*) (Merrem, 1820), Indian or common krait (*Bungarus caeruleus*) (Schneider, 1801), Sri Lankan cobra (*Naja polyocellata*) (Mehrtens, 1987) and saw-scaled viper (*Echis carinatus*) (Schneider, 1801) are considered as snakes of highest medical importance (Silva *et al.*, 2023). While *Echis carinatus* is classified as a medically important snakes in Sri Lanka, instances of envenomation by this species are not considered life-threatening, with no reported fatalities (Gnanathasan *et al.*, 2012). There is a divergence of opinions regarding whether *E. carinatus* merits classification in a higher tier of medical significance. Its current status will be upheld until substantiated data are published (Snakebite Expert Committee of Sri Lanka Medical Association, 2021).

Intraspecific taxonomic classification of the *E. carinatus* clade has been unstable (Escoriza *et al.*, 2010),

\* Corresponding author (murugananthan@univ.jfn.ac.lk;  <https://orcid.org/0000-0001-5609-8543>)



This article is published under the Creative Commons CC-BY-ND License (<http://creativecommons.org/licenses/by-nd/4.0/>). This license permits use, distribution and reproduction, commercial and non-commercial, provided that the original work is properly cited and is not changed in anyway.



due to its complex taxonomic history, with several subspecies being recognized in the past (David & Ineich, 1999), including as *E. c. carinatus* (Schneider, 1801), *E. c. sochureki* (Stemmler, 1969), *E. c. multisquamatus* (Cherlin, 1981), *E. c. astolae* (Mertens, 1970), *E. c. sinhaleyus* (Deraniyagala, 1951).

The Sri Lankan population is distributed mainly in the dry and sandy coastal plains in North-Western (Kalpitiya, Wilpattu National Park), Northern (Mannar, Jaffna, Kilinochchi) and Eastern Provinces (Kularatne *et al.*, 2011). It is responsible for more than 50% of the snakebite in the Jaffna District (Kularatne *et al.*, 2011; Sivansuthan, 2011). Deraniyagala (1955) elevated the Sri Lankan population of *E. carinatus* as an endemic sub species based on its morphological variations compared to the northern races of India. But the taxonomic position was not studied in detail after the initial reporting by Deraniyagala in 1955. Conversely, the sub species *E. c. carinatus* and *E. c. sochureki* are considered as a medically important viperine species in South Asia. Among them the *E. c. sochureki* causes numerous bites in northern India whereas *E. c. carinatus* is regionally highly abundant and causes many bites in the western and southern India (Alirol *et al.*, 2010). However, in Sri Lanka, *E. carinatus* envenomation is reported to be mild with zero mortality. Hence, it was hypothesized that the Sri Lankan saw-scaled viper could be a different subspecies that could be genetically distinct from the *E. c. carinatus* existing in India. Therefore, it was strongly suggested to explore this possibility further with genetic studies, morphological assessments, and venom profiling (Kularatne *et al.*, 2011; Peranantharajah *et al.*, 2012). The present phylogenetic study was designed to explore the phylogenetic relationships of the Sri Lankan population of *E. carinatus* using gene sequences from four mitochondrial gene markers.

## MATERIALS AND METHODS

Tissue samples were obtained from a total number of 12 adult saw-scaled viper specimens, representing different geographical locations of the Northern Province, Sri Lanka (Supplementary Figure S1), which were collected in 2020 and 2021 and deposited in the laboratory of the Department of Parasitology, Faculty of Medicine, University of Jaffna. In addition to the sequences originated from this study, an additional 89 sequences derived from the previous studies were retrieved from GenBank and included in the analyses (Pook *et al.*, 2009;

Rhadi *et al.*, 2016). For the out-group taxa, sequences of *Cerastes cerastes* (Linnaeus, 1758) (Wüster *et al.*, 2008) were retrieved from GenBank. Information on all the samples and sequences used in this study is presented under the Supplementary Table S1. The animal study protocol was approved by the Institutional Animal Ethics Review Committee of University of Jaffna.

### DNA extraction and PCR amplification

DNA was extracted from the tissue samples obtained from freshly dissected snake specimens using the Qiagen Tissue DNA extraction kit. Four mitochondrial gene markers, namely, cytochrome b (*Cytb*), NADH dehydrogenase subunit 4 (*NADH4*), small subunit of 12S rRNA (*12S*) and small subunit of 16S rRNA (*16S*) were amplified by polymerase chain reaction (PCR) using the primers and the thermal conditions as described by Pook *et al.* (2009). Details of the primers are provided in the Supplementary Table S2. Finally, the amplified PCR products were purified using the QIAquick PCR Purification Kit (Qiagen) and submitted for direct cycle sequencing by Macrogen Inc. (Korea) using the same primers. A total of 48 sequences derived from this study were deposited at the Gen-Bank under the Accession numbers of OP7339076 - OP739099 (*Cytb* and *NADH4*), OP737787 - OP737798 (*12S*) and OP737803 – OP737814 (*16S*).

### Alignment of sequences

A total of 102 consensus DNA sequences (including out-group sequence) were aligned using the multiple sequence alignment tool, ClustalW (Larkin *et al.*, 2007) in Mega 11 (Tamura *et al.*, 2021) with default parameters. The two protein coding genes (*Cytb* and *NADH4*) were translated into amino acid sequences as per the vertebrate mitochondrial genetic code using the translation tool available in ExPASy server (Gasteiger, 2003) to check for stop codons or frameshift mutations.

### Phylogeny

Phylogenetic analysis was carried out separately for individual mitochondrial markers and for the concatenated dataset (790 bp of *Cytb*, 645 bp of *NADH4*, 364 bp of *12S*, and 481 bp of *16S*) of 102 taxa using Bayesian inference (BI) and maximum likelihood (ML) methods through MrBayes v3.2.7 (Ronquist *et al.*, 2012) and IQ-TREE (Nguyen *et al.*, 2015), respectively. The best-fit nucleotide substitution model and the partitioning schemes

for phylogenetic inference analysis were determined for each methodological approach using PartitionFinder v2.1 (Lanfear *et al.*, 2016). Each codon position of each gene was provided as the starting subset for the analyses. The optimal nucleotide substitution model and the nucleotide partitioning scheme for the Bayesian phylogenetic inference (BI) were determined by setting branch lengths as 'unlinked', model as MrBayes, model selection under the Akaike information criterion corrected (AICc) and search method as the 'greedy' algorithm. While, the optimal nucleotide substitution model and the nucleotide partitioning scheme for the maximum likelihood analysis were determined by setting branch lengths as 'linked', model selection under the AICc and search method as the 'greedy' algorithm (Lanfear *et al.*, 2012).

Maximum likelihood analysis was conducted in IQ-TREE with 1,000 ultrafast bootstrap (BP) iterations (Minh *et al.*, 2013), implementing the partitioning scheme obtained from PartitionFinder 2. BI was carried out in MrBayes v3.2.7 using Markov Chain Monte Carlo (MCMC) randomization in three parallel runs of four chains (3 heated and 1 cold) of 20 million generations with a sampling frequency of 200 and a diagnostic frequency of 5000. Convergence of the three parallel runs was assessed using Tracer v1.7.2, and the first 25% of trees were discarded as burn-in. BI statistical support for respective clades was evaluated as Bayesian posterior probability values (PP) (Huelsenbeck *et al.*, 2001) using the trees remaining after the burn-in. Posterior probability (PP) values of 0.95 above were considered as an indication for strong support (Mulcahy *et al.*, 2011; Wilcox, 2002). FigTree software (V1.4.4) was used to visualize the trees resulting from the BI and ML (<http://tree.bio.ed.ac.uk/software/figtree/>). Intraspecific genetic distances between each pair of individual taxa within the *E. carinatus* clade and interspecific genetic distances between each pair of *Echis* species were calculated using Mega 11 (Tamura *et al.*, 2021) based on uncorrected p-distances parameters.

### Genetic diversity and haplotype network

The genetic diversity and demographic history of population of *E. carinatus*, were explored by estimating the number of haplotypes (h), polymorphic sites (S), parsimony informative sites (P), Nucleotide diversity ( $\pi$ ), haplotype diversity (Hd), Tajima's D (Tajima, 1989), and Fu and Li's F (Fu and Li, 1993) neutrality tests using DnaSP v. 6.12 (Rozas *et al.*, 2017). The haplotype

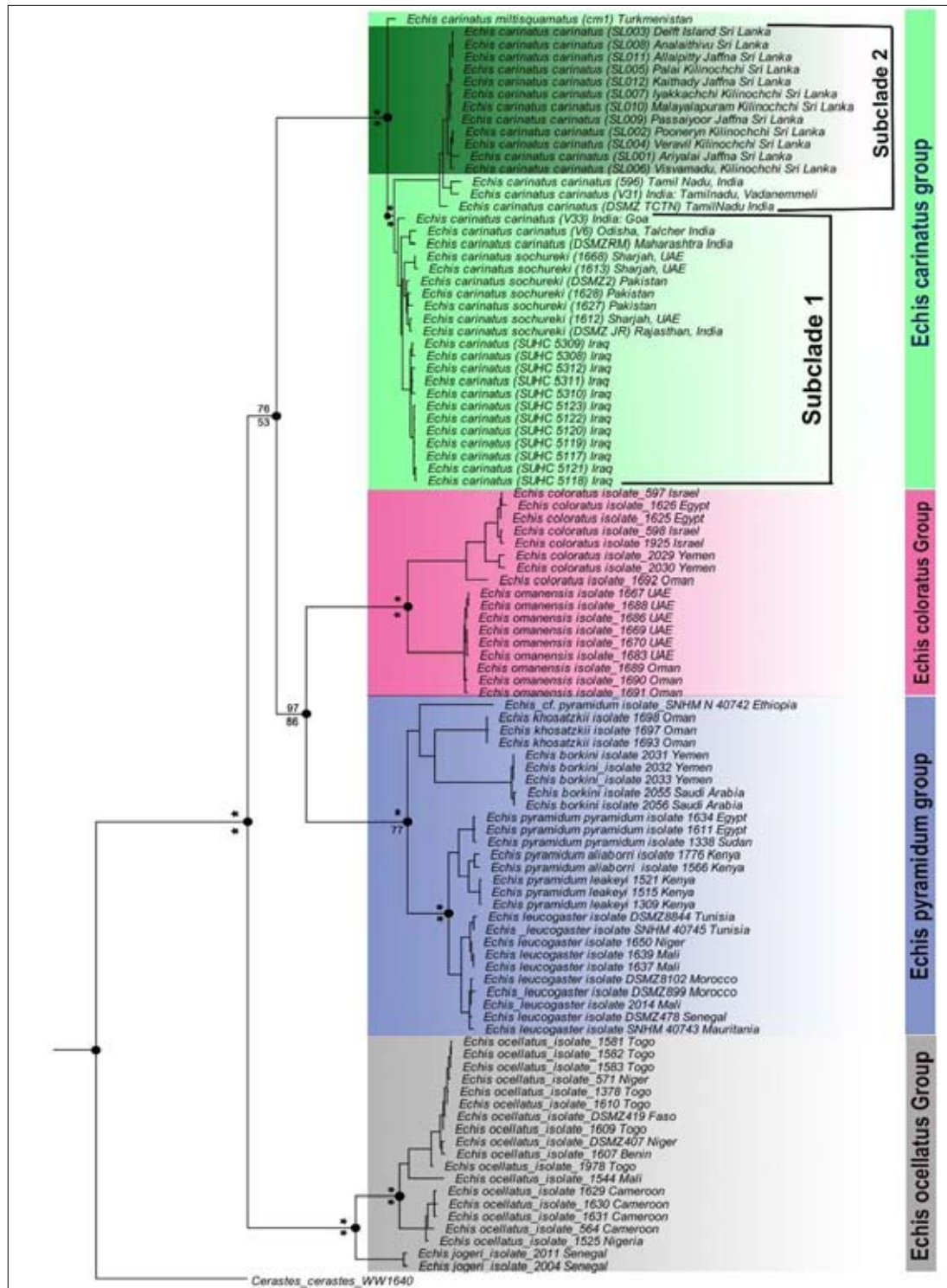
network for *Cytb* and *NADH4* of the *E. carinatus* clade were constructed using PopART v1.7 (Leigh and Bryant, 2015) through a median-joining network (Bandelt *et al.*, 1999). The sequences generated by the current study and the sequences obtained from the GenBank (Table 1) were used for this analysis.

### Divergence-time estimation

The combined mtDNA dataset was used in BEAST 2 (Bouckaert *et al.*, 2014) to estimate the divergence timings among major lineages of *Echis*. The divergence-timing analyses were carried out separately on both the concatenated dataset of 120 taxa and condensed dataset of 24 taxa, representing unique haplotypes. Each lineage was represented by a selected individual referred to as a Molecular Operational Taxonomic Unit (MOTU) (Blaxter *et al.*, 2005)). The optimal substitution model for each gene subset was determined using PartitionFinder 2. The model selection was based on the 'Akaike information criterion corrected (AICc)' using 'unlinked' branched length, and 'Beast' with 'greedy' schemes search. The pairwise measures of divergence, assuming a Poisson distribution for the accumulation rate of substitutions, indicate that the substitution rate of *Cytb* and *ND4* in Viperidae is 1.4%  $\text{my}^{-1}$  (with 95% confidence limits ranging between 1.09-1.77%  $\text{my}^{-1}$ ). Therefore, the mean substitution rate per lineage per site for *Cytb* was calculated as 0.7%  $\text{my}^{-1}$ . To calibrate the *Cytb* clock rate, the average Viperidae *Cytb* substitution rate of 0.007 substitutions per site per million years was employed (Ursenbacher *et al.*, 2006; Wüster *et al.*, 2002). A Yule pure-birth model and a relaxed clock under lognormal distribution were used as the tree and clock prior, respectively. The substitution rate for other markers was estimated relative to *Cytb*. Two independent runs, each consisting of 20 million generations, were conducted with a sampling interval of 5000 generations for the Markov Chain Monte Carlo (MCMC) method. Convergence of the runs and an effective sample size (ESS) greater than 200 for the combined run were assessed using Tracer. The post-run diagnostic parameters of Tracer v1.7.2 revealed significantly high effective sample sizes (ESS). The initial 10% of generations were discarded as burn-in (Pook *et al.*, 2009), and the LogCombiner tool was used to combine the two runs. Finally, maximum clade credibility (MCC) tree was constructed from the posterior sample of trees using TreeAnnotator and visualized with FigTree v1.4.4.

**Table 1:** Details of the samples considered for construction of haplotype network in PopART v1.7, including localities, voucher reference, and GeneBank accession numbers.

Species	Sample/Isolate	Locality	CYTB	Haplotype	NADH4	Haplotype
<i>E. c. sochureki</i>	DSMZ JR	Jaisalmer, Rajasthan, India	GQ359434	H13	GQ359522	H12
<i>E. c. carinatus</i>	DSMZ TCTN	Tuticorin, Tamil Nadu, India	GQ359435	H1	GQ359523	H5
<i>E. c. carinatus</i>	DSMZ RM	Ratnagiri, Maharashtra, India	GQ359439	H18	GQ359527	H11
<i>E. c. multisquamatus</i>	AJ Coll. Göran Nilson.,	Turkmenistan (cm1)	AJ275702	H20	-	
<i>E. c. sochureki</i>	DSMZ 2	Pakistan	GQ359441	H15	GQ359528	H11
<i>E. c. carinatus</i>	WW 596	Chennai, Tamil Nadu, India	GQ359433	H10	GQ359521	H6
<i>E. c. sochureki</i>	WW 1612	Sharjah, UAE (cs1)	GQ359436/		GQ359524/	H9
<i>E. c. sochureki</i>	WW 1613	Sharjah, UAE (cs1)	GQ359437/	H12	GQ359525/	H8
<i>E. c. sochureki</i>	WW 1627	Pakistan	GQ359440/	H16	EU624223/	H10
<i>E. c. sochureki</i>	WW 1628	Pakistan	GQ359438/	H17	GQ359526/	H7
<i>E. c. sochureki</i>	WW 1668	Al Wasit, Sharjah, UAE (cs1)	EU852295/	H12	EU852301/	H8
<i>E. carinatus</i>	V31	Tamil Nadu, Vadanemmeli	MG995822.1	H11	MG995837.1	H5
<i>E. carinatus</i>	V6	India: Odisha, Talcher	MG995811.1	H19	MG995831.1	H11
<i>E. carinatus</i>	isolate 5117	Iraq	KX233705.1	H22	-	
<i>E. carinatus</i>	isolate 5118	Iraq	KX233707.1	H23	-	
<i>E. carinatus</i>	isolate 5119	Iraq	KX233706.1	H21	-	
<i>E. carinatus</i>	isolate 5120	Iraq	KX233708.1	H21	-	
<i>E. carinatus</i>	isolate 5121	Iraq	KX233712.1	H23	-	
<i>E. carinatus</i>	isolate 5122	Iraq	KX233709.1	H24	-	
<i>E. carinatus</i>	isolate 5123	Iraq	KX233710.1	H21	-	
<i>E. carinatus</i>	isolate 5308	Iraq	KX233719.1	H25	-	
<i>E. carinatus</i>	Isolate 5309	Iraq	KX233718.1	H23	-	
<i>E. carinatus</i>	Isolate 5310	Iraq	KX233717.1	H28	-	
<i>E. carinatus</i>	isolate 5311	Iraq	KX233716.1	H27	-	
<i>E. carinatus</i>	isolate 5312	Iraq	KX233715.1	H26	-	
<i>E. carinatus</i>	SL001	Jaffna, Ariyala	OP739076	H9	OP739088	H1
<i>E. carinatus</i>	SL002	Kilinochchi, Pooneryn	OP739077	H6	OP739089	H2
<i>E. carinatus</i>	SL003	Delft Island	OP739078	H4	OP739090	H3
<i>E. carinatus</i>	SL004	Kilinochchi, Veravil	OP739079	H7	OP739091	H2
<i>E. carinatus</i>	SL005	Kilinochchi, Palai	OP739080	H3	OP739092	H4
<i>E. carinatus</i>	SL006	Kilinochchi, Visvamaadu	OP739081	H8	OP739093	H2
<i>E. carinatus</i>	SL007	Kilinochchi, Iyakkachchi	OP739082	H2	OP739094	H4
<i>E. carinatus</i>	SL008	Jaffna, Analaitivu	OP739083	H4	OP739095	H3
<i>E. carinatus</i>	SL009	Jaffna, Passaiyoor	OP739084	H5	OP739096	H4
<i>E. carinatus</i>	SL010	Kilinochchi, Malayalapuram	OP739085	H5	OP739097	H4
<i>E. carinatus</i>	SL011	Jaffna, Allaipitty	OP739086	H4	OP739098	H3
<i>E. carinatus</i>	SL012	Jaffna, Kaithady	OP739087	H3	OP739099	H4



**Figure 1:** Molecular phylogenetic relationships of genus *Echis*, based on Bayesian inference of the concatenated sequence alignment of the *Cytb* + *NADH4* + *16S* + *12S* (2280 bp) mitochondrial-gene markers. Asterisks (\*) above and below nodes indicate Bayesian posterior probabilities (BPP)  $\geq 0.95$  and maximum likelihood ultrafast bootstrap values, respectively. Values below 60% for both Bayesian posterior probabilities and maximum likelihood ultrafast bootstrap are not displayed. The scale bar corresponds to the number of substitutions per site. The numbers in parentheses correspond to the sample isolate numbers listed in Supplementary Table S1.

## RESULTS AND DISCUSSION

Saw-scaled vipers are considered as one of the most dangerous snakes in the world and kill more people than any other venomous snakes in their vicinity (Einterz and Bates, 2003; Pitman, 1972; Warrel *et al.*, 1977). However, the taxonomic classification of genus *Echis* has a long history of controversy and confusions (Escoriza *et al.*, 2010; Lenk *et al.*, 2001; Rhadi *et al.*, 2015). Based on the morphological analysis, there were many species and sub species reported in the past, under this genus (Babocsay, 2003, 2004; Cherlin, 1990). Among them, a wide range of morphological variation was reported in *E. carinatus* clade (Arnold *et al.*, 2009). Although there were five subspecies reported under the *E. carinatus* clade, the subsequent molecular phylogenetic studies have focused primarily on three subspecies. Two sub species namely *E. c. astolae* and *E. c. sinhaleys* (Deraniyagala, 1951) were not incorporated in any of these studies due to the unavailability of specimens from the Astolae Island in Pakistan and Sri Lanka (Pook *et al.*, 2009). This phylogenetic analysis primarily relied on the 2280 bp of concatenated *mt*-DNA sequence alignment (*Cytb*: 790 bp; *NADH4*: 645 bp; 12S: 364 bp; 16S: 481 bp) from 102 individual DNA sequences. The final alignment of mitochondrial protein-coding genes was free of indels, frameshifts or non-sense codons. The optimal nucleotide substitution models for each partition, as determined by PartitionFinder 2, are presented in Supplementary Table S3.

The ML (Supplementary Figure S2) and BI (Figure 1) analysis of this molecular phylogenetic analysis retrieved mostly concordant trees with similar topologies. However, the topological differences, which were observed among the trees belong to individual mitochondrial (*Cytb* [Supplementary Figure S3], *NADH4* [Supplementary Figure S4]) and concatenated (*Cytb*+*NADH4*+*16S*+*12S* [Figure 1]) datasets. Discrepancies between the individual data sets and the concatenated data set are highlighted where necessary. The main emphasis of these results lies on the concatenated data set, which exhibited the well-resolved tree, displaying high node support in both the BI and ML methods. Phylogenetic analysis retrieved four main monophyletic clades within genus *Echis*, with strong node support. The main four clades correspond to the *E. ocellatus*, *E. carinatus*, *E. coloratus*, and *E. pyramidum* (Figure 1). For convenient reference, aforementioned clades are introduced as (a) 'ocellatus clade', (b) 'carinatus clade', (c) 'coloratus clade', and (d) 'pyramidum clade'. The ocellatus clade was recovered as the sister group to the carinatus + coloratus + pyramidum group with strong node support (PP = 1.00, BP = 100) in

both BI tree and ML tree for concatenated dataset (Figure 1 and Supplementary Figure S1). The pyramidum clade is recovered as the sister group to rest of the three groups with strong node support (PP=1.00) in BI tree for *Cytb* (Supplementary Figure S3). However, the BI tree for *NADH4* (Supplementary Figure S4) retrieved carinatus clade as the sister group for rest of the three clades with high node support (PP = 1.00).

The carinatus clade is recovered as the sister group to the remaining monophyletic clade of coloratus + pyramidum group, with mixed and weak node support (PP = 0.76, BP = 53, Figure 1) implying the unresolved evolutionary relationship. The BI for *Cytb* also recovered an unresolved sister group relationship between carinatus clade and coloratus clade, with weak node support (PP= 0.72, Supplementary Figure S3). The intra sister-group relationships of carinatus clade are not congruent among phylogenetic analyses. The *E. c. multisquamatus* is recovered as the sister group to the remaining species in the carinatus clade with high node support (PP = 1.00, BP = 100, Figure 1), in the BI and ML for *Cytb* + *NADH4* + *16S* + *12S*, while the BI for *Cytb* did not recover *E. c. multisquamatus* as the sister group to the remaining species (Supplementary Figure S3) of the carinatus clade. The monophyletic clade comprising both Sri Lankan and Tamil Nadu populations of *E. c. carinatus* is recognized as the sister group to remaining terminal taxa of *E. carinatus*, *E. c. carinatus*, and *E. c. sochureki* (PP = 1.00, BP = 100). But the sub species *E. c. carinatus* is recovered as a paraphyletic group within the carinatus clade among all the phylogenetic analysis.

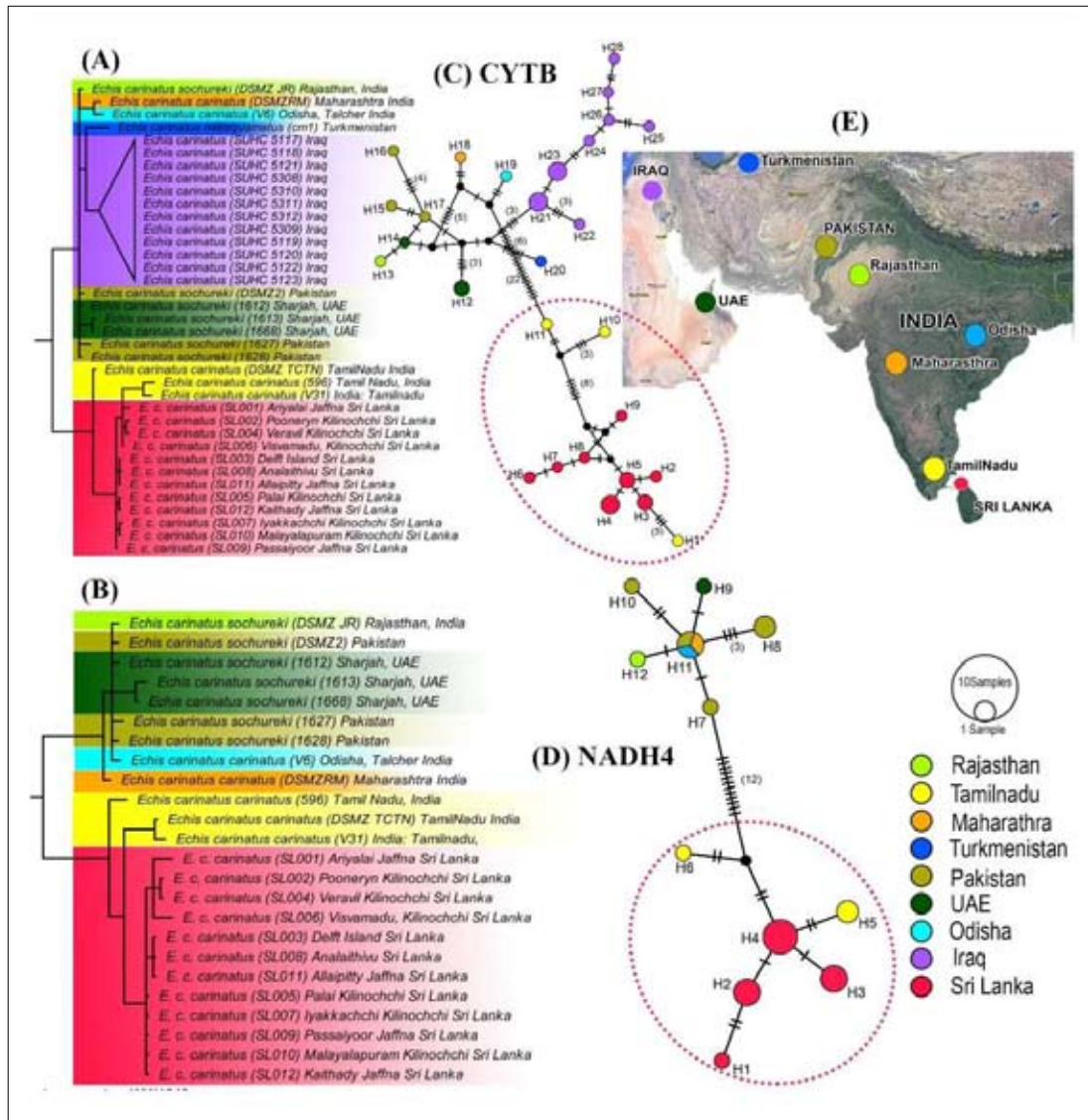
The pyramidum and coloratus clades were recovered as sister groups to each other with strong node support in both BI for *Cytb* + *NADH4* + *16S* + *12S* (PP = 1.00, Figure 1) and BI for *NADH4* (PP = 0.97, Supplementary Figure S4). However, weak node support (BP = 86, Figure 1) observed in ML for *Cytb* + *NADH4* + *16S* + *12S* suggests the unresolved evolutionary relationship. In contrast to that, coloratus and carinatus are recognized as sister groups for each other with unresolved sister group relationships due to the weak node support in BI for *Cytb* (PP = 0.73, Supplementary Figure S3).

The genetic diversity data for *E. carinatus* are shown in Supplementary Table S4. According to analysis conducted to examine the genetic diversity and phylogeographic structure for *Echis carinatus*, the numbers of haplotypes (h) between Sri Lanka and the populations of Maharashtra + Odisha + Rajasthan + Pakistan + UAE were similar for *Cytb* gene, while numbers of haplotypes between Sri Lanka + Tamil Nadu and Maharashtra + Odisha +



Rajasthan + Pakistan + UAE populations were similar for *NADH4* gene. The number of parsimony informative sites was similar between Sri Lankan and Maharashtra + Odisha + Rajasthan + Pakistan + UAE, while haplotype diversity between Iraq and Pakistan + UAE populations

was similar for *NADH4* gene. None of the neutrality tests were significant in any of the populations except the Fu and Li's F test which retrieved a negative significant value for *Cytb* in the Pakistan+UAE population.



**Figure 2:** The haplotype network using PopART v1.7, for *Cytb* and *NADH4* mitochondrial-gene markers of *Echis carinatus*. (A) *Echis carinatus* clade of Bayesian inference tree for *Cytb* (790 bp) gene marker. (B) *Echis carinatus* clade of Bayesian inference tree for *NADH4* (645 bp) gene marker. (C) Median joining haplotype network for 790bp fragment of *Cytb* gene marker. (D) Median joining haplotype network for 645 bp fragment of *NADH4* gene marker. (E) Map showing the sample localities. The areas of the circles proportionally represent the number of individuals sharing a specific haplotype, while the number of mutational steps > 3 is shown in parentheses. Hypothetical nodes are represented by black circles. The colors in the legend correspond to different sample localities. Haplotypes depicted by red circles represent subclade 2 in Figure 1, while the remaining haplotypes represent subclade 1 in Figure 1.

The median-joining network splits *E. carinatus* populations into two phylogeographic structures through 22 mutational steps in the *Cytb* haplotype network (Figure 2C) and 12 mutational steps in the *NADH4* haplotype network (Figure 2D). The separation of these two phylogeographic structures corresponds to the subclade 1 and subclade 2, in Figure 1. The Sri Lankan and Tamil Nadu populations of *E. c. carinatus* are confined to the corresponding phylogeographic structure of subclade 2 (Figure 1, Figure 2C and 2D), while the remaining part corresponds to the subclade 1, which comprises *E. carinatus*, *E. c. carinatus*, *E. c. sochureki* from various geographical regions (Figure 1, Figure 2C and 2D). In the median-joining network for *NADH4*, a single haplotype is shared among each of the samples from Rajasthan (DSMZ JR), Odisha (V6), and Pakistan (DSMZ 2) (Figure 2). However, in both the median-joining network for *Cytb* and *NADH4*, no haplotypes are shared between the Sri Lankan and Tamil Nadu populations of *E. c. carinatus* (Figure 2C and 2D).

Uncorrected pairwise genetic distances (p-distance) computed for the *Cytb* and *NADH4* genes for the species of *Echis* are presented in (Table 2). Inter-specific p-distances among ten different species of *Echis* are given in Table 2A. The minimum interspecific p-distance for the *Cytb* was 2 – 6, while for *NADH4* it was 8 – 10. The maximum interspecific p-distance was observed for *Cytb* and *NADH4* were 16 – 18 and 25 – 27 respectively. The intra-specific p-distance within *E. carinatus* was estimated using the sequences of eight different populations from various geographical locations (Table 2B). The maximum intra-specific p-distances were 5.2 – 6.4 for *Cytb* and 6.2 – 7.3 for *NADH4*, while the minimum values were 0.4 – 1.4 for *Cytb* and 0.3 for *NADH4*. The p-distances for Sri Lankan population of *E. carinatus* showed genetically distant relationship to the UAE population (5.2 – 5.6 for *Cytb* and 5.3 – 7.0 for *NADH4*) and genetically proximal affinity to the Tamil Nadu population (1.5 – 2.3 for *Cytb* and 1.8 – 2.9 for *NADH4*).

**Table 2:** Inter-specific and intra-specific uncorrected pairwise genetic distances for *Cytb* and *NADH4* gene markers in *Echis* species.

(A) Inter-specific uncorrected pairwise genetic distances (%)

	<i>Cytb</i> <i>NADH4</i>	1							
1	<i>E. ocellatus</i>	-							
		-	2						
2	<i>E. pyramidum</i>	15 - 17	-						
		25 - 27	-	3					
3	<i>E. leucogaster</i>	14 - 17	3 - 10	-					
		24 - 28	4 - 27	-	4				
4	<i>E. omanensis</i>	16 - 17	14 - 16	15 - 16	-				
		21 - 25	18 - 24	18 - 20	-	5			
5	<i>E. khosatzkii</i>	14 - 16	9	9 - 10	14 - 14	-			
		24 - 27	10 - 27	11 - 13	19 - 21	-	6		
6	<i>E. jogeri</i>	8 - 10	15 - 17	14 - 15	17	16	-		
		11 - 12	10 - 24	23 - 25	23	23 - 24	-	7	
7	<i>E. coloratus</i>	16 - 17	15 - 17	15 - 16	7 - 9	14 - 15	16 - 17	-	
		19 - 24	17 - 23	18 - 20	8 - 10	18 - 21	20 - 22	-	8
8	<i>E. borkini</i>	15 - 18	10 - 16	10 - 15	14 - 16	9 - 15	16 - 18	15 - 17	-
		24 - 26	10 - 25	10 - 11	19 - 20	10 - 11	22 - 23	17	-
9	<i>E. carinatus</i>	15 - 18	14 - 17	14 - 17	14 - 17	15 - 16	15 - 16	14 - 17	19
		20 - 25	15 - 24	18 - 25	15 - 19	19 - 23	18 - 21	12 - 16	18 - 21
10	<i>E. multisquamatus</i>	16 - 18	16 - 16	16	16	15	15 - 16	15 - 17	2 - 18
		-	-	-	-	-	-	-	-



(B) Intra-specific uncorrected pairwise genetic distances (%).

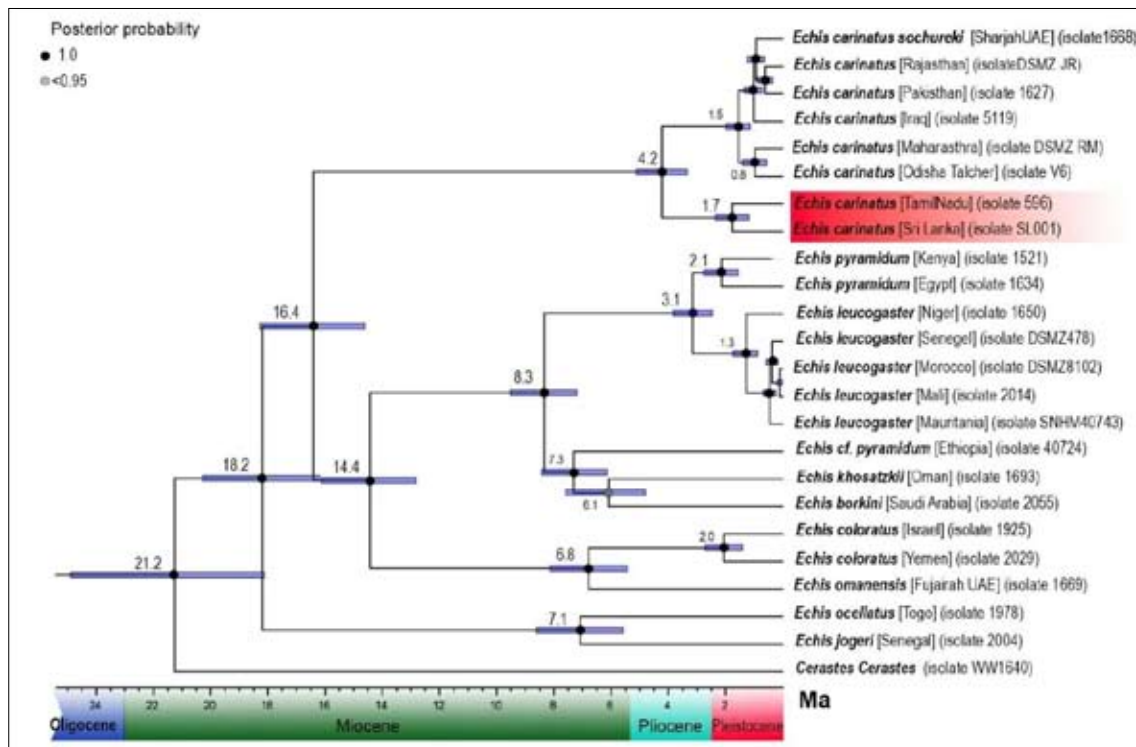
	<i>Cytb</i>						
	<i>NADH4</i>	1					
1	UAE	-					
		-	2				
2	Pakistan	0.5 - 1.9	-				
		0.3 - 1.5	-	3			
3	Iraq	1.5 - 2.4	1.7 - 2.2	-			
		-	-	-	4		
4	Rajasthan	0.4 - 1.4	0.6 - 1.3	1.5 - 2.4	-		
		0.3 - 1.3	0.3 - 0.5	-	-	5	
5	Odisha	1.8 - 2.3	1.8 - 2.3	2.5 - 3.2	1.9	-	
		0.5 - 1.5	0.5 - 0.6	-	0.5	-	6
6	Maharastra	1.6 - 2.2	1.6 - 2.3	2.2 - 3.0	1.8	1.0	-
		0.5 - 1.3	0.3 - 0.8	-	0.5	0.3	-
7	Tamil Nadu	4.2 - 5.5	1.8 - 2.3	5.1 - 6.1	3.4 - 4.7	3.8 - 4.1	4.1 - 4.7
		5.0 - 7.0	4.7 - 6.4	-	5.0 - 6.3	4.8 - 6.1	6.2 - 7.3
8	Sri Lanka	5.2 - 5.6	3.5 - 5.1	5.2 - 6.4	4.7 - 5.1	3.8 - 4.7	4.8 - 5.2
		5.3 - 7.0	5.1 - 6.2	-	5.3 - 5.9	5.1 - 5.9	6.2 - 7.3
							1.5 - 2.3
							1.8 - 2.9

The divergence-timing analysis for the concatenated dataset (*Cytb* + *NADH4* + *16S* + *12S*) of 102-taxa, using a *Cytb* substitution rate in BEAST yielded congruent tree topologies (Supplementary Figure S5) with those from the BEAST analysis of 24-taxa concatenated dataset (*Cytb* + *NADH4* + *16S* + *12S*, is shown in Figure 3). The estimated ages for the selected nodes in both analyses were remarkably similar. The phylogenetic relationships among *Echis* species, as well as the resulting topology, in the BEAST analysis predominantly consistent with those obtained from the BI and ML analyses of the same dataset. Estimated divergence timing based on the 24-taxa dataset is presented here. The basal split between *Echis* and *Cerastes cerastes* was dated at 21.2 million years ago (Ma), in the early Miocene (95% highest posterior density (HPD): 18.1-24.9 Ma), while the split between ocellatus clade and its sister group was dated at in the early Miocene 18.2 (95% HPD: 16.2-20.2) Ma, (Figure 3). The divergence between carinatus clade and its sister group was dated at 16.4 in the mid Miocene (96% HPD: 14.6-18.3) Ma, while the split between pyramidum and coloratus clades was dated at 14.4 Ma in

the mid Miocene. The first split within the ocellatus clade, pyramidum clade and coloratus clade was estimated to have occurred 7.6 Ma (95% HPD: 5.6-8.6 Ma), 8.2 Ma (95% HPD: 7.2-9.5 Ma), and 6.8 Ma (95% HPD: 5.4-8.1 Ma), respectively in the late Miocene (Figure 3). The subsequent diversification of *E. carinatus* was estimated to have occurred 4.2 Ma (95% HPD: 3.3-5.1 Ma) in the early Pliocene. The divergence between Sri Lankan and South Indian (Tamil Nadu) population of *E. carinatus* was dated at 1.7 Ma (95% HPD: 1.2-2.3 Ma), in the mid Pleistocene epoch (Figure 3). However, a discrepancy in the absolute values of estimated divergence times was observed in the current study compared to the study by Pook *et al.* (2009), particularly towards the root of the tree. Pook *et al.* (2009) estimated cladogenesis within the *Echis* using fossil calibration information with secondary calibration points. The divergence-timing estimates in the current study were primarily based on the *Cytb* substitution rate for Viperidae and are younger than those estimated by Pook *et al.* (2009). The primary focus of the present study was to understand the sequence of divergence of the Sri Lankan population of *E. c. carinatus*

from their most recent common ancestor. Although fossil evidence-based divergence-timing estimates result in older divergence estimates towards the root of the tree,

current study retrieved approximately similar estimations for *Echis carinatus* lineage compared to the study of Pook et al. (2009).



**Figure 3:** Bayesian divergence-times estimated for *Echis* (saw-scaled viper) for 24-taxon dataset using *Cytb* substitution rate in Beast v2.6.7 for the concatenated sequence alignment of the *Cytb*+*NADH4*+*16S*+*12S* (2280bp) mitochondrial-gene markers. The node values represent the mean ages for divergence estimates and node bars denote 95% highest probability density for divergence-time estimates.

Mitochondrial markers are maternally inherited and typically lack recombination. They possess a faster substitution rate than nuclear markers, making them valuable for revealing genetic variations within closely related species and populations. Therefore, these markers find extensive application in studies focused on the geographic distribution of evolutionary lineages, known as phylogeography (Pethiyagoda & Sudasinghe, 2021). Considering the potential of using mitochondrial markers to determine the genetic structure, Pook et al. (2009) conducted the first comprehensive molecular investigation based on four mitochondrial gene sequences and recognized all the species and subspecies of the genus *Echis* under four distinct clades. The subsequent molecular phylogenetic studies have also identified the same four major clades, despite some variation among the sister group relationships of these four clades (Arnold et al., 2009; Barlow et al., 2009; Escoriza et al.,

2010; Rhadi et al., 2016; Ashraf et al., 2020). Aligned with the prior reporting, the present phylogenetic study also retrieved the genus *Echis* into four distinct clades and clustering all the Sri Lankan samples under the monophyletic clade of *E. carinatus*. Given the enhanced resolution achievable through the inclusion of a sufficient number of phylogenetic informative sites supporting a consistent tree structure (Rokas and Carroll, 2005; Heath et al., 2008) in addition to the samples incorporated by Pook et al. (2009), new samples from distinct geographical areas of Iraq, Tamil Nadu, Odisha, and Goa, with the twelve samples from Sri Lanka, were integrated to increase the resolution of the *E. carinatus* clade in the present study.

However, discrepancies can be observed between traditional subspecies classification and resulted phylogenetic tree. According to the traditional

classification of *E. carinatus*, Indian population of saw-scaled viper belongs to the subspecies *E. c. carinatus*, except the Rajasthan population which belongs to the *E. c. sochureki*. Our phylogenetic analysis has resolved this misclassification, exclusively by introducing new sequences to the carinatus clade which was not incorporated by Pook *et al.* (2009). The resulted phylogenetic analysis retrieved two distinct monophyletic subclades for the carinatus clade, implying that the clade of *E. carinatus* is composed mainly of two sub species namely, *E. c. carinatus* and *E. c. sochureki*. The subclade *E. c. carinatus* comprises Sri Lankan and South Indian (Tamil Nadu) populations, while the subclade *E. c. sochureki* comprises rest of the populations (Iraq, Pakistan, UAE, Rajasthan, Maharashtra, Odisha, and Goa) of *E. carinatus* (Figure 3). The resulted classification based on the phylogenetic tree is also supported by the results obtained from the analysis; uncorrected pairwise genetic distance of *E. carinatus* clade (Supplementary Table S5) and the median-joining network split for *E. carinatus* populations (Figure 2C, D).

Indeed, the clear demarcation in the geographical boundaries of these two sub clades, a very close genetic distance between the Sri Lanka and Tamil Nadu populations within the subclade 2 compared to more northern species (subclade 1), and the divergence time estimation indicating a split of Tamil Nadu and Sri Lankan population during the Pleistocene around 1.7 Ma, strongly support the dispersal of the Sri Lankan saw-scaled viper from South India. Sri Lanka is separated from the southern tip of the Indian mainland by the narrow and shallow Gulf of Mannar and Palk Strait (Chauhan, 2008). The separation of Sri Lanka from the main land of India occurred, as a result of marine transgression event in early Miocene (Cooray, 1991; Sahni and Mitra, 1980). In the Messinian age, the two landmasses were reconnected during the Miocene-lower Pliocene possibly due to the reduced sea level in the Messinian age (Aharon *et al.*, 1993). This landmass connectivity occurred from time to time in the Quaternary period and the most recent separation occurred during the early Holocene sea level rise between 12000 and 7000 years ago (Gunatilaka, 2000). Based on the fossil evidence, it was assumed that during the past 1 million years, the two lands were one landmass for most of the time. The recurrent connections between these landmasses facilitated biotic exchange between India and Sri Lanka (Biswas and Pawar, 2006). The subsequent complete separation of Sri Lanka from the Indian mainland enabled the island to have a combination of many uniquely endemic species as well as a subset of the Indian biota within its overall fauna (Bossuyt *et al.*, 2004).

The divergence time estimated by the molecular dating in the present study also suggests that the subclade 2 including the South Indian (TN) and Sri Lankan *E. carinatus* populations diverged from the Northern *Echis carinatus* population (Subclade1) during Pliocene around 4.2 Ma. Subsequently, the split between Sri Lankan and India occurred around 1.7 Ma during the Pleistocene epoch. Furthermore, it's worth noting that the dry zones in northwestern Sri Lanka and southeastern India share similar arid climates where the *E. c. carinatus* subspecies is found (Pethiyagoda and Sudasinha, 2021). The closely connected landmass of South India and Northern Sri Lanka for longer period which experience the same tropical humid and dry climate may facilitate the dispersal of this subspecies into the Northern Sri Lanka and eventually to the west and east coastal directions, which is evident by the relatively high number of saw-scaled viper bite cases reported from these areas (Kularatne *et al.*, 2011; Pirasath *et al.*, 2021; Sivansuthan, 2011). Previously, Sri Lankan saw-scaled viper was designated as a distinct subspecies endemic to Sri Lanka based on phenotypic characters by Deraniyagala (1955). However, the present study on the genotypic characters shows shallow genetic divergence between the *Echis* populations of Sri Lanka and Southern India along with the observations from phylogenetic analysis, uncorrected p-distance analysis, haplotype network, and divergence-timing analysis. Therefore, the Sri Lankan saw-scaled viper population can be synonymized with the *E. c. carinatus* of South Indian (Tamil Nadu) population.

On the other hand, based on the clinical observations, Sri Lankan saw-scaled viper population was hypothesized to be a different subspecies of *E. carinatus* (Kularatne *et al.*, 2011; Peranantharajah *et al.*, 2012). But the present study and the clinical observations of the envenomed victims reported from the different geographical locations of India strongly suggest that, *E. c. carinatus* distributed in Sri Lanka and the Southern India is less toxic when compared with *E. c. sochureki* distributed in North India, which leads to more severe form of hemo-nephrotoxic envenomation (Chauhan and Thakur, 2016). Kochar *et al.* (2007) has also described the ineffectiveness of the polyvalent antivenom produced in Southern India for neutralizing the envenoming effect of saw-scaled viper bites of Northern Indian population suggesting a genetically distant relationship between Southern Indian and Northern Indian populations of saw-scaled viper. Therefore, our phylogeographic study infers that the Sri Lankan population of *E. carinatus* has a genetically proximal affinity to the South Indian population, suggesting that the Sri Lankan saw-scaled viper population dispersed from South India during the Plio-Pleistocene dispersal event.

## CONCLUSION

In conclusion we suggest that the *E. carinatus* clade comprises two distinct subspecies; *E. c. carinatus* and *E. c. sochureki*. The Sri Lankan and South Indian populations belong to the subspecies *E. c. carinatus*, whereas *E. c. sochureki* comprises the remaining populations of *E. carinatus* distributed across northern, western, and eastern India (Rajasthan, Maharashtra, Odisha, and Goa), as well as in Pakistan, Sharjah, and Iraq.

## Acknowledgements

We are deeply grateful to the colleagues Mr. P.T. Amalraj, Mr. K. Anpu, and Mr. G. Raheethan, who assisted us in collecting the specimens. We extend our sincere gratitude to Mr. H. Sudasinghe for providing expert guidance in molecular phylogenetic analysis. We also would like to specially thank Dr. (Mrs.) A. Sivaruban and Mr. N.D. Abeyaweera from the Department of Zoology, Faculty of Science, University of Jaffna for their assistance in examining the specimens.

This study was funded by the National Research Council, Sri Lanka, under the Investigator Driven Research Grants No.20-078.

## REFERENCES

- Aharon, P., Goldstein, S. L., Wheeler, C. W., & Jacobson, G. (1993). Sea-level events in the South Pacific linked with the Messinian salinity crisis. *Geology*, 21(9), 771. [https://doi.org/10.1130/0091-7613\(1993\)021<0771:SLEITS>2.3.CO;2](https://doi.org/10.1130/0091-7613(1993)021<0771:SLEITS>2.3.CO;2)
- Alirol, E., Sharma, S. K., Bawaskar, H. S., Kuch, U., & Chappuis, F. (2010). Snake bite in South Asia: A review. *PLoS Neglected Tropical Diseases*, 4(1), e603. <https://doi.org/10.1371/journal.pntd.0000603>
- Arnold, N., Robinson, M., & Carranza, S. (2009). A preliminary analysis of phylogenetic relationships and biogeography of the dangerously venomous Carpet Vipers, *Echis* (Squamata, Serpentes, Viperidae) based on mitochondrial DNA sequences. *Amphibia-Reptilia*, 30(2), 273–282. <https://doi.org/10.1163/156853809788201090>
- Ashraf, M. R., Nadeem, A., Smith, E. N., Javed, M., Smart, U., Yaqub, T., Hashmi, A. S., & Thammachoti, P. (2020). Molecular phylogenetics of saw-scaled viper (*Echis carinatus*) from Pakistan. *Pakistan Journal of Zoology*, 52(4). <https://doi.org/10.17582/journal.pjz/20190710090746>
- Babocsay, G. (2003). Geographic variation in *Echis coloratus* (Viperidae, Ophidia) in the Levant with the description of a new subspecies. *Zoology in the Middle East*, 29(1), 13–32. <https://doi.org/10.1080/09397140.2003.10637966>
- Babocsay, G. (2004). A new species of saw-scaled viper of the *Echis coloratus* complex (Ophidia: Viperidae) from Oman, Eastern Arabia. *Systematics and Biodiversity*, 1(4), 503–514. <https://doi.org/10.1017/S1477200003001294>
- Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16(1), 37–48. <https://doi.org/10.1093/oxfordjournals.molbev.a026036>
- Barlow, A., Pook, C. E., Harrison, R. A., & Wüster, W. (2009). Coevolution of diet and prey-specific venom activity supports the role of selection in snake venom evolution. *Proceedings of the Royal Society B: Biological Sciences*, 276(1666), 2443–2449. <https://doi.org/10.1098/rspb.2009.0048>
- Biswas, S., & Pawar, S. S. (2006). Phylogenetic tests of distribution patterns in South Asia: towards an integrative approach. *Journal of Biosciences*, 31(1), 95–113. <https://doi.org/10.1007/BF02705240>
- Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., & Abebe, E. (2005). Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 360(1462), 1935–1943. <https://doi.org/10.1098/rstb.2005.1725>
- Bossuyt, F., Meegaskumbura, M., Beenaerts, N., Gower, D. J., Pethiyagoda, R., Roelants, K., Mannaert, A., Wilkinson, M., Bahir, M. M., Manamendra-Arachchi, K., Ng, P. K. L., Schneider, C. J., Oommen, O. V., & Milinkovitch, M. C. (2004). Local endemism within the western ghats-Sri Lanka biodiversity hotspot. *Science*, 306(5695), 479–481. <https://doi.org/10.1126/science.1100167>
- Botejue, M. (2020). An overview of snake fauna of Sri Lanka: a snake hotspot in Asia. *Towards the Excellence in Military Healthcare Delivery*, 22.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 10(4), e1003537. <https://doi.org/10.1371/journal.pcbi.1003537>
- Chauhan, P. R. (2008). Large mammal fossil occurrences and associated archaeological evidence in Pleistocene contexts of peninsular India and Sri Lanka. *Quaternary International*, 192(1), 20–42. <https://doi.org/10.1016/j.quaint.2007.06.034>
- Chauhan, V., & Thakur, S. (2016). The North-South divide in snake bite envenomation in India. *Journal of Emergencies, Trauma, and Shock*, 9(4), 151. <https://doi.org/10.4103/0974-2700.193350>
- Cherlin, V. (1990). Taxonomic revision of the snake genus *Echis* (Viperidae). II. An analysis of taxonomy and description of new forms [in Russian]. *Proc. Zool. Inst. Leningr*, 207, 193–223.
- Cooray, P. G. (1991). An Introduction to the geology of Sri Lanka (Ceylon). In *National Museums of Sri Lanka* (2nd (rev.))
- David, P., & Ineich, I. (1999). *Les Serpents Venimeux Du Monde: Systématique et Repartition* (P. Laboratoire des Reptiles & Amphibiens, MNHN (ed.); 3rd ed., Issue January). Dumerilia, Paris (MNHN).

- Deraniyagala, P. (1955). *Echis Carinatus* Sinhaleya (Deraniyagala). In *In A coloured atlas of some vertebrates from Ceylon* (pp. 92–94). Ceylon National Museums.
- Einterz, E. M., & Bates, M. E. (2003). Snakebite in northern Cameroon: 134 victims of bites by the saw-scaled orcarpet viper, *Echis ocellatus*. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 97(6), 693–696. [https://doi.org/10.1016/S0035-9203\(03\)80105-0](https://doi.org/10.1016/S0035-9203(03)80105-0)
- Escoriza, D., Metallinou, M., Donaire-Barroso, D., Amat, F., & Carranza, S. (2010). Biogeography of the white-bellied carpet viper *Echis leucogaster* Roman, 1972 in Morocco, a study combining mitochondrial DNA data and ecological niche modeling. *Butlletí de La Societat Catalana d'Herpetologia*, 18, 55–68. [http://molevol.cmima.csic.es/carranza/pdf/Echis\\_Butlleti-SCH\\_18\\_2009.pdf](http://molevol.cmima.csic.es/carranza/pdf/Echis_Butlleti-SCH_18_2009.pdf)
- Fu, Y. X., & Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics*, 133(3), 693–709. <https://doi.org/10.1093/genetics/133.3.693>
- Gasteiger, E. (2003). ExpASY: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 31(13), 3784–3788. <https://doi.org/10.1093/nar/gkg563>
- Gnanathasan, A., Rodrigo, C., Peranantharajah, T., & Coonghe, A. (2012). Case report: Saw-scaled viper bites in Sri Lanka: Is it a different subspecies? Clinical evidence from an authenticated case series. *American Journal of Tropical Medicine and Hygiene*, 86(2), 254–257. <https://doi.org/10.4269/ajtmh.2012.11-0447>
- Gunatilaka, A. (2000). Sea-levels as historical time-markers in prehistoric studies. *Journal of the Royal Asiatic Society of Sri Lanka*, 45, 19–34.
- Heath, T. A., Hedtke, S. M., & Hillis, D. M. (2008). Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution*, 46(3), 239–257. <https://doi.org/10.3724/SPJ.1002.2008.08016>
- Huelsenbeck, J. P., Ronquist, F., Nielsen, R., & Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science (New York)*, 294(5550), 2310–2314. <https://doi.org/10.1126/science.1065889>
- Kochar, D. K., Tanwar, P. D., Norris, R. L., Sabir, M., Nayak, K. C., Agrawal, T. D., Purohit, V. P., Kochar, A., & Simpson, I. D. (2007). Rediscovery of severe saw-scaled viper (*Echis sochureki*) envenoming in the Thar Desert Region of Rajasthan, India. *Wilderness and Environmental Medicine*, 18(2), 75–85. <https://doi.org/10.1580/06-WEME-OR-078R.1>
- Kularatne, S. A. M., Sivansuthan, S., Medagedara, S. C., Maduwage, K., & de Silva, A. (2011). Revisiting saw-scaled viper (*Echis carinatus*) bites in the Jaffna Peninsula of Sri Lanka: Distribution, epidemiology and clinical manifestations. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 105(10), 591–597. <https://doi.org/10.1016/j.trstmh.2011.07.010>
- Kumazawa, Y., Ota, H., Nishida, M., & Ozawa, T. (1998). The Complete Nucleotide Sequence of a Snake (*Dinodon semicarinatus*) Mitochondrial Genome With Two Identical Control Regions. *Genetics* 150, 313–329. <https://doi.org/10.1093/genetics/150.1.313>
- Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, 29(6), 1695–1701. <https://doi.org/10.1093/molbev/mss020>
- Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2016). PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, msw260. <https://doi.org/10.1093/molbev/msw260>
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>
- Leigh, J. W., & Bryant, D. (2015). <sc>popart</sc> : full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6(9), 1110–1116. <https://doi.org/10.1111/2041-210X.12410>
- Lenk, P., Kalyabina, S., Wink, M., & Joger, U. (2001). Evolutionary Relationships among the True Vipers (Reptilia: Viperidae) Inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, 19(1), 94–104. <https://doi.org/10.1006/mpev.2001.0912>
- Meegaskumbura, M., Bossuyt, F., Pethiyagoda, R., Manamendra-Arachchi, K., Bahir, M., Milinkovitch, M. C., & Schneider, C. J. (2002). Sri Lanka: An amphibian hot spot. *Science*, 298(5592), 379–379. <https://doi.org/10.1126/science.298.5592.379>
- Mendis Wickramasinghe, L. J., Bandara, I. N., Vidanapathirana, D. R., & Wickramasinghe, N. (2019). A new species of *Aspidura* Wagler, 1830 (Squamata: Colubridae: Natricinae) from Knuckles, World Heritage Site, Sri Lanka. *Zootaxa*, 4559(2), 265–280. <https://doi.org/10.11646/zootaxa.4559.2.3>
- Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, 30(5), 1188–1195. <https://doi.org/10.1093/molbev/mst024>
- Mulcahy, D. G., Beckstead, T. H., & Sites, J. W. (2011). Molecular systematics of the Leptodeirini (Colubroidea: Dipsadidae) revisited: Species-tree analyses and multi-locus data. *Copeia*, 2011(3), 407–417. <https://doi.org/10.1643/CH-10-058>
- Murugananthan, A., Gnanathasan, C. A., Kumanan, T., Amarasinghe, K. P., & Pirasath, S. (2024). Morphological characterization and sexual dimorphism of saw-scaled viper (*Echis*: viperidae: ophidia) population in Sri Lanka. *Ceylon Journal of Science*, 53(4), 477–489. <https://doi.org/10.4038/cjs.v53i4.8442>
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. <https://doi.org/10.1093/molbev/msu300>
- Pethiyagoda, R., & Sudasinha, H. (2021). *The ecology and*



- biogeography of Sri Lanka: a context for freshwater fishes. WHT Publications (Private) Limited. [https://www.researchgate.net/publication/356128655\\_The\\_ecology\\_and\\_biogeography\\_of\\_Sri\\_Lanka\\_a\\_context\\_for\\_freshwater\\_fishes](https://www.researchgate.net/publication/356128655_The_ecology_and_biogeography_of_Sri_Lanka_a_context_for_freshwater_fishes)
- Pirasath, S., Gajan, D., Guruparan, M., Murugananthan, A., & Gnanathan, A. (2021). Saw-scaled viper envenoming complicated with acute myocardial infarction. *SAGE Open Medical Case Reports*, 9, 2050313X2110077. <https://doi.org/10.1177/2050313X21100775>
- Pitman, C. R. S. (1972). The saw-scaled viper or carpet viper, (*Echis carinatus*) in Africa and its bite. *The Journal of the Herpetological Association of Africa*, 9(1), 6–34. <https://doi.org/10.1080/04416651.1972.9650822>
- Pook, C.E., Joger, U., Stümpel, N. & Wüster, W. (2009) When continents collide: Phylogeny, historical biogeography and systematics of the medically important viper genus *Echis* (Squamata: Serpentes: Viperidae). *Molecular Phylogenetics and Evolution* 53, 792–807. <https://doi.org/10.1016/j.ympev.2009.08.002>
- Rhadi, F. A., Pouyani, N. R., Karamiani, R., & Mohammed, R. G. (2015). First record and range extension of the saw scaled viper, *Echis carinatus sochureki*, Stemmler 1969, (Squamata: Viperidae), from AL - Basra, Southern. 9(2), 6–9.
- Rhadi, F. A., Rastegar-Pouyani, E., Rastegar-Pouyani, N., Ghaleb Mohammed, R., & Yousefkhani, S. (2016). Phylogenetic affinities of the Iraqi populations of Saw-scaled vipers of the genus *Echis* (Serpentes: Viperidae), revealed by sequences of mtDNA genes. *Zoology in the Middle East*, 62(4), 299–305. <https://doi.org/10.1080/09397140.2016.1257402>
- Rokas, A., & Carroll, S. B. (2005). More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Molecular Biology and Evolution*, 22(5), 1337–1344. <https://doi.org/10.1093/molbev/msi121>
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), 539–542. <https://doi.org/10.1093/sysbio/sys029>
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, 34(12), 3299–3302. <https://doi.org/10.1093/molbev/msx248>
- Sahni, A., & Mitra, H. C. (1980). Lower miocene (Aquitaniens-Burdigalian) palaeobiogeography of the Indian subcontinent. *Geologische Rundschau*, 69(3), 824–848. <https://doi.org/10.1007/BF02104649>
- Silva, A., De Alwis, T., Wijesekara, S., & Somaweera, R. (2023). Minimising misidentification of common medically important snakes of Sri Lanka in the hospital setting. *Anuradhapura Medical Journal*, 17(2), 50–57. <https://doi.org/10.4038/amj.v17i2.7788>
- Sivansuthan, S. (2011). A Descriptive Study of Offending Species and Epidemiology of Snake Bites of Two Areas in the Dry Zone of Sri Lanka: Anuradhapura and Jaffna. S. Sivansuthan 1, S.A.M. Kularatne", K.P. Maduwage 3, H.M.C.D. Ratnayake", R. 16(November), 2011.
- Snakebite Expert Committee of Sri Lanka Medical Association. (2021). *SLMA Guidelines for the Management of Snakebite in Hospital - 2021* (M. Fernando (ed.). Sri Lanka Medical Association. [https://dghs.gov.bd/images/docs/Guideline/NCDC\\_guidelines\\_01\\_08\\_18.pdf](https://dghs.gov.bd/images/docs/Guideline/NCDC_guidelines_01_08_18.pdf)
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585–595. <https://doi.org/10.1093/genetics/123.3.585>
- Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, 38(7), 3022–3027. <https://doi.org/10.1093/molbev/msab120>
- Ursenbacher, S., Conelli, A., Golay, P., Monney, J. C., Zuffi, M. A. L., Thiery, G., Durand, T., & Fumagalli, L. (2006). Phylogeography of the asp viper (*Vipera aspis*) inferred from mitochondrial DNA sequence data: Evidence for multiple mediterranean refugial areas. *Molecular Phylogenetics and Evolution*, 38(2), 546–552. <https://doi.org/10.1016/j.ympev.2005.08.004>
- Warrel, D. A., Davidson, N. M., Greenwood, B. M., Ormerod, L. D., Pope, H. M., Watkins, B. J., & Prentice, C. R. M. (1977). Poisoning by bites of the saw-scaled or carpet viper (*Echis carinatus*) in Nigeria. *QJM: An International Journal of Medicine*, 46(1), 33–62. <https://doi.org/10.1093/oxfordjournals.qjmed.a067493>
- Wilcox, T. (2002). Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Molecular Phylogenetics and Evolution*, 25(2), 361–371. [https://doi.org/10.1016/S1055-7903\(02\)00244-0](https://doi.org/10.1016/S1055-7903(02)00244-0)
- Wüster, W., Peppin, L., Pook, C. E., & Walker, D. E. (2008). A nesting of vipers: Phylogeny and historical biogeography of the Viperidae (Squamata: Serpentes). *Molecular Phylogenetics and Evolution*, 49(2), 445–459. <https://doi.org/10.1016/j.ympev.2008.08.019>
- Wüster, W., Salomão, M. D. G., Quijada-Mascareñas, J. A., Thorpe, R. S., & BBBSP. (2002). Origin and evolution of the South American pitviper fauna: evidence from mitochondrial DNA sequence analysis. In: *Biology of the Vipers*, pp. 111–128. Eagle Mountain Publishing.

## Supplementary Materials

**Supplementary Table S1:** Primers used in the amplification of mitochondrial gene sequences (Pook *et al.* 2009)

Primer name	Position*	Sequence 5'- 3'
CYTB	14902	CTGAAAAACCACCGTTGT
GludgMod2 EchR	15787	GCTCCDCCBAGTTTTRTT
ND4		
NADH4	11677	CACCTATGACTACCAAAAGCTCATGTAGAAGC
HIS12763V	12594	TTCTATCACTTGGATTTGCACCA
12S		
L1091	478	AAACTGGGATTAGATACCCCACTAT
H1557	980	GTACACTTACCTTGTTACGACTT
16S		
L2510	1828	CGCCTGTTTATCAAAAACAT
H3059	2376	CCGGTCTGAACTCAGATCACGT

\* The position refers to the position in the mitochondrial genome of *Dinoden semicarinatus* (Kumazawa *et al.* 1998) at which the 5' end of the primer aligns.

**Supplementary Table S2:** The best-fit nucleotide substitution model and the partitioning schemes used for the phylogenetic inference analysis as determined by PartitionFinder 2

Analysis	Number of sequences	Number of partitions (Subsets)	Partitions	The best model
Bayesian inference: MrBayes		1	Cytb cp1, Cytb cp2, NADH4 cp2, NADH4 cp2, 16S, 12S	GTR+I+G
		2	Cytb cp3, NADH4 cp1	GTR+I+G
Maximum likelihood inference: IQ-TREE		1	Cytb cp1, NADH4 cp2, 12S	TVM+I+G
		2	Cytb cp2	HKY+I
		3	Cytb cp3	TIM+I+G
		4	NADH4 cp1	GTR+I+G
		5	NADH4 cp3	TRN+I+G
		6	16S	GTR+I+G
Divergence timing analysis based on Cytb substitution rate: BEAST2		1	Cytb, NADH4	GTR+I+G+X
		2	16S, 12S	GTR+I+G+X

cp, codon position





(a) Intra-specific uncorrected pairwise genetic distances.

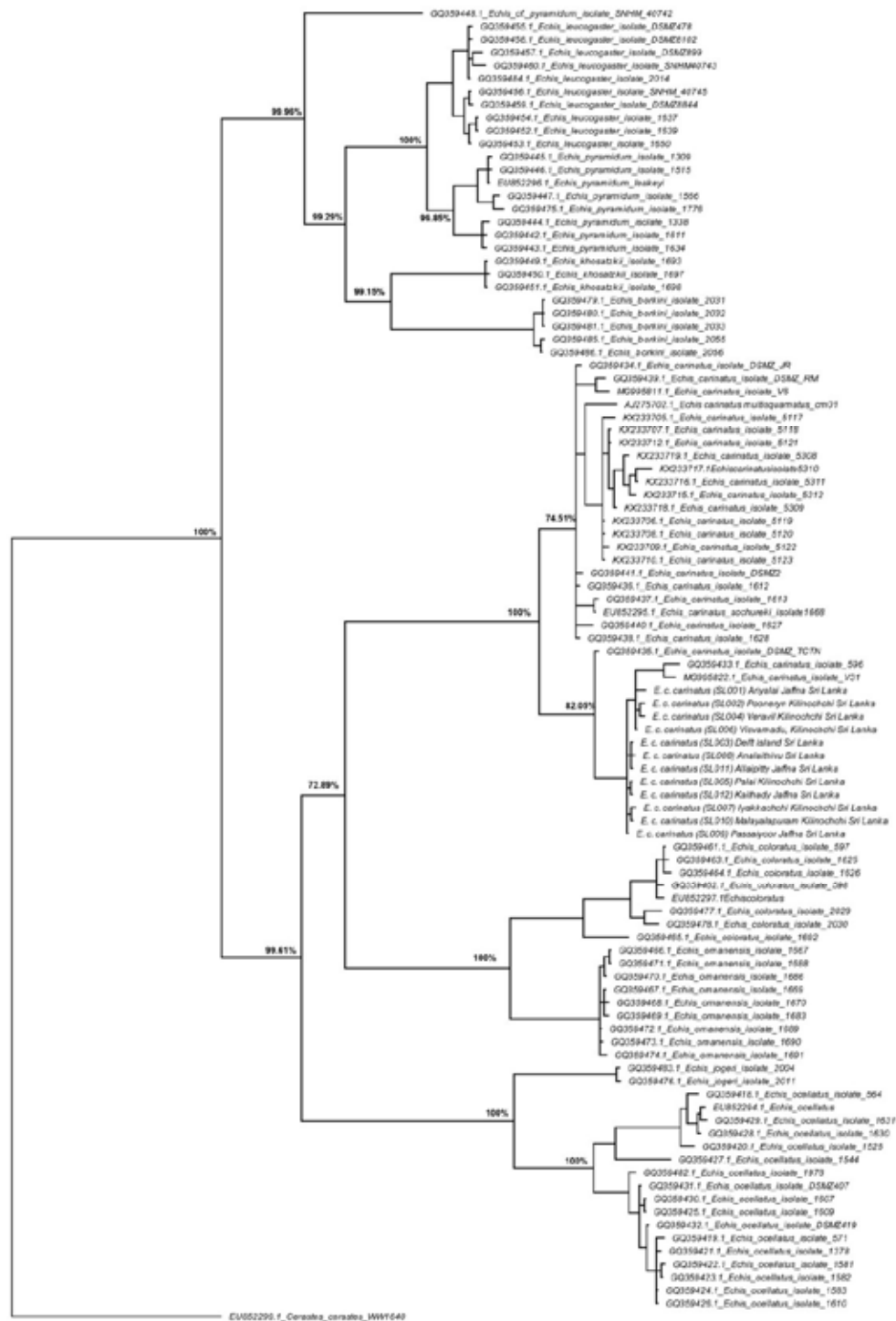
<i>Cytb</i>								
<i>NADH4</i>		1						
1	UAE	-						
		-	2					
2	Pakistan	0.005-0.019	-					
		0.003-0.015	-	3				
3	Iraq	0.015-0.024	0.017-0.22	-				
		-	-	-	4			
4	Rajasthan	0.004-0.014	0.006-0.013	0.015-0.024	-			
		0.003-0.013	0.003-0.005	-	-	5		
5	Odisha	0.018-0.023	0.018-0.023	0.025-0.032	0.019	-		
		0.005-0.015	0.005-0.006	-	0.005	-	6	
6	Maharastra	0.016-0.022	0.016-0.023	0.022-0.030	0.018	0.010	-	7
		0.005-0.013	0.003-0.008	-	0.005	0.003	-	
7	Tamil Nadu	0.042-0.055	0.018-0.023	0.051-0.061	0.034-0.047	0.038-0.041	0.041-0.047	-
		0.050-0.070	0.047-0.064	-	0.050-0.063	0.048-0.061	0.062-0.073	-
8	Sri Lanka	0.052-0.056	0.035-0.051	0.052-0.064	0.047-0.051	0.038-0.047	0.048-0.052	0.015-0.023
		0.053-0.070	0.051-0.062	-	0.053-0.059	0.051-0.059	0.062-0.073	0.018-0.029

## Supplementary Figures

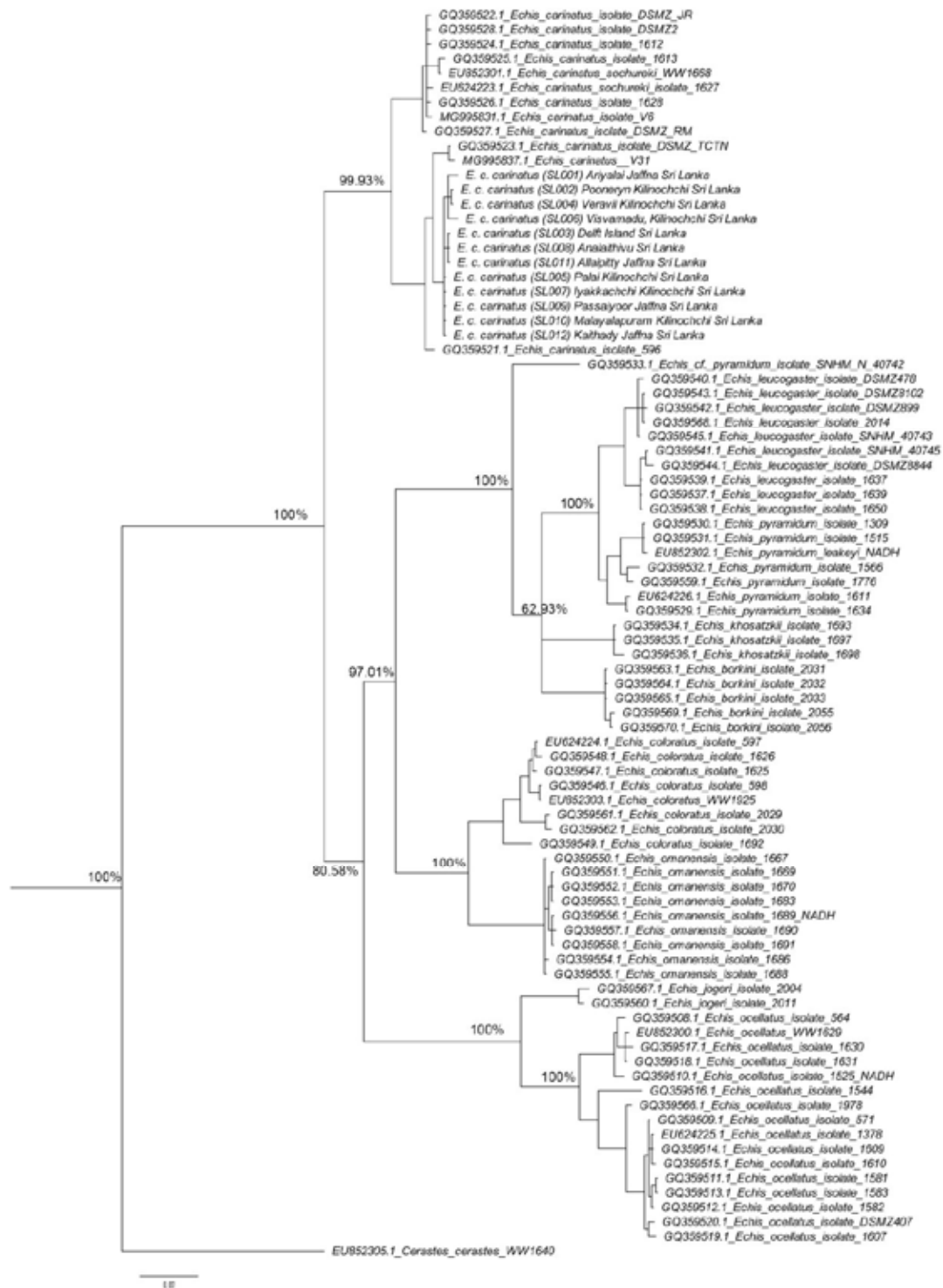
**Supplementary Figure S1:** Molecular phylogenetic relationships of genus *Echis*, built upon Maximum Likelihood (ML) of the concatenated sequence alignment of the *Cytb* + *NADH4* + *16S* + *12S* (2280 bp) mitochondrial-gene markers in IQ-TREE. The node values indicate 100% maximum likelihood ultrafast bootstrap values. Values below 60% maximum likelihood ultrafast bootstrap are not displayed. The scale bar corresponds to the number of genetic changes per site. The numbers in parentheses correspond to the sample isolates numbers listed in Table 1.



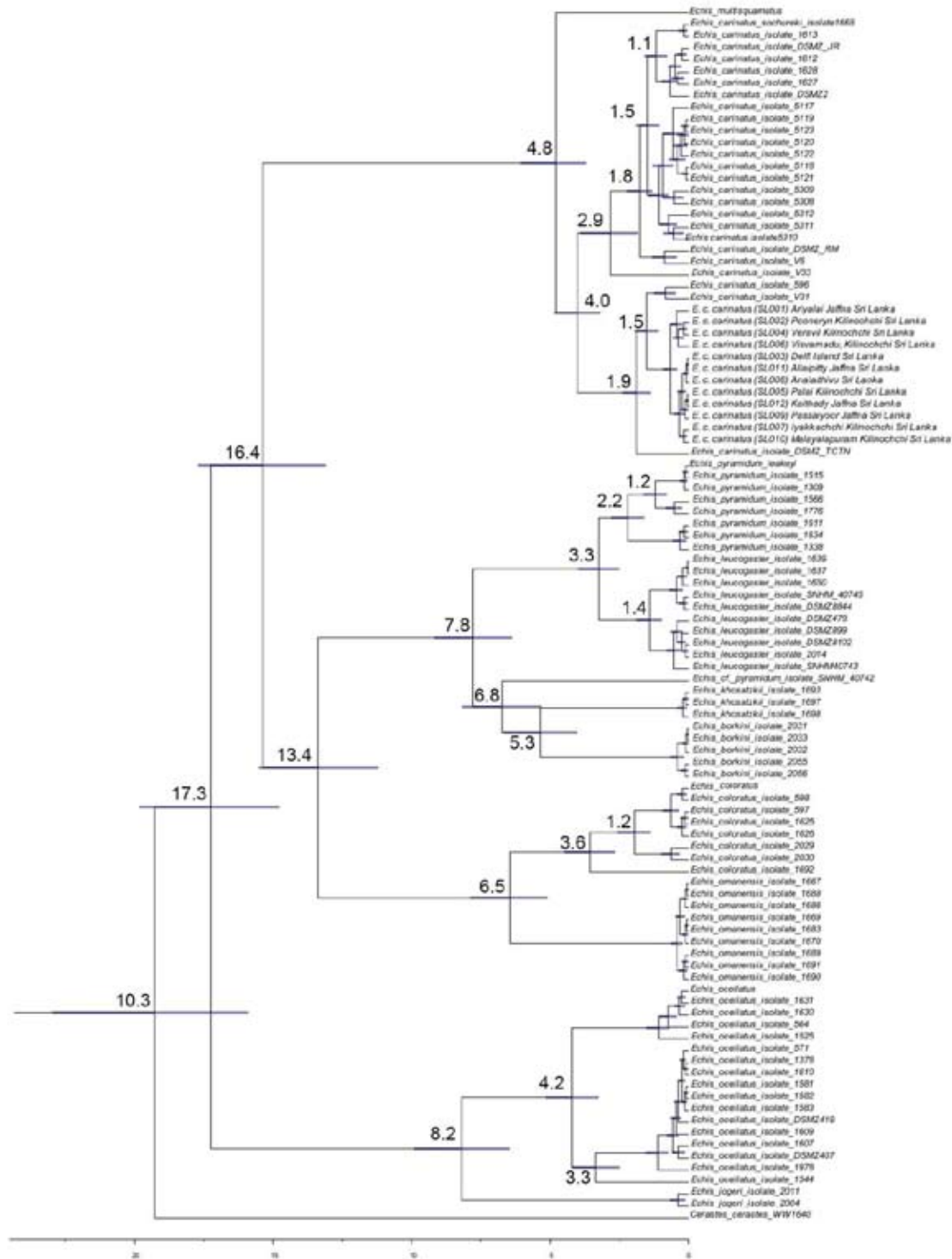
**Supplementary Figure S2:** Molecular phylogenetic relationships of genus *Echis*, built upon Bayesian inference (BI) of the sequence alignment of the *Cytb* (790 bp) mitochondrial-gene markers in MrBayes. The node values indicate 100% Bayesian posterior probabilities values. Values below 60% Bayesian posterior probabilities are not displayed. The scale bar corresponds to the number of genetic changes per site. The numbers in parentheses correspond to the sample isolates numbrs listed in Table 1.



**Supplementary Figure S3:** Molecular phylogenetic relationships of genus *Echis*, built upon Bayesian inference (BI) of the sequence alignment of the *NADH4* (645 bp) mitochondrial-gene markers in MrBayes. The node values indicate 100% indicate 100% Bayesian posterior probabilities values. Values below 60% Bayesian posterior probabilities are not displayed. The scale bar corresponds to the number of genetic changes per site. The numbers in parentheses correspond to the sample isolates numbrs listed in Table 1.



**Supplementary Figure S4:** The complete Bayesian divergent-time estimated tree of *Echis* (saw-scaled viper) for 102-taxon dataset, using *Cytb* substitution rate in Beast v2.6.7 for the concatenated sequence alignment of the *Cytb*+*NADH4*+*16S*+*12S* (2280bp) mitochondrial-gene markers. The node values represent the mean ages for divergence estimates and node bars denote 95% highest probability density for divergence-time estimates.









# JOURNAL OF THE NATIONAL SCIENCE FOUNDATION OF SRI LANKA

## GUIDANCE TO CONTRIBUTORS

### GENERAL INFORMATION

#### Scope

The Journal of the National Science Foundation of Sri Lanka publishes the results of research in all aspects of Science and Technology. It is open for publication of Research Articles, Reviews, Research Communications and Correspondence.

#### IT related and other non-empirical articles

The JNSF is a journal primarily devoted to natural sciences. It also considers for publication significant and novel contributions from formal sciences. Authors of emerging sub-disciplines of Computing and related areas such as Machine Learning, Artificial Intelligence and Data Sciences are requested to carefully adhere to the following guidelines when submitting manuscripts for this journal.

- Clear formulation of outcome-oriented **Research Objective/s** for targeted knowledge (sub)domain/s or (sub)discipline/s.
- Selection and comprehensive summarization of **appropriate Research Method/s** adopted to achieve the stated Research Objective/s.
- Reporting a sound (**Empirical**) **Evaluation** of the research finding/s thereby arguing reliability, validity, and generalizability of research claim/s.

#### Categories of manuscripts

**Research Articles:** Research Articles are papers that present complete descriptions of original research. Research Articles should include an Abstract, Keywords, Introduction, Methodology, Results and Discussion, Conclusion and Recommendations where relevant. References should be prepared according to the “Guidelines for the preparation of manuscripts”. Maximum length of the article should be limited to 30 pages with a word count less than 10,000 including references, figures and tables. Any articles above this limit will be returned. Please refer author guidelines for further details (<https://jnsfsl.sljol.info/about/submissions#author-guidelines>)

**Reviews:** Reviews are critical presentations on selected topics of Science or Technology. They should be well focused and organized and avoid general “textbook” style. As reviews are intended to be critical presentations on selected topics, the author (or the principal author in a multi-author review) need to have had substantial leadership in research supported by a publication track record in the areas covered by the review. A person/s wishing to submit a Review Article should obtain prior approval from the Editorial Board by submitting a concise summary of the intended article, along with a list of the author’s publications in the related area ([jnsf@nsf.gov.lk](mailto:jnsf@nsf.gov.lk)). Maximum length of the article should be limited to 40 pages with a word count of 12,000 including references, figures and tables. Any articles beyond this limit will be returned.

**Research Communications:** Research Communications are intended to communicate important new findings in a specific area of limited scope that are worthy of rapid dissemination among the scientific community. Authors are required to provide a statement justifying the suitability of the submission for a Research Communication. The article should include an Abstract, Keywords, Introduction, Methodology, Results & Discussion, Conclusion and References. Maximum length of the article should be limited to 10 pages with a word count of 2,500 including references, figures and tables. Any articles beyond this limit will be returned.

**Correspondence:** Correspondence will be accepted regarding one or more articles in the preceding four issues of the Journal, as well as Letters to the Editor. Articles covering important scientific events or any other news of interest to scientists, reviews of books of scientific nature, articles presenting views on issues related to science and scientific activity will also be considered. Publication will be made at the discretion of the Editor-in-Chief. Maximum length of the article should be limited to 05 pages with a word count of 1,500 including references, figures and tables. Any articles beyond this limit will be returned.

### SUBMISSION OF MANUSCRIPT

Authors submitting articles to the JNSF should first create an account in the Sri Lanka Journals Online System (<https://jnsfsl.sljol.info/>). All manuscripts should be in MS Word format and must be submitted to the journal’s online platform at <https://jnsfsl.sljol.info/submit/start/>. Submissions via emails are not encouraged. Please make sure that **no** author information is mentioned in the article submitted. Complete names and details of affiliations of all authors must be fed into the system during the online submission process. Authors are required to provide their personal, validated ORCID ID (by obtaining an ORCID ID from <https://orcid.org/>) when submitting the manuscript. No change to the authors or order of authors will be accepted after the submission. All those who have made significant contributions should be listed as co-authors. The corresponding author should ensure that all contributing co-authors are included in the author list and have approved the final version of the paper and have agreed to its submission for publication.

All submissions should be in English. If the manuscript conforms to the guidelines specified, the date received will be the date that the manuscript was submitted to the online system.

Submissions are accepted for processing on the understanding that they will be reviewed and that they have not been submitted for publication elsewhere (including publication as a full paper or extended abstract as a part of Conference Proceedings).

#### Articles deposited in pre-print repositories

Authors should agree to remove the manuscript from the pre-print repositories if the article is accepted for publication in the JNSF.

### **Suggesting potential reviewers by authors**

The authors are requested to suggest three names of referees when submitting their manuscript, in the Cover Letter space provided at the bottom of the page in the first stage of online submission. Referees should not be from the institution where the work was carried out and should not have been co-authors in previous publications. The address, institutional affiliation and e-mail of the suggested referees should be provided. Please note that the JNSF is not bound to select all or any of the suggested referees for sending the manuscript for reviewing

### **Authorship**

All authors designated as authors should be eligible for authorship. Those who have made a substantial contribution to the concept or design of the work; or acquisition, analysis or interpretation of data are recognized as Authors.

The corresponding author should be prompt and ensure adherence to timelines when responding to requests, queries and recommendation of reviewers conveyed by or on behalf of the Editor-in Chief and Editorial Board.

### **Supplementary materials**

Any experimental data necessary to evaluate the claims made in the paper but not included in the paper should be provided as supplementary materials. Supplementary materials will be sent to the reviewers and published online with the manuscript if accepted. The supplementary materials should conform to Journal guidelines and should be uploaded as separate files. Authors should number Supplementary Tables and Figures as, for example, 'Supplementary Table S1'. Refer to each piece of supplementary material at the appropriate point(s) in the main article. Supplementary Materials may include description of the materials and methods, controls, or tabulated data presented in Tables or Figures, and programming codes.

### **Peer review**

The manuscripts submitted to the JNSF will initially be screened by the Editorial Board and, if suitable, will be referred to at least two subject experts in the relevant field. The peer-review process of the JNSF is double-blind.

When revision of a manuscript has been requested, the revised manuscript should be submitted on or before the stated deadline. The authors' response to the comments of referees should be tabulated with the comment and response. The decision of the Editorial Board shall be final.

Accepted papers are subject to editing. The date of acceptance will be the date the Editorial Board accept the paper for publication.

### **Article processing fee**

Article processing fee of US\$ 250 will be levied for each manuscript in two stages, except when the corresponding author is affiliated with a Sri Lankan institution,

- An initial processing fee of US\$ 20 will be levied for each manuscript at the peer-review stage.
- The remaining US\$ 230 will be charged for accepted articles at the time of publication.

Payments can be made online via NSF Payment Portal (<http://pg.nsf.gov.lk/>)

### **Authors' declaration**

The authors are required to accept the conditions indicated in the online author declaration statement.

### **Copyright**

Articles in JNSF are published under the Creative Commons License CC-BY-ND. This license permits use, distribution and reproduction of articles for commercial and non-commercial purposes, provided that the original work is properly cited and is not changed in anyway. The copyright of the article is with the National Science Foundation of Sri Lanka. Therefore, authors are requested to check with institution's copyright and publication policy before submitting an article to the JNSF. Authors secure the right to reproduce any material that has already been published or copyrighted elsewhere. When an article is accepted for publication, the authors are required to submit the Transfer of Copyright document signed by all the authors.

### **Post-publication corrections**

The Editorial Board reserves the right to take action on publishing an erratum or corrigendum. If serious errors are identified in a published article, the Journal may consider a retraction or publishing a correction.

## **STRUCTURE OF MANUSCRIPT**

### **Manuscript**

The manuscript should be free of errors and prepared in single column, using double-spaced text of Times New Roman 12 font leaving 1 inch margins. Pages should be numbered consecutively.

#### **a. Style**

The paper should be written clearly and concisely. The style of writing should conform to scholarly writing. Slang, jargon, unauthorized abbreviations, abbreviated phrasings should not be used. In general, the impersonal form should be used. Poor usage of language will result in rejection of the manuscript during initial screening.

#### **b. Layout**

Manuscripts other than review articles should be generally organized as follows: Title, Abstract, Keywords, Introduction, Methodology, Results and Discussion, Conclusions and Recommendations (where relevant), Acknowledgements and References. Pages should be arranged in the following order:

**First page** should include the title of the manuscript. **Author information should not be mentioned anywhere in the manuscripts.** Any statement (including acknowledgment) which can reveal author identity should be removed. If a major part of the research has been published as an abstract in conference proceedings, it should be mentioned with citation in the space provided for "Comments for Editor". Authors must also indicate the **general and specific research area** of the manuscript in the title page.

**Title:** Should accurately and concisely reflect the contents of the article.

**Running title:** Should be a shortened title (limited to a maximum of 50 characters) that could be printed at the top of every other page of the Journal article.

**Abstract:** Should be between 200 - 250 for research articles and 200 - 300 for reviews. It should not contain any references and should be able to stand on its own. It should outline objectives and methodology together with important results and conclusions.

**Keywords:** Include a maximum of six keywords, which may include the names of organisms (common or scientific), methods or other important words or phrases specific to the study.

**Introduction:** This should state the reasons for performing the work with a brief review of related research studies in the context of the work described in the paper. Objectives of the study should be clearly stated.

**Materials and Methods:** This section should give the details of how you conducted your study. New methods may be described in detail with an indication of their limitations. Established methods can be mentioned with appropriate references. Sufficient details should be included to allow direct repetition of the work by others. Where human subjects are involved, they should be referred to by numbers or fictitious names. A paper reporting the results of investigations on human subjects or on animals must include a statement to the effect that the relevant national or other administrative and ethical guidelines have been adhered to, and a copy of the ethical clearance certificate should be submitted. Methods of statistical analyses used should be mentioned where relevant.

### **Results and Discussion**

Results: the results should be concisely and logically presented. Repetition of the same results in figures, tables or text should be avoided.

Discussion: data essential for the conclusions emerging from the study should be discussed. Long, rambling discussions should be avoided. The discussion should deal with the interpretation of results. It should logically relate new findings to earlier ones. Unqualified statements and conclusions not completely supported by data should be avoided.

Molecular sequence data, such as gene or rDNA sequences, genome sequences, metagenomic sequences etc. must be deposited in a public molecular sequence repository, such as GenBank, that is part of the International Nucleotide Sequence Database Collaboration (INSDC). The accession numbers obtained must be cited in the text, Table or on Figures of phylogenetic trees of the manuscript.

**Conclusion:** The conclusion should be brief, highlight the outcomes of the study and should be aligned with the objectives of the study. It should not contain references.

**Competing Interest statement:** The authors should include a statement on conflict of interest disclosing any financial or other substantive conflicts of interest that may influence the results or interpretation of the research in the space provided in the online article submission form. All sources of financial support for the project should also be disclosed.

**Acknowledgement:** Should be brief and made for specific scientific, financial and technical assistance only. If a significant part of the research was performed in an institution other than the authors' affiliations should be acknowledged. All those who have made substantial contribution to the research but do not qualify to be authors should be acknowledged.

### **References :**

The JNSF uses APA ( 7<sup>th</sup> Edition) reference style

All research work of other authors, when used or referred to or cited, should be correctly acknowledged in the text and in the References.

All the references in the text should be in the list and vice versa

Citing references in the text:

- References to the literature must be indicated in the text and tables as per the Author-Year System, by the author's last name and year, in parenthesis (i.e. Able, 1997) or (Able & Thompson, 1998).
- Citation to work by more than two authors should be abbreviated with the use of et al. (i.e. Able *et al.*, 1997).
- Multiple publications by the same first author in the same year should be coded by letters, (i.e. Thompson, 1991a, 1991b, 1992, 1993).
- Multiple citations of different authors should be made in chronological order and separated by a semicolon, (i.e. Zimmerman *et al.*, 1986; Able *et al.*, 1997).

Citing references in the List of references:

- The list of References should be arranged in alphabetical order based on the last name of the first author.
- In APA 7th ed., **up to 20 authors** should be included in a reference list entry. Write out the last name and first initial(s) for each contributor.

#### **Example for 2–20 authors:**

Wright, A., Komal, G., Siddharth, D., Boyd, G., Cayson, N., Beverley, K., Travers, K., Begum, A., Redmond, M., Mills, M., Cherry, D., Finley, B., Fox, M., Ferry, F., Almond, B., Howell, E., Gould, T., Berger, B., Bostock, T., & Fountain, A. (2020). Styling royalty. London Bridge Press.

- For references with more than 20 authors, after listing the 19th author replace any additional author names with an ellipsis ( ... ) followed by the final listed author's last name and first initial(s).

#### **Example for 21+ authors:**

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woolen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetmaa, A., ... Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3), 437-471. <http://doi.org/fg6rf9>

- All the initials of the author must be given after the last name and the year of publication should follow in parentheses.
- This should be followed by the full title of the referred publication.
- When journal articles are listed, the journal name should be given in full and in italics and followed by the volume number, issue number in parentheses and then the inclusive pages.
- Where there are several publications by the same author(s) and published in the same year they should be differentiated by adding a lower-case letter after the year.

#### **Example**

Clarke, P. N., & Fawcett, J. (2014a). Life as a mentor. *Nursing Science Quarterly*, 27(3), 213-215. <https://doi.org/10.1177/0894318414534492>

Clarke, P. N., & Fawcett, J. (2014b). Life as a nurse researcher. *Nursing Science Quarterly*, 27(1), 37-41. <https://doi.org/10.1177/0894318413509708>

- Digital object identifiers (DOIs) should be included for all references where available.

Details about this reference style can be obtained from below links

- <https://apastyle.apa.org/style-grammar-guidelines/references>
- <https://apastyle.apa.org/style-grammar-guidelines/references/examples>
- <https://libguides.jcu.edu.au/apa>

**Abbreviations and Symbols:** Unless common, these should be defined when first used, and not included in the abstract. The SI System of units should be used wherever possible. If measurements were made in units other than SI, the data should be reported in the same units followed by SI units in brackets, e.g. 5290 ft (1610 m).

**Formulae and Equations:** Equations should be typewritten and quadruple spaced. They should be started on the left margin and the number placed in parentheses to the right of the equation.

**Nomenclature:** Scientific names of plants and animals should be printed in italics. In the first citation, genus, species and authority must be given. e.g. *Borassus flabellifer* Linn. In latter citations, the generic name may be abbreviated, for example, *B. flabellifer* L.

**Tables and figures:** Tables and Figures should be clear and intelligible and kept to a minimum, and should not repeat data available elsewhere in the paper. Any reproduction of illustrations, tabulations, pictures etc. in the manuscript should be acknowledged.

**Tables:** Tables should be numbered consecutively with Arabic numerals and placed at the appropriate position in the manuscript. If a Table must be continued, a second sheet should be used and all the headings repeated. The number of columns or rows in each Table should be minimized. Each Table should have a title, which makes its general meaning clear, without reference to the text. All Table columns should have explanatory headings. Units of measurement, if any, should be indicated in parentheses in the heading of each column. Vertical lines should not be used and horizontal lines should be used only in the heading and at the bottom of the table. Footnotes to Tables should be placed directly below the Table and should be indicated by superscript lower case italic letters (<sup>a</sup>, <sup>b</sup>, <sup>c</sup>, etc.).

**Figures:** All illustrations are considered as figures, and each graph, drawing or photograph should be numbered consecutively with Arabic numerals and placed at the appropriate position in the manuscript. Any lettering to appear on the illustrations should be of a suitable size for reproduction and uniform lettering should be used in all the Figures of the manuscript. Scanned figures or photographs should be of high quality (**300 dpi**), to fit the proportions of the printed page (12 × 17 cm). Each figure should carry a legend so that the general meaning of the figure can be understood without reference to the text. Where magnifications are used, they should be stated.

#### Units of measurement

Length: km, m, mm, µm, nm

Area: ha, km<sup>2</sup>, m<sup>2</sup>

Capacity: kL, L, mL, µL

Volume: km<sup>3</sup>, m<sup>3</sup>, cm<sup>3</sup>

Mass: t, kg, g, mg, µg

Time: year(s), month(s), wk(s),

d(s), h, min, s

Concentration: M, mM, N, %,

g/L, mg/L, ppm

Temperature: °C, K

Gravity: x g

Molecular weight: mol wt

Others: Radio-isotopes: 32P

Radiation dose: Bq

Oxidation-reduction potential: rH

Hydrogen ion concentration: pH