Print ISSN 1391-4588
E - ISSN 2362-0161
Volume 49 No. 3
September 2021

Journal of the National Science Foundation of Sri Lanka

www.nsf.gov.lk

NATIONAL SCIENCE FOUNDATION

JOURNAL OF THE NATIONAL SCIENCE FOUNDATION OF SRI LANKA

Impact Factor : 0.515

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

Subscription : Foreign : US $ 75 (SAARC countries)
US $ 150 (other countries)
Local : Rs. 2000
} inclusive of airmail postage

Accepted as standing orders or on a calendar year basis.
Single issues - Available on request;
Rs. 600 per copy

Payment must accompany all orders. Remittances in favour of the National Science Foundation of Sri Lanka.

Manuscripts : Research Articles, Research Communications, Reviews and Correspondence in all fields of Science and Technology may 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.

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

Manuscripts and all correspondence relating to them should be sent to the Editorial Office, National Science Foundation, 47/5, Maitland Place, Colombo 07, Sri Lanka.

Fax: 94-11- 2694754
E-mail: jnsf@nsf.gov.lk
JNSF home page: *http://www.nsf.gov.lk/index.php/nsf-science-magazine*

Publication fee : A publication fee of US$ 150 will be levied for each manuscript except when the corresponding author is affiliated to a Sri Lankan institution.

Indexing : **The JNSF is indexed in Science Citation Index Expanded, Journal Citation Reports/Science Edition, BIOSIS Previews, Zoological Record, Biological Abstracts, Chemical Abstracts, Scopus, TEEAL, Ulrich's, AGRICOLA, EBSCOhost and CAB Abstracts.**

JOURNAL OF THE
NATIONAL SCIENCE FOUNDATION
OF SRI LANKA

Volume 49 Number 3 September 2021

CONTENTS

Guidelines for contributors

Cover: Floral pictures of some *Alpinia* species found in Sri Lanka; (A) *A. abundiflora,* (B) *A. fax,* (C) *A. galanga,* (D) *A. calcarata,* (E) *A. maleccensis,* (F) *A. zerumbet*.

See *J.Natn.Sci. Foundation Sri Lanka* 2021 49 (3): 321 – 465

J.Natn.Sci.Foundation Sri Lanka 2021 49 (3) : 321
DOI: http://dx.doi.org/10.4038/jnsfsr.v49i3.10815

EDITORIAL

Pitfalls in the blackbox approach to chemical analysis of plant extracts and environmental samples

Many multidisciplinary research projects, particularly those related to biological activities of plant extracts and environmental studies, involve chemical analysis at some stage. There is a growing tendency to treat this aspect of the research as a routine 'black box' affair. Often, a sample is given to the analyst who subjects it to some form of instrumental analytical method (with varying degrees of automation) and the output is given back to the researcher, who then treats the data as accurate and derives various conclusions; sometimes quite erroneous.

Errors can arise due to lack of awareness of the researcher of the science behind the analytical method. For example, in GCMS analysis for the identification of unknown compounds, the reliability of the results depends not only on instrumental factors but also on data processing methods. It is surprising how often compounds with very low match factors are reported as being present in the sample under study, without any qualification. A more serious source of error, is in not relating the analytical result to the nature of the sample being analysed. As an example, it is not uncommon to find common plasticizers (which probably would have originated from impure solvents used in the extraction process) being reported as plant metabolites. (It is a reflection of the current trend towards compartmentalization of knowledge that consideration of possible bio-synthetic pathways, to novel compounds considered to be isolated from plant materials is no longer fashionable.)

In quantitative analysis, the need for method validation when non-standard methods are used is also often neglected, which brings into question the validity of the conclusions based on the analytical results. With environmental samples in particular, the pre-analytical steps need to carefully chosen so as to avoid artefact formation and matrix effects on the analyte signals. Undoubtably, some of these problems could be avoided if a competent analytical chemist is involved in the project, not only at the point of analysis but also at its planning stage.

It is worthwhile to note that in multidisciplinary research requiring chemical analysis, a black box approach coupled with artificial intelligence which can rapidly generate a large amount of data, while being convenient, cannot substitute for the human intelligence, knowledge and judgement, required to provide a meaningful interpretation of that data.

Ajit Abeysekera

J.Natn.Sci.Foundation Sri Lanka 2021 49 (3): 323 - 336
DOI: http://dx.doi.org/10.4038/jnsfsr.v49i3.10023

RESEARCH ARTICLE

Nitrogen fertiliser replacement by single and multi-strain rhizobial inoculants for black gram, green gram and soybean cultivation in Sri Lanka

CS Hettiarachchi[1*], CL Abayasekara[2], P Saravana Kumar[2], S Rajapakse[3], SA Kulasooriya[4], EMHGS Ekanayake[4], RKGK Kumara[4] and HMAC Gunaratne[5]

[1] *Department of Geography, Faculty of Arts,University of Peradeniya, Peradeniya.*
[2] *Department of Botany, Faculty of Science, University of Peradeniya, Peradeniya.*
[3] *Department of Molecular Biology and Biotechnology,University of Peradeniya, Peradeniya.*
[4] *National Institute of Fundamental Studies, Hantana Road, Kandy.*
[5] *Plenty Foods PLC, No 19, 3rd Lane, Rathmalana.*

Submitted: 16 June 2020; Revised: 23 November 2020; Accepted: 22 January 2021

Abstract: Various environmental, economic and health problems have arisen in the world due to the continuous application of N-fertilisers for crop production, especially in the third world countries. The current study was undertaken to develop effective rhizobial inoculants for three major legume crops in Sri Lanka, namely *Vigna mungo*, *Vigna radiata* and *Glycine max* to replace the application of nitrogen fertilisers. Rhizobial isolates were obtained from the root nodules of different cultivars of edible legumes, non-edible wild legumes and wild relatives of *Vigna* spp. Authentication and screening for effectiveness of the isolates were carried out, and five strains were selected as effective isolates and cross inoculated with the three legumes along with a stress tolerant strain, which was previously screened. A pot experiment was followed by a field trial in the dry zone of Sri Lanka under farmers' conditions as single and multi-strain inoculations. The results of the pot experiment indicated that the addition of rhizobial inoculants increased the growth performance in all treatments. In the field trial, both single and multi-strain inoculants gave significant increases in yield, compared to N-fertiliser application in all three crops, *viz*; an increase of the yield from 3 % to 39 % in *V. mungo*, 5 % to 14 % in *V. radiata*, and 4 % to 13 % in *G. max*. In conclusion, the current study has shown that single and multi-strain rhizobial inoculants are capable of completely replacing urea application to *V. mungo, V. radiata* and *G. max,* in Sri Lanka without any yield reduction.

Keywords: *Glycine max*, multi strain, N-fertiliser rhizobial inoculants, single-strain rhizobial inoculants, *Vigna mungo, Vigna radiata.*

INTRODUCTION

Rudimentary inoculation practices, such as moving soil from the previously cultivated fields, with well nodulated legumes, were recommended soon after Hellriegel's 1886 report on 'legumes could fix N_2'. The commercial use of pure cultures of rhizobia as inoculants was then patented in 1896 by Nobbe and Hiltner (Fred *et al*., 1932). Innovations in inoculant product formulations has led the establishment of inoculant manufacturing industries in Europe, North America and Australia. Despite the long history of legume inoculant use and development, only a few farmers in the developing countries of Asia have adopted inoculation into their legume cultivation practices. With time legume inoculation with rhizobia has gained importance in agricultural biotechnology (Aurora *et al*., 2017; Santos *et al*., 2019).

Examination of the legume inoculation technology identifies a potential for rapid penetration into the Asian agricultural input markets. The use of legume inoculation is inexpensive and requires little technical knowledge. The economic and environmental risk associated with inoculant use is minimal (Hettiarachchi *et al*., 2014). Biological nitrogen fixation by legumes is an essential important component of small-holder and low-input cropping systems commonly practiced in Asia. The

* Corresponding author (shantha@pdn.ac.lk; https://:orcid.org/0000-0002-5308-8785)

limited use of inoculant in Asia is not surprising because rhizobial inoculant production and marketing in many parts of Asia have been hampered by problems common to many developing countries. However, evidence is available to ascertain that many farmers could derive economic benefit by using inoculants (Ojiem *et al*., 2014; Chekanai *et al*., 2018; Rurangwa *et al*., 2018).

The overuse of chemical nitrogen fertilisers has caused an imbalance in the nitrogen cycle and aggravated the pollution in surface and groundwater. Increased loads of nitrogen fertiliser to freshwater and marine ecosystems have caused eutrophication (Khan *et al*., 2017). There is no doubt that the excessive use of chemical fertiliser and other agrochemicals is not healthy for the ecosystem, environment and human beings. Despite these problems, many farmers continue to use excessive amounts of fertilisers and toxic agrochemicals (pesticides and herbicides), with the false expectation of higher crop yield. Overuse or abuse of artificial fertilisers and other agrochemicals has increased during the last two decades and is widely used by farmers in Sri Lanka. Both the agrochemicals and petrochemicals act as environmental pollutants as well as a contributor of certain diseases, such as chronic kidney disease of unknown etiology (CKDu); a disease mainly affecting the North Central Province and spreading to other areas in Sri Lanka (Rajapakse *et al*., 2016), and cancer which is prevailing in several parts of Sri Lanka. Approximately 3.5 % of the government budget is currently being spent on the agro-subsidies.

In Sri Lanka, the Rhizobial Inoculant Production Facility, initiated at the Department of Botany, University of Peradeniya which has later moved to the National Institute of Fundamental Studies (NIFS), Kandy has reported that the rhizobial inoculants prepared with ground coir dust as the carrier material, could produce high yields of soybean compared to 50 kg per hectare of urea (maximum level recommended by the Department of Agriculture) (Kulasooriya *et al.*,2017). The selling price of a 250 g inoculum packet recommended for use for one acre is Rs. 400 (≈ $ 2). The usual rate of application of urea among soybean farmers is 40 kg/acre and is applied in two doses; 20 kg as basal and 20 kg as a top dressing. As of current market price, this costs Rs. 4250 (≈ $ 22) for a farmer, in addition to the cost of Rs. 3000 (≈ $15) for 5 labourers required for the application of fertiliser. In addition, the farmer has to bear transport cost of urea from the purchasing point to his field. It has also been observed that weed growth is reduced at least by 50 % in crop cultivations inoculated with rhizobial inoculants (Kulasooriya *et al*., 2017) compared to those fertilised with urea. A farmer could save another Rs. 4000 (≈ $ 20) per acre on the cost of agrochemicals needed for weed and pest control. Therefore, the use of the inoculants could make a total saving of Rs. 11,000 (≈ $ 55) per acre to a farmer compared to a field receiving the complete package of recommended agro-chemicals. The use of rhizobial inoculants produced by the NIFS is gaining popularity in Sri Lanka (Kulasooriya *et al*., 2017).

It is a challenge to develop a novel rhizobial inoculant that can promote higher levels of nitrogen fixation under practical field conditions. When the factors such as moisture, temperature, soil pH and salinity become extreme, improving yield becomes complicated (Giddens *et al*., 1982). Natural rhizobia of wild legumes growing under adverse conditions, such as salt stress, elevated temperatures and drought, generally exhibit higher tolerance to such conditions. The rhizobia of wild legumes in arid zones exhibit higher tolerance to prevailing adverse conditions, such as salt stress, elevated temperatures and desiccation (Zahran, 2001). Studies are needed to test the possibility of using effective strains isolated from black gram, green gram and soybean along with stress-tolerant rhizobia as inoculants to increase N-content of these edible legumes under stress conditions and the yield. Arora *et al.* (2017) reported that stress tolerant rhizobial species can be incorporated in developing bioformulations that can withstand salinity, drought and high temperatures. Although research has been undertaken on these lines, only a few contributed in developing inoculants. Ahmad *et al.* (2013) stated that halo-tolerant, auxin producing *Rhizobium* strains improve osmotic stress tolerance in mung bean. Further, Tewari and Arora (2014) recorded the use of exopolysaccharides (EPS) in bioformulation, as EPS protects inoculated rhizobial cells from stress factors such as salinity, desiccation and pH.

Multi-strain rhizobial inoculation of African Acacias under nursery conditions showed a significant increase in total nitrogen than that of the control plants in six out of the seven species (Sutherland *et al*., 2000). Significant increases in dry weight and total nitrogen over controls ranged from 19 % to 75 % and 11 % to 89 %, respectively. On the other hand, studies conducted on common bean (*Phaseolus vulgaris* L.) showed that three rhizobial strains evaluated were equally effective in the accumulation of total shoot N and that the multi-strain inoculant offered no consistent advantage over the single-strain inoculants. In the United States multi-strain inoculants are produced commercially (Burton *et al*., 1980) to provide a compensatory mechanism to theoretically meet the constraints imposed by the host-

strain-environment interactions, which is impossible or limited with single-strain inoculants. A very few systematic studies have been carried out to evaluate the performances of single-strain and multi-strain inoculants and limited information is available demonstrating the effects of different rhizobial strains for efficient nitrogen fixation.

There is evidence indicating that differential competition for nodule occupancy between strains of *Bradyrhizobium japonicum* in the presence of nitrate in sand cultures (McNeil *et al*., 1982). According to Somasegaran and Bohlool (1990) the nitrogen-fixing effectiveness of multistrain inoculants was found to be determined by both the effectiveness of the component strains and the percentage of the nodules occupied by them. Multistrain formulations were equally effective as most effective single-strain inoculants (Kyei-Boahen *et al*., 2005) or intermediate between the most and the least effective. The percentage of nodules occupied and the amount of nitrogen fixed by the component strains of a multi-strain inoculant showed highly significant linear correlation.

Black gram (*Vigna mungo* L. Hepper), a member of the Asian *Vigna* crop group is an annual pulse crop native to Central Asia. It is the staple crop in Central and south East Asia. This crop plays an important role in daily diets because of its high protein content (20–25 %), which is double of wheat and three times of rice (FAO, 1994). Green gram (*Vigna radiata*) seed is more palatable, nutritive, digestible and non-flatulent than other pulses grown. Its seeds contain 24.2 % protein, 1.3 % fat and 60.4 % carbohydrates and 118 mg and 340 mg of calcium and phosphorous, respectively per 100 g of seeds. It is rich in vitamin A and considered as a substitute for animal protein and forms a balanced diet when used with cereals (Consideine, 1992). Soybean (*Glycine max* L.) is the most important grain legume crop in the world in terms of total production and international trade. Soybean seeds contain 18 % to 23 % oil and about 38 % to 44 % protein (Hymowizt *et al*., 1998). These three crops are the most important grain legumes cultivated in the rain fed farming systems in dry and intermediate zones of Sri Lanka.

Comparative studies between single-strain and multi-strain inoculants have not been reported in Sri Lanka. A combination of effective high nitrogen fixing rhizobial strains with stress tolerant strains, if shown to be superior and applicable to a wider range of habitats, could be advantageous to any commercial inoculant producer. Therefore, the main objective of the current study was to evaluate the efficiency of single and multistrain rhizobial inoculants on black gram, green gram and soybean in Sri Lanka.

METHODOLOGY

Pot experiment under semi aseptic conditions

A pot experiment was carried out in a plant house under semi aseptic conditions, in the Department of Botany, Faculty of Science, University of Peradeniya, Sri Lanka (7°15'60.00" N 80°35'59.99" E). Three crop legumes, namely, *Vigna mungo* (black gram), *Vigna radiata* (mung bean) and *Glycine max* (soybean), were used in the study.

Preparation of pots

All the pots (black plastic: 43 cm diameter) were surface sterilised using bleaching powder (4 g) dissolved in water (100 mL), followed by rinsing with hot water. A rigifoam (polystyrene foam) disc of 1 cm thickness was placed at the bottom of a pot and filled with washed autoclaved river sand till the level of sand reached 1 inch from the top of the pot. Pots were filled with boiling water and drained 2 to 3 times to remove any soluble nitrogen present.

Seed preparation and germination

Vigna mungo, *V. radiata* and *G. max* seeds were surface sterilised separately by immersing in 70 % alcohol for 1–2 min followed by rinsing 5 times with sterilised water. Seeds were planted after sterilisation. A sterilised plastic tube (diameter ½ inch) was inserted in the middle to supply nutrients and water during the experiment. After planting of seeds the pots were covered using cling wrap prior to inoculation to prevent air borne contamination.

Strain selection

Five rhizobial strains (C8, C10, M5, M6 and VD1) isolated from different host plants (Table 1), which were previously screened (Hettiarachchi *et al*., 2014) as effective nitrogen fixing, and high nodulating strains along with K7 (isolated from *Vigna trilobata*), which was previously screened (Hettiarachchi *et al*., 2013) as a stress tolerant strain were used for inoculation as single and multi-strain inoculant combinations as described in

Table 1: Rhizobial strain isolated host plant, Gene Bank accession number of rhizobial strains, rhizobia cross inoculated crop plant and different combinations of rhizobia used for treatment

Host Plant	Rhizobial strains	Gene Bank Accession No	Crop species	Rhizobial combinations used
Mimosa pudica	M5 (ef)	KF008230	*V. mungo*	M5VD1K7, M5VD1, M5K7, VD1K7, M5, VD1, K7
Vigna dalzelliana	VD1 (ef)	KF008232		
Crotalaria brownie	C10 (ef)	KF008228	*V. radiata*	C10M6K7, C10M6, C10K7, M6K7, C10, M6, K7
Mimosa pudica	M6 (ef)	KF008231		
Crotalaria brownie	C8 (ef)	KF008227	*G. max*	C8VD1K7, C8VD1, C8K7, VD1K7, C8, VD1, K7
Vigna trilobata	K7 (st)			

* ef - effective st - stress tolerant

Table 1. Sequencing of 16S rRNA region was carried out for all effective isolates and submitted to the Gene Bank and Accession numbers were obtained. For K7, DNA fingerprinting was carried out using ERIC primer.

Inoculant preparation

Single-strain rhizobial inocula were cultured separately on ½ Lupin agar (LA) medium. The inocula were separately obtained using 1 % sucrose solution and the resulting rhizobial broths were separately transferred into autoclaved beakers. Multi-strain inoculants were prepared by adding equal volumes from each single-strain, resulting in equal volumes for each treatment.

Inoculation

After 3 days of seeding of the host plant, using a micropipette, 1 mL of the rhizobial inoculum was inoculated directly on the seedlings. As controls, the uninoculated and nitrogen controls were injected with 1 mL of distilled water. Pots were covered with sterilised gravel to prevent air borne contamination.

Experimental design

Four replicate pots were used with three plants per pot (12 plants) for each *Rhizobium* inoculum with nitrogen positive and negative controls. These were arranged in a complete randomised design (CRD).

Plant house conditions, nutrients and watering

Autoclaved water and nutrient solutions (60 mL, devoid of nitrogen) (N^-) were added to all the plants on designated days of the week. In addition, 5 mL of 1 % KNO_3 was added weekly to the nitrogen positive control (N^+) plants (Master class in Rhizobial Technology, 2012).

Plant and nodule assessment

After 8 weeks, the plants were visually rated (scale of 0–10, based on plant growth performance) and then harvested. The roots were carefully washed, and the nodules were detached, and wrapped with absorbent tissue paper and allowed to dry at room temperature. The nodules, shoots and roots were oven-dried at 70 °C for 48 h and weighed using an analytical balance (KeRn ABS - 220-4 No. WB 1210059).

Strain effectiveness

The effectiveness of the strains was calculated using the equation below (Fernando *et al.*, 2005). Dry mass (DM) of the strain-inoculated plant was compared with the N^- and N^+ controls.

$$\% \text{ Strain effectiveness} = \frac{[\text{DM of the inoculated plant} - \text{DM of the N- control}] \times 100}{[\text{DM of the N+ control} - \text{DM of the N- control}]}$$

Field trials in the dry zone

Field description

Field trials were carried out in the dry zone with the same combinations of rhizobial inoculants, as used for the pot experiments. *Vigna mungo* and *V. radiata* fields were in Bulagala dry-zone (7°54'0''N80°37'60''E) and *G. max* fields were located at in Galnawa dry zone (8°02' 02 " N80 °28' 45" E).

Experimental design

A randomised complete block design (RCBD) was used with three replicate blocks per treatment. The plot sizes were 2.74 × 1.52 m for *V. mungo*, 3.66 × 1.52 m for V. *radiata* and 4.27 × 1.83 m for *G. max*. The three rows of plants adjacent to the edge of each plot on all four sides were not considered when taking readings, in order to minimise the edge effect from all sides of each plot. Nitrogen positives and negatives were used as controls.

Single-strain and multi-strain rhizobial inoculum preparation

Single and multi-strain inoculants were prepared similar to the pot experiment. Rhizobial broth cultures were injected into autoclaved, powdered and packeted coir dust (Kulasooriya *et al*., 2007).

Seeding and agronomic practices

Seeds were mixed with the coir dust-based inoculum and sown (Kulasooriya *et al*., 2008). The recommended seed requirement was used. Plots were irrigated once a week and recommended agronomic practices were applied (Table 2). Before the addition of the basal dressing the plants were thinned according to the recommendations. A basal dressing of fertiliser urea (only for the N^+ control), Triple Super Phosphate (TSP) and Muriate of Potash (MOP) (to all the treatments) was added 12 days after sowing. A top dressing of urea was added at flowering, only to the nitrogen positive control.

Nodule and dry weight assessment

Before harvesting (after 8 weeks), 30 plants were visually rated (scale of 0–10, based on plant growth performance). The roots were carefully washed and the nodules were detached and wrapped with absorbent tissue paper and dried at room temperature. The nodules, shoots and roots were oven-dried at 70 °C for 48 h and weighed.

Yield data

Yield and yield component data of remaining plants (*V. mungo* ~ 80 plants, *V. radiata* ~ 100 plants and *G. max* ~ 200 plants) were recorded after 80–90 days of planting.

Statistical analysis

Data on nodule number (NN), nodule dry weight (ND), shoot dry weight (SD), root dry weight (RD), total dry weight (TD), yield and yield component data were subjected to analysis of variance (ANOVA) followed by Duncan's Multiple Range Tests (DMRT) to separate the treatment means.

RESULTS AND DISCUSSION

Pot experiment

Vigna mungo

Morphological differences were observed within treatments showing variations in the effectiveness of the rhizobial treatment. The highest average visual rate (AVR) of nine was observed with multi-strain M5K7 treatment, whereas the lowest (seven) was with the N minus control (N^- control) and rest of the treatments showed a value of eight (Supplementary Table 1). A significant increase in nodulation (ANN) and nodule dry weight (AND) was observed with the multi-strain M5K7 compared to the rest of the treatments. All the rhizobial inoculated treatments, including the N+ control, showed no significant difference in average shoot dry weight (ASD), Average Root Dry weight (ARD) and total dry matter production. These values were significantly higher than the N^- control.

Vigna radiata

Single-strain C10 showed the highest value for AVR (nine), ANN, AND, ASD and ARD (Supplementary Table 2). When considering the total dry matter production, single-strain C10 gave the highest value, which was not significantly different from the value obtained by K7 treatment.

Glycine max

Multi-strain VD1K7 showed the highest value (nine) and the N- showed the lowest value (six) with respect to AVR. With respect to ANN, AND, ASD, ARD and total dry weight, multi-strain VD1K7 gave significantly higher values than all other treatments (Supplementary Table 3).

Table 2: Seedling and agronomic practices: requirements, recommendations

Host Plant	Seed requirement	Spacing between plants	Basal fertiliser (kg/h)	Top dressing (only to N+ control)
V. mungo	30 kg/ha	40 cm × 10 cm	TSP 100, MOP 75, Urea 30 (only to N+ control)	Urea 35 kg/h
V. radiata	30 kg/ha	40 cm × 10 cm	TSP 100, MOP 75, Urea 30 (only to N+ control)	Urea 35 kg/h
G. max	55 kg / ha	40 cm × 5 cm	TSP 150, MOP 75, Urea 50 (only to N+ control)	Urea 50 kg/h

The Strain Effectiveness (SE) values indicated that all the treatments resulted in a high level of effectiveness in *Vigna mungo* (Table 3), according to Beck *et al*. (1993). SE values of the isolates were rated as highly effective (> 80 %), effective (50 – 80 %) and ineffective (< 35%). *Vigna radiata* SE values showed that other than three treatments, all other treatments had high effectiveness. With regards to *G. max*, except C8VD1K7 treatment, all other treatments showed high effectiveness.

The pot experiment was mainly done in order to evaluate the infective ability of the single-strain (SS) and multi strain (MS) inoculants. All the tested inoculants produced nodules in the three crops under investigation. There was no nodulation in the uninoculated controls in the pot experiment, demonstrating that aseptic conditions were met in the experimental set up and maintenance of the plants in the greenhouse was adequate (Bala *et al.*, 2003). Overall, the multi-strain inoculants produced more nodules than the single-strain inoculants in the pot experiments

Similar to previous studies, host biomass production in the pot experiment was used as the criterion for strain effectiveness in N_2 fixation. Compared to N^- application, in all three crops, dry weights increased in inoculated treatments. Hoben (1994) and Peoples *et al.* (2002) explained that shoot dry matter is a good indicator of relative isolate effectiveness. The current results show that the same or higher dry matter could be obtained by rhizobial inoculation without adding N fertilisers.

Field trial

The pot experiments were conducted under aseptic conditions in a greenhouse whereas in the field trials, the inoculant strains had to compete with the indigenous rhizobia and other indigenous microbes present in the soil. Although some strains are effective in N_2 fixation, they may not be able to compete with the indigenous rhizobia and other soil microorganisms for substrates and space in most locations (Santos *et al.*, 2019). Also, the naturally occurring rhizobium populations often occur in high numbers in soil and can compete strongly with the introduced rhizobium inoculants. Better N_2 fixation can be achieved by selecting superior rhizobia. However, the selection of these rhizobia would need to take into consideration not only their N_2 fixing capacity but also the competitive ability against native rhizobia, which are frequently ineffective in N_2 fixation (Hettiarachchi *et al*., 2014). Superior N_2-fixing strains have to outcompete native rhizobia and occupy a significant proportion in nodules. For this to be achieved, rhizobia have to be selected under natural conditions in competition with the native rhizobia (Rengel, 2002) emphasizing the necessity of conducting a field trial.

Plant growth parameters

Vigna mungo

With respect to AVR values, single-strain VD1 and multi-strain M5K7 gave the highest value (nine), whereas the rest of the treatments gave a similar value (eight). With respect to ANN, multi-strain M5VD1K7 and K7 gave significantly higher values than both N controls. When considering the AND values, M5VD1K7 and M5VD1 gave significantly higher values than the N controls (Supplementary Table 4). Multi-strain M5K7 gave a significantly higher ASD value than the N controls. When looking at the ARD values, no significant difference was observed within the treatments, whereas multi-strain M5K7 and single-strain VD1 gave the highest values. Although the multi-strain M5K7 and the single-strain

Table 3: Percentage strain effectiveness (SE) based on total dry matter production of the targeted legume crops in comparison with $^{N+}$ and $^{N-}$ controls (pot experiment).

V. mungo		*V. radiata*		*G. max*	
Strain	%SE	Strain	%SE	Strain	%SE
M5VD1K7	95.71	C10M6K7	91.37	C8VD1K7	39.53
M5VD1	114.87	C10M6	19.32	C8K7	198.97
M5K7	148.58	C10K7	116.87	C8VD1	126.27
VD1K7	95.22	M6K7	26.67	VD1K7	462.18
M5	89.29	C10	245.49	C8	93.84
VD1	103.36	M6	72.83	VD1	219.08
K7	106.28	K7	172.16	K7	198.91
N+	100	N+	100	N+	100
N -	0	N-	0	N-	0

VD1 gave comparatively higher values with respect to total dry matter production, no significant difference was observed within the treatments.

Vigna radiata

Except the multi-strain C10M6, the rest of the rhizobial strain inoculated treatments gave higher AVR values than both N controls, in which multi-strain C10M6K7, C10K7 and M6K7 and single-strain M6 gave the highest value of nine (Supplementary Table 5). Multi-strain C10M6K7 gave the highest values for both ANN and AND, which were significantly higher than both N controls. Multi-strain C10K7 gave the highest value for ASD and ARD and is again significantly higher than both N controls. No significant difference was observed in the total dry matter production, in which multi-strains C10M6K7 and C10K7 gave the highest values.

Glycine max

In *G. max*, multi-strain C8K7 and single-strain VD1 gave the highest values for AVR (9 and 8.5 respectively). Multi-strain VD1K7 gave the highest value for ANN, which was significantly higher than the rest of the treatments (Supplementary Table 6). Some of the rhizobial strain inoculated treatments gave significantly higher values than both controls with respect to AND values. Only the multi-strain C8K7 gave significantly higher values for ASD, ARD and total dry matter production than the rest of the treatments. Others showed significantly similar values.

In the current study, differences in Average Number of Nodules (ANN) were observed for the same treatment in the pot and field experiments. ANN results showed that only two rhizobial inoculants gave lower values than the control values, indicating that these were not able to compete with the native soil rhizobial population. Similar to the previous field trials, with N^+ treatment, only a few number of nodules were observed indicating the negative effect of nitrogen fertiliser application on nodulation of the legume plants (Crews *et al.*, 2004; Xuan *et al*., 2017). Nodule dry weight is essential in strain evaluation, as it serves as an indicator for symbiotic efficiency (Graham *et al.*, 2004).

With reference to AVR values, most of the time, multi- strain inoculants were superior to the single-strain inoculants in both pot and field experiments. This could be attributed to the multi-strain's ability to effectively nodulate and enhance solubilisation of other essential soil minerals, such as phosphorus (Koskey *et al*., 2017). The rhizobial strain inoculated treatments always showed higher AVR values compared to the N^- control. In some instances, even higher than the N^+ control. Since AVR values correlate with other growth and yield parameters investigated, it was found that the growth performance has increased with the addition of both single-strain and multi-strain rhizobial inoculants. Zablotowicz *et al*. (1991) found that increasing rhizobial diversity enhances the shoot dry weight in bean plants. In some trials, there were no significant differences among single-strain and multi strain treatments with respect to the total dry matter production. When considering the infectivity and effectivity in these inoculants with respect to the pot experiment, in *V. mungo* although the infectivity varied within the inoculants the effectivity was fairly high in all the inoculants. In *V. radiata* one single-strain inoculant and in *G. max* one multi strain inoculant showed high infectivity as well as high effectivity. In the field trial, in *V. mungo* both single-strain and multi strain inoculants seems to be effective. In *V. radiata* although the single-strain inoculant performed better in the pot experiment, the multi strain inoculants performed better in the field trial. With respect to *G. max* the multi strain inoculants seem to be better than the single-strain inoculants. A proper combination of different infective and effective Rhizobium strains could enhance nodule, occupancy hence biological fixation of nitrogen. Hungria *et al*. (2000) noted that a combination of specific rhizobia strains, performs better in promoting N^-fixation and growth of different bean cultivars as compared to the use of individual rhizobia strains. These results support the claims made by Kawaka *et al*. (2014) Korir *et al*., (2017) and Koskey *et al*. (2017), that there is a direct association among nodulation, plant growth and nitrogen accumulation in legume plants.

Yield data

Vigna mungo

Multi-strain M5K7 and the single-strain VD1 gave significantly higher NP values compared to the N^+ treatment (Table 4). Most of the rhizobial strain inoculated treatments other than M5 gave significantly higher PL values than the N^- control. Although the values were not significantly higher, most of the rhizobial strain inoculated treatments gave higher number of seeds per pod than both controls. All the rhizobial strain inoculated treatments gave a higher yield than both controls (N^+ and N^-). Significantly higher yield values were obtained from multi-strains M5VD1 and M5K7, and single-strains VD1 and K7.

Vigna radiata

Other than C10, all other rhizobial strain inoculated treatments gave significantly higher values than both controls with respect to NP values (Table 4). When considering PL values, only multi-strain M6K7 and the single-strain K7 showed significantly higher values than the N^- control. When considering the seed yield, all except single-strain C10 gave numerically higher values than the N^+ control. However, only the multi-strain C10K7 and single-strain K7 gave statistically significant higher values.

Glycine max

All the treatments gave significantly higher values than the N- control with respect to NP values, among which multi-strains C8K7 and VD1K7 gave significantly higher values (Table 4). With respect to PL all the treatments gave similar values, which were significantly higher than the N^- control. The multi-strain C8K7 showed a significantly higher number of seeds/pods than the rest of the treatments. All the treatments showed significantly higher value than the N- control, with respect to 100 seed weight.

Table 4: Yield and yield component data

Strain	NP	PL(cm)	No. seeds per Pod	100 seed weight (g)	Seed yield /plant (g)	Estimated seed yield (kg/ha)
A) *V. mungo*						
M5VD1K7	32.71^{bc}	4.74^{a}	7.10^{a}	4.78^{ab}	9.25^{c}	2398.25^{c}
M5VD1	39.29^{ab}	4.85^{a}	7.20^{a}	5.41^{a}	11.96^{a}	3101.44^{a}
M5K7	41.43^{a}	5.17^{a}	7.21^{a}	4.99^{ab}	11.05^{a}	2864.92^{a}
VD1K7	36.86^{ab}	5.04^{a}	7.12^{a}	4.55^{ab}	9.43^{b}	2444.50^{b}
VD1	41.14^{a}	5.13^{a}	6.34^{ab}	5.61^{a}	11.48^{a}	2976.22^{a}
M5	32.29^{bc}	4.40^{ab}	7.50^{a}	4.42^{ab}	8.82^{c}	2285.81^{c}
K7	36.00^{b}	5.089^{a}	7.45^{a}	4.92^{ab}	10.72^{a}	2779.86^{a}
N+	31.56^{bc}	4.74^{a}	6.58^{ab}	4.13^{b}	8.56^{c}	2218.82^{c}
N-	28.11^{c}	4.23^{b}	6.17^{ab}	3.93^{b}	7.48^{d}	1938.53^{d}
B) *V. radiata*						
C10M6K7	21.86^{ab}	7.06^{ab}	7.25^{ab}	5.57^{a}	8.73^{ab}	2183.00^{ab}
C10M6	21.29^{ab}	7.05^{ab}	6.44^{b}	5.00^{ab}	8.43^{ab}	2108.82^{ab}
C10K7	22.63^{ab}	6.89^{ab}	6.22^{b}	5.41^{a}	9.01^{a}	2252.00^{a}
M6K7	21.37^{ab}	7.87^{a}	8.45^{a}	5.03^{ab}	8.58^{ab}	2144.60^{ab}
C10	20.14^{bc}	6.03^{b}	6.57^{b}	4.51^{b}	7.74^{bc}	1935.00^{bc}
M6	22.71^{a}	6.66^{ab}	6.68^{b}	5.00^{ab}	8.44^{ab}	2109.18^{ab}
K7	22.29^{ab}	7.68^{a}	7.60^{a}	5.87^{a}	9.21^{a}	2303.73^{a}
N+	19.67^{c}	7.12^{ab}	8.01^{a}	4.40^{b}	8.02^{bc}	2006.00^{bc}
N-	17.44^{d}	6.57^{b}	6.52^{b}	3.51^{c}	6.80^{d}	1699.50^{d}
C) *G.max*						
C8VD1K7	49.33^{c}	3.25^{a}	2.10^{c}	14.45^{ab}	16.64^{e}	4402.80^{e}
C8VD1	47.22^{d}	3.24^{a}	2.19^{a}	14.95^{ab}	16.86^{e}	4459.55^{e}
C8K7	62.66^{a}	3.03^{a}	2.18^{a}	16.92^{a}	20.42^{a}	5401.56^{a}
VD1K7	60.22^{a}	3.24^{a}	2.10^{c}	16.37^{a}	20.31^{a}	5373.60^{a}
C8	57.77^{b}	3.10^{a}	2.10^{c}	16.04^{a}	18.94^{b}	5011.32^{b}
VD1	52.77^{bc}	3.13^{a}	2.05^{d}	15.91^{a}	17.63^{d}	4663.47^{d}
K7	50.75^{c}	3.11^{a}	2.15^{b}	15.98^{a}	18.84^{b}	4985.03^{b}
N+	55.00^{b}	3.03^{a}	2.11^{c}	15.03^{ab}	18.07^{cd}	4779.52^{cd}
N-	44.29^{e}	2.10^{b}	2.03^{d}	13.07^{C}	14.60^{f}	3859.89^{f}

NP = Average number of pods; PL = average pod length values in the same column [separately for sections A), B) and C)] followed by the same letter are not significantly different at 5 % probability level.

Multi-strain C8K7 and VD1K7, and single-strain C8 and K7 gave significantly higher yield values compared to the N^+ control.

Comparison of yield performance considering all three crops investigated under inoculation with fertiliser applications ascertains that certain strains (both single and multi) had given higher responses for the three crops tested. For *V. mungo*, all the inoculants gave higher values than the N^+ fertiliser application in which two multi strain inoculants (M5VD1 and M5K7) and two SS inoculants (VD1 and K7) gave comparatively higher values than the N^+ fertiliser application. In *V. radiata*, only one single-strain inoculant (VD1) gave a lower value than N^+ fertiliser application. In *G. max*, two multi-strain inoculants (M5VD1K7 and M5VD1) and one single-strain inoculant gave a lower value than N^+ fertiliser application (Figure 1).

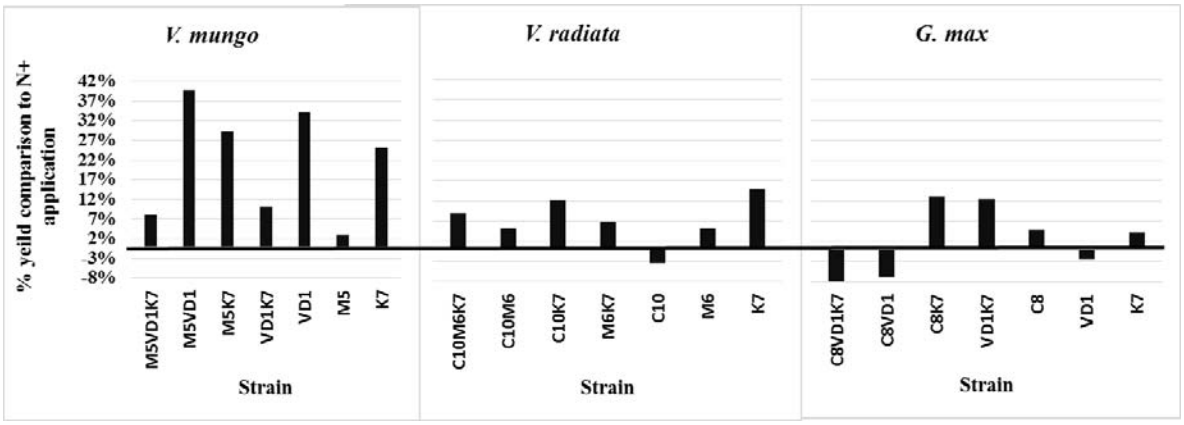


Figure 1: Percentage yield increase/decrease in comparison to N^+ fertiliser application (zero line indicates N^+ fertilizer)

According to the results obtained, both the single-strain and multi-strain inoculants are capable of completely replacing N^- fertiliser application. However, multi-strain inoculants seem to be superior to single-strain inoculants (Figure 2).

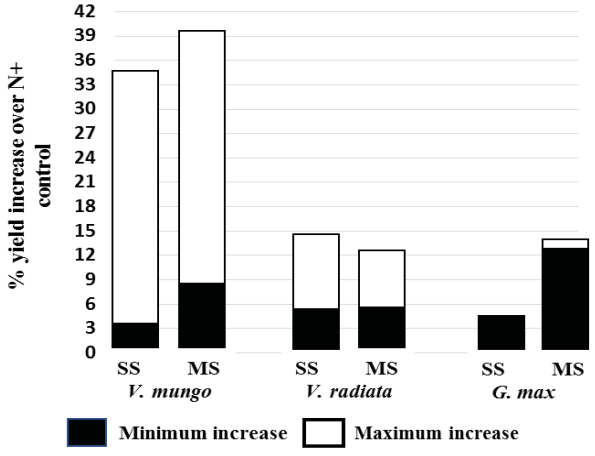


Figure 2: Yield comparison between single-strain and multi strain rhizobial inoculants

In Australia, in early studies with rhizobial inoculation, only single-strain inoculants were used as inoculants to prevent possible dominance and antagonistic effects of a particular strain in the mixture (Schwinghamer *et al*., 1979) to diagnose loss of effectiveness, and to facilitate quality control (Thompson *et al*., 1980). However, in the United States, multi strain inoculants were produced commercially (Burton *et al*., 1980) to provide a compensatory mechanism to theoretically meet the constraints imposed by the host-strain-environment interactions, which is impossible with single-strain inoculants.

Unlike many free-living diazotrophs, rhizobia are able to exhibit highly efficient nitrogen fixation, only when they are in the host nodule cells in endosymbiotic form as bacteroid (Hakoyama *et al*., 2009). Rhizobial nitrogen-fixing activity is restricted to symbiotic bacteroids, and free-living rhizobia do not fix atmospheric nitrogen, a feature unique to legume/rhizobium symbiosis (Kneip *et al*., 2007).

It appears that the effectiveness of mixed inoculants is dependent on the effectiveness and competitiveness of the strains in the mixture. Although much work has been documented in rhizobial ecology in evaluating the success of inoculant strains (measured as percent nodule occupancy) in numerous competition experiments, meaningful interpretations of nodule occupancies in relation to the nitrogen fixed by the nodule occupants have apparently not been quantified. Furthermore, it is noted that little information has been reported on rhizobial inoculants in the past decade (Aurora *et al*., 2017) and this limits a thorough discussion on this area of study.

Alhough the increase of nodulation through rhizobial inoculation resulted in elevated growth performances and enhanced yield, multi locational field testing over two or more seasons is needed before these inoculants are recommended for farmer use. Santos *et al*. (2019) states that farmers are more receptive to use of inoculants due to high-quality products and the availability of multi-strains, which cost less than chemical fertilisers. In the context of sustainable agriculture, microbial inoculants play a major role to alleviate the negative environmental impact caused by chemical fertilisers (Santos *et al*., 2019).

CONCLUSIONS

In conclusion, the current study has shown that the addition of rhizobial inoculants, both single-strain and multi-strain has completely replaced urea applications to *V. mungo*, *V. radiata* and *G. max* crop cultivation in Sri Lanka while increasing the yields of all three crops investigated, significantly.

Acknowledgements

National Science Foundation of Sri Lanka (Research grant no. RG/ 2008/SUNR/01) and the National Research Council of Sri Lanka (Research grant no. NRC 11-121) gratefully acknowledged for their financial assistance.

REFERENCES

Ahmad M., Zahir Z.A. & Nazli F. (2013). Effectiveness of halo-tolerant, auxin producing Pseudomonas and Rhizobium strains to improve osmotic stress tolerance in mung bean (*Vigna radiata* L.). *Brazilian Journal of Microbiology* **44**(4): 1341–1348.
DOI: https://doi.org/10.1590/S1517-83822013000400045

Arora N.K., Verma M. & Mishra J. (2017). Rhizobial bioformulations: past, present and future. In: *Microorganisms for Sustainability* (ed. S. Mehnaz), pp. 69–75. Springer Nature Singapore Pte Ltd., Singapore.
DOI: https://doi.org/10.1007/978-981-10-4862-3_4

Bala A., Murphy P. & Giller K.E. (2003). Distribution and diversity of rhizobia nodulating agroforestry legumes in soils from three continents in the tropics. *Molecular Ecology* **12**(4): 917–930.
DOI: https://doi.org/10.1046/j.1365-294X.2003.01754.x

Beck D.P., Materon L.A. & Afandi F. (1993). *Practical Rhizobium sp. Legume Technology Manual,* pp. 1–54. International Center for Agricultural Research in the Dry Areas (ICARDA), Aleppo, Syria.

Burton J.C. & Martinez C.J. (1980). *Rhizobial Inoculants for Various Leguminous Species*. Nitragin Company Technology Bulletine. No. 101. Nitragin Co., Milwaukee, Wisconsin, USA.

Chekanai V., Chikowo R. & Vanlauwe B. (2018). Response of common bean (Phaseolus vulgaris L.) to nitrogen, phosphorus and rhizobia inoculation across variable soils in Zimbabwe. *Agriculture, Ecosystems & Environment* **266**(01): 167–173.
DOI: https://doi.org/10.1016/j.agee.2018.08.010

Consideine M. (1992). Effect of urea on photosynthesis and yield in Mung bean. *Journal of Agronomy and Crop Science* **168**(01): 91–94.
DOI:https://doi.org/10.1111/j.1439-037X.1992.tb00983.x

Crews T.E. & Peoples M.B. (2004). Legumes versus fertilizer sources of nitrogen: ecological trade-offs and human needs. *Agriculture, Ecosystems and Environment* **102**(3): 279–297.
DOI: https://doi.org/10.1016/j.agee.2003.09.018

FAO (1983). *Technical Hand Book on Symbiotic Nitrogen Fixation: Legume-Rhizobium*, pp. 3–4. The Food & Agricultural Organization of the United Nations, Rome, Italy.

Fernando G.A. (2005). The rhizobia nodulating shrubs for revegetation of arid lands: isolation of native strains and specificity of the plant - rhizobia interaction by cross inoculation tests. *Arid Land Research and Management* **19**(01): 307–326.
DOI: https://doi.org/10.1080/15324980500299649

Fred E.B., Baldwin I.L. & McCoy E. (1932). *Root Nodule Bacteria and Leguminous Plants*, pp. 343. University of Wisconsin Press, Madison, WI, USA.

Giddens J. E., Dunigan E.P. & Weaver R.W. (1982). Legume inoculation in the Southeastern U.S.A. *The University of Georgia, College of Agriculture Experimental Station Southern Cooperative Service Bulletin* **283**: 1–38.

Graham P.H., Hungria M. & Tlusty B. (2004). Breeding for better nitrogen fixation in grain legumes: where do the rhizobia fit in? *Crop Management* Review **3**(1): 1–6.
DOI: https://doi.org/10.1094/CM-2004-0301-02-RV

Hakoyama T. *et al*., (19 authors) (2009). Host plant genome overcomes the lack of a bacterial gene for symbiotic nitrogen fixation. *Nature* **462**: 514–517.
DOI: https://doi.org/10.1038/nature08594

Hellriegel H. & Wilfarth H. (1886). Erfolgt die assimilation des freien Stick & offs durch die Leguminose nunter

Mitwirkung Niederer Organismen. *Berichte der Deutschen Botanischen Gesellschaft* **7**: 138–143.

Hettiarachchi C.S., Abayasekara C.L., Rajapakse S., Kulasooriya S.A. & Saravana Kumar P. (2013). Screening of stress tolerant rhizobial isolates from wild legumes growing in the southern coastal region of Sri Lanka. *Proceedings of the 33rd Annual Sessions of the Institute of Biology, Sri Lanka.* Colombo, Sri Lanka, September, pp. 49.

Hettiarachchi C.S., Kumar P.S., Abayasekara C.L., Rajapakse S., Kulasooriya S.A., Ekanayake E.M.H.G.S. & Kosala R.K.G.K. (2014). Response of *Glycine max* to inoculation with rhizobial strains isolated from crop wild relatives of *Vigna* spp., *Crotalaria* spp. and *Mimosa* spp. *Jaffna University International Research Conference* (JUICE) *Full Paper Proceedings publications*. *Jaffna University, Jaffna, Sri Lanka.* March, pp. 258–262.

Hettiarachchi C.S., Abayasekara C.L., Kumar P.S., Rajapakse S., Kulasooriya S.A., Ekanayake E.M.H.G.S. Kosala, R.K.G.K. & Gunaratna H.M.A.C. (2014). Single-strain and multi-strain rhizobial inoculants for black gram (*Vigna mungo*) cultivation. *Proceedings of the Postgraduate Institute of Science, Research Congress* (PIGS). University of Peradeniya, Sri Lanka, 10–11 October, pp. 18.

Hoben H.J., Somasegara, P., Boonkerd N. & Gaur Y.D. (1994). Polyclonal antisera production by immunization with mixed cell antigens of different rhizobial species. *World Journal of Microbiology and Biotechnology* **10**: 538–542. DOI: https://doi.org/10.1007/BF00367662

Hungria M., Andrade D.S., Chueire L.M.O., Probanza A., Guttierrez-Mañero F.J. & Megías M. (2000). Isolation and characterization of new efficient and competitive bean (*Phaseolus vulgaris* L.) rhizobia from Brazil. *Soil Biology and Biochemistry* **32**(11-12): 1515–1528. DOI: https://doi.org/10.1016/S0038-0717(00)00063-8

Hymowitz T., Singh R.J. & Kollipara K.P. (1998). The genomes of the *Glycine*. In: *Plant Breeding Reviews* (ed. J. Janick), pp. 289–311. Wiely Online Library. DOI: https://doi.org/10.1002/9780470650110.ch8

Kawaka F., Dida M. M., Opala P.A., Ombori O., Maingi J., Osoro N. & Muoma J. (2014). Symbiotic efficiency of native rhizobia nodulating common bean (*Phaseolus vulgaris* L.) in soils of Western Kenya. *International Scholarly Research Notices* **2014**(1): 1–8. DOI: https://doi.org/10.1155/2014/258497

Kyei-Boahen S., Nleya T., Hynes R., & Walley F.L. (2005). Single and multistrain rhizobial inocula for pinto and black bean cultivars. *Journal of Plant Nutrition* **28**(10): 1679–1692. DOI: https://doi.org/10.1080/01904160500250664

Khan M.N., Mobin M. & Abbas Z.K. (2017). Fertilizers and their contaminants in soils, surface and groundwater. *Earth Systems and Environmental Sciences* **5**: 225–240. DOI: https://doi.org/10.1016/B978-0-12-809665-9.09888-8

Kneip C., Lockhart P., Voß C. & Maier U.G. (2007). Nitrogen fixation in eukaryotes - new models for symbiosis. *BMC Evolutionary Biology* **7**: 55–66. DOI: https://doi.org/10.1186/1471-2148-7-55

Korir H., Mungai N. W., Thuita M., Hamba Y. & Masso C. (2017). Co-inoculation effect of rhizobia and plant growth promoting rhizobacteria on common bean growth in a low phosphorus soil. *Frontiers in Plant Science* **8**:141. DOI: https://doi.org/10.3389/fpls.2017.00141.

Koskey G., Mburu S.W., Njeru E.M., Kimiti J.M., Ombori O. & Maingi J.M. (2017). Potential of native rhizobia in enhancing nitrogen fixation and yields of climbing beans (*Phaseolus vulgaris* L.) in contrasting environments of Eastern Kenya. *Frontiers in Plant Science* **8**: 443. DOI: https://doi.org/10.3389/fpls.2017.00443

Kulasooriya S.A., Seneviratne G. & Ekanayake E.M.H.G. S. (2017). Soil microbial diversity and its utilization in agriculture in Sri Lanka. In: *Microbial Biotechnology, Applications, Agriculture and Environment* (eds J. K. Patra, C. N. Vishnuprasad & G Das), pp. 203-224. Springer Nature Publishers, Singapore. DOI: https://doi.org/10.1007/978-981-10-6847-8_9

Kulasooriya S.A. (2008). *Biological Nitrogen Fixation: Fundamentals and Utilization.* Peradeniya Science Publication No 27. pp. 1-143. Science Education Unit, Faculty of Science, University of Peradeniya, Sri Lanka.

Kulasooriya S.A., Ekanayake E.M.H.G.S. & Kosala Kumara, R.K.G. (2007). *Application of Rhizobial Inoculants for Pulse Crop Cultivation in Sri Lanka*. Workshop on Agricultural Technologies, Council for Agricultural Research Policy May 20, CARP Headquarters, Colombo, Sri Lanka.

Kyei-Boahen S., Nleya T., Hynes R.K. & Walley F.L. (2005). Single and multi-strain rhizobial inocula for pinto and black bean cultivars. *Journal of Plant Nutrition* **28**(10):1679–1692. DOI: https://doi.org/10.1080/01904160500250664.

Master class in rhizobial technology (2012). *The Isolation, Identification and Utilization of Root Nodule Bacteria (rhizobia) in Promoting Sustainable Agricultural Productivity*. Institute of Fundamental Studies Kandy, Sri Lanka and Murdoch University, Western Australia.

McNeil D. L. (1982). Variations in ability of *Rhizobium japonicum* strains to nodulate soybeans and maintain fixation in the presence of nitrate. *Applied and Environmental Microbiology* **44**:647–652. DOI: https://doi.org/10.1128/aem.44.3.647-652.1982

Ojiem J. O., Franke A. C., Vanlauwe B., De Ridder N., & Giller K. E. (2014). Benefits of legume–maize rotations: Assessing the impact of diversity on the productivity of smallholders in Western Kenya. *Field Crops Research* **168**: 75–85. DOI: https://doi.org/10.1016/j.fcr.2014.08.004

Peoples M.B., Herridge D.F. & Ladha J.K.(2002). Biological nitrogen fixation: An efficient source of nitrogen for sustainable agricultural production. *Plant and Soil* **174**: 3–28. DOI: https://doi.org/10.1007/BF00032239

Rajapakse S., Mitrakrishnan C.S. & Selvarajah M. (2016). Chronic kidney disease of unknown etiology in Sri Lanka. *International Journal of Occupational and Environmental Health* **22**(3):259-264. DOI: https://doi.org/10.1080/10773525.2016.1203097

Rengel Z. (2002). Breeding for better symbiosis. *Plant and Soil* **245**: 147–162.
DOI: https://doi.org/10.1023/A:1020692715291

Rurangwa E., Vanlauwe B. & Giller K. E. (2018). Benefits of inoculation, P fertilizer and manure on yields of common bean and soybean also increase yield of subsequent maize. *Agriculture, Ecosystems and Environment* **261**: 219–229.
DOI: https://doi.org/10.1016/j.agee.2017.08.015

Santos M.S., Nogueira M.A. & Hungria M. (2019). Microbial inoculants: reviewing the past, discussing the present and previewing an outstanding future for the use of beneficial bacteria in agriculture. *AMB Express* **9**(1): 205.
DOI: https://doi.org/10.1186/s13568-019-0932-0

Schwinghamer E. A. & Brockwell J. (1978). Competitive advantage of bacteriocin- and phage-producing strains of Rhizobium trifohi in mixed culture. *Soil Biology and Biochemistry* **10**:383–387.
DOI: https://doi.org/10.1016/0038-0717(78)90062-7

Somasegaran P. & Ben Bohlool B. (1990). Single-Strain versus multi strain inoculation: effect of soil mineral N availability on rhizobial strain effectiveness and competition for nodulation on chick-pea, soybean, and dry bean. *Applied and Environmental Microbiology* **56**(11): 3298–3303.
DOI: https://doi.org/10.1128/aem.56.11.3298-3303.1990

Sutherland J.M., Odee D.W., Muluvi G.M., McInroy S.G. & Patel A. (2000). Single and multi-strain rhizobial inoculation of African acacias in nursery conditions. *Soil Biology and Biochemistry* **32**(3): 323–333.
DOI: https://doi.org/10.1016/S0038-0717(99)00157-1

Tewari S. & Arora N.K. (2014). Talc based exopolysaccharides formulation enhancing growth and production of *Hellianthus annuus* under saline conditions. *Cellular and Molecular Biology* **60**(5):73–81.

Thompson J. A. (1980). Production and quality control of legume inoculants. In: *Methods for Evaluating Biological Nitrogen Fixation* (ed. F. J. Bergersen), pp. 489-533. John Wiley & Sons, Inc., New York USA.

Xuan X., Chunmei M., Shoukun D., Yao X. & Zhenping G. (2017). Effects of nitrogen concentrations on nodulation and nitrogenase activity in dual root systems of soybean plants. *Soil Science and Plant Nutrition* **63**(5): 470–482.
DOI: 10.1080/00380768.2017.1370960.

Zablotowicz R. M., Tipping E.M., Lifshitz R.& Kloepper J. W. (1991). Plant growth promotion mediated by bacterial rhizosphere colonizers. In: *The rhizosphere and plant growth*, pp. 315–326. Springer, Dordrecht.
DOI: https://doi.org/10.1007/978-94-011-3336-4_70

Zahran H.H. (2001). Rhizobia from wild legumes: diversity, taxonomy, ecology, nitrogen fixation and biotechnology. *Journal of Biotechnology* **91**(2): 143–153.
DOI: https://doi.org/10.1016/S0168-1656(01)00342-X

Supplementary data

Supplementary Table 1: Plants visual rate, nodule number, nodule dry weight, shoot dry weight and root dry weight of *V. mungo*

STRAIN	AVR	ANN	AND	ASD	ARD
M5VD1K7	8.00	5.9701^{c}	0.0097^{bc}	0.6560^{a}	0.0937^{a}
M5VD1	8.00	11.6667^{b}	0.0133^{b}	0.6939^{a}	0.0991^{a}
M5K7	9.00	18.3333^{a}	0.0198^{a}	0.7605^{a}	0.1087^{a}
VD1K7	8.00	8.0000^{bc}	0.0122^{b}	0.6550^{a}	0.0936^{a}
M5	8.00	4.2812^{c}	0.0095^{bc}	0.6433^{a}	0.0919^{a}
VD1	8.00	11.7252^{b}	0.0130^{b}	0.6711^{a}	0.0958^{a}
K7	8.00	9.3521b^{c}	0.1200^{b}	0.6769^{a}	0.0967^{a}
N+	8.00	0.0000^{d}	0.0000^{d}	0.6645^{a}	0.0949^{a}
N-	7.00	0.0000^{d}	0.0000^{d}	0.4668^{b}	0.0667^{b}

AVR = average visual rate, ANN = average no. nodules/plant, AND = average nodule dry weight (g/plant), ASD = average shoot dry weight (g/plant), ARD = average Root dry weight (g/plant). Values in the same column followed by the same letter are not significantly different at 5 % probability level.

Supplementary Table 1: Plants visual rate, nodule number, nodule dry weight, shoot dry weight and root dry weight of *V. radiata*

STRAIN	AVR	ANN	AND	ASD	ARD
C10M6K7	8	12.8333^{c}	0.0133^{b}	0.6571^{b}	0.0939^{b}
C10M6	8	6.4721^{d}	0.0073^{c}	0.5439^{bc}	0.0777^{bc}
C10K7	8.5	17.3562^{b}	0.0153^{b}	0.6972^{b}	0.0996^{b}
M6K7	8	13.2542^{c}	0.0134^{b}	0.5545^{bc}	0.0794^{bc}
C10	9	21.0000^{a}	0.0196^{a}	0.8993^{a}	0.1285^{a}
M6	8	6.7522^{d}	0.0075^{c}	0.6280^{b}	0.0897^{b}
K7	8.5	16.6452^{b}	0.0169^{ab}	0.7841^{ab}	0.1120^{ab}
N+	8	0.0000^{e}	0.0000^{d}	0.6707^{b}	0.0958^{b}
N-	7	0.0000^{e}	0.0000^{d}	0.5135^{bc}	0.0734^{bc}

AVR = average visual rate, ANN = average number of nodules, AND = average nodule dry weight (g/plant), ASH = average shoot height (cm), ARL= average root length (cm), ASD = average shoot dry weight (g/plant), ARD = average root dry weight (g/plant). Values in the same column followed by the same letter are not significantly different at 5 % probability level.

Supplementary Table 3: Plants visual rate, nodule number, nodule dry weight, shoot dry weight and root dry weight of *G. max*

STRAIN	AVR	ANN	AND	ASD	ARD
C8VD1K7	8	5.7210^{d}	0.0062^{d}	0.8856^{b}	0.1265^{b}
C8K7	7.5	18.4572^{b}	0.0162^{b}	1.0896^{b}	0.1557^{b}
C8VD1	7.5	6.0000^{d}	0.0078^{d}	0.9965^{b}	0.1425^{b}
VD1K7	9	27.4251^{a}	0.0211^{a}	1.4262^{a}	0.2038^{a}
C8	8	16.8521^{bc}	0.0129^{bc}	0.9551^{b}	0.1364^{b}
VD1	8	19.7542^{b}	0.0139^{bc}	1.1152^{b}	0.1593^{b}
K7	8	13.3342^{c}	0.0132^{bc}	1.0895^{b}	0.1556^{b}
N+	8	0.0000^{e}	0.0000^{e}	0.9630^{b}	0.1376^{b}
N-	6	0.0000^{e}	0.0000^{e}	0.8350^{bc}	0.1193^{b}

AVR = average visual rate, ANN = average number of nodules, AND = average nodule dry weight (g/plant), ASH = average shoot height (cm), ARL= average root length (cm), ASD = average shoot dry weight (g/plant), ARD = average root dry weight (g/plant). Values in the same column followed by the same letter are not significantly different at 5 % probability level.

Field experiment

Supplementary Table 4: Plants visual rate, nodule number, nodule dry weight, shoot dry weight and root dry weight of *V. mungo*

STRAIN	AVR	ANN	AND	ASD	ARD
M5VD1K7	8	51.00^{a}	0.1386^{a}	11.6545^{bc}	1.6649^{ab}
M5VD1	8	47.33^{ab}	0.1111^{b}	13.0484^{b}	1.8641^{ab}
M5K7	9	29.29^{c}	0.0878^{c}	15.7730^{a}	2.2533^{a}
VD1K7	8	38.88^{b}	0.0857^{c}	12.7199^{b}	1.8171^{ab}
M5	8	33.50^{c}	0.0806^{c}	12.5093^{b}	1.7871^{ab}
VD1	9	46.22^{ab}	0.0904^{c}	15.2458^{ab}	2.1779^{a}
K7	8	48.67^{a}	0.0901^{c}	12.7244^{b}	1.8177^{ab}
N+	8	42.86^{b}	0.0803^{c}	11.5312^{bc}	1.6473^{ab}
N-	8	39.67^{b}	0.0700^{c}	11.6473^{bc}	1.6341^{ab}

AVR = average visual rate, ANN = average number of nodules, AND = average nodule dry weight (g/plant), ASH = average shoot height (cm), ARL= average root length (cm), ASD = average shoot dry weight (g/plant), ARD = average root dry weight (g/plant) Values in the same column followed by the same letter are not significantly different at 5% probability level.

Supplementary Table 5: Plants visual rate, nodule number, nodule dry weight, shoot dry weight and Root dry weight of *V. radiata*

STRAIN	AVR	ANN	AND	ASD	ARD
C10M6K7	9	47.5^{a}	0.0580^{a}	4.2858^{ab}	0.6123^{ab}
C10M6	7	44.33^{ab}	0.0378^{bc}	3.2323^{b}	0.4618^{b}
C10K7	9	39.21^{b}	0.0465^{b}	4.4582^{a}	0.6369^{a}
M6K7	9	43.16^{ab}	0.0443^{b}	3.8511^{b}	0.5502^{b}
C10	8	40.67^{b}	0.0331^{c}	3.4288^{b}	0.4898^{b}
M6	9	40.89^{b}	0.0413^{b}	3.7989^{b}	0.5427^{b}
K7	8.5	39.56^{b}	0.0489^{b}	4.1168^{ab}	0.5881^{b}
N+	8	34.57^{bc}	0.0283^{c}	3.6981^{b}	0.5283^{b}
N-	7.5	29.75^{c}	0.0277^{c}	3.2204^{b}	0.4601^{b}

AVR = average visual rate, ANN = average number of nodules, AND = average nodule dry weight (g/plant), ASH = average shoot height (cm), ARL= average root length (cm), ASD = average shoot dry weight (g/plant), ARD = average root dry weight (g/plant). Values in the same column followed by the same letter are not significantly different at 5 % probability level.

Supplementary Table 6: Plants visual rate, nodule number, nodule dry weight, shoot dry weight and root dry weight of *G. max*

STRAIN	AVR	ANN	AND	ASD	ARD
C8VD1K7	7.5	50.71^{b}	0.2462^{c}	6.8668^{b}	0.9810^{b}
C8K7	9	53.71^{b}	0.3927^{ab}	8.9506^{a}	1.2787^{a}
C8VD1	7	53.63^{b}	0.2254^{c}	6.6441^{b}	0.9492^{b}
VD1K7	8	73.00^{a}	0.4262^{a}	7.2959^{b}	1.0423^{b}
C8	8	55.38^{b}	0.3117^{b}	6.9658^{b}	0.9951^{b}
VD1	8.5	59.57^{b}	0.2865^{b}	7.2616^{b}	1.0374^{b}
K7	8	56.33^{b}	0.3152^{b}	7.2634^{b}	1.0376^{b}
N+	8	43.43^{c}	0.2205^{c}	6.8877^{b}	0.9840^{b}
N-	7	30.00^{cd}	0.2065^{c}	6.4960^{b}	0.9280^{b}

AVR = average visual rate, ANN = average number of nodules, AND = average nodule dry weight (g/plant), ASH = average shoot height (cm), ARL= average root length (cm), ASD = average shoot dry weight (g/plant), ARD = average root dry weight (g/plant). Values in the same column followed by the same letter are not significantly different at 5 % probability level.

J.Natn.Sci.Foundation Sri Lanka 2021 49 (3): 337 - 350
DOI: http://dx.doi.org/10.4038/jnsfsr.v49i3.9820

RESEARCH ARTICLE

Fruit morphology helps identifying evolutionary groups in Alpinieae (Zingiberaceae): inferences from phylogenetic analysis of gingers in Sri Lanka

P Karunarathne[1,2,3*], D Yakandawala[2] and P Samaraweera[2,3]
[1] *Department of Life Sciences, Texas A&M University, Corpus Christi, Corpus Christi, Texas, USA.*
[2] *Department of Botany, Faculty of Science, University of Peradeniya, Peradeniya.*
[3] *Postgraduate Institute of Science, University of Peradeniya, Peradeniya.*

Submitted: 10 Mar 2020; Revised: 01 September 2020; Accepted: 26 March 2021

Abstract: To assess the systematics and the phylogenetic placement of the members of the two important genera *Alpinia* Roxb. and *Amomum* Roxb. in Sri Lanka, molecular data of twelve in-group species together with three out-group taxa of the family Zingiberaceae were extensively analysed for phylogenetic significance. The current analysis of the evolutionary relationships of the Sri Lankan members of the genera of interest, utilising DNA sequence data of the chloroplast genome regions *trn* L-F and *trn* S-fM, has resolved four groups of *Alpinia* and two major clades of *Amomum* with substantial parsimony analysis and Bayesian inferences consistency values. With new accessions from the entire native range of the family, this result points to the need for an inevitable re-circumscription of the genus *Alpinia*, and shows congruence to the recent reshuffling of the genus *Amomum* and the family Zingiberaceae, except for the placement of each genus as monophyletic groups in the context of our study. Here, we suggest swapping the group defining species of an *Alpinia* clade observed in previous studies. Finally, our study suggests the use of fruit morphology to distinguish among recognised groups for Sri Lankan species as they exhibit positive correspondence.

Keywords: *Alpinia*, *Amomum*, phylogeny, Sri Lanka, Zingiberaceae.

INTRODUCTION

The two largest genera *Alpinia* and *Amomum* of the family Zingiberaceae, not only show a wide geographical distribution (Kiew, 1982; Smith, 1990; Larsen *et al*., 1998; Wu & Larsen, 2000; Sirirugsa, 2001) but also possess an array of homoplasious and plesiomorphic characters that make the classification of the genera problematic (Kress *et al*., 2002; Lamxay & Newman, 2012). Recent phylogenetic analyses of *Alpinia* and *Amomum* revealed the polyphyletic nature of the genera (Rangsiruji *et al*., 2000a; Xia *et al*., 2004; Kress *et al*. 2005; Boer *et al*., 2018) rejecting the previous classification which was based solely on morphological characters (Schumann, 1904; Smith, 1990). Boer *et al*. (2018) presented a more thorough phylogenetic analysis of Alpiniae, concluding the presence of 10 *Amomum* clades and six *Alpinia* clades. Further, they proceeded with re-circumscription of the *Amomum* clade by resurrecting three genera (*Conamomum, Meistera,* and *Wurfbania*) and describing three new genera (*Epiamomum, Lanxangia,* and *Sundamomum*). However, the current phylogenetic and evolutionary analyses of the two genera are far from being completed as they lack many representative species (e.g. species from South and Southeast Asia). Furthermore, only a few of the Sri Lankan species have been used in the aforementioned studies [*Amomum pterocarpum, Meistera echonocarpa* (formerly *Amomum echinocarpum*)*, Alpinia abundiflora* and *Alpinia fax*] despite the fact that there are several taxonomically paramount species occurring in Sri Lanka (Burtt & Smith, 1983). In this paper, phylogenetic placement of

* Corresponding author (piyal.karunarathne@tamucc.edu; https://orcid.org/0000-0002-1934-145X)

the Sri Lankan members of the two genera of interest is comparatively and comprehensively discussed, using the inferences drawn from the DNA sequence analysis.

With the recognition of the polyphyletic origin of the previously recognised tribes of the family Zingiberaceae (Sakai & Nagamasu, 1998; Rangsiruji *et al*., 2000a; Kress *et al*., 2002; Harris *et al*., 2003; Kress & Specht, 2005), several generic-level phylogenetic studies addressed the evolutionary relationships of the genera *Alpinia* (Rangsiruji *et al*., 2000b; Kress *et al*., 2005) and *Amomum* (Kaewsri, 2006; Kaewsri *et al*., 2007; Droop, 2012) as well as the tribe Alpinieae (Kaewsri *et al*., 2007; Kress *et al*., 2007, Boer *et al*., 2018) nesting two genera in comparative assessments. A common conclusion of these studies was that the members of the tribe do not follow the previous classification system and instead they tend to show different inter- and infra-generic evolutionary relationships indicating the polyphyletic nature of *Alpinia* and *Amomum.*

Nuclear and plastid DNA sequence analysis of *Alpinia* indicated that taxa described under this genus are closer to other genera of Alpinieae than they are to each other making classification of the tribe difficult (Rangsiruji *et al*., 2000a; Kress *et al*., 2005) due to its polyphyletic origin. Rangsiruji *et al*. (2000a) identified four separate clades of *Alpinia* whereas Kress *et al*. (2005) and Boer *et al*. (2018) described the separation of the genus into six groups. However, none of these studies agreed with the comprehensive infra-generic classification of *Alpinia* by Smith (1990), where she identified two subgenera within the genus encompassing 221 species. Similarly, during an analysis to assess the species boundaries of *Aframomum,* Harris *et al*. (2000) suggested that *Amomum* might be paraphyletic. Later studies by Xia *et al*. (2004) demonstrated the segregation of the genus *Amomum* into three major clades and they used fruit morphology to explain the clades observed. In another study by Kaewsri *et al*. (2007) using amplified fragment length polymorphism (AFLP) makers to analyse the phylogenetic relationships of Thai *Amomum* showed the positive correspondence of fruit morphology to the clades observed. Droop (2012) explained the presence of at least seven groups of *Amomum,* inferred by the results of the phylogenetic analysis using Internal transcribed spacer (ITS) and *mat*K data, whereas Boer *et al*. (2018) found 10 clades of *Amomum* using the same gene regions and they used both anther crest and fruit morphology in combination to discriminate the clades.

Although *Alpinia* and *Amomum* add the highest number of endemic species to the ginger family in Sri Lanka (Burtt & Smith, 1983), taxonomy and phylogenetic relationships of the members of these two genera has been poorly studied. The last taxonomic revision of the family Zingiberaceae in Sri Lanka was in 1983 (Burtt & Smith, 1983). Since then, several major revisions have taken place in the world utilising new phylogenetic tools, which resulted in major changes in the classification of the family (Rangsiruji *et al*., 2000a; Xia *et al*., 2004; Kress *et al*., 2005; de Boer *et al*., 2018). More importantly, none of the Sri Lankan endemic species had been used in any of the previous phylogenetic analyses. This observation justifies the need for a thorough phylogenetic account of Sri Lankan members of the family Zingiberaceae. As a pioneering step, the present study employed molecular data to determine evolutionary relationships among the members of the two genera *Alpinia* and *Amomum* in Sri Lanka and to compare the results with the current major revisions.

The goals of this study are: to construct phylogenetic hypothesis of interspecific and intergeneric relationships of the two genera primarily focusing on Sri Lankan species and compare the results with currently existing classifications; and to assess the congruence of several morphological characters with the molecular analysis results.

METHODOLOGY

Sampling

Ingroup taxa

Representative samples from 12 species of both genera were collected for the study (Figure 1 and 2). These included two native *Alpinia* species, four species of non-native *Alpinia*, four native *Amomum* species, one endemic *Amomum* species, and a new record of *Amomum* from Sri Lanka [*Wurfbainia villosa var. zeylanicus* (formerly *Amomum villosum zeylanicus*)] - Karunarathne *et al*. (2014). Unfortunately, four species of endemic *Amomum* (*A. nemorale, A. graminifolium, A. benthamianum,* and *A. trichostachyum*), one species of endemic *Alpinia* (*A. rufescens*) and one native *Amomum* species (*A. hypoleucum*) were not found during comprehensive field sampling. Most of these species have not been collected for more than 100 years and almost all types of localities are under intense human activities. Therefore, we present

Table 1: List of species of Sri Lankan *Alpinia* Roxb. and *Amomum* Roxb. showing the latest National Red List (2012) status of the native and endemic species and the GenBank accession numbers for the two sequences *trn* L-F and *trn* S-fM of each studied species.

Species	Suggested infra-generic clade (Xia *et al*., 2004; Kress *et al*., 2005)	National Red List status (MOE, 2012)	Last collection	Collected in this study	GenBank accession numbers	
					trnL-trnF	*trnS-trnfM*
Alpinia abundiflora	*A. fax* clade	LC	2008	Yes	KF748154	KF748169
Alpinia fax	*A. fax* clade	VU	1996	Yes	KF748153	KF748168
Alpinia rufescens‡	*A. fax* clade	CR(PE)	1862	No		
Amomum pterocarpum	-	EN	1987	Yes	KF748162	KF748177
Amomum masticatorium	-	EN	1988	Yes	KF748160	KF748175
Amomum fulviceps	-	VU	1973	Yes	KF748159	KF748174
Amomum hypoleucum	-	CR(PE)	1861	No		
Meistera echinocarpa	-	VU	1989	Yes	KF748161	KF748176
Amomum graminifolium‡	-	EN	1994	No		
Amomum acuminatum‡	-	CR(PE)	1855	Yes	KF748157	KF748172
Amomum trichostachyum‡	-	EN	1973	No		
Amomum nemorale‡	-	CR(PE)	1861	No		
Amomum benthamianum‡	-	CR(PE)	1864	No		
Wurfbainia villosa zeylanicus‡	-	-	-		KF748158	KF748173
Non-native species						
Alpinia galanga	*A. galanga* clade	-	-	Yes	KF748152	KF748167
Alpinia calcarata	*A. zerumbet* clade	-	-	Yes	KF748151	KF748166
Alpinia malaccensis	*A. zerumbet* clade	-	-	Yes	KF748155	KF748170
Alpinia zerumbet	*A. zerumbet* clade	-	-	Yes	KF748156	KF748171
Out-group species						
Zingiber cylindricum	-	-	-	Yes	KF748148	KF748163
Zingiber sp.	-	-	-	Yes	KF748149	KF748164
Hedychium coronarium	-	-	-	Yes	KF748150	KF748165

LC - least concerned; VU - vulnerable; EN - endangered; CR (PE) - critically endangered (possibly extinct)
‡: endemic species

an adequate sampling of extant taxa of the two genera in Sri Lanka (Table 1).

Outgroup taxa

Two representative species of *Zingiber* (*Z. cylindricum* and *Zingiber* sp.) and one species of *Hedychium* (*H. coronarium*) were collected during the field sampling for out-group taxa. For a better comparison of the overall phylogeny and the geographical distribution of the phylogenetic clusters, DNA sequences of *trn*L-*trn*F intergenic spacer region of four *Aframomum*, three *Renealmia*, and one *Elettariopsis* accessions were also downloaded, following a *nucleotide blast* from GeneBank for outgroup taxa (supplementary Table 1). These three genera are sister groups to the studied two genera in the present study (Kress & Specht, 2005; Kress *et al.*, 2007).

DNA extraction and sequencing

Fresh leaf samples dried with silica beads were used for DNA extractions. Genome DNA was extracted using a modified Doyle and Doyle (1987) CTAB protocol with

few additional purification steps: suspended DNA in TE was treated with ProteinaseK and RNase followed by phenol chloroform extraction and sodium acetate precipitation of DNA. PCR amplification was done for two chloroplast gene regions *trn* L-F and *trn* S-*f*M, using *trn* L-F (Taberlet *et al*., 1991) and *trn* S-*f*M (Shaw *et al*., 2005) primer pairs, respectively. The chloroplast genome regions *mat*K, and *trn* L-F together with nuclear ribosomal ITS regions have been explicitly used in evolutionary studies of the family Zingiberaceae (Johnson & Soltis, 1994; Baldwin *et al*., 1995; Kress *et al*., 2002). However, only the chloroplast intergenic spacer *trn* L-F and the non-coding region *trn* S-fM were used because in the initial steps, the primers that were used for ITS and *mat*K regions failed to amplify the desired region for several species (*A. echinocarpum, A. pterocarpum,* and *A. fax*) and two species produced multiple bands for ITS region (*A. acuminatum* and *A. fulviceps*). The region *trn* S-fM has been documented to be informative in phylogenetic studies (Shaw *et al*., 2005; Minami *et al*., 2009).

According to the manufacturers' recommendations all amplifications used Promega GoTaq® Flexi DNA polymerase (Madison, WI, USA). Eppendorf Mastercycler® thermal cycler (Hauppauge, NY, USA) with the following programs was used for PCR amplifications.

trn L-F region: preheat at 80 °C for 3 min, 2 cycles with 3 temperature segments (94 °C for 2 min, 50 °C for 40 s and 72 °C for 2 min), 30 cycles with 3 temperature segments (94 °C for 30 s, 50 °C for 30 s and 72 for °C 1 min) and final extension 72 °C for 8 min.

trn S-fM region: preheat at 94 °C for 3 min, 30 cycles with 3 temperature segments (94 °C for 30 s, 50 °C for 30 s, 72 °C for 1 min) final extension 72 °C 8 min.

PCR product purification and DNA sequencing were done by Macrogene Inc. Korea (Sanger sequencing: 96-capillary 3730XL DNA Analyzer: Applied Biosystems®). Same primer pairs as in PCR amplifications were used in sequencing for the respective gene region. Raw sequences were screened and assembled in SeqMan Pro, DNAStar® Lasergene 8.1 (DNASTAR, Inc., 2009). Assembled contigs were edited using BioEdit (ver. 7.1.9) (Tom Hall, Ibis Biosciences, California, USA). Edited sequences were then aligned using ClustalX 2.1 (Larkin *et al*., 2007). Alignment files were further edited in BioEdit to process unaligned bases. This was repeated for all sequences of the studied two DNAs separately. A combined analysis of the two regions was also performed to obtain better resolved trees and thereby to estimate phylogeny with the strongly supported results (Nixon & Carpenter, 1996).

Furthermore, to conduct a more inclusive sampling of the two genera, sequences of both *Alpinia* and *Amomum* were downloaded from GeneBank to analyse with the accessions from the present study. DNA sequences of *trn*L-*trn*F intergenic spacer region of six *Alpinia* and seven *Amomum* accessions were downloaded following a *nucleotide blast* from GeneBank (Supplementary Table 1).

Phylogenetic analysis

Maximum Parsimony Analysis (MPA)

MPA of the aligned sequence data was accomplished using PAUP* 4.0 (Swofford, 2003) on a power Macintosh G4. Heuristic search strategies were used in 1000 Tree Bisection Reconnection (TBR) branch swapping replicates of random taxon addition with MULPARS in effect. The majority-rule strict consensus tree assembly was used to build the most parsimonious trees. All positions were weighted equally; gaps were treated as missing values. The Consistency Index (CI) and Retention Index (RI) values for tree topologies were also calculated with PAUP. Bootstrap analyses (Felsenstein, 1985) were conducted using PAUP with a simple addition of 1000 fast-swap replicates to 100 tree topologies with a 50 % confidence.

Bayesian Inferences

Bayesian analysis was used to draw model based phylogenetic inferences (Yang & Rannala, 1997; Larget & Simon, 1999) from the DNA sequence data. MrBayes (Huelsenbeck & Ronquist, 2001) was used for the analysis to obtain the posterior probabilities (PP). Models for the sequence evolution was determined using *jModeltest* (Posada, 2008) and the following parameters were used in the analysis; GTR, nst = 6, rates = inverse gamma, statefreqpr = fixed (equal). The analysis comprised generations of 10,000,000 Markov Chain Monte Carlo (MCMC) chains with a sampling frequency of every 10,000 generations. The initial 25 % samples from each run were discarded as burn-in, after which the run was summarised as a majority rule consensus tree and the posterior probability values were saved in MrBayes. Separate analysis of the two regions and the combined analysis produced a similar tree topology for the strict consensus. Therefore, only the combined tree is presented here.

RESULTS AND DISCUSSION

Analysis of *trn* L-F intergenic spacer data

A total of 449 parsimony-informative characters were present in the *trn* L-F region, which produced 5473 equally parsimonious trees in more than 270,000 rearrangement steps in parsimony analysis. Strict consensus of these trees (Supplementary Figure 1) placed all the in-group taxa in a monophyletic clade (bootstrap 91 %) separating out-group taxa. *Alpinia* and *Amomum* were placed in two separate clades (*A* and *B in* Supplementary Figure 1) with moderate and strong bootstrap values: 65 and 100, respectively. Hence, genera *Alpinia* and *Amomum* in Sri Lanka are strongly supported as monophyletic. Bayesian inferences also supported this observation with a high PP value of one.

Figure 1: Floral pictures of some *Alpinia* species found in Sri Lanka; (A) *A. abundiflora,* (B) *A. fax,* (C) *A. galanga,* (D) *A. calcarata,* (E) *A. maleccensis,* (f) *A. zerumbet*

Two major clades were observed in the *Amomum* group: *Am I, A. pterocarpum* (bootstrap 100 %) and *Am II, A. echinocarpum, A. villosum zeylanicus*, *A. masticatorium, A. fulviceps* and *A. acuminatum* (bootstrap 93 %). Separation of these two clades were also observed in the Bayesian analysis with PP values of 1.00 and 0.98, respectively. Four clades were identified in the Alpinia group: *Alp I, Alp II, Alp III, and Alp IV. A. abundiflora* and *A. fax* formed a separate group (*Alp I*) basal to the rest of the *Alpinia* with moderate bootstrap values (69 %) and with high Bayesian support (PP=0.98). The other three clades (*Alp II- A. galanga; Alp III- A. calcarata*; Alp *IV- A. malaccensis* and *A. zerumbet*), possess moderate support values.

.

Analysis of *trn* S-fM intergenic spacer data

This region resulted in 3673 equally parsimonious trees in more than 670,000 rearrangement steps in the MPA. This region harboured 140 parsimony-informative characters (Supplementary Figure 2). Strict consensus of these trees also produced two major groups of in-group species with moderate bootstrap values and high PP values: A) *Alpinia* clade (bootstrap = 69 %, PP = 1); B) *Amomum* clade (bootstrap = 50 %, PP = 1) (Supplementary Figure 2). However, all in-group species formed a single monophyletic clade providing higher support values (87 %, PP = 1). Same three clades that were observed with *trn* L-F for the *Amomum* group was also observed with this analysis but with different bootstrap values. *A. fax* and *A. abundiflora* formed the basal clade (*Alp I*) of the *Alpinia* group with strong support values (bootstrap 80 %, PP = 1). Despite the low PP value (0.58), the rest of the *Alpinia* further split into three distinctive groups separating *A. galanga* (*Alp II*); *A. calcarata* (*Alp III*); *A. zerumbet* and *A. malaccensis* (*Alp IV*).

Figure 2: Floral pictures of few Sri Lankan *Amomum* species; (A) *A. masticatorium,* (B) *A. echinocarpum,* (C). *A. pterocarpum,* (D). *A. fulviceps*

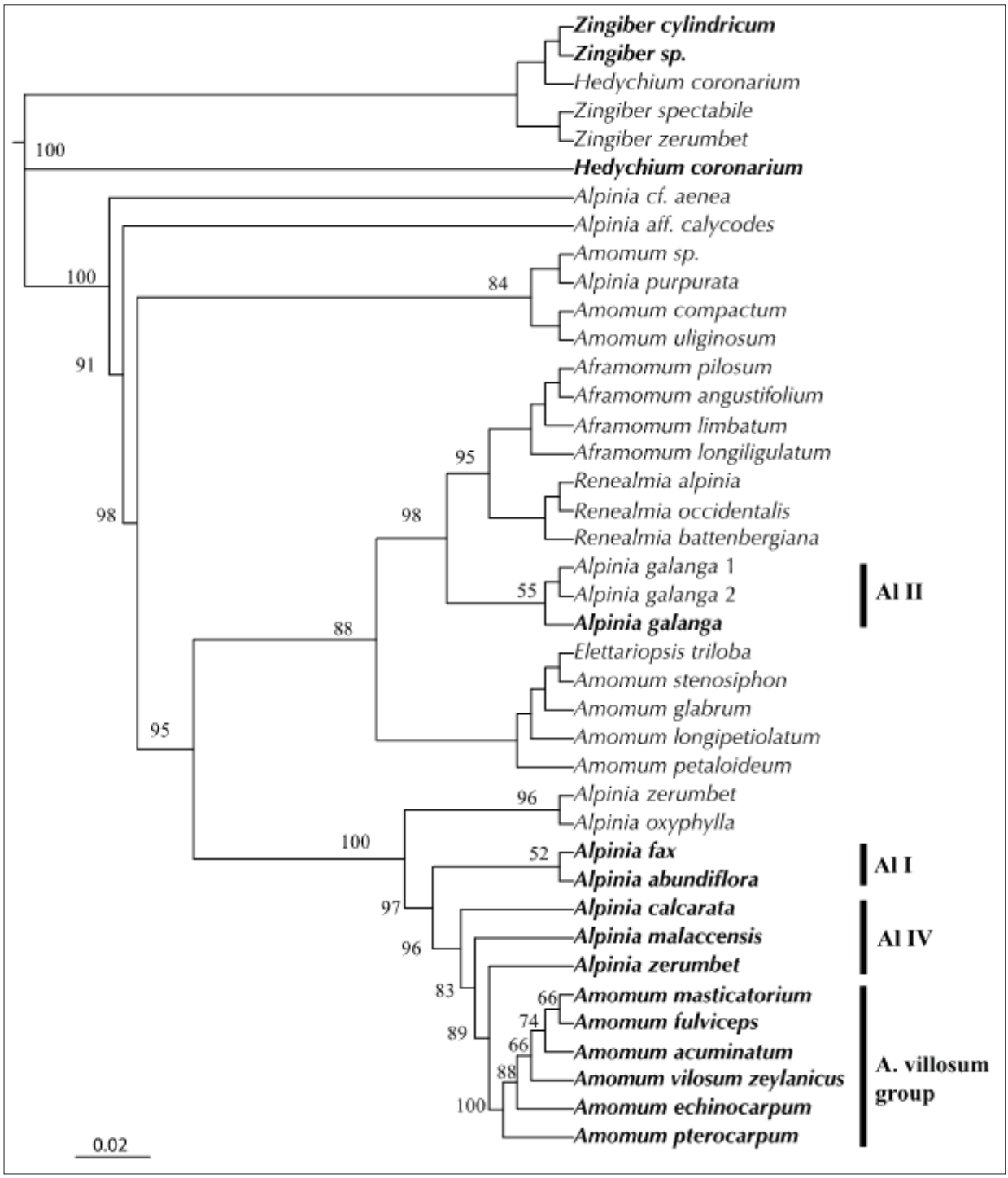


Figure 3: Best trees resulting from the parsimony analysis of the GeneBank accessions and the data from the present study of *trn L-F* intergenic spacer region (length = 510; consistency index = 0.901; retention index = 0.966; rescaled consistency index = 0.900); branch labels: above - bootstrap values, Al I, II, IV – *Alpinia* groups that corresponds to Kress *et al*. (2005), *A. villosum* group - *Amomum* group that corresponds to Xia *et al*. (2004) and *Wurfbainia* group in Boer *et al*. (2018).

Analysis of the combined dataset

Combined analysis resolved Sri Lankan Alpinias into three major clades (i.e. Al I, Al II, Al IV), and *Amomum* into one group (i.e. *A. villosum* group *sensu* Xia *et al*., 2004 and *Wurfbainia* group *sensu* Boer *et al*., 2018) (Figure 3, also see Supplementary Figure 3). *A. galanga* grouped with GeneBank accessions in a separate group while all the other *Alpinia* were resolved in sister clades. Interestingly, *A. fax* and *A. abundiflora* resolved into one clade in the latter while the rest of the species were separated in individual branches with strong statistical support (Figure 3). Within the *Amomum* clade, we see the same inner groups as were resolved in separate analyses. Nevertheless, all the clades were strongly supported by bootstrap values except for two branches within the *Amomum* clade.

Overall, *Alpinia* and *Amomum* in Sri Lanka form a monophyletic group with strong phylogenetic support values (bootstrap 100 %, PP = 1). All the *Amomum* species formed a monophyletic clade (bootstrap 100 %, PP = 1) where they were further divided into two monophyletic groups separating basal clade *Am I*: *A. pterocarpum* (bootstrap: 100 %, PP = 1) and the second clade *Am II* nesting all the other species (bootstrap 100, PP = 0.98) (Supplementary Figure 3). *A. fax* and *A. abundiflora* were placed separately (*Al I*) from other Alpinias with high support values (bootstrap 97 %, PP = 1). Although the other three *Alpinia* accessions nested together with all the other Sri Lankan species, *A. galanga* clustered with other accessions of the species from the GeneBank (Al II). *Al IV* clade nested non-native - *A. zerumbet* and *A. malaccensis* (bootstrap 100, PP = 1).

Bayesian analysis of the *trn* L-F and *trn* S-fM sequence data resulted in cladograms similar to that of the MPA. All the clades observed in the maximum parsimony analysis were supported with higher posterior probability (PP) values in the Bayesian analysis. The two genera are monophyletic with strong PP value of 1.00. The *Alpinia* group formed two clades in the analysis separating *A. abundiflora* and *A. fax* from the rest with a PP value of 0.98. Overall, the Bayesian inferences of the phylogenetic trees is congruence with the results of MPA.

The complexity of the classification of the two polyphyletic genera *Alpinia* and *Amomum* has been acknowledged by many authors (Smith 1990; Larsen *et al*., 1998; Kress *et al*., 2002). As a result, many researchers who were interested in the family Zingiberaceae have dealt with taxonomy, phylogeny and the classification of these two genera. Remarkably, the phylogenetic relationships of the members of the family in Sri Lanka have never been assessed and none of the endemic species have been assessed in any of the recent phylogenetic studies of the family (e.g. Kress *et al*., 2002, Xia *et al*., 2005, Boer *et al,*. 2018, Droop 2019). Hence, the current study attempted phylogenetic analysis of the Sri Lankan members of the two genera *Alpinia* and *Amomum*. Results of the present analysis provides better insight to the evolutionary relationships of the studied species in Sri Lanka.

According to the latest phylogenetic analysis of *Amomum* group (de Boer *et al*., 2018), Sri Lankan taxa of the *Amomum* group consist of three genera (i.e. *Amomum, Wurfbainia* and *Meistera*). However, the present study has not resulted in similar groups, and was rather congruent with the previous studies of the family Zingiberaceae (Rangsiruji *et al*., 2000a; Kress *et al*., 2002) by resolving only two clades. This may be due to the limited sampling in our study representing all the species of *Amomum* group. Although the present analysis does not include accessions representing all the genera of the family, our study highlighted two *Amomum* clades with strong statistical support (Figure 3). Therefore, in congruence with infra-generic relationships that were illuminated in recent studies (Kress *et al*., 2007; Droop, 2012), it can be concluded that Sri Lankan members of the two genera of interest make two evolutionary groups. Reduced staminodes, capsule: globose, echinate or irregular shaped, and relatively short stamen can be listed as a combination of features that characterise the separation of the clade.

Alpinia

Although all the *Alpinia* species studied in this analysis formed a monophyletic group, the separation of the clades within the genus (i.e. *Alp I*, *Alp II*, *Alp III* and *Alp IV*) agrees with the broader studies of the family (Rangsiruji *et al*., 2000a; Kress *et al*., 2002). These groups are in congruence with the recent phylogenetic studies on the family (Rangsiruji *et al*., 2000a; Kress *et al*., 2002, 2005, 2007) and also confirm the incongruence with earlier morphological treatments (Schumann, 1904; Smith, 1990).

A. abundiflora and *A. fax* were placed as the basal group (*Alp I*), separately from all the other members of the genus with strong support values in both MPA and Bayesian analyses of the molecular data. Recognition of this clade was also described in phylogenetic analyses of *Alpinia* by Kress *et al*. (2005) and Rangsiruji *et al*. (2000b). As the latter study explained, these species (namely *A. fax* clade) are characterised by the presence of capitate, usually radical inflorescence surrounded by a world of sterile bracts; fertile bracts subtending a cincinnus of several flowers. Furthermore, this observation reinforces the suggestion by Kress *et al.* (2007) that these species require a new generic name. Kress *et al.* (2005) also suggested that the inclusion of *A. rufescens* in this clade is appropriate since this species also shares this distinctive morphological character combination. *A. rufescens* has been recorded only in Sri Lanka. It is only known from the type locality and has never been recollected since its type gathering (1862). During our extensive field sampling efforts (comprised of more than 25 locations) it was observed that the type locality of *A. rufescens* is no longer a potential area to support these species since the area is under intense human activities. This species was not found from any other location either. After close and careful observation of the characters of this clade, it could be suggested that *A. abundiflora* be used as the clade defining species instead of *A. fax* because *A. abundiflora* possesses typical characters such as capitate inflorescence, occurrence of the inflorescence either (rarely) on the vegetative stem or on a separate basal peduncle and typically a cincinnus of seven (7) flowers. These are also shared with *A. rufescens* (except for the number of flowers in a cincinnus). On the other hand, the inflorescence of *A. fax* rather elongates with age and it bears a lesser number of flowers in a cincinnus (seldom up to six flowers). Further, Karunarathne *et al*. (2015) reported the confusion of morphological characters of *A. fax* that was also presumably recorded in South India (Kumar *et al*., 2002).

Although the inclusion of *Alpinia galanga* with the rest of Sri Lankan *Alpinia* (except *Alp* I) showed weak support values (bootstrap 56 %, PP = 0.58: combined analysis), the separation of groups *Alp I*, through *Alp IV* was strongly supported. The clade *Alp I* separated *A. abundiflora* and *A. fax* from the rest with strong support (bootstrap 75 %, PP = 0.98). Next to the *A. abundiflora* clade, *Alp II* isolated *A. galanga* basal to the *A1* group. *A. galanga* clade is characterised by higher number of flowers in the branched inflorescence, smaller size of flowers, clawed labellum and the glabrous ovary. Inclusion of additional species such as *A. nigra* (not found in our field observations) might better explain these characters. Although *A. calcarata* was included as a member of the *A. zerumbet* clade in previous studies, the present results placed it as a separate clade (*Alp III*). Linear-lanceolate shape of the lamina and the reduced lamina width along with the smaller sized fruits can be listed as deviating characters. *A. zerumbet* and *A. malaccensis* formed the terminal clade with strong bootstrap support. Besides the robustness of the habit of the plant, relatively large, globose fruits are characteristic to this clade.

Amomum

All the studied Sri Lankan *Amomum* species resolved into two major clades (*Am I* and *Am II*). This exhibits congruence with recent evolutionary analyses of the genus (Xia *et al*., 2004; Kaewsri *et al*., 2007; Droop, 2012, Boer *et al*., 2018). Results of the current analysis also do not correspond to earlier infra-generic classification of the genus (e.g. Schumann 1904; Smith 1990).

A. pterocarpum was placed as the basal most lineage (clade *Am I*) of the *Amomum* group. The clade is characterised by the presence of ribbed capsule with wings, sub-globose inflorescence with quickly deciduous bract and bracteoles. *Am II* clade resembles the *A. maximum* clade which was observed in the study by Xia *et al*. (2004) and other studies (Kress *et al*., 2007; Droop, 2012). The main characteristic feature of this group is the presence of smooth or winged fruit coat (Xia *et al*., 2004). *Am II* placed all the other members of the genus (*A. echinocarpum, A. masticatorium A. fulviceps* and *A. acuminatum, A. villosum zeylanicus*) in a single clade with strong branch support values. The most conspicuous morphological characters that shared among these species are slender basal peduncle covered with scaly sheaths, ovoid to elongated inflorescence, brown-red bracts and bracteoles, tri-lobed anther appendage, bifid labellum and echinate fruits. This is in congruence with the recent evolutionary analysis of the genus. Xia *et al.* (2004) identified this group as *A. villosum* clade with echinate fruits being a distinctive character and later studies confirmed this placement (Xia *et al*., 2004; Kress *et al*., 2007; Droop, 2012). The new record of species *Amomum villosum* var. *zeylanicus* shows the typical characters that Xia *et al*. (2004) highlighted for the group. In contrast to this, Boer *et al*. (2018) separated these species into two separate genera: i) *Meistera* – *M. echinocarpa, M. masticatorum*; ii) *Wurfbainia* – *W. villosa* (var. *zeylanicus*), respectively. They described echinate fruits and semilunar anther crest for *Meistera* and echinate fruits and eared anther crest for *Wurfbainia* as conspicuous characters of the two genera. However,

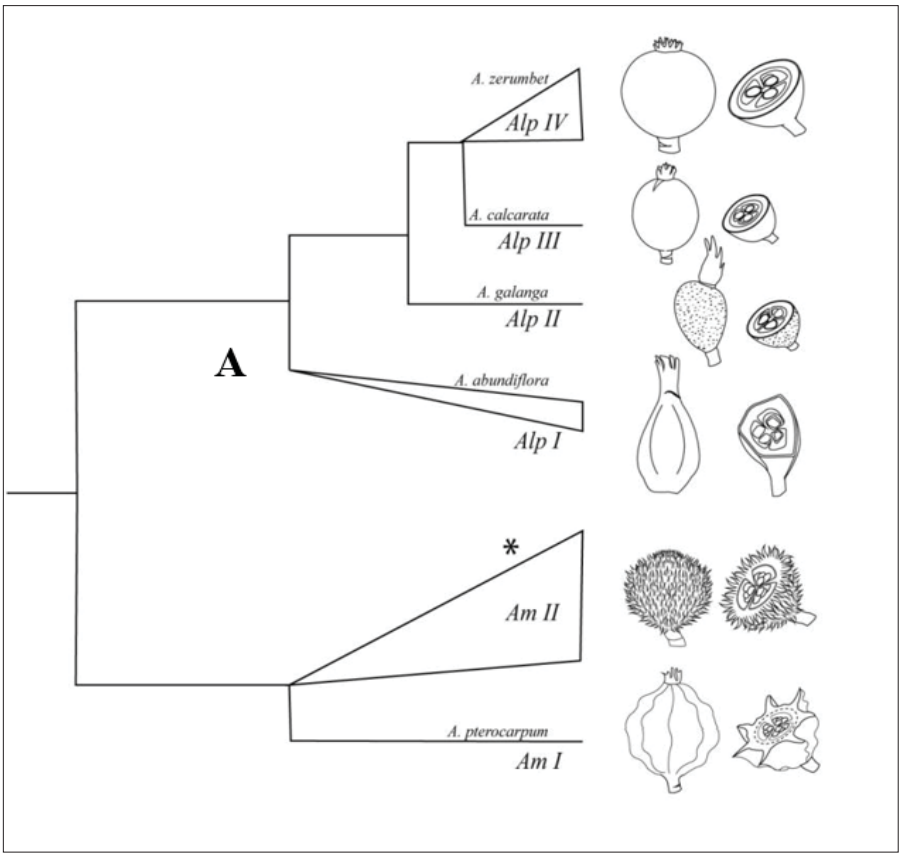


Figure 4: Tree depicting the fruit morphology attribution for the characterisation of the *Alpinia* and *Amomum* clades observed in Sri Lanka, obtained from the molecular data analysis; A. *Alpinia* clade, B. *Amomum* clade; *. *Am II* clade: *Amomum* group of species bearing echinate fruits *(A. echinocarpum, A. masticatorium, A. fulviceps, A. acuminatum,* and *A. villosum zeylanicus.*) (Drawings are according to their relative size)

this separation of polyphyletic clades was not observed in the present study.

Moreover, after comprehensive observation of the morphological characters of each clade resulted in the current phylogenetic analysis, besides the above-mentioned group specific characters, fruit morphology of the studied species can be effectively utilised to separate different clades of the Sri Lankan members of *Alpinia* and *Amomum*. Figure 4 shows a detailed attribution of the fruit morphology of the groups for this context. However, we admit that we are not the first to use the fruit morphology to separate phylogenetic clades in the family Zingiberacea (Kress *et al*., 2002; Xia *et al*., 2005; Droop *et al*., 2012; Boer *et al*., 2018). Nevertheless, the use of the fruit morphology for Sri Lankan species presents an efficient strategy.

Sri Lankan gingers in a global phylogeny

The three *Alpinia* groups observed in our analysis are in close congruence with the most recent thorough analysis of the genus (Kress *et al*., 2005). Most importantly the analysis shows that Alpinias are not monophyletic, contrary to our analysis only with Sri Lankan species (Figure 4). This is expected as several other evolutionary groups in the family included in the analysis, and as observed in the phylogenetic analysis of the family (Kress *et al*., 2002; 2005; de Boer 2018). In contrast, the *Amomum* species in Sri Lanka formed a monophyletic group with strong bootstrap support (100 %). Furthermore, this resolution of one *Amomum* clade in Sri Lanka corresponds to the same *Amomum villosum* clade as in Xia *et al*. (2004) of the *Amomum* phylogeny. This suggests that all the Sri Lankan species of the genus

evolved from one common ancestor although the genus itself in the region is not monophyletic. However, none of the other species of *Amomum* in this group were included in the previous evolutionary analyses of the genus. Therefore, a comparison of the results of this study is not appropriate. Most importantly, even though the evolutionary groups within the two genera slightly changed with the global phylogeny, our use of the fruit morphology still applies to the evolutionary groups found in Sri Lanka (Figure 4). Moreover, this observation can further be extended to identify corresponding clades of the two genera in other regions as well, as they show high congruence to previous studies by Xia *et al*. (2004), Kress *et al*. (2005) and Boer *et al*. (2018).

Overall, the current analysis of the evolutionary relationships of Sri Lankan *Alpinia* and *Amomum* utilising DNA sequence data of the chloroplast genome regions *trn L-F* and *trn S-fM* and the combined dataset resolved four groups of *Alpinia* and two major clades of *Amomum*. This result is in congruence and consistent with recent phylogenetic analyses of the two genera using molecular data to a certain extent, except for the placement of each genus as monophyletic groups in the context in the present analysis. This discrepancy in comparison to other recent studies seems to arise from the fact that none of the endemic Sri Lankan species were used in previous studies. Finally, the present study suggests using the fruit morphology to distinguish the observed groups for Sri Lankan species because they exhibit positive correspondence with the fruit morphology, which is also consistent.

Acknowledgement

This research project was supported by the National Science Foundation of Sri Lanka (Grant SIDA/2007/BT/02). The support and services provided by the Forest Department of Sri Lanka, Department of Wildlife Conservation Sri Lanka and the National Herbarium of Sri Lanka is also appreciated.

Conflict of interest

All the authors of this manuscript declare that no conflict of interest has occurred among authors pertaining to this study.

REFERENCES

Baldwin B.G., Sanderson M.J., Porter J.M., Wojciechowski M.F., Campbell C.S. & Donoghue M.J. (1995). The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* **82**(2): 247–277. DOI: https://doi.org/10.2307/2399880

de Boer H. *et al*. (11 authors) (2018). Convergent morphology in Alpinieae (Zingiberaceae): recircumscribing *Amomum* as a monophyletic genus. *TAXON* **67**(1): 6–36. DOI: https://doi.org/10.12705/671.2

Burtt B.L. & Smith R.M. (1983). Zingiberaceae. In: *Revised Handbook to the Flora of Ceylon* (eds. M.D. Dassanayake & E.R. Fossberg), pp. 488–532. Amerind Publication Co. Pvt. Ltd., Colombo.

Droop A.J. (2012). Systematic and biogeographic studies in the genus *Amomum* Roxb.(Zingiberaceace) in Sumatra. *PhD thesis*, University of Aberdeen, UK.

Felsenstein J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**(4): 783–791. DOI: https://doi.org/10.1111/j.1558-5646.1985.tb00420.x.

Harris D.J., Newman M.F., Hollingsworth M.L., Möller M. & Clark A. (2003). The phylogenetic position of *Aulotandra* (Zingiberaceae). *Nordic Journal of Botany* **23**(6): 725–734. DOI: https://doi.org/10.1111/j.1756-1051.2003.tb00451.x.

Harris D.J., Poulsen A.D., Frimodt-Møller C., Preston J. & Cronk Q.C.B. (2000). Rapid radiation in *Aframomum* (Zingiberaceae): evidence from nuclear ribosomal DNA internal transcribed spacer (ITS) sequences. *Edinburgh Journal of Botany* **57**(03): 377–395. DOI: https://doi.org/10.1017/S0960428600000378.

Huelsenbeck J.P. & Ronquist F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**(8): 754–755. DOI: https://doi.org/10.1093/bioinformatics/17.8.754

Johnson L.A. & Soltis D.E. (1994). matK DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. *Systematic Botany* **19**(1): 143–156. DOI: https://doi.org/10.2307/2419718

Kaewsri K. (2006). Systematic studies of the genus *Amomum* Roxb.(Zingiberaceae) in Thailand. *PhD thesis*, Kasetsart University, Bangkok, Thailand.

Kaewsri W., Paisooksantivatana Y., Veesommai U., Eiadthong W. & Vajrodaya S. (2007). Phylogenetic analysis of Thai *Amomum* (Alpinioideae: Zingiberaceae) using AFLP markers. *Kasetsart Journal (Natural Science)* **41**: 213 – 226.

Karunarathne P., Yakandawala D. & Samaraweera P. (2015). Important notes on the identity of *Alpinia fax* (Thwaites) B.L. Burtt & R.M. Sm. *Taiwania* **60**(4): P204–207. DOI: https://doi.org/10.6165/tai.2015.60.204

Karunarathne P., Yakandawala D. & Samaraweera P. (2014). On the occurrence of a new variety of *Amomum villosum* (Family Zingiberaceae) in Central Hills of Sri Lanka. *Phytotaxa* **172**(2): 129. DOI: https://doi.org/10.11646/phytotaxa.172.2.9

Kiew K.Y. (1982). The genus *Elettariopsis* (Zingiberaceae) in Malaya. *Notes from the Royal Botanic Garden, Edinburgh* **40**(1): 139–152.

Kress W.J., Liu A.-Z., Newman M. & Li Q.-J. (2005). The molecular phylogeny of *Alpinia* (Zingiberaceae): a

complex and polyphyletic genus of gingers. *American Journal of Botany* **92**(1): 167–178.
DOI: https://doi.org/10.3732/ajb.92.1.167

Kress W.J., Newman M.F., Poulsen A.D. & Specht C. (2007). An analysis of generic circumscriptions in tribe Alpinieae (Alpinioideae: Zingiberaceae). *The Gardens' Bulletin Singapore* **59**(1 & 2): 113–128.

Kress W.J., Prince L.M. & Williams K.J. (2002). The phylogeny and a new classification of the gingers (Zingiberaceae): evidence from molecular data. *American Journal of Botany* **89**(10): 1682–1696.
DOI: https://doi.org/10.3732/ajb.89.10.1682

Kress W.J. & Specht C.D. (2005). Between cancer and capricorn: phylogeny, evolution and ecology of the primarily tropical Zingiberales. *Biologiske Skrifter* **55**: 459–478.

Kumar P., Sabu M. & Jayasree S. (2002). Alpinia fax Burtt & Smith. A new record for South India. *Rheedea* **12**(2): 179–183.

Lamxay V. & Newman M.F. (2012). A revision of *Amomum* (Zingiberaceae) in cambodia, laos and vietnam. *Edinburgh Journal of Botany* **69**(1): 99–206.
DOI: https://doi.org/10.1017/S0960428611000436.

Larget B. & Simon D.L. (1999). Markov Chasin Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Molecular Biology and Evolution* **16**(6): 750–759.
DOI: https://doi.org/10.1093/oxfordjournals.molbev.a026160.

Larkin M.A. *et al*. (13 authors) (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**(21): 2947–2948.
DOI: https://doi.org/10.1093/bioinformatics/btm404.

Larsen K., Lock J.M., Maas H. & Maas P.J.M. (1998). Zingiberaceae. In: *Flowering Plants · Monocotyledons* (ed. K. Kubitzki), pp. 474–495. Springer Berlin Heidelberg, Berlin, Germany.
DOI: https://doi.org/10.1007/978-3-662-03531-3_49

Minami M., Nishio K., Ajioka Y., Kyushima H., Shigeki K., Kinjo K., Yamada K., Nagai M., Satoh K. & Sakurai Y. (2009). Identification of Curcuma plants and curcumin content level by DNA polymorphisms in the trnS-trnfM intergenic spacer in chloroplast DNA. *Journal of Natural Medicines* **63**(1): 75–79.

Ministry of Environment (MOE) (2012). *Threatened List. National Red List 2012 of Sri Lanka. Conservation Status of the Fauna and Flora*, pp. viii + 476. Ministry of Environment, Colombo, Sri Lanka.

Nixon K. C. & Carpenter J. M. (1996). On simultaneous analysis. *Cladistics* **12**(3): 221–241.
DOI: https://doi.org/10.1111/j.1096-0031.1996.tb00010.x.

Posada D. (2008). jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**(7): 1253–1256.
DOI: https://doi.org/10.1093/molbev/msn083

Rangsiruji A., Newman M. F. & Cronk Q. C. B. (2000a). A study of the infrageneric classification of *Alpinia* (Zingiberaceae) based on the ITS region of nuclear rDNA and the trnL-F spacer of chloroplast DNA. Monocots: systematics and evolution, pp 695–709. CSIRO, Collingwood, Australia.

Rangsiruji A., Newman M. F. & Cronk Q.C.B. (2000b). Origin and relationships of *Alpinia galanga* (Zingiberaceae) based on molecular data. *Edinburgh Journal of Botany* **57**(1): 9–37.
DOI: https://doi.org/10.1017/S0960428600000020.

Sakai S. & Nagamasu H. (1998). Systematic studies of Bornean Zingiberaceae I. *Amomum* in Lambir Hills, Sarawak. *Edinburgh Journal of Botany* **55**(01): 45.
DOI: https://doi.org/10.1017/S0960428600004352

Schumann K. (1904). Zingiberaceae. In: *Das Pflanzenreich* (ed A. Engler.), pp. 1–458. Wilhelm Engelmann, Leipzig, Germany.

Shaw J. *et al.* (10 authors) (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* **92**(1): 142–166.
DOI: https://doi.org/10.3732/ajb.92.1.142

Sirirugsa P. (2001). *Zingiberaceae of Thailand*. Biodiversity Research and Training Program, Bangkok, Thailand, pp 63–77. Jirawat Express Co., Ltd, Bangkok, Thailand.

Smith R.M. (1990). Alpinia (Zingiberaceae): a proposed new infrageneric classification. *Edinburgh Journal of Botany* **47**(1): 1–75.
DOI: https://doi.org/10.1017/S0960428600003140

Swofford D.L. (2003). {PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4.}.

Taberlet P., Gielly L., Pautou G. & Bouvet J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**(5): 1105–1109.
DOI: https://doi.org/10.1007/BF00037152

Wu T.-L. & Larsen K. (2000). Zingiberaceae. In: *Flora of China* (eds Z.-Y. Wu & P.H. Raven), pp. 322–377. Sci. Press, Beijing, China.

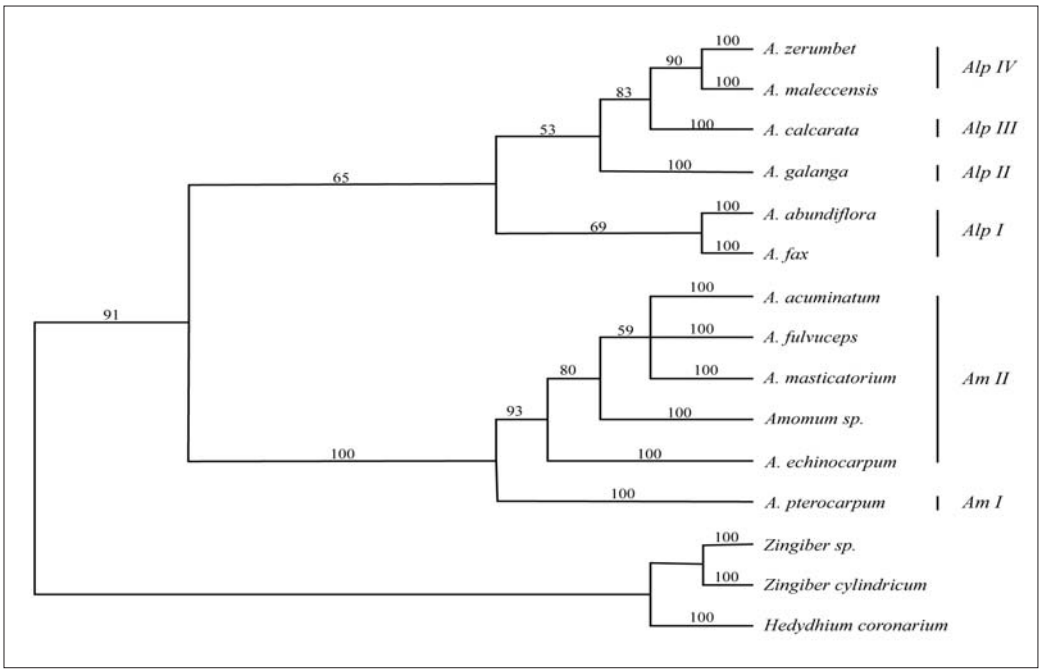
Xia Y.-M., Kress W.J. & Prince L.M. (2004). Phylogenetic analyses of *Amomum* (Alpinioideae: Zingiberaceae) Using ITS and matK DNA Sequence Data. *Systematic Botany* **29**(2): 334–344.
DOI: https://doi.org/10.1600/036364404774195520

Yang Z. & Rannala B. (1997). Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. *Molecular Biology and Evolution* **14**(7): 717–724.
DOI: https://doi.org/10.1093/oxfordjournals.molbev.a025811

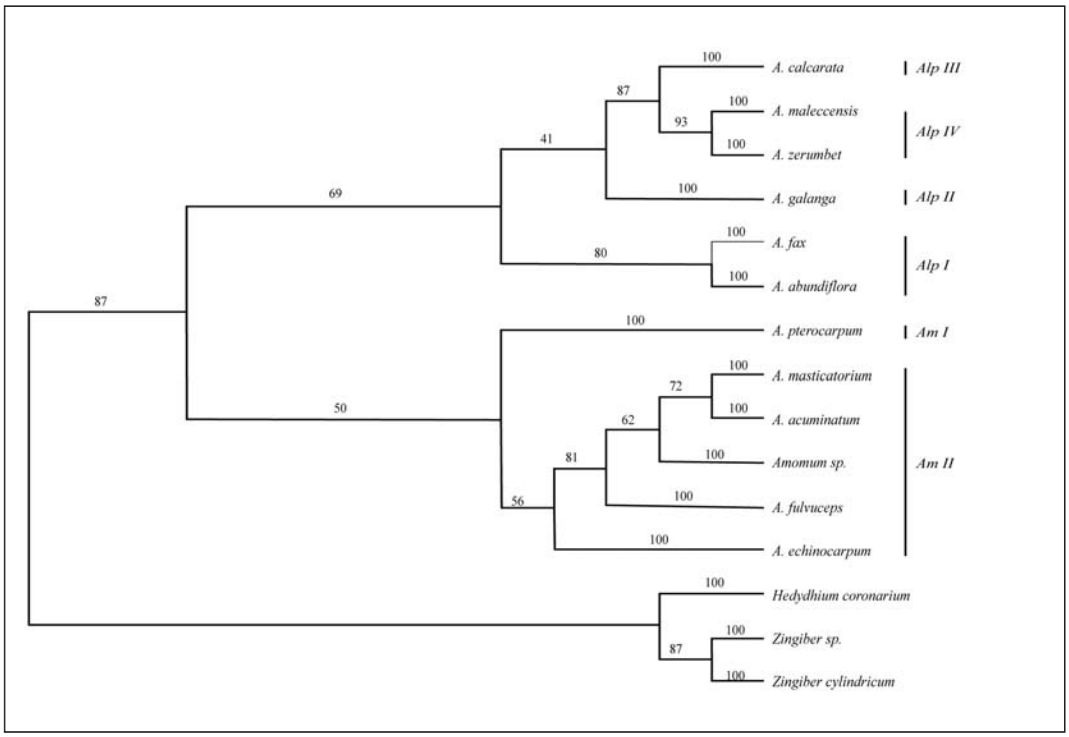
Supplementary data

Supplementary Table 1: GeneBank accessions used in the global phylogeny analysis of gingers in comparison with Sri Lankan species.

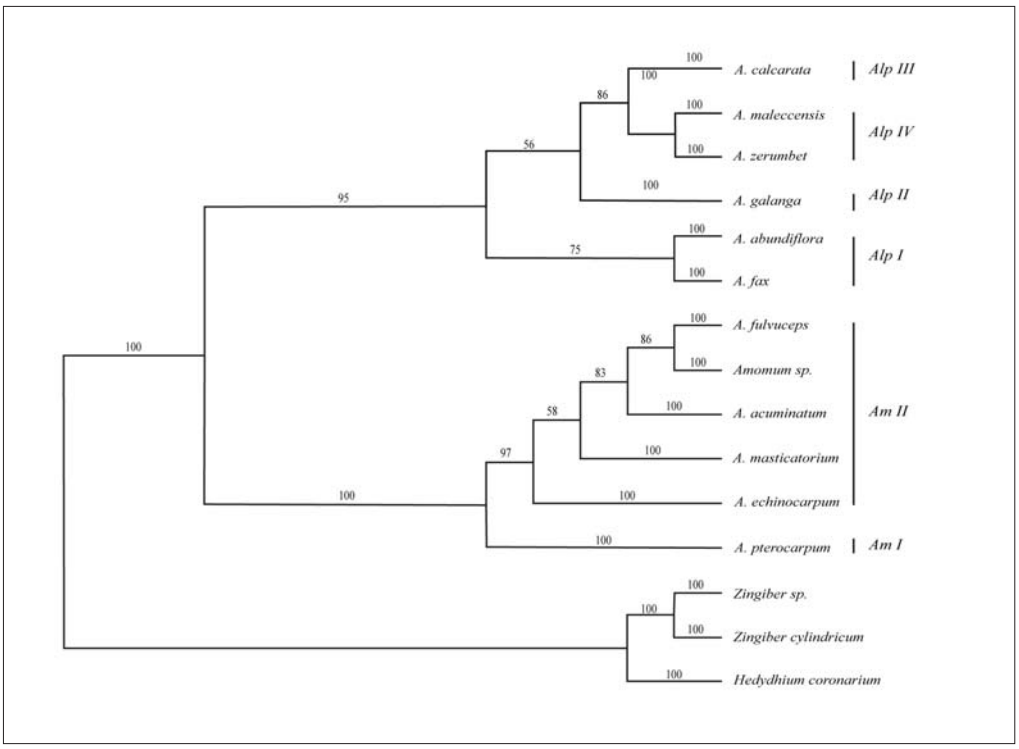
Species	Accession no.	Author(s)
Amomum stenosiphon	MH603415.1	Givnish *et al*., 2018
Elettariopsis triloba	AY769794.1	Harris *et al*., 2005
Amomum longipetiolatum	AY769788.1	Harris *et al*., 2005
Alpinia aff. *calycodes*	AY769797.1	Harris *et al*., 2005
Amomum petaloideum	AY769789.1	Harris *et al*., 2005
Alpinia cf. *aenea*	AY769796.1	Harris *et al*., 2005
Alpinia galanga	KY412470.1	Anju & Yusuf, 2016
Amomum uliginosum	AY769790.1	Harris *et al*., 2005
Zingiber zerumbet	KY412469.1	Anju & Yusuf, 2016
Amomum sp.	AY769793.1	Harris *et al*., 2005
Alpinia purpurata	MH603399.1	Givnish *et al*., 2018
Zingiber spectabile	MH603449.1	Givnish *et al*., 2018
Amomum compactum	NC_036992.1	Wu, 2018
Alpinia oxyphylla	KY985237.1	Gao *et al*., 2017
Alpinia galanga	KJ609030.1	Mathew *et al*., 2014
Amomum glabrum	FJ848631.1	Auvray *et al*., 2009
Alpinia zerumbet	MH603400.1	Givnish *et al*., 2018
Hedychium coronarium	MH603422.1	Givnish *et al*., 2018
Aframomum limbatum	FJ848663.1	Auvray *et al*., 2009
Aframomum pilosum	FJ848652.1	Auvray *et al*., 2009
Aframomum angustifolium	MH603398.1	Givenish *et al*., 2018
Aframomum longiligulatum	FJ848639.1	Auvray *et al*., 2009
Renealmia battenbergiana	AY769802.1	Harris *et al*., 2004
Renealmia occidentalis	AY769801.1	Harris *et al*., 2004
Renealmia alpinia	DQ444491.1	Sarkinen *et al*., 2006



Supplementary Figure 1: One of the best trees resulted from the parsimony analysis of the *trn L-F* sequence data (length = 579; consistency index = 0.936; retention index = 0.982; rescaled consistency index = 0.920) showing bootstrap values; A – *Alpinia* group, B – *Amomum* group.



Supplementary Figure 2: One of the best trees resulted from the parsimony analysis of the *trn S-fM* sequence data (length = 743; consistency index = 0.860; retention index = 0.576; rescaled consistency index = 0.495) showing bootstrap values; **A** – *Alpinia* group, **B** – *Amomum* group.



Supplementary Figure 3: One of the best trees resulted from the parsimony analysis of the combined sequence data of *trn L-F* and *trn S-fM* (length = 1289; consistency index = 0.887; retention index = 0.939; rescaled consistency index = 0.833); branch labels: above - bootstrap values, below – posterior probability; A – *Alpinia* group, B – *Amomum* group, A1- *Alpinia* group without the basal clade of the A group.

J.Natn.Sci.Foundation Sri Lanka 2021 49 (3): 351 - 360
DOI: http://dx.doi.org/10.4038/jnsfsr.v49i3.10231

RESEARCH ARTICLE

Characterisation of clay mineralogy of the major soils in the Northern region of Sri Lanka

P Gowthamy[1], SP Indraratne[2], R Weerasooriya[3] and RB Mapa[1*]
[1] *Department of Soil Science, Faculty of Agriculture, University of Peradeniya, Peradeniya.*
[2] *Department of Environmental Studies and Sciences, University of Winnipeg, Canada.*
[3] *National Institute of Fundamental Studies, Hanthana, Kandy.*

Submitted: 28 September 2020; Revised: 07 January 2021;Accepted: 28 February 2021

Abstract: Clay content and mineralogy affect the behaviour of soils by influencing many soil properties. The soils of Sri Lanka were classified into 23 great soil groups (GSG) in 1972 and presently reclassified according to the international methods of soil taxonomy. Recently, the GSGs of latosols, namely, calcic red latosols, calcic yellow latosols (CRL and CYL), red latosols and yellow latosols (RL and YL) from the Northern region of Sri Lanka were classified under entisol soil order of soil taxonomy in contrast to earlier classification of oxisol soil order. Therefore, the objectives of this study were to characterise clay mineralogy of the major soil series in the Northern region of Sri Lanka and to confirm the soil classification of latosols. Soil samples from nine soil series were collected and analysed for clay mineralogy using infra-red (IR) spectroscopy. Dominated clay minerals were montmorillonite (70 %) and kaolinite (20 %) for vertisol soil order, kaolinite (50−70 %) and montmorillonite (40−20 %) for alfisol soil order and kaolinite (80−90 %) for entisol soil order. All soils have shown evidence for the presence of quartz and feldspar. Latosol GSG showed distinctive IR peaks at 3695 cm^{-1}, 3670 cm^{-1}, 3650 cm^{-1}, and 3620 cm^{-1} representing well-crystallised kaolinite with no evidence for oxide minerals. Latosols did not have mineralogical signatures to classify under oxisols. Mineralogical composition of soil orders of alfisols and vertisols was confirmed with their respective classification. Clay mineralogical information confirmed that latosol GSG could be classified under entisol order though the latosols are at an advanced stage of soil development.

Keywords: Clay minerals, infra-red spectroscopy, latosols, oxisols.

INTRODUCTION

Clay content and mineralogy plays a vital role in governing soil properties in agricultural and engineering applications. Clay particles (< 2 μm), being the finest mineral particles found in soils offer the highest specific surface influencing soil behaviour due to retention of nutrients and water. Clay particles constitutes variety of minerals, namely, micas (illites), vermiculites, smectites (montmorillonites), kaolins (kaolinites), chlorite and other interlayer minerals that differ widely in structure and composition (Ito & Wagai, 2017). In addition to amount of clay, the clay mineralogy mediates many biological and physical properties in soils; formation of stable aggregates (Kraemer *et al*., 2019), persistence of organic matter in soils (Zhao *et al*., 2020), soil moisture characteristic (Williams *et al*., 1983) and the buffering capacity of soil (Indraratne, 2006). Hence, the importance of knowing clay mineralogy in soil management is well documented. In soil classification, dominance of clay types can be used as a tool to confirm soil orders under soil taxonomy (Soil Survey Staff, 2014); oxisols soil order consists of 1:1 clay minerals (kaolinite) and Fe and Al oxyhydroxides, such as gibbsite, hematite and goethite (Schaefer *et al*., 2008), Vertisols soil order dominate 2:1 expandable smectite clay minerals (Pal, 2017) and alfisols soil order

* Corresponding author (maparb@yahoo.com; https://:orcid.org/0000-0002-2223-2588)

is dominated by kaolinite and illite (Zhang *et al*., 2016). Soil family level informationin soil taxonomy is directly related to dominant soil mineral types (e.g., kaolinitic, smectitic, gibbsitic) and particle-size distribution (e.g., clayey, loamy, sandy) (Soil Survey Staff, 2014).Soils of Sri Lanka were classified in 1972 into 23 great soil groups (GSG) by de Alwis and Panabokke (1972). With the global advancement of soil science , a need arisen to classify the soils of Sri Lanka according to the international methods of soil taxonomy and to develop a national soil database (Mapa, 2020).

Clay mineralogy of Sri Lankan soils has been studied by many researchers (Panabokke, 1958; Kalpage *et al*., 1963; De Alwis & Pluth, 1976; Indraratne 2006;), mostly using x-ray diffraction technique covering major rainfall zones (wet, intermediate and dry zones). In general, wet zone consists of kaolinite, illite, gibbsite, hydroxy-interlayered vermiculite, and traces of boehmit; intermediate zone consists of kaolinite, smectites, vermiculites, and mica; dry zone consists of kaolinite, smectite, vermiculite, and mica (Indraratne, 2020). Most of these studies did not include soils of the Northern region of Sri Lanka. De Alwis and Panabokke (1972) identified nine GSG in the Northern region consisting of calcic red latosols (CRL), calcic yellow latosols (CYL), red latosols (RL), yellow latosols (YL), grumusols, solodized solonetz (SS), reddish brown earth soils (RBE), low humic gley soils (LHG), and sandy regosols (SR). According to the latest soil classification of soil taxonomy (Mapa *et al*., 2010), these GSGs belong to entisols (CRL, CYL, RL, YL and SR), alfisols (RBE, LHG, and SS) and vertisols (Grumusols) soil orders. In an earlier classification, latosols in the Northern region (CRL, CYL, RL, and YL) were classified under oxisols (de Alwis & Panabokke, 1972). Though the other GSG are found elsewhere in Sri Lanka, latosols are unique only to the Northern region of the country. Miocene limestone of Sri Lanka lying on the Precambrian basement of the North, Northern and South-western coastal belts (Cooray, 1984; Katupotha & Dias, 2001) is the parent materials mainly responsible for the development of soils in the region. The latosols are developed on transported material that overlies Miocene limestone and mainly occurs in the Jaffna Peninsular (Panabokke, 1996; Chandrajith, 2020). Hence, there is a controversy in classifying latosols according to soil taxonomy; De Alwis and Panabokke (1972) placed them under oxisols while Mapa *et al*. (2010), placed them under entisols. Deeply weathered soils with no distinct horizon boundaries occurring in dense rainfall areas are classified as latosols in many parts of the world (Ruivo & Cunha 2003). In latosols the soil profile is uniform in chemical and mineral composition, kaoline making up >50 % of the clay fraction together with iron oxides and gibbsites (Sherman & Alexander, 1959). Most of these latosols reported globally were classified under oxisols in Soil Taxonomy (Schaefer *et al*., 2008). Therefore, clay mineralogical analysis of latosols of the Northern region will provide sufficient information to a certain degree to solve this controversy on soil classification.

Fourier transform infrared (FTIR) spectroscopy provide valuable information on mineralogical composition as each mineral has a unique absorption pattern in the mid-IR range (Hahn *et al*., 2018). Clay mineral units generally constitute hydroxyl groups, tetrahedral silicate/aluminate anions, octahedral metal cations, and interlayer cations. FTIR-spectral signatures of minerals enable to identify individual minerals as well as non-crystalline admixtures using unique absorption patterns (Vaculicova & Plevova, 2005). IR spectra of clay minerals are usually characterise by three main areas: (i) the stretching and bending vibrations of the inner surface –OH groups observed in the region of 3700 to 3600 cm^{-1}, (ii) the stretching and bending vibrations of the Si–O groups, and (iii) the Si–O–M (with M = Al, Mg, etc.) vibrations extending from 1200 to 400 cm^{-1} (Müller *et al*., 2014).

The technical difficulty in quantifying clay-sized minerals is a well-known fact (Środoń, 2013), which leads to limited published soil datasets, globally (Ito & Wagai, 2017). Soil mineralogical characterisation has not been done to date for the soils in Northern region of Sri Lanka. The peculiar climatic and geological characteristics of the Northern region offer a unique opportunity to unravel the provenance of clay types in the region. Identifying the mineralogical composition of the soils of the Northern region is essential in terms of filling a knowledge gap on soil mineralogy which could be used for planning agricultural and engineering applications. Therefore, we characterised the clays and other minerals in major soils in the Northern region using Fourier transform infrared spectroscopy (FTIR). The FTIR spectra provide spectral signatures that can be unequivocally used to identify different minerals in soils. We hypothesised that the presence of Fe and Al hydroxides together with well-crystallised kaolinite in latosols would give an insight to confirm latosols under oxisols.

METHODOLOGY

Study location and collection of soils

Soils of the Northern part of Sri Lanka were collected from pre-identified benchmark sites representing major soil series (Mapa, 2016) during 2013 and 2014 covering two agro-ecological regions (DL3 and DL4) in low-country dry zone. The major series of soil of Northern

region were identified based on soil, topography, parent material, climate, vegetation and previous information with the assistance of the Land Use Division of the Irrigation Department of Sri Lanka. Out of 12 recently geo-referenced benchmark soil profiles of Northern region of Sri Lanka (Mapa, 2016), nine different major soil series were selected for mineralogical analysis (Figure 1). Details of identified benchmark soil profiles and their locations were described elsewhere (Vitharana *et al.*, 2019). The soil series names, GSG names according to classification in 1972 (De Alwis & Panabokke, 1972) and their Soil Taxonomic equivalents (Soil Survey Staff, 2014) to sub-group level are given in Table 1. These include GSGs of CRL, CYL, SS and SR from Jaffna, RBE and LHG from Vavuniya, RL and YL from Vishwamadu, and grumusols from Mannar. The exact location / benchmark site for each soil series was recorded using GPS (Figure 1, Table 1). A soil pit was dug to a depth of 1m or until hit by the parent rock at each location. Composite soil samples collected from the surface horizon (0-30 cm depth) of benchmark sites were subjected to the separation of clay fraction and mineralogical analysis.

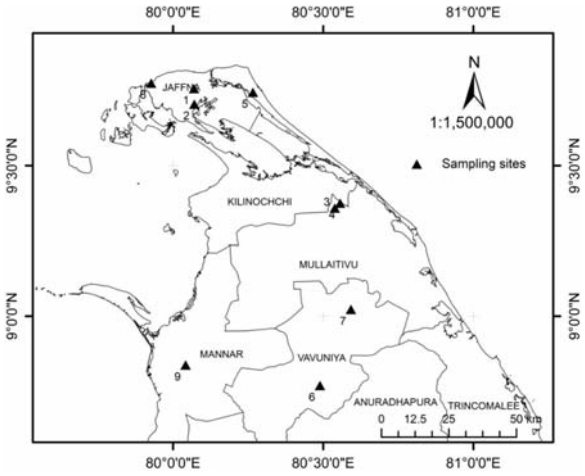


Figure 1: Sampling locations for clay mineralogy analysis of the study

Table 1: The sample locations, great soil groups, soil series names and equivalent soil taxonomic names of soils of the Northern region of Sri Lanka

Sample location	Soil series	Great soil group*	Abbreviation	Soil taxonomy
Lat : 090 45'25.5" N, Lon: 0800 04' 11.4"E	Inuvil	Calcic red latosols	CRL	Typic Ustorthents
Lat : 090 42'18.3"N, Lon: 0800 04'15.6"E	Chankanai	Calcic yellow Latosols	CYL	Aquic Ustorthents
Lat : 090 21' 31.6" N, Lon: 0800 32' 19.1" E	Mulathivu	Red latosols	RL	Typic Ustorthents
Lat : 090 22' 30" N, Lon: 0800 33' 22.6" E	Thanniyuttu	Yellow latosols	YL	Oxyaquic Ustorthents
Lat : 090 44' 40.1" N, Lon: 0800 15' 57.4" E	Colombatharaj	Sandy regosol	SR	Aquic Quartzipsamments
Lat : 080 46'8.90" N, Lon: 0800 29' 21.11" E	Vavunia	Low humic gley	LHG	Typic Endoaqualfs
Lat : 090 01'22.30" N, Lon: 0800 35' 32.34" E	Nadunkanei	Reddish brown earth	RBE	Typic Rhodustalfs
Lat : 090 46' 29.4" N, Lon: 0790 55' 37.3" E	Tondaimanar	Soladized solonetz	SS	Albic Natraqualfs
Lat: 080 50' 14.1" N, Lon: 0800 02' 30.1" E	Murunkkan	Grumusols	Grumusols	Aridic Endoaquerts

*source de Alwis & Panabokke, 1972; Lat=latitude;=Lonlongitude

Separation of clay fraction and mineralogical analysis

Soil samples were washed with distilled water to remove soluble salts and treated with 30 % H_2O_2 for organic matter removal and sodium citrate-bicarbonate-dithionite at 80 ^{0}C for iron oxide removal before separation of the clay fraction (< 2 μm).

Clay fraction was separated by sedimentation under gravity according to Stoke's Law (Gee & Or, 2002). Clay samples were air-dried, and soil mineralogy was characterised using infrared spectrometry (IR) (White & Roth, 1986). We used IR spectrums of kaolinite, illite feldspar, and quartz reported by Djomgoue and Njopwouo (2013), Hanhn *et al*. (2018) and Müller *et al*. (2014) to interpret data. These were selected as the probable minerals present in the dry zone soil, based on our prior knowledge of types of clay and other minerals of Sri Lanka (Indraratne, 2020).

FT-IR spectroscopic studies

For the FT-IR studies, all spectra were acquired with a Thermo Nicolet Magna FTIR Si 50 (Madison, WI) with a DTGS KBr detector at the Department of Soil Science, University of Peradeniya. The wavenumber range was 4000 to 400 cm^{-1}, and each spectrum consisted of 64 scans; the resolution was 4 cm^{-1}, and wavenumber accuracy was 1 cm^{-1}. About 20 mg of finely ground soil samples were mixed with 12 mg IR grade KBr and pressed into a 3-mm-diameter pellet using paper inserts (Spectra-Tech, Shelton, CT). Clay IR spectra were compared against standard kaolinite, montmorillonite, illite, quartz and feldspar spectra to attest their signatures.

RESULTS AND DISCUSSION

Qualitative analysis of clay minerals

The IR spectrum of kaolinite is readily identifiable and definable over other clay minerals due to a unique pattern in the spectral region of the inner surface -OH vibrations. Kaolinite shows four clear distinctive peaks at 3695 cm^{-1}, 3670 cm^{-1}, 3650 cm^{-1}, and 3620 cm^{-1}, which are attributed to the inner phase and outer phase motion modes, and the stretching vibration of inner surface -OH groups (Balan *et al*., 2005). Out of the four well resolved (-OH) bands in the IR spectrum of kaolinite, the fourth band (3620 cm^{-1}) is attributed to the vibrations of inner hydroxyl groups (Dijomgoue & Njopwouo, 2013). The IR spectra of reddish brown earth (RBE) and low humic gley (LHG) showed four distinct peaks due to kaolinite at 3695 cm^{-1}, 3670 cm^{-1}, 3650 cm^{-1}, and 3620 cm^{-1}. In FTIR of RBE showed a clear, distinct, and equal intensity peaks at 3695 cm^{-1} and 3620 cm^{-1}, whereas LHG showed two sharp peaks but with a stronger peak at 3620 cm^{-1} than at 3695 cm^{-1} (Figure 2). The characteristic vibration peaks for smectite (O-H stretching) were at 3628 cm^{-1} (Madejová, 2003). Generally, the 3620 cm^{-1} stretching mode can be presented in both kaolinite and 2:1 clay mineral [dioctahedral illite and (or) illite-vermiculite] (Szymański *et al*., 2014). When the bond stretching intensity of 3620 cm^{-1} was stronger than that of 3695 cm^{-1}, that indicated the presence of a large amounts of 2:1 type clay mineral (Zhang *et al*., 2016). LHG showed a stronger peak at 3620 cm^{-1} than at 3695 cm^{-1} indicating that LHG constitutes higher montmorillonite and lower kaolinite than RBE. Broad peaks can observe represented water-absorption bands in the region (2800–3700 cm^{-1}) of smectite, illite and chlorite (Hahn *et al*., 2018; Madejová, 2003). Wavelength peaks at 3624 cm^{-1} (Al-OH) and 3422 cm^{-1} (Water) can be attributed to montmorillonite (Long *et al*., 2013). Water-absorption peaks at 3422 cm^{-1} in both RBE and LHG indicates the presence of montmorillonite, but the sharpness of the peak is greater in LHG than in RBE. Hydrated phyllosilicates (e.g. chlorite, illite, smectite) show peaks around 1630 cm^{-1}, related to the H-O-H bonds of absorbed water (Angaji *et al*., 2013). These peaks, however, are not diagnostic features of specific clay minerals, but indicative of the presence of water-absorbed minerals like montmorillonite. A peak was observed at 1620 cm^{-1} in both RBE and LHG, and the LHG peak is stronger than the RBE. Illite and montmorillonite have similar spectra, except a medium spectral bending vibration peak of OH group at 1635 cm^{-1} and a weak shoulder peak at 885 cm^{-1} attributed to the Fe-Al-OH vibration, which is unique only to montmorillonite (Müller *et al*., 2014). The changes in the structures of kaolinite and 2:1 type clay minerals (illites and smectites) also occurred in the range of 1200 - 400 cm^{-1} (Zhang *et al*., 2016). Well-crystalline kaolinite would present four groups of dual vibration modes near 1100−1120, 1000−1040, 910−940 and 753−795 cm^{-1}, which were assigned to Si-O bands (Madejová, 2003). All these peaks in the range of 1200 - 400 cm^{-1} are visible in RBE and LHG, where RBE showed more clear and distinct peaks than LHG. Furthermore, a shoulder peak arising from the Al-OH-Al bending vibration is evident at 915 cm^{-1} for illite, montmorillonite and kaolinite (Srasra *et al*., 1994). We conclude that alfisols consisted of kaolinite and montmorillonite (and may also be with illite) as clay minerals. The sharpness and prominence of the peaks revealed that the LHG is having an equal amounts of kaolinite and montmorillonite, where as RBE is predominating with kaolinite. The major absorption features of quartz are the peaks in between

1200 to 900 cm^{-1} assigned to the asymmetric stretching vibrations of the Si–O groups with a peak maximum at 1080 cm^{-1}, the symmetric stretch at 800 cm^{-1} and 780 cm^{-1}, and the symmetric and asymmetric Si–O bending modes at 695, 520, and 450 cm^{-1}, respectively (Müller *et al*., 2014; Saikia *et al*., 2008). RBE and LHG showed peaks at 1002 cm^{-1}, 800 cm^{-1}, 780 cm^{-1}, 695 cm^{-1} and 520 cm^{-1}, confirming presence of quartz. The IR spectrum of SS, which is a saline soil, was very complicated and undefined probably due to interferences from other minerals or free ions and oxides in the soil. We could not predict the presence of clay minerals, because there were no clear identifiable peaks for kaolinite, montmorillonite or illite in the absorption spectra in SS. At the same time, in SS, there was a prominent peak at 1742 cm^{-1}, which was non-definable in terms of major clay minerals.

The carbonates have the fundamental vibrations due to $(CO_3)_2$- ion assigned to the asymmetric stretch at 1400 cm^{-1} and the outer plane bending vibration at 875 cm^{-1} (Chester & Elderfield, 1967). There was a broad peak at 1400 cm^{-1} and unidentifiable peaks at 875 cm^{-1}, 727 cm^{-1} and 712 cm^{-1} in the spectra of SS signifying the probable presence of calcite. Hence, out of the three alfisols, RBE and LHG consisted of kaolinite, montmorillonite (illite) and quartz, LHG having higher montmorillonite than RBE. SS showed no strong evidence for clay minerals due to broad peaks. The smectite, illite and kaolinite were the dominant clay minerals present in alfisols found in many parts of the dry zone of Sri Lanka (Indraratne, 2010). Clay mineralogy of RBE and LHG collected from North Central Province of Sri Lanka also reported kaolinite as the dominant and smectite (montmorillonite)

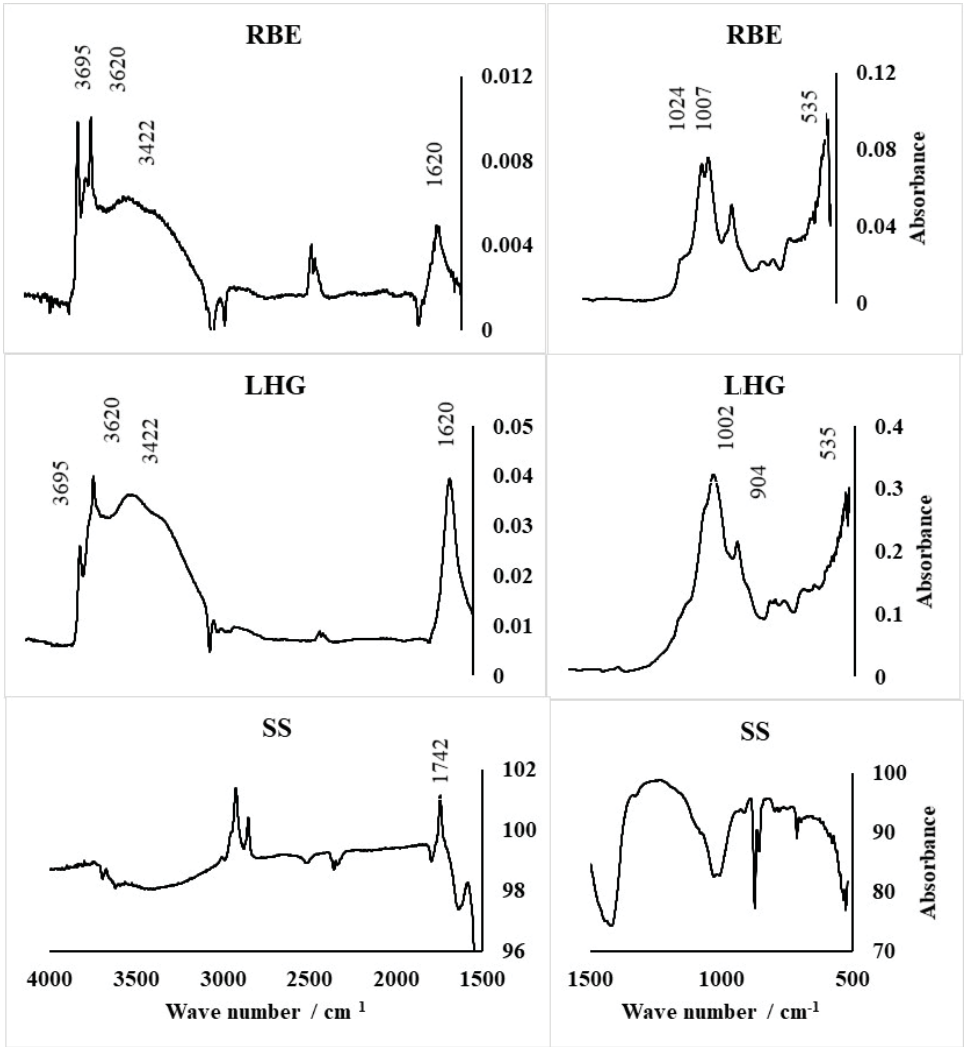


Figure 2: Infrared spectra for the wave number range 4000 to 400 cm^{-1} for clay fraction of great soil group of reddish brown earth (RBE), low humic gley (LHG) and solodized solonetz (SS) classified under alfisol order

as the accessory minerals (Kalpage *et al*., 1963 Mapa 1992:). Clay mineralogical changes have been reported for RBE and LHG where RBE at a well-drained crest position showed a greater level of kaolinite in clay fraction, while LHG in the poorly drained valley showed significant amount of montmorillonite clay (Panabokke, 1958 Mapa, 1992;). SS collected from DL1a agro-ecological zone of Sri Lanka (dry zone), previously indicated clay mineralogy of 50 % kaolinite, 20 % montmorillonite and 30 % illite (Indraratne, 2020). Kaolinite and montmorillonite were present in the grumusols as indicated by the peaks at 3695 cm^{-1} and 3620 cm^{-1} , and the latter is stronger than the former peak (Figure 3). The clay fraction of the grumusols had obvious and distinct peaks at 3422 cm^{-1} and 1635 cm^{-1} (Figure 3), related to the H-O-H bonds of absorbed water, confirming the dominance of montmorillonite clay mineral. Grumusols showed a broad distinctive peak in the range of 1200 cm^{-1} to 840 cm^{-1}, indicating the presence of kaolinite, montmorillonite and quartz mixture. There were no distinctive kaolinite peaks present as in RBE, and the dominance of montmorillonite is quite visible in grumusols. Grumusols classified into vertisols had montmorillonite as the dominant clay mineral and kaolinite as the accessory mineral. In a previous study, a grumusols belonged to vertisols, showed 20−30 % of kaolinite and 70–80 % of montmorillonite and vermiculite in the clay composition (Indraratne, 2020).

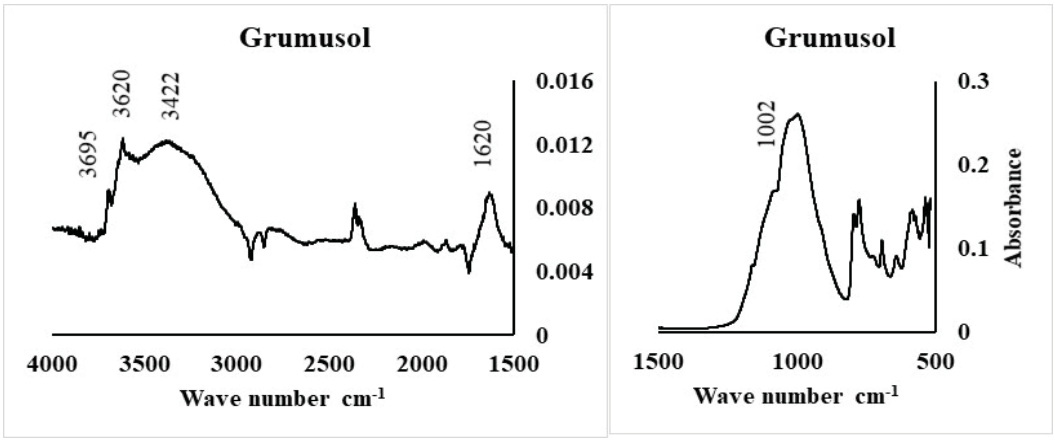


Figure 3:. Infrared spectra for the wave number range 4000 to 400 cm^{-1} for clay fraction of great soil group of grumusols classified to vertisol order

Sandy regosols (SR), classified under entisols had distinct, clear peaks for kaolinite at 3695 cm^{-1}, and 3620 cm^{-1} (Figure 4) with the indication of the presence of both kaolinite and montmorillonite as discussed before. Montmorillonite signatures in SR were not strong as in grumusols, LHG or RBE, indicating the presence of trace amounts. Peaks present in between 1200 and 900 cm^{-1}, with a very clear peak at 1002 cm^{-1}, signified the presence of quartz, and four distinctive peaks in the region of 1100 cm^{-1} to 795 cm^{-1} indicated the presence of well-crystallised kaolinites. Therefore, kaolinite is the dominant clay mineral in SR

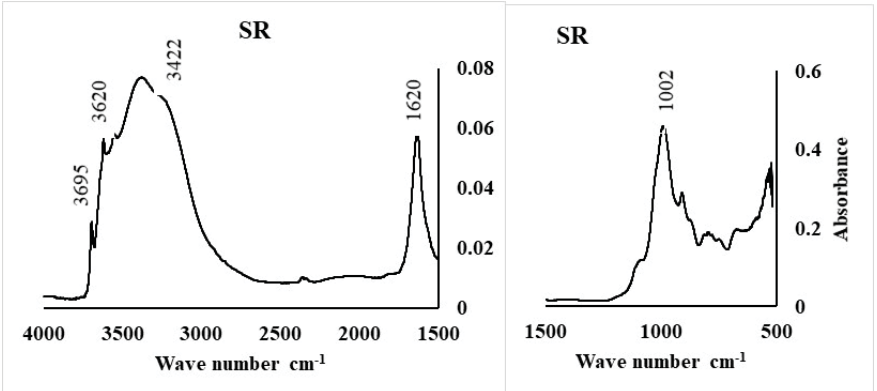


Figure 4:. Infrared spectrafor the wave number range 4000 to 400 cm^{-1} for clay fraction of great soil group of sandy regosols (SR) classified to entisol order

with traces of montmorillonite. The infrared spectra of the clay minerals of CRL, CYL, RL, and YL are shown in Figure 5. All four great soil groups had distinct, clear peaks at 3695 cm^{-1} and 3620 cm^{-1}, a clear indication of the presence of kaolinite clay mineral. In CYL and CRL, a stronger peak at 3620 cm^{-1} than at 3695 cm^{-1}, and peak at 1620 cm^{-1} indicated the presence of montmorillonite as an accessory mineral. RL and YL showed similar spectral fingerprints and kaolinite as the dominant clay mineral showed equally strong, distinctive peaks at 3695 cm^{-1} and 3620 cm^{-1}. A clear peak at 1002 cm^{-1} and the symmetric and asymmetric Si–O bending mode at 695 cm^{-1} and 520 cm^{-1} signify the presence of quartz mineral. The prominence of kaolinite in RL and YL further confirmed with the clear kaolinite peaks from 1100 cm^{-1} to 795 cm^{-1} range. Kaolinite is the dominant clay mineral present in these four latosols classified under entisols order. X-ray diffraction studies confirmed that RL consists of dominantly kaolinite and traces of illite (Indraratne, 2020). IR spectra should report four strong absorption OH-stretching bands at 3397, 3467, 3529 and 3623 cm^{-1} wavenumbers to confirm the presence of gibbsite (Balan *et al*., 2005; Favaro *et al*., 2010) and there were no such peaks visible for any of the soils studied.

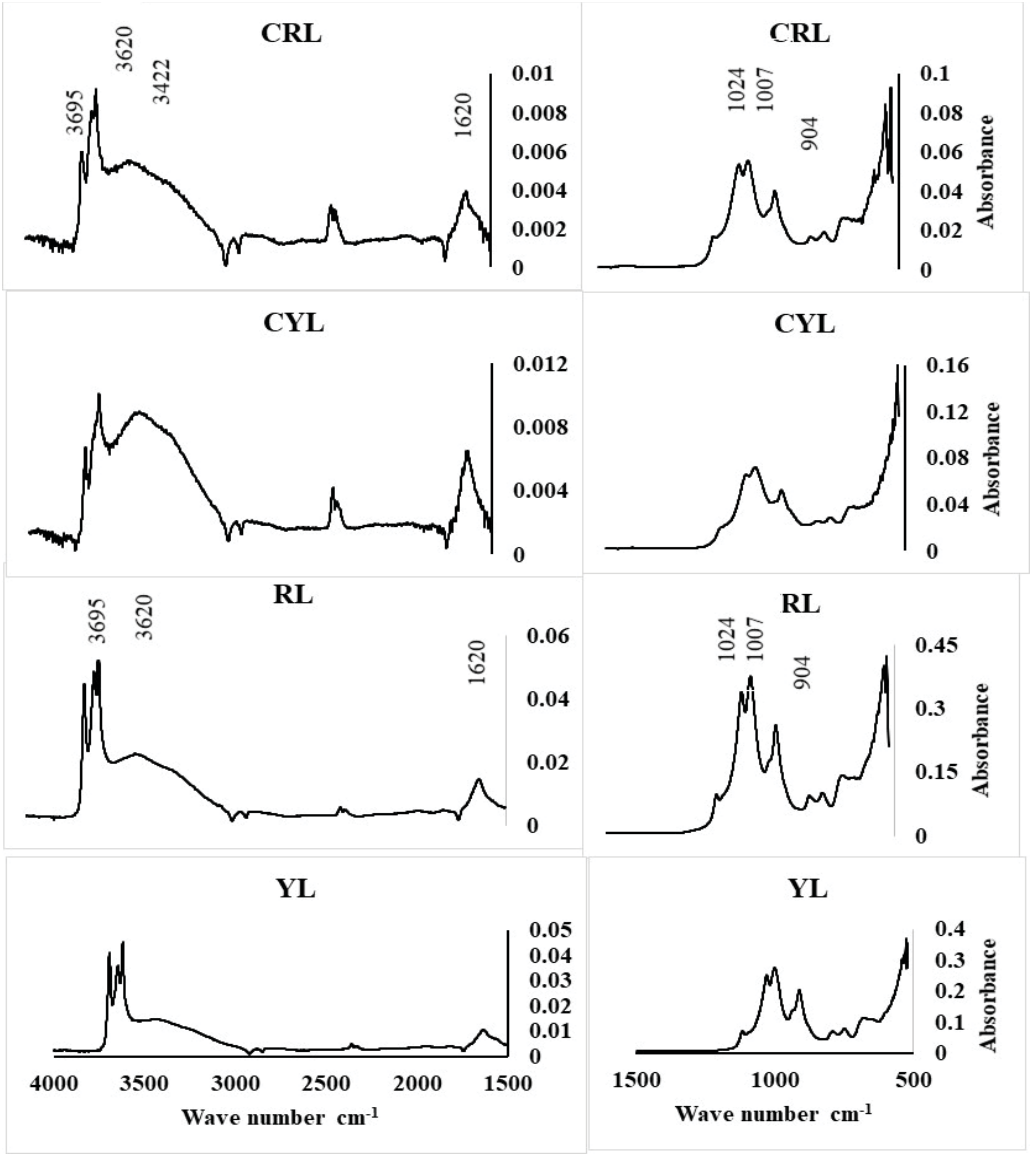


Figure 5:. Infrared spectra for the wave number range 4000 to 400 cm^{-1} for clay fraction of great soil groups of calcic red latosol (CRL), calcic yellow latosol (CYL), red latosol (RL) and yellow latosol (YL) classified under entisol order

The kaolinite group with 1:1 type layer structure is predominant in the humid tropics, and the illite/mica group is abundant in arid and high-latitude regions (Ito & Wagai, 2017). Kaolinite, smectite, vermiculite, and mica were the mineralogical make-up for dry zone soils of Sri Lanka (Indraratne, 2020). RBE and LHG consisted of kaolinite, montmorillonite and quartz confirming mineralogical signature for Alfisols. Vertisols contain a high proportion of smectite in the topsoil, which are known to have localised distributions in the tropics (Ito & Wagai, 2017). Grumusols classified under vertisols had montmorillonite as the dominant clay mineral and kaolinite as the accessory mineral. Well-crystallised kaolinite was the dominant clay mineral in SR, CYL, CRL, RL and YL classified under entisols order.

Clay mineralogy to predict relative development stages of great soil groups

Using IR spectra of clay minerals given in Figures 2−5 for the soils of the Northern region of Sri Lanka (except for SS), the relative abundance was estimated by sharpness and prominence of peaks for kaolinite and montmorillonite (Table 2). Relative abundance of clay minerals for SS was derived from the work of Indraratne (2010) in order to compare it with other soils. Kaolinite was the dominant clay mineral in all soils except grumusols. Soils can be arranged according to the dominance (sharpness of the peak) of kaolinite among soils in ascending order as grumusols < LHG < RBE < CRL < CYL < SR < RL, YL. Montmorillonite content of the soil is varied among soils from very high (grumusols) to trace amounts (RL, YL, and SR). Clear and distinct montmorillonite peaks was observed in grumusols LHG and RBE. Weak montmorillonite signatures were observed in CYL, CRL SR and RL and YL with trace amounts. According to the sharpness and dominance of frequency, peaks of montmorillonite soils could be arranged in ascending order as grumusols >> LHG > > RBE >> CYL, CRL > SR > RL, YL. The LHG occurred in a poorly drained (aqualfs) part of the catena compared to RBE, showed a high content of montmorillonite clays than in the RBE, which is a well-drained (ustalfs) soil. The reason for this difference is the lower weathering rates experienced in poorly drained conditions which affect the mineralogy even if the soils are formed from the same parent martial.

Generally, evolutionary sequence of clay minerals is in the order primary weatherable minerals (biotite, muscovite, feldpar) illite smectite (and or vermiculite) kaolinite (Djomgoue & Njopwouo, 2013). Weerasuriya *et al* (1991) predicted weathering stage of Sri Lankan soils using geochemical characteristics such as residual primary minerals, secondary clay minerals and silica content. The relative abundance of phyllosilicates, one of the indices of weathering intensity, can be used to predict the relative weathering stage of studied soils (Table 2). Grumusols, the vertisolic soil, qualified as the youngest soil among the studied soils due to the presence of 70% 2:1 montmorillonite clays. Alfisols soils, SS, RBE and LHG, were the second youngest soil, LHG being younger than SS and RBE due to presence of a higher proportion of montmorillonite. When compared with alfisols in the Northern region (RBE, LHG), order entisols (CRL, CYL, RL, YL and SR) had low to traces of montmorillonites, and high proportions (> 80 %) of kaolinite indicating these entisols are at an advanced stage of weathering compared to other soil orders. There was a limited information on mineralogy of latosolic soils in Sri Lanka. According to this study, kaolinite is the dominant clay mineral with or without traces of montmorillonite in latosols. With the help of the clay mineralogy, soils of the Northern region can be arranged from young to mature; grumusols < LHG < RBE, SS < < CYL, CRL < SR < RL, YL. Distinct variability of clay mineral proportions can be observed in some soil orders; in alfisols, kaolinite > illite/mica > smectite / vermiculite, in vertisols, smectite / vermiculite >> kaolinite, in entisols, illite/mica > smectite / vermiculite > kaolinite, and in oxisols, kaolinite > gibbsite > Fe-oxides (Ito and Wagai 2017). According to the mineralogical composition alfisols (RBE, SS and LHG) and vertisols (Grumusols) confirmed their respective classification. Latosolic soils do not fall under oxisols, due to absence of oxide minerals. Advanced weathering stage of latosols indicates that these soils are fallen under entisols because of lack of diagnostic horizons, but not necessarily they are young at soil development.

Table 2: Estimate of relative abundance of clay minerals (%) in the soils of the northern region

	Great soil group								
Clay mineral	Grumusol	LHG	SS*	RBE	CYL	CRL	RL	YL	SR
Kaolinite	20	50	50	70	80	80	90	90	80
Montmorillonite	70	40	20	20	10	10	0	0	10
Other clay-size minerals	10	10	30	10	10	10	10	10	10

*source (Indraratne, 2020). low humic gley soils (LHG), solodized solonetz (SS), reddish brown earth soils (RBE), calcic yellow latosols (CYL), Calcic red latosols (CRL), red latosols (RL), yellow latosols (YL), and sandy regosols (SR)

CONCLUSIONS

Great soil groups belonging to alfisols soil order (RBE, SS and LHG) confirmed mineralogical signature of kaolinite and montmorillonite as dominant clay minerals and quartz as an accessory mineral. Grumusols classified under vertisols soil order had montmorillonite as the dominant clay mineral and kaolinite as the accessory mineral. Well-crystallised kaolinite is the dominant clay mineral in SR. Kaolinite is the dominant clay mineral present in CYL, CRL, RL and YL classified under entisols order. With the help of the clay mineralogy analysis, soils of the Northern region can be arranged from young to mature; grumusols < LHG , SS < RBE < < CYL, CRL < SR < RL, YL. The absence of oxide clay minerals in CYL, CRL, RL and YL indicates that these latosols cannot be justified classify under oxisols soil order and reclassified under entisol order.

Acknowledgements

This work was supported by the National Research Council of Sri Lanka (NRC Grant 12-122). The assistance given by Mr. A.R. Dassanayake, specialist officer at the Land Use Division of the Irrigation Department in identifying the great soil groups of the Northern region of Sri Lanka and Dr. H.A. Kadupitiya of the Natural Resource Management Center, Department of Agriculture, Peradeniya in preparation of the sampling site map is appreciated.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

Angaji M.T., Zinali A.Z. & Qazvini N.T. (2013). Study of physical, chemical and morphological alterations of smectite clay upon activation and functionalization via the acid treatment. *World Journal of Nano Science and Engineering* **03**(4): 8.
DOI: https://doi.org/10.4236/wjnse.2013.34019

Balan E., Lazzeri M., Saitta A.M., Allard T., Fuchs Y. & Mauri F. (2005). First-principles study of OH-stretching modes in kaolinite, dickite, and nacrite. *Journal of American Mineralogist* **90**: 50–60.
DOI: https://doi.org/10.2138/am.2005.1675

Chester R. & Elderfield H. (1967). The Application of Infra-Red absorption spectroscopy to sheet in mixed-layer illite/smectite from bentonites. *European Journal of Mineralogy* **10**: 111–124.

Chandrajith R. (2020). Geology and geomorphology. In: *The Soils of Sri Lanka* (ed. R.B. Mapa), pp. 23–34. Springer International, Switzerland.
DOI: https://doi.org/10.1007/978-3-030-44144-9

Cooray P.G. (1984). Geology, with special reference to the Precambrian. In: *Ecology and Biogeography in Sri Lanka,* pp. 1–34. Springer, Dordrecht, Germany.
DOI: https://doi.org/10.1007/978-94-009-6545-4_1

De Alwis K.A. & Panabokke C.R. (1972). Handbook of the soils of Sri Lanka. *Journal of the Soil Science Society of Sri Lanka* **2**: 21–52.

De Alwis K.A. & Pluth D.J. (1976). The Red Latosols of Sri Lanka II. Mineralogy and weathering. *Journal of the Soil Science Society of America* **40**: 920–928.
DOI:https://doi.org/10.2136/sssaj1976.03615995004000060032x

Djomgoue P. & Njopwouo D. (2013). FT-IR spectroscopy applied for surface clays characterization. *Journal of Surface Engineered Materials and Advanced Technology* **3**(04): 275.
DOI: http://dx.doi.org/10.4236/jsemat.2013.34037

FAO (2014). *World Reference Base for Soil Resources. International Soil Classification System for Naming Soils and Creating Legends for Soil Maps*. Food and Agriculture Organization, Rome, Italy.

Favaro L., Boumaza A., Roy P., Lédion, J., Sattonnay G., Brubach J.B., Huntz A.M. & Tétot R. (2010). Experimental and ab initio infrared study of χ-, κ-and α-aluminas formed from gibbsite. *Journal of Solid State Chemistry* **183**(4): 901–908.
DOI: https://doi.org/10.1016/j.jssc.2010.02.010

Gee G.W. & Or D. (2002). *Particle Size Analysis. Methods of Soil Analysis.* Part 4. Physical Methods (eds. H. Dane & C. Topp), pp. 255–289. Soil Science Society of America, Madison, Wisconsin, USA.

Hahn A., Vogel H., Andó S., Garzanti E., Kuhn G., Lantzsch H., Schüürman C., Vogt C. & Zabel M. (2018). Using Fourier transform infrared spectroscopy to determine mineral phases in sediments. *Sedimentary Geology* **375**: 27–35.
DOI: https://doi.org/10.1016/j.sedgeo.2018.03.010

Indarratne S.P. (2020). Soil mineralogy. *In: Soils of Sri Lanka* (ed. R.B. Mapa). Springer International, Switzerland.
DOI: https://doi.org/10.1007/978-3-030-44144-9

Indraratne S.P. (2010). Mineralogy of the soils of Sri Lanka. In: *Soils of the Dry Zone of Sri Lanka. Morphology, Characterization and Classification* (eds. R.B. Mapa, S. Somasiri & A.R. Dassanayake). Special publication No. 7, pp. 48–67. Soil Science Society of Sri Lanka, Survodaya Publishers, Colombo.

Indraratne S.P. (2006). Occurrence of organo-mineral complexes in relation to clay mineralogy of some Sri Lankan soils. *Journal of the National Science Foundation of Sri Lanka* **34**: 29–36.

Ito A. & Wagai R. (2017). Global distribution of clay-size minerals on land surface for biogeochemical and climatological studies. *Scientific data* **4**: 170103.
DOI: https://doi.org/10.1038/sdata.2017.103

Kalpage F.S.C.P., Mitchell B.B. & Mitchell W.A. (1963).

The mineralogy of some Ceylon soils. *Clay Minerals Bulletin* **5**(30): 308–318.
DOI: https://doi.org/10.1180/claymin.1963.005.30.07

Katupotha J. & Dias P. (2001). The geological evolution correlated to the stratigraphy of the Kalpitiya. *Journal of Indian Association Sedimentologists* **20**: 21–37.

Kraemer F.B., Hallett P. D., Morrás H., Garibaldi L., Cosentino D., Duval M. & Galantini J. (2019). Soil stabilisation by water repellency under no-till management for soils with contrasting mineralogy and carbon quality. *Geoderma* **355**: 113902.
DOI: https://doi.org/10.1016/j.geoderma.2019.113902

Long H., Wu P. & Zhu N. (2013). Evaluation of Cs+ removal from aqueous solution by adsorption on ethylamine-modified montmorillonite, *Chemical Engineering Journal* **225**: 237–244
DOI: https://doi.org/10.1016/j.cej.2013.03.088

Madejová J. (2003). FTIRS techniques in clay mineral studies. *Vibrational Spectroscopy* **31**(1): 1–10.
DOI: https://doi.org/10.1016/S0924-2031(02)00065-6

Mapa R.B. (2020). Soil research and soil mapping history. In: *The Soils of Sri Lanka* (ed. R.B. Mapa), pp. 1–13. Springer International. Switzerland.
DOI: https://doi.org/10.1007/978-3-030-44144-9

Mapa R.B. (2016). Characterization of Soils in the Northern Region of Sri Lanka to Develop a Soil Data Base for Land Use Planning and Environmental Applications. Final Report Submitted to National Research Council of Sri Lanka. Grant No. 12–122.

Mapa R.B., Somasiri S. & Dassanayake A.R. (2010). *Soils of the Dry Zone of Sri Lanka. Morphology, Characterization and Classification,* Special publication No. 7, pp. 288. Soil Science Society of Sri Lanka, Survodaya Publishers, Colombo.

Mapa R.B. (1992). Clay mineralogy of six Sri Lankan soils. *Journal of Geological Society of Sri Lanka* **4**: 45–47.

Müller C.M., Pejcic B., Esteban L., Delle L., Piane C., Raven M. & Mizaikoff B. (2014). Infrared attenuated total reflectance spectroscopy: an innovative strategy for analyzing mineral components in energy relevant systems. *Scientific reports* **4**: 6764.
DOI: https://doi.org/10.1038/srep06764

Pal D.K. (2017). Cracking clay soils (vertisols): pedology, mineralogy and taxonomy. In: *A Treatise of Indian and Tropical Soils pp.* 9–42. Springer, Germany.
DOI: https://doi.org/10.1007/978-3-319-49439-5_2

Panabokke C.R. (1958). A pedology study of dry zone soils. Tropical Agriculturist **64**: 151–174.

Panabokke C.R. (1996). *Soils and Agro-Ecological Environments of Sri Lanka.* Natural Resources Series No. 2, pp. 219. Natural Resource, Energy and Science Authority of Sri Lanka, Colombo 07.

Ruivo M.L.P. & Cunha E.S. (2003). Mineral and organic components. In: *Archaeological Black Earth and Yellow Latosol In Caxiuang, Amazon, Brazil.* WIT Transactions on Ecology and the Environment, volume. 64. WIT Press, UK.
DOI: https://doi.org/10.2495/ECO030342

Saikia B.J., Parthasarathy G. & Sarmah N.C. (2008). Fourier transform infrared spectroscopic estimation of crystallinity in SiO_2 based rocks. J*ournal of Bulletin of Materials Science* **31**: 775–779.
DOI: https://doi.org/10.1007/s12034-008-0123-0

Schaefer C.E.G.R., Fabris J.D. & Ker J.C. (2008). Minerals in the clay fraction of Brazilian Latosols (Oxisols): a review. *Clay Minerals* **43**(1): 137–154.
DOI: https://doi.org/10.1180/claymin.2008.043.1.11

Sherman G.D. & Alexander L.T. (1959). Characteristics and genesis of low humic latosols. *Soil Science Society of America Journal* **23**(2): 168–170.
DOI: https://doi.org/10.2136/sssaj1959.03615995002300020025x

Soil Survey Staff (2014). *Keys to Soil Taxonomy*, 12th edition, pp. 372. United States Department of Agriculture, USA.

Srasra E., Bergaya F. & Fripiat J.J. (1994). Infrared spectroscopy study of tetrahedral and octahedral substitutions in interstratified illite-smectite clay. *Journal of Clays and Clay Minerals* **42**: 237–241.

Środoń J. (2013). Identification and quantitative analysis of clay minerals. In: *Developments in Clay Science*, volume 5, pp. 25–49. Elsevier, Netherlands.
DOI: https://doi.org/10.1016/B978-0-08-098259-5.00004-4

Szymański W., Skiba M., Nikorych V.A. & Kuligiewicz A. (2014). Nature and formation of interlayer fillings in clay minerals in Albeluvisols from the Carpathian Foothills, Poland. *Geoderma* **235**–**236**: 396–409.
DOI: https://doi.org/10.1016/j.geoderma.2014.08.001

Vaculikova L. & Plevova E. (2005). Identification of clay minerals and micas in sedimentary rocks. *Acta Geodynamica et Geomaterialia* **2**(2): 163.

Vitharana U.W.A., Mishra U. & Mapa R.B. (2019). National soil organic carbon estimates can improve global estimates. *Geoderma* **337**: 55–64.
DOI: https://doi.org/10.1016/j.geoderma.2018.09.005

Weerasuriya T., Nesbitt H.W. & Fyfe W.S. (1991). Geochemical characteristics of some Sri Lankan soils. *Journal of the Soil Science Society of Sri Lanka* **7**: 54–75

White J.L. & Roth C.S. (1986). *Infrared Spectrometry. Methods of Soil Analysis. Part I. Physical and Mineralogical Methods* (ed. A. Klute) pp. 291–326 . Soil Science Society of America, Madison, Wisconsin, USA.

Williams J., Prebble R.E., Williams W.T. & Hignett C.T. (1983). The influence of texture, structure and clay mineralogy on the soil moisture characteristic. *Soil Research* **21**(1): 15–32.
DOI: https://doi.org/10.1071/SR9830015

Zhao Q. *et al*. (12 authors) (2020). Strong mineralogic control of soil organic matter composition in response to nutrient addition across diverse grassland sites. *Science of the Total Environment* **736**: 137839.
DOI: https://doi.org/10.1016/j.scitotenv.2020.137839

Zhang Z.Y., Huang L., Liu F., Wang M.K., Fu Q.L. & Zhu J. (2016). Characteristics of clay minerals in soil particles of two Alfisols in China.
Applied Clay Science **120**: 51–60.
DOI: https://doi.org/10.1016/j.clay.2015.11.018

J.Natn.Sci.Foundation Sri Lanka 2021 49 (3): 361 - 368
DOI: http://dx.doi.org/10.4038/jnsfsr. v49i3.10269

RESEARCH ARTICLE

Solar luminance distribution in the principal plane for different wavelengths at two locations in Sri Lanka

CP Abayaratne[1*], AVUA Wickramarathna[1*] and MK Jayananda[2*]
[1]Department of Physics, Faculty of Applied Sciences, University of Sri Jayewardenepura, Nugegoda.
[2]Department of Physics, Faculty of Science, University of Colombo, Colombo 3.

Submitted: 28 September 2020; Revised: 07 January 2021; Accepted: 28 February 2021

Abstract: The colour of the sky has been the subject of many studies related to solar luminance distribution. This paper presents the angular sky luminance distributions measured at the ground level for white light and three spectral bands centred at blue (450 nm), green (550 nm) and red (650 nm) in the principal plane of the sun at two locations in Sri Lanka. The luminance measurements were taken by an LDR detector along with three colour filters. Results reveal that the ground-level angular luminance distributions of white, green, and red light are similar with high peak levels of direct sun luminance which sharply decrease away from the solar disc followed by a relatively small gradual rise for angular solar distances exceeding 90º. In contrast, the direct sun luminance for blue light shows a small peak directly under the sun which too decreases away from the solar disc and remains with little variation up to 90º but rises to comparatively high levels of luminance for angular solar distances above 90º. The measured blue peak luminance under direct sunlight is higher in Kandy which is at a higher altitude with a shorter air column above the site compared to Mahiyanganaya where the peak luminance is smaller due to more light being scattered by the longer air column at the lower altitude in Mahiyanganaya.

Keywords: Atmospheric scattering, sky luminance, skylight scattering, solar radiance distribution.

INTRODUCTION

The colour of the sky, which depends on the spectral composition, has been of interest to many scientists and laymen for centuries. According to the famous Rayleigh theory of scattering of solar radiation by atmospheric air molecules, the extent of scattering is inversely proportional to the fourth power of the wavelength and explains the blue colour of the sky (Coulson, 1988; Bohren, 2007; Mani, 2008). Any deviations to the composition of the atmosphere such as the presence of anisotropic molecules, aerosols, their sizes and surface albedo can cause measurable differences in the spectral intensity distribution across the sky, although it may not be noticeable to the naked eye. Details of atmospheric conditions such as concentration and size distribution of aerosols can be determined with the use of solar intensity measurements (Rangarajan & Mani, 1984; Kaskaoutis & Kambezidis, 2006; Rosairo *et al*., 2011; Djafer & Irbah, 2013). Scattering by particles with dimensions comparable to or larger than the wavelength of the visible light is described by Mie scattering, which does not depend on the wavelength but strongly depends on conditions such as air pollution and cloudiness (Wald, 2018). Clouds appear in white because of Mie scattering by water droplets which are large compared to the wavelengths of the visible light (Wald, 2018). In contrast to Rayleigh scattering, Mie scattering is more complicated and it has a strong forward scattering where the diameter of the particles is the same or larger than the incident wavelength. Mostly, dust particles have size distributions comparable with the wavelength of the visible light, water droplets, ice crystals and

* Corresponding author (cmapala@sjp.ac.lk https: //orcid.org/ : 0000-0001-6762-6444)

aerosols. When the particle diameter is much larger than the wavelength of visible light, a theoretically more complex type of atmospheric scattering independent of the wavelength known as non selective scattering explains the haze seen in the lower atmosphere caused by water droplets and large aerosols (Wald, 2018). All these factors as well as ground reflections contribute to the spectral composition of skylight.

Visible spectrum together with near ultraviolet and near infrared radiation attributes to 99 % solar radiation while the visible region attributes to 40 % of the solar radiation received at the earth's surface on clear days (Mani, 2008). The amount of short wavelength solar radiation that reaches the earth's surface is limited due to absorption by oxygen and ozone at high altitudes of the atmosphere, and spectral distribution of solar radiation gives a peak around 550 nm wavelength at the mean sea level (Mani, 2008). Although many scientific articles about the spectral intensity distribution of sunlight along with images are available in the literature, not many such articles can be found on the angular distribution of skylight which is sunlight scattered by the constituents of the atmosphere. Some measured relative skylight intensity distributions carried out in several locations in USA using radiometers based on photomultipliers have been presented by Coulson (1988). Contour plots of measured radiance distribution using a CCD camera system over certain regions in USA have been reported by Liu and Voss (1997). Hisdal (1986) gives spectral irradiance distributions under different sky conditions. However, such measurements have not been carried out in Sri Lanka and the purpose of this work is to present the angular distribution of sky luminance for three spectral bands centred at 450, 550 and 650 nm. The angular distribution of sky luminance found would be of interest to better understand the processes that take place in the atmosphere through the interaction of direct and scattered sunlight with regional atmospheric constituents. For easy reference, the formula for the intensity I_R for the light of wavelength λ scattered by an angle ϕ in Rayleigh scattering is given in Equation 1, where ε and n are the relative permittivity and the refractive index of the droplets, respectively, while ε_0 and n_0 are those of the surrounding medium (Wald, 2018).

$$I_R = \frac{1}{2}\left(\frac{2\pi n_0}{\lambda}\right)^4 r^6 \left[\frac{(\varepsilon-\varepsilon_0)}{(\varepsilon+2\varepsilon_0)}\right]^2 (1+cos^2\phi) \qquad ...(1)$$

The length of the direct geometrical pathway (known as relative air mass) through the atmosphere affects the attenuation of the solar radiation, and is given by the formula of Equation 2 (Luo, 2016);

$$Air\ mass = \frac{1}{\sin\theta}\left[\frac{P}{P_0}\right], \qquad ...(2)$$

where θ is the solar elevation angle, P is the pressure at a given altitude (in Pa) and P_0 is the pressure at sea level (101325 Pa). If the direct path way is shorter as during midday, attenuation is less and if the path way is longer as at sunrise and sunset, more attenuation can take place. The main focus of this paper is to present the observations with possible implications that can be extended into the UV region, based on the results of angular spectral distribution of the sky luminance on the principal plane of the sun, for some solar elevation angles during the morning hours between 0700–0930 h. A brief description of the instrument used for measurements and the procedure followed are given in the Methodology section. Results are presented in the following section with a qualitative discussion and some implications of the results to be continued in a future investigation.

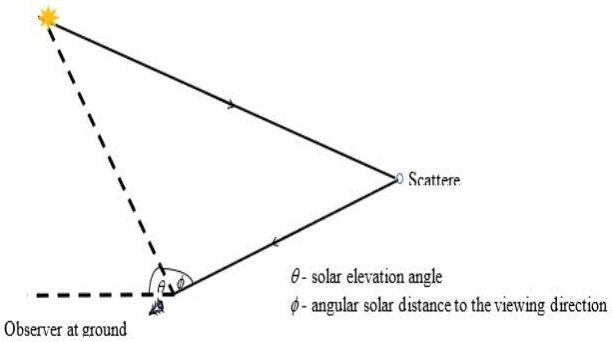


Figure 1 : Solar elevation angle θ and the angular solar distance ø to viewing direction in the principal plane

The principal plane carries more information than any other plane across the sky (Coulson, 1988; Liu & Voss, 1997; Dahlberg *et al*., 2011). The vertical plane passing through the sun, a scatter particle and an observer is known as the principal plane of the sun and is illustrated in Figure 1 when the Sun is at a solar elevation angle θ. Figure 1 also shows a light ray scattered towards the eye of an observer at the ground level having an angular distance ϕ with the sun. Measurements were taken using an LDR detector and were later calibrated to find the sky luminance in Lux.

METHODOLOGY

Luminance measurements at the ground level were obtained using a fully automated polarimeter (Figure 2) constructed to find the angular distribution of the skylight polarization in the principal plane. This

polarimeter is an improvement over a partially automated polarimeter constructed earlier (Abayaratne *et al*., 2016)

A window comparator unit was used as a sun-tracking unit and three hybrid motors were used for movements of the polarimeter. To observe the temperature and humidity there are inbuilt humidity and temperature sensors (DHT 11). The LDR was placed at the bottom of a PVC tube with one end of the tube fitted with an end-cap having a small aperture (~1 cm) in the centre to use as a collimator, allowing a narrow beam of light to fall on the detector with an acceptance angle of 2º. Three optical filters of centre wavelengths of 450 nm, 550 nm, and 650 nm and each of half power bandwidth ± 40 nm was placed on the aperture consecutively to select the appropriate wavelength band or to allow white light to enter through the collimator. Each filter had a transmissivity of 60 %, for which a correction was made for the luminance measurements.

A polarizer was placed between the aperture and the detector, since the instrument was initially constructed to find the degree of polarization, which is not required in measuring the total sky luminance. First, the instrument tracks the sun through rotations around a horizontal axis and a vertical axis and records the position of the sun. At this point, the collimator of the instrument scans the principal plane when it rotates around the horizontal axis. Next, it goes back to the starting point (horizontal direction in the principal plane of the sun) and moves two degrees at a time in the principal plane of the sun and at each such point rotates the polarizer in three degree steps 60 times and reads the corresponding voltage across the LDR sensor. After that, it chooses the maximum and minimum values of the voltage readings in two mutually perpendicular directions. These voltages were later converted to luminance (in Lux) after calibrating with the Lux sensor (BH1750FVI) and polarizers to control the intensity of light emitted by an LED source.

Due to limitations of the source intensity, the luminance received by the detector was limited to 1000 Lux. Therefore, measured intensities exceeding 1000 Lux were obtained by extrapolating the calibration curve, which imposed a limitation on the accuracy of intensities above 1000 Lux. Furthermore, intensities of blue and red light were much lower compared to the green light due to the lower source emissions in blue and red spectral regions and lower detector sensitivity, which makes it difficult to calibrate the voltages with red and blue light. Due to this reason, the spectral dependence of the detector response was not taken into consideration. After reading the required voltages, the collimator rotates by two degrees automatically as programmed by an Arduino platform until the entire principal plane is scanned.

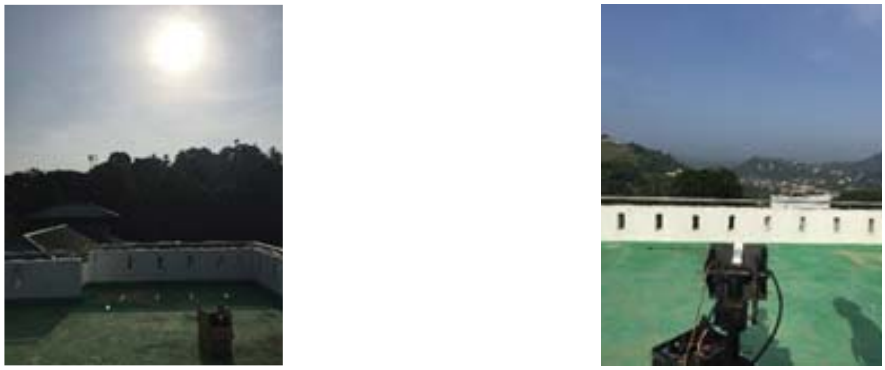
The luminance values corresponding to the two polarizations, I_{max} and I_{min} can be added to obtain the total intensity $I(\theta)$. ($I(\theta) = I_{max} + I_{min}$). Measurements were taken in two locations in Sri Lanka, Kandy and Mahiyanganaya. As far as the content of this paper is concerned, the difference between the two sites is the altitude difference. Geographical details of the locations and the dates of data collection are given in Table 1.

Table 1: Geographical details of the measurement sites and the dates of data collection

Location	Latitude	Longitude	Altitude	Data collection dates
Kandy	7.29 N	80.63 E	586 m	01, 02, 03, 04, 05 May 2018
Mahiyanganaya	7.33 N	80.99 E	90 m	20, 21, 22, 23 July 2018

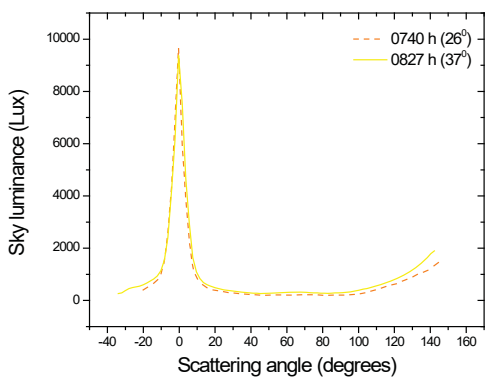
RESULTS AND DISCUSSION

Figures 3 (a) and (b) show two images taken on 01.05.2018 at 0750 h of the East and West sky respectively, in Kandy. Figures 4 (a), 4(c), 4(e) and 4(g) represent the variation of the average solar luminance (measured in Lux) over 4 days around the same time (nearly the same solar elevation angle θ) at the ground level as a function of ϕ on the principal plane in Kandy for white (without filters), blue, green and red light respectively while Figures 4 (b), 4(d), 4(f) and 4(h) represent the same for Mahiyanganaya. Each data set was collected within a period of about 8 minutes. The respective mean time and mean solar elevation angle for each data set is given in the inset of each figure.

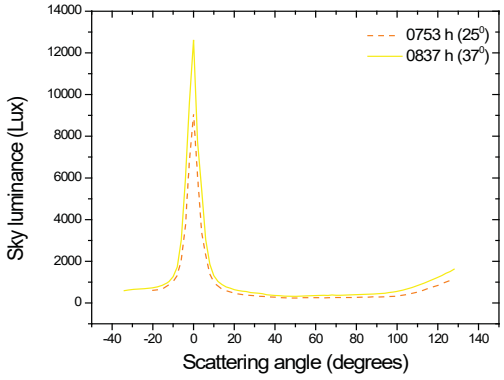


(a) (b)

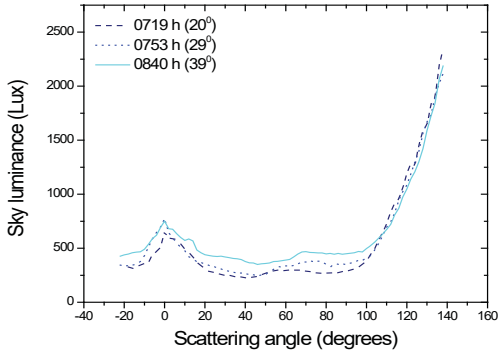
Figure 3: Images of the (a) East and (b) West sky in Kandy taken on 01.05.2018 around 0750 h.



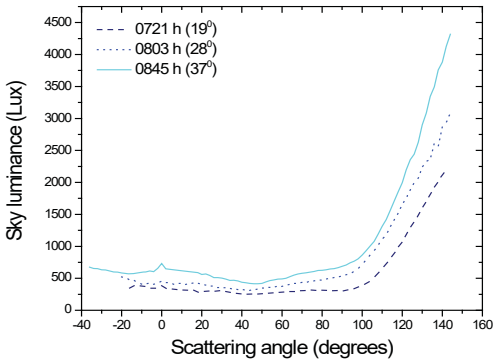
(a): Kandy - White light



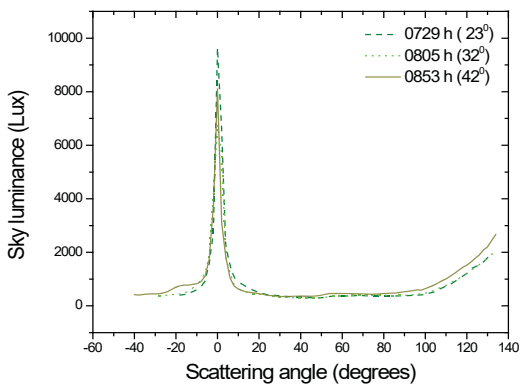
(b): Mahiyanganaya – White light



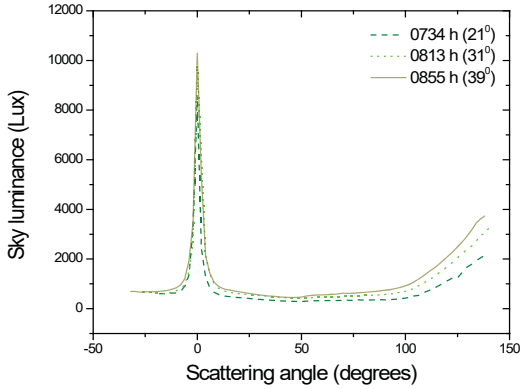
(c): Kandy – Blue light (450 nm)



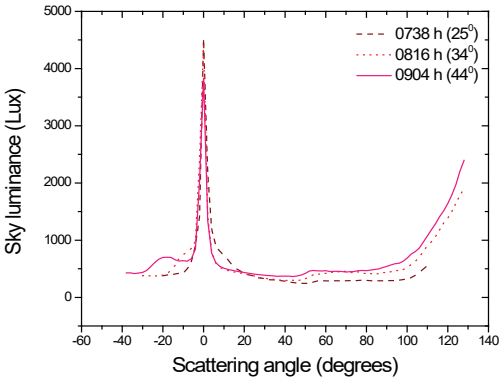
(d): Mahiyanganaya - Blue light (450 nm)



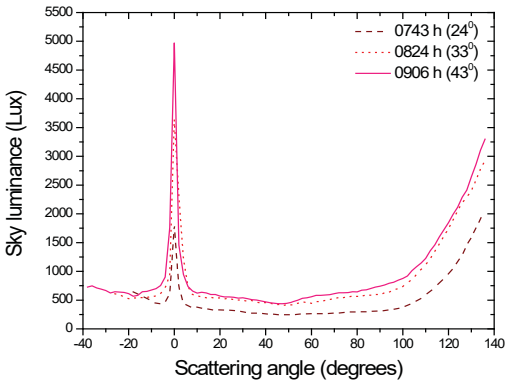
(e): Kandy - Green light (550 nm)



(f): Mahiyanganays - Green light (550 nm)



(g): Kandy - Red light (650 nm)



(h): Mahiyanganaya - Red light (650 nm)

Figure 4: Variations of solar luminance with angular solar distance at the mean times and solar elevation angles given in the inset for white light and the three different wavelengths for Kandy (a, c, e, g) and Mahiyanganaya (b, d, f, h)

Some results expected from theory can be verified using the graphs of Figure 4. It can be easily seen by comparing the three peaks in each graph of Figure 4 separately for each location that the direct sun luminance increases with increasing solar elevation angle since the path travelled through the atmosphere decreases with the rising sun (and decreasing relative air mass), in accordance with the expected result. (It should be noted that there was a time gap of about 10 minutes and a difference of about 10º in the solar elevation angle between measurements for each graph in the order white, blue, green and red.)

Moving away from the sun through the principal plane, the solar intensity gradually decreases until the angular distance is about 6º, and remains without significant variation until the angular solar distance ϕ reaches about 90º. For values of ϕ above 90º, all graphs show a gradual increase in the solar luminance. Out of all these graphs those for blue light are of particular interest, which is the most important finding of this work. While the graphs for white, green and red light look quite similar in the angular distribution of solar luminance in the principal plane, all graphs for blue light (450 nm) show higher levels of luminance for angular solar distances exceeding 100º relative to the direct Sun luminance. This is qualitatively explained by the higher scattering undergone by shorter wavelengths as predicted by Rayleigh theory.

Table 2: Direct solar luminance, angular solar distance when the scattered light intensity is equal to the direct solar radiance and the sky luminance at an angular solar distance of 135º for blue light

Location	Solar elevation angle (ϕ)	Time	Direct solar luminance (Lux)	Angular solar distance (θ) to a direction with luminance equal to that directly under the Sun	Luminance at an angular solar distance of 135º (Lux)
Kandy	20º	0719 h	652	108º	2047
	29º	0753 h	751	113º	1964
	39º	0840 h	751	114º	1978
	19º	0721 h	385	100º	1869
	28º	0803 h	456	74º	2483
	37º	0845 h	723	94º	3420

Despite the sky being overcast with a cloud cover during the morning hours when data was collected in Kandy after rain the previous night on each of the four days compared to Mahiyanganaya, it can be seen that the direct solar radiance at nearly equal solar elevation angles (approximately 20º, 28º, 38º) are higher for blue light in Kandy where the air column above the ground level is shorter than that in Mahiyanganaya due to the higher altitude of the Kandy site while the scattered light seen at an angular solar distance of 135º is higher in Mahiyanganaya. As can be expected from Rayleigh theory of scattering that blue light is scattered more in Mahiyanganaya during the longer travel path through the atmosphere thus displaying a reduced direct solar luminance and a higher sky luminance for higher angular solar distances. In a similar manner, the scattered skylight luminance can be observed to reach a value equal to the direct solar luminance sooner (at a lower value of ϕ) in Mahiyanganaya compared to Kandy. (Note also that this is not observed for the other two colours or white light possibly due to the different atmospheric conditions.)

The observation of higher luminances in directions away from the sun for blue light can be expected from UV light as well due to the high degree of scattering of UV light according to the Rayleigh Theory. Although the luminance directly under the sun is low for shorter wavelengths such as those of UV light, a high luminance can be expected in the backscattered light reaching an observer from the direction opposite to the sun. In order to prevent being exposed to UVA radiation during the morning hours, we may need to shield ourselves from solar radiation reaching us from angular solar distances above 100º rather than from the direct sun. In other words, the radiation coming from directions opposite to the sun carries more of blue (and possibly also UVA light) than the light coming directly from the sun at solar elevation angles below 40º considered in this work. As a future extension of this work, this fact can be investigated using a suitable UV sensor to measure UV intensity. This would be of importance for Sri Lanka which is a tropical country with high values of UV index and no reported investigations.

Health hazards such as skin cancers and cataract in the eyes caused by exposure to UV radiation, have been published by the World Health Organization (WHO, 2020). The intensity of UV radiation is indicated by the UV Index, which is the cumulative effect of UV radiation coming from all directions (Vanicek *et al*., 2000). When measuring the UV index, the usual practice is to place the UV sensor horizontally and measure the UV radiation incident on the sensor from all directions (Koepke *et al*., 1998). Scanning the principal plane could be used as an alternative method which also provides additional information such as the angular distribution of UV radiation.

Figure 5 shows the angular luminance distributions obtained by averaging the data taken on four days in each location for solar elevation angles of approximately 40º for white light, blue, green and red bands studied here in the same figure for Kandy and Mahiyanganaya indicating the decreasing levels of luminance in the order green, red and blue as expected. However it is noteworthy that the luminance values in Figures 4 and 5 have limited accuracy especially those exceeding 1000 Lux as a result of calibration errors and disregarding the exact spectral dependence of the LDR. It would be possible to improve the accuracy if the LDR sensor is replaced with a Lux sensor and the spectral dependence of the sensor is taken into account. However, the main conclusions of the paper are not affected by this inaccuracy.

Throughout the periods that data was collected, a fair amount of clouds were present in the sky in both sites which could be the reason for the irregularities

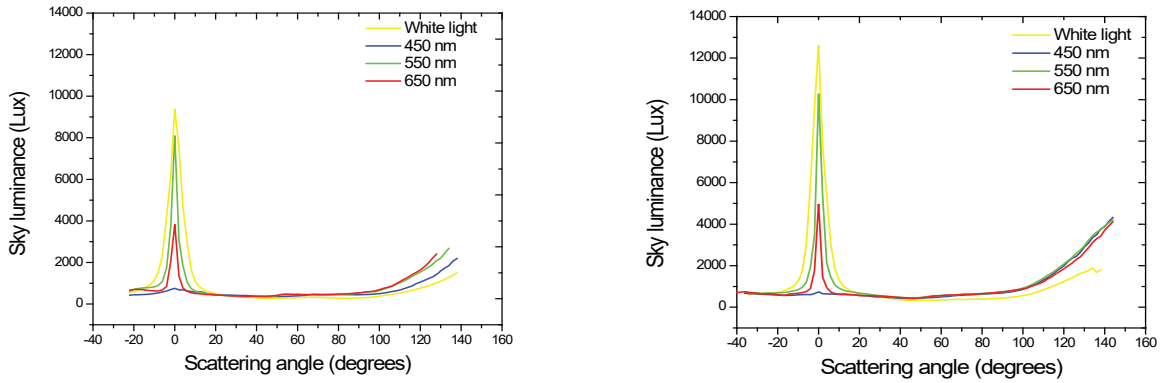


Figure 5: Variation of sky luminance with angular solar distance for different wavelengths given in the inset at approximately 40º ±10º at (a) Kandy (b) Mahiyanganaya

observed in all graphs. According to the cloud cover data obtained from the Climate Division, Department of Meteorology, Sri Lanka, on 04.07.2019, in Katugastota (Kandy) and Badulla (near Mahiyanganaya), given in Table 3, a crude average taken for the relevant data indicates thicker cloud cover in Kandy. The temperatures observed in Kandy ranged from 24 ºC - 32 ºC, while the humidity varied between 43 % - 80 %. The temperatures in Mahiyanganaya were between 28 ºC - 40 ºC and the

Table 3: Cloud cover data obtained from the Climate Division, Department of Meteorology, Sri Lanka on 04.07.2019.

(Higher values indicate thicker cloud covers)

Katugastota, Kandy		Badulla	
Date	Cloud cover	Date	Cloud cover
01.05.2018	4	20.07.2018	3
02.05.2018	4	21.07.2018	4
03.05.2018	4	22.07.2018	2
05.05.2018	3	23.07.2018	3
Average	3.75		3

CONCLUSION

Angular sky luminance distribution in the principal plane of the sun obtained using sky luminance measurements made with the use of an LDR sensor and an observation hitherto not reported in the literature to the best of our knowledge has been presented in this paper. Graphs illustrate that the peak luminance increases with the solar elevation angle θ for all colour bands, blue, green and red as well as white light in agreement with our common experience. The key finding of this work, which can be expected from the Rayleigh theory of scattering but often overlooked in practice, is the presence of more intense blue light in the sky opposite the sun compared to the direct solar luminance. The directional spectral dependence of the brightness of the sky could be useful in understanding the scattering processes in the atmosphere further and to extract details of the atmospheric constituents. These observations made with a low cost instrument can be refined using more sophisticated instruments. As a further extension of this work, the experiment could be performed with a UV sensor to determine the angular intensity distribution of UV light to check whether significant UV radiation arrives from directions other than that of the Sun.

Acknowledgements

The authors thank the University of Sri Jayewardenepura for providing funds through the university grant ASP/06/RE/SCI/2013/016. The authors also acknowledge Prof. D.A. Tantrigoda for his advice..

REFERENCES

Abayaratne C.P., Wickramarathna A.V.U.A., Rodrigo D.D. & Mannathunga K.S. (2016). A low-cost partially automated polarimeter for investigating skylight polarization. *Journal of the National Science Foundation of Sri Lanka* **44**(1): 95–103.
DOI: http://dx.doi.org/10.4038/jnsfsr.v44i1.7986

Bohren C.F. (2007). Atmospheric optics. In: *The Optics Encyclopaedia: Basic Foundations and Practical Applications* (eds. T.G. Brown, K. Creath, H. Kogelnik, M.A. Kriss, J. Schmit & M.J. Weber). Wiley-VCH Verlag GmbH & Co., Germany.
DOI: https://doi.org/10.1002/9783527600441.oe004

Coulson K.L. (1988). *Polarization and Intensity of Light in the Atmosphere*, pp. 375–391. A. Deepak Publishers. Hampton, USA.

Dahlberg A.R., Pust N.J. & Shaw J.A. (2011). Effects of surface reflectance on skylight polarization measurements at the Mauna Loa Observatory. *Optics Express* **19**(17): 16008–16021.

do Rosário N.E., Yamasoe M.A. & Longo K.M. (2009). Aerosol optical depth and ångström coefficient retrievals over the Amazon forest during 2007 biomass burning season. *AIP Conference Proceedings, American Institute of Physics* **1100**(01): 494–497.

Djafer D. & Irbah A. (2013). Estimation of atmospheric turbidity over Ghardaïa city. *Atmospheric Research* **128**: 76–84.
DOI: https://doi.org/10.1016/j.atmosres.2013.03.009

Hisdal V. (1986). Spectral distribution of global and diffuse solar radiation in Ny-Alesund, Spitsbergen. *Polar Research* **5**: 1–27.

Kaskaoutis D.G. & Kambezidis H.D. (2006). Investigation into the wavelength dependence of the aerosol optical depth in the Athens area. *Quarterly Journal of the Royal Meteorological Society: A Journal of the Atmospheric Sciences, Applied Meteorology and Physical Oceanography* **132**(620): 2217–2234.
DOI: https://doi.org/10.1256/qj.05.183

Koepke P. *et al.* (24 authors) (1998). Comparison of models used for UV index calculations. *Photochemistry and Photobiology* **67**(6): 657–662.

Liu Y. & Voss K. (1997). Polarized radiance distribution measurement of skylight. II. Experiment and data. *Applied Optics* **36**(33): 8753–8764.

Luo R. (ed.) (2016). Encyclopedia of *Color Science and Technology*. Springer-Verlag, New York, USA.

Mani A. (2008). *Handbook of Solar Radiation Data for India. Resonance* November 2008: 1082–1086. Available at *https://www.ias.ac.in/article/fulltext/reso/013/11/1082-1086*

Rangarajan S. & Mani A. (1984). A new method for the determination of atmospheric turbidity. Tellus B: Chemical and Physical Meteorology 36(1): 50–54.
DOI: https://doi.org/10.3402/tellusb.v36i1.14804

Vanicek K., Frei T., Litynska Z. & Schmalwiese A. (2000). UV-Index for the Public. A guide for publication and interpretation of solar *UV Index forecasts for the public* prepared by the Working Group 4 of the COST-713 Action 'UVB Forecasting'. European Communities, Brussels, Belgium.

Wald L. (2018). Basics in solar radiation at Earth surface. Available at *https://hal-mines-paristech.archives-ouvertes.fr/hal-01676634*

WHO (2020). Ultraviolet radiation and health, Available at *https://www.who.int/uv/uv_and_health/en/, accessed on 16 October 2020.*

J.Natn.Sci.Foundation Sri Lanka 2021 49 (3): 369 - 382
DOI: http://dx.doi.org/10.4038/jnsfsr.v49i3.10223

RESEARCH ARTICLE

Functional trait diversity of wild rice species in Sri Lanka: implications for field identification and application

S Sandamal[1], A Tennakoon[1], DABN Amarasekera[2], B Marambe[3] and D Ratnasekera[1*]
[1]*Department of Agricultural Biology, Faculty of Agriculture, University of Ruhuna. Mapalana, Kamburupitiya.*
[2]*Department of Crop Science, Faculty of Agriculture, University of Ruhuna. Mapalana, Kamburupitiya.*
[3]*Department of Crop Science, Faculty of Agriculture, University of Peradeniya, Peradeniya.*

Submitted: 22 September 2020; Revised: 16 February 2021; Accepted 26 March 2021

Abstract: The study assessed the variations in morphologically, physiologically and anatomically distinct traits of wild rice species in Sri Lanka; *O. nivara*, *O. rufipogon*, *O. eichingeri*, *O. rhizomatis* and *O. granulata*, which could be useful in rice breeding. The wild rice species were grown in a common garden, and the morphological traits were measured soon after heading. The results showed qualitative parameters such as the panicle type, awning, stigma colour, lemma and palea pubescence, seed coat colour, blade pubescence and ligule shape, are distinctive among the five species and are promising characters in their field identification. ANOVA revealed that the quantitative traits, such as flag leaf length, flag leaf width, culm length, culm diameter, panicle length, 100 grain weight and plant height are useful for further confirmation of species. The highest net photosynthetic rate (5.86 μmol m^{-2} s^{-1}), high cluster width of the base (61.4 μm), and trichome density (184.33 per 25 mm^2 area) were observed in *O. rufipogon* compared to the rest, and such desirable traits are effective in rice breeding. Moreover, transpiration rates, stomatal conductance and sub-stomatal CO_2 concentration are ideal physiological traits to be considered in super rice breeding. Significant correlations were observed between transpiration and photosynthesis processes. Thus, our study provides a clear picture on habitat preferences, life cycle, distinctive morphologies and diverse functional traits to be effectively used in field identification and future utilisation of wild relatives of rice in the plant breeding programmes.

Keywords: : Field identification, functional traits, genetic resources, species divergence, variation, wild rice.

INTRODUCTION

The global rice (*Oryza sativa*) production is expected to increase in the next few decades, with special focus on productivity enhancement, owing to limited land and increase in demand (Lim *et al*., 2013; Tan & Norhaizan, 2020). Improving the yield potential of rice varieties has been the main breeding objective in many countries for several decades to meet this challenge. The ideotype breeding is a key approach for crop improvement. 'Crop ideotype' is an idealised plant type with a specific combination of characteristics favourable for photosynthesis, growth, and grain production based on knowledge of plant and crop physiology and morphology (Khan *et al*., 2015). In this context, wild species of rice provides a wide range of favourable characters and is a valuable reservoir of genetic resources (Khush, 1997). Moreover, improving rice varieties by incorporating desirable traits from wild relatives may lead to advances in rice breeding, as the wild species of rice seems to harbour significantly higher genetic and phenotypic diversity than the cultivated rice (Sarla *et al*., 2003). Consequently, the knowledge of functional trait diversity among wild relatives will largely enhance their efficient utilisation, in addition to effective conservation (Lu *et al*., 2002; Ren *et al*., 2003).

* Corresponding author (disnaratnasekera@gmail.com https: //orcid.org/ : 0000-0003-3176-6888)

Consequently, the knowledge of functional trait diversity among wild relatives will largely enhance their efficient utilisation, in addition to effective conservation (Lu *et al*., 2002; Ren *et al*., 2003). However, our understanding of functional trait diversity, particularly in wild *Oryza* species, is still limited (Duan *et al*., 2007; Rathore *et al*., 2016). Identifying traits and characterising their variation under different environmental factors is important to understand the functional trait diversity among species (Micol & Hake, 2003; Itoh *et al*., 2005; Kadioglu & Terzi, 2007; Alvarez *et al*., 2008; Tian *et al*., 2012). The variability of functional traits among species is indicative of their important physiological processes including photosynthesis. Of the functional traits of wild rice, net photosynthetic rate (net assimilation rate), transpiration, and stomatal conductance are significant parameters to regulate the plant growth and development (Rathore *et al*., 2016). Determination of diversity among the wild relatives of cultivated *Oryza* spp. based on functional traits will be needed in directing future efforts to discover desirable traits and thus facilitate effective germplasm conservation and utilisation in rice breeding (Zhu *et al*., 2014).

The genus *Oryza* contains approximately 24 species distributed in Asia, Africa, Australia, and America. Of the species, only two are cultivated species and the remaining 22 species are wild relatives of rice (Vaughan, 1989; Khush, 1997). Moreover, Sri Lanka is one of the secondary diversity centres for rice genetic resources (Ikeda & Vaughan, 1991). Five wild species of *Oryza viz*. *O. nivara* (AA genome), *O. rufipogon* (AA genome), *O. eichingeri* (CC genome), O. *rhizomatis* (CC genome) and *O. granulata* (GG genome) are known in Sri Lanka and *O. rhizomatis* is considered endemic to the country. Among these, *O. nivara* and *O. rufipogon* are closely related to Asian cultivated rice (Banaticla-Hilario, 2012). The pest and disease resistance in these wild rice species are well-documented (Liyanage & Senanayake, 2010; Madurangi *et al*., 2012). Their distribution, habitats, pollination and flowering patterns are also well described under Sri Lankan conditions (Liyanage *et al*., 2002; Liyanage & Senanayake, 2010; Rajkumar *et al*., 2015). Such information directs rice breeders to identify genetically diverse parents to gain desired traits when developing new rice cultivars.

However, the understanding of morphological diversity, particularly in wild *Oryza* species is still limited and sometimes contradictory. Thus, identification of wild species in their natural habitat is difficult or misleading to draw proper conclusions. In this regard, distinct structures in different plants of the same or different species need to be examined in detail and compared (Sattler & Hall, 1994). The characters that are used in plant identification, classification and description should be diagnostic or key characters that can be either qualitative or quantitative or both qualitative and quantitative. Plant growth habitat, growth patterns, seedling characters, leaf characters, inflorescence and flowers, fruit characters and seed characters are the major traits considered in the proper identification process. However, plant descriptions are often limited, and the morphological, physiological and anatomical distinctions among these species are often vague and not clear enough for field identification and species differentiation. Thus, the field identification of these species is difficult and often confusing based on the available information. Therefore, this study aimed at characterisation of Sri Lankan wild rice species to identify morphologically distinct traits to support field identification, and physiologically and anatomically distinct traits that could be useful in rice breeding.

METHODOLOGY

Sample collection and field establishment

Five wild rice species, namely, *O. eichingeri*, *O. granulata*, *O. nivara*, *O. rhizomatis* and *O. rufipogon* found in Sri Lanka were collected from their typical natural habitats (S1, S2, S3, S4 and S5, respectively) (Table 1).

Mature seeds or root stocks of ten individual plants of each wild rice species were collected from the naturally occurring populations keeping a minimum distance of 5 m between plants to prevent the collection of ramets from a single genet. Exact locations of samples collected were recorded by a Global Positioning System (GPS; Garmin Oregon 550). Thereafter, ten individuals from each wild rice population of the respective wild rice species were established in cement pots (40 cm length × 40 cm width × 45 cm height), using seeds or rootstocks having 10 pots per population, in a common garden at the Faculty of Agriculture of the University of Ruhuna (latitude 06.060337°N and longitude 80.5681455°E), Sri Lanka, from January to December 2016. Each pot per population was considered as a replicate. Pots were arranged in a completely randomised design and the morphological, physiological and leaf anatomical features were characterised.

Table 1: Geographical and ecological information of sampling sites of five wild rice species in Sri Lanka

Species name	Location	SL accession code	GPS values		Habitat Description
			Latitude	Longitude	
O. eichingeri	Wawulpane-Rathnapura	S1	06°25' 59.88"	80°43' 59.87"	Forest, stream banks
O. granulata	Walakada-Urubokka	S2	06°21' 10.44"	80° 41' 31.55"	Secondary forests,shade
O. nivara	Vellavelly-Batticaloa	S3	07°30' 33.10"	81°43' 55.10"	Shallow-water lake
O. rhizomatis	Sirnawatkulum-Mannar	S4	08°54' 36.40"	79°57' 42.30"	Shrubs and weeds
O. rufipogon	Thihagoda-Matara	S5	06°00' 01.60"	80°33' 43.90"	Stream

Morphological characterisation

Morphological characters (qualitative and quantitative) from seedling to mature stage were measured as described in the list of descriptors for wild and cultivated rice (*Oryza* spp.) published by the Biodiversity International, International Rice Research Institute and West Africa Rice Development Association (BI-IRRI-WARDA, 2007). Morphological diversity was measured by 11 quantitative (Table 2) and 28 qualitative traits (Table 3). For each character, average measurements taken from three randomly selected tillers per plant, including the main culm was considered. As time of planting was same for all five species, measurements were taken soon after heading (except seedling height).

Functional trait characterisation

Photosynthetically active radiation (PAR), leaf transpiration rate (Evap), stomatal conductance (GS), leaf surface temperature (LT), net photosynthetic rate (PN) and sub-stomatal CO_2 concentration (C Int) were determined using TPS-2 (MA 01913, Portable Photosynthesis System, Amesbury, USA). The measurements were taken randomly from the fully expanded top five leaves of the main culm and matured tillers for each selected plant, and repeated in all replicates. The same leaves were sampled for anatomical investigations. The mid portion of each leaf blade was inserted in the leaf chamber for gas-exchange measurements. For all *Oryza* species except *O. granulata*, two leaf blades were used to fully cover the cuvette luminal surface area. The leaf width of *O. granulata* was higher than that of the cuvette diameter and thus, a single leaf blade was assembled. Data were obtained between 10 a.m. to 2 p.m. with an air temperature of around 30 °C. Measurements were taken after the plants were exposed to sunlight for approximately 1 h and the leaf functional traits that were given by the leaf gas exchange were recorded.

Fixing, staining and observation of leaf anatomy

A 3 cm long section of the first fully expanded leaf blade from each sampled plant was separated for the leaf structural studies. The leaf sections were cleared and fixed as described by Huckelhoven & Kogel (1998) with slight modifications. The leaf sections were placed in a clearance solution [0.15 % trichloroacetate, (w/v) in ethanol:chloroform (4:1; v/v)] for 48 h while the solution were changed twice in between the time. Then the samples were washed twice (15 min each) with 50 % ethanol, twice (15 min each) with 50 mM NaOH, thrice (10 min each) with MilliQ H_2O and finally 30 min incubation in 0.1 M Tris/HCl (pH 8.5). Samples were then stained using 0.1 % (w/v) safranin for 5 min, washed thoroughly 2–3 times with MilliQ H_2O followed by staining with 0.5 % (w/v) aniline blue for 2 h. After several rinses with MilliQ H_2O, tissues underwent microscopic assessments. Leaf cross sections were made using a microtome knife to observe bulliform cells and were observed at 10×40 magnification with the Olympus BH-2 light microscope fitted with inbuilt digital camera, and were quantified by Image J software. Minimum of 5 bulliform cell clusters of five different leaves from five different plants were measured. Structural measurements were made only on the bulliform cell cluster. Leaf and cell structural traits were determined from light microscopy on leaf sections. Number of cells per one cluster (A), distance between two clusters (µm) (B), middle cell width (µm) (C), cluster width of the base (µm) (D), area of the cluster (μm^2) (E), vein density (F), number of stomata (10×40 magnification) (G) and trichome density per 25 mm^2 area (10×4 magnification) (H), were the traits measured to characterise leaf anatomy.

Statistical analysis

Statistical analysis was performed using SAS version 9.2 (SAS Institute) and Minitab version 17 (Minitab, 2014). First, ANOVA was carried out to describe the variability of each structural and functional leaf trait, based on the entire three (morphology, physiology and anatomy) datasets for five wild *Oryza* species. Quantitative traits of different wild rice species were statistically described using means and standard error of the means of particular traits to figure out the general information related to different species. Quantitative variables were subjected to Pearson's correlation analysis at $p = 0.05$. Pearson correlation matrices were calculated based on the mean values of each structural and functional trait of each leaf of each *Oryza* species, to evaluate the trait-to-trait associations.

RESULTS AND DISCUSSION

Qualitative and quantitative traits for field identification of wild rice species

The results revealed that the quantitative traits evaluated showed a distinct variation among the species (Table 2). The variation and unique morphological traits that could be useful in species identification in field level are illustrated in Figure 1.

Table 2: Means and standard error of means for the quantitative traits of five wild rice species

Variable	*O. eichingeri*	*O. granulata*	*O. nivara*	*O. rhizomatis*	*O. rufipogon*	LSD*
Seedling height (cm)	28.25 ± 0.323	33.2 ± 1.34	67.13 ± 1.53	42.63 ± 3.67	100.0 ± 2.65	8.2783
Ligule length (mm)	2.125 ± 0.125	1.75 ± 0.25	30.5 ± 1.04	3.75 ± 0.25	28.25 ± 0.854	1.7274
Flag leaf length (cm)	28.5 ± 0.736	15.25 ± 0.323	69.5 ± 2.18	35.25 ± 0.878	68.625 ± 0.826	3.6984
Flag leaf width (cm)	1.325 ± 0.0629	2.05 ± 0.0289	0.75 ± 0.0289	1.425 ± 0.0479	0.775 ± 0.025	0.1867
Culm length (cm)	42.0 ± 2.35	55.8 ± 1.29	97.88 ± 2.93	67.33 ± 1.14	152.25 ± 2.39	6.8981
Culm diameter (mm)	5.825 ± 0.118	3.0 ± 0.314	6.475 ± 0.197	5.475 ± 0.197	7.3 ± 0.332	0.4855
Plant height (cm)	88.25 ± 3.75	60.25 ± 1.96	123.0 ± 3.88	95.45 ± 2.07	184.75 ± 2.54	9.822
Panicle length (cm)	20.75 ± 1.01	8.85 ± 0.194	25.13 ± 1.16	29.13 ± 1.98	31.375 ± 0.898	5.7191
100-grain weight (g)	0.6625 ± 0.0175	1.08 ± 0.0129	1.955 ± 0.0646	0.7225 ± 0.0342	1.6875 ± 0.0427	0.1133
Grain length (mm)	6.3625 ± 0.0239	6.3625 ± 0.0315	8.288 ± 0.114	5.5625 ± 0.0125	8.2 ± 0.0645	0.1914
Grain width (mm)	2.125 ± 0.0323	2.325 ± 0.0323	2.625 ± 0.0433	2.125 ± 0.025	2.4125 ± 0.0239	0.0851

The values presented are the means ± standard error of means of 10 replicates in each of the five species. One way ANOVA was used to compare the mean values. Traits among species are significantly different at $p < 0.0001$. *LSD (least significant difference) at $p = 0.05$.

Among the quantitative traits, flag leaf length (FLL) and flag leaf width (FLW), culm length (CL) and culm diameter (CD), panicle length (PL), 100-grain weight (SW) and plant height (PH) are distinctive parameters in identifying species in the field. Among the species, *O. granulata* showed the lowest ($p < 0.0001$) PH (60.3 cm), FLL (15.3 cm) and PL (8.9 cm) and a higher FLW (2.1 cm) indicating that this species is more appropriate for shade environments (Table 2). *O. rufipogon* recorded the highest CL (152.3 cm), CD (7.3 mm) and PH (184.8 cm) ($p < 0.0001$) compared to other species (Table 2) indicating that the species has developed higher lodging resistance than the rest of the species. Further, the presence of such characteristics may help survival in permanently inundated habitats (Banaticla-Hilario *et al*., 2013). Previous studies have indicated that the culm-related traits such as a wider culm diameter and less number of tillers, are directly associated with crop physiology and yield due to increase in lodging resistance of the plant (Chuanren *et al*., 2004). *O. nivara* showed the highest 100-grain weight ($p < 0.0001$) among studied species. The 100-grain weight was also positively correlated ($p < 0.001$) with the LL ($r = 0.929$), FLL ($r = 0.827$), CL ($r = 0.745$) and PH ($r = 0.696$) (data not shown). The quantitative traits reported significant variations among species are distinctive indicators that could be used for field identification of wild rice species.

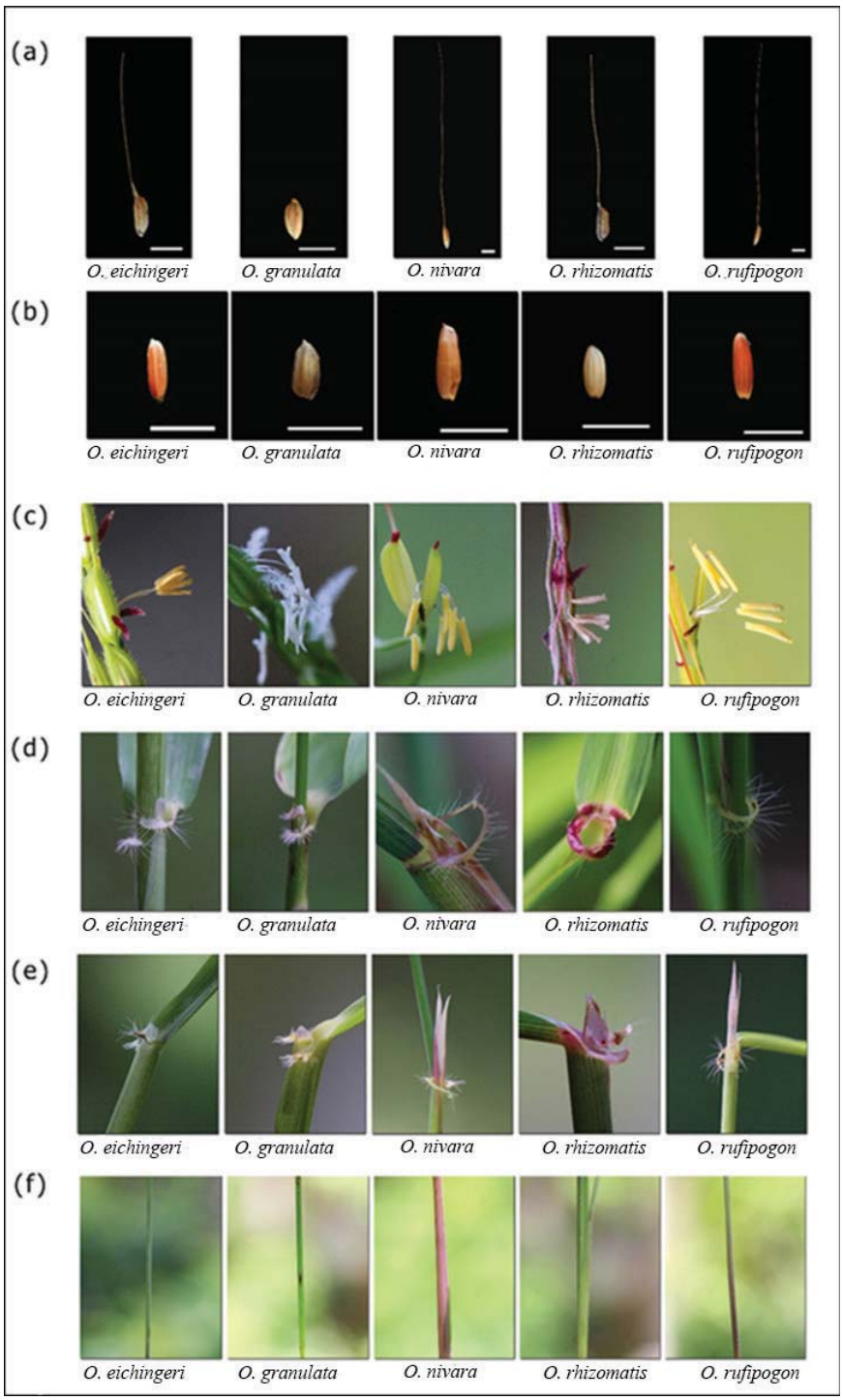


Figure 1: Variation Characteristic morphological traits found in five species of wild rice.
(a) seed shape and awn; (b) pericarp colour; (c) androecium; (d) auricle shape and colour;
(e) ligule shape and colour; (f) internode colour

The morphological differences among wild rice species are directly correlated to their natural habitats, life cycle and breeding system (Banaticla-Hilario *et al*., 2013). Ammiraju *et al*. (2010) reported that the genus *Oryza* has experienced a rapid diversification within a short evolutionary time period. In Sri Lanka, wild rice species are niched to diverse eco-geographic environments (Liyanage & Senanayake, 2010; Sandamal *et al*., 2018b). Most of the morphological traits are influenced by the environmental factors and thus, we evaluated them in the common garden under same environmental conditions as reported by Abhayagunasekara *et al*. (2018). Among them, traits linked with reproductive parts of the plant are the most important for identification and classification of wild *Oryza* spp. in Sri Lanka. Qualitative parameters such as the panicle type, awning, stigma colour, lemma and palea pubescence, seed coat colour, blade pubescence and ligule shape showed vast differences among the five species. Both *O. granulata* and *O. nivara* had compact panicles while open panicles were observed in *O. rufipogon*, *O. eichingeri* and *O. rhizomatis* (Figure 2).

Though *O. granulata* showed compact panicle, it is well exserted. Moreover, *O. rufipogon*, and *O. rhizomatis* had light secondary branching in the panicle, while there was no branching in *O. eichingeri.* An erect panicle was observed in all species except in *O. rhizomatis.* Open panicles may help to increase out crossing than self-pollination (Banaticla-Hilario *et al*., 2013). *O. granulata* had no awns while other species had awns with different lengths (Figure 1a). Comparatively long awns were detected in *O. nivara* and *O. rufipogon*. The floral morphology among species showed many characteristic differences in stamens. *O. granulata* had plumose type stigma with white colour stamens, which was clearly divided into two parts at the base of the stamen (Figure 1c). *O. eichingeri* and *O. rhizomatis* had stamens of the same size (length and width) and shape but differed in colour, i.e. dark yellow and pale purple, respectively (*field observation*). Both *O. nivara* and *O. rufipogon* produced anthers of the same shape and colour (yellow) but *O. rufipogon* had longer anthers than *O. nivara* when compared to those reported by Banaticla-Hilario

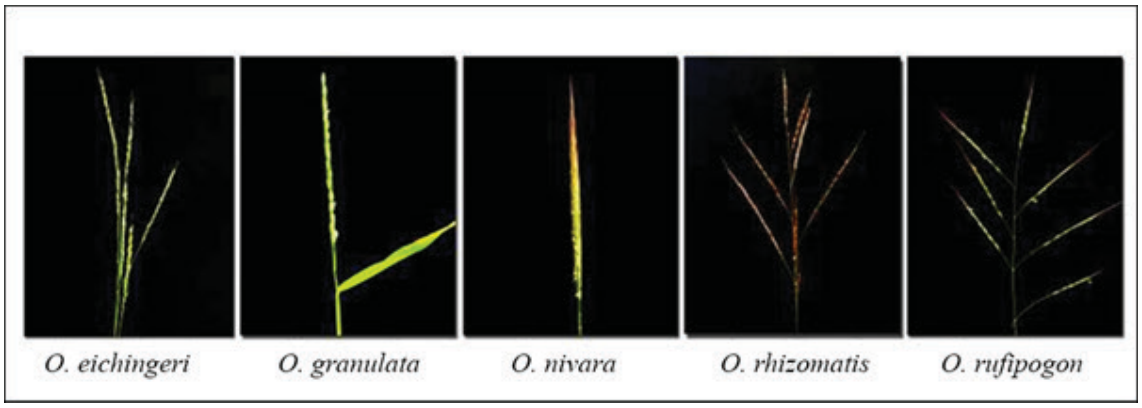


Figure 2: Panicle morphology of five wild rice species

et al. (2013). Presence of pubescence and the colour of lemma and palea, and other seed characteristics are good indicators to differentiate among wild rice species. Seeds of *O. granulata* were glabrous while the seeds of other species had pubescence (Table 3). Five wild rice species revealed large variations in seed coat colour, *viz*. red, light brown, brown and white (Figure 1b). Moreover, the dark green colour of lemma and palea was observed in the immature panicles of *O. granulata.* The size of the seeds (length and width) is one of the most stable characteristics (Table 2), which has a high heritability and therefore, can be used to distinguish species (Jackson, 1995).

Marginal differences were observed in the shapes and colour of auricle and ligule (Figure 1d) in wild rice species, which is one of the key characteristics to identify *Oryza* species from other species in the family Poaceae. Further, a two-cleft ligule was observed only in *O. nivara* and *O. rufipogon* among all five wild *Oryza* species (Figure 1e).

Table 3: Diversity of qualitative traits among five wild rice species. All the measurements were taken according to the rice descriptor.

Variable	*O. eichingeri*	*O. granulata*	*O. nivara*	*O. rhizomatis*	*O. rufipogon*
Blade pubescence	Intermediate	Glabrous	Pubescent	Glabrous	Pubescent
Blade colour	Green	Green	Green	Green	Green
Basal leaf sheath colour	Green	Green	Purple lines	Green	Purple lines
Leaf angle	Intermediate	Horizontal	Intermediate	Horizontal	Intermediate
Ligule shape	Truncate	Truncate	2 - Cleft	Acute to acuminate	2 - Cleft
Ligule colour	White	White	Purple lines	Purple lines	Purple line - 50 %,white - 50 %
Collar colour	Pale green	Pale green	Pale green	Pale green	Pale green
Auricle colour	Pale green	Pale green - 50 %, Purple - 50 %	Purple - 75 % Pale green - 25 %	Pale green	Pale green - 75 %, White - 25 %
Culm angle	Spreading	Procumbent	Open	Spreading	Intermediate
Internode colour	Green	Green	Purple lines	Green	Purple
Flag leaf angle	Horizontal	Horizontal	Intermediate	Horizontal	Intermediate
Panicle type	Intermediate	Compact	Compact	Intermediate	Open
Secondary branching	Absent	Absent	Light - 75 %, Absent - 25 %	Light-50 %, Absent-50%	Light - 75 %, Absent - 25 %
Panicle exsertion	Well exserted	Well exserted	Moderately well exserted	Well exserted	Moderately well exserted
Panicle axis	Straight	Straight	Straight	Straight	Straight
Awning	Short and fully awned	Absent	Long and fully awned	Short and fully awned	Long and fully awned
Stigma colour	Purple	White	Black	Light purple	Purple
Sterile lemma colour	Straw (Yellow)	Straw (Yellow)	Straw (Yellow)	Straw (Yellow)	Straw (Yellow)
Sterile lemma length	Short	Short	Long	Short	Long
Panicle shattering	High (More than 50 %)	High (More than 50 %)	High (More than 50 %)	High (More than 50 %)	High (More than 50 %)
Leaf senescence	Late and slow	Late and slow	Early	Late and slow	Intermediate
Spikelet fertility	Partly sterile	Fertile	Partly sterile	Highly sterile	Partly sterile
Panicle threshability	Easy	Easy	Easy	Easy	Easy
Apiculus colour	Brown	Straw (Tawny)	Red	Straw	Brown (Tawny)
Lemma and palea colour	Brown spots on straw	Straw	Brown (Tawny)	Brown furrows on straw	Brown (Tawny)
Lemma and palea pubescence	Short hairs	Glabrous	Hairs on upper portion	Short hairs	Hairs on upper portion
Seed coat colour	Red	Brown	Light brown	White	Red
Endosperm type	Non-glutinous	Non-glutinous	Glutinous	Non-glutinous	Non-glutinous

Generally, most rice varieties cultivated in Asia have pubescent leaves, and those in Africa and America are glabrous (Khush, 2001). Our observations indicated that both *O. nivara* and *O. rufipogon* had pubescent leaf blades while *O. granulata* and *O. rhizomatis* had glabrous leaf blades (Table 3). Glabrous trait may be selectively neutral in rice. However, trichomes are thought to be vital for plant defence against biotic and abiotic stresses. Thus, breeding for pubescent rice varieties is mainly targeted at the practical advantages of paddy production. Except for these qualitative traits, others showed minor variations among the species (Table 3). In contrast, no variations were observed for six qualitative parameters, *viz*. blade colour, collar colour, panicle axis, sterile lemma colour, panicle shattering and panicle threshability.

Functional trait diversity of five wild rice species

The physiological functions of the five wild rice species used in this study varied from each other, indicating the potential of using such functional traits in rice breeding. Photosynthesis forms an essential aspect of plant metabolism and the balance sheet of growth and development, which is sensitive to different abiotic stresses (Gupta *et al*., 2002; Panda *et al*., 2008; Gauthami *et al*., 2014). Under the same environmental conditions, a remarkably high net photosynthetic rate was observed in *O. rufipogon* compared to the rest of the species (Table 4). The reduction in photosynthetic rates may be a result of the changes in stomatal and non-stomatal factors (Panda *et al*., 2008; Mathobo *et al*., 2017). Generally, a lower photosynthetic efficiency occurs due to the inhibition of photosynthetic enzymatic activity, and the decrease in chlorophyll and oxidative loads (Hayat *et al*., 2012). The five wild rice species had transpiration rates ranging from 0.7 to 2.1 mmol H_2O $m^{-2}s^{-1}$. Although the majority of wild rice species (*O. granulata*, *O. nivara* and *O. rhizomatis*) had transpiration rates ranging from 0.722 – 0.778 mmol H_2O $m^{-2}s^{-1}$ similar to that of cultivated rice, *O. eichingeri* and *O. rufipogon* showed higher ($p < 0.0001$) transpiration rates (1.618 – 2.06 mmol H_2O $m^{-2}s^{-1}$). Transpiration is an important process for plants to create a negative pressure gradient that helps water and nutrient absorption from its roots. Results of the present study revealed a positive correlation ($p < 0.001$) between the stomatal conductance and transpiration ($r = 0.914$) (Supplementary table 01). Stomatal conductance directly influences regulation of gas flow and is known for its strong correlation with transpiration (Giuliani *et al*., 2013). Further, it helps cooling a plant and promote cell enlargement (Crawford *et al*., 2012). The highest stomatal conductance ($p < 0.0001$) in the present study was detected in *O. rufipogon.* The decrease of stomatal conductance was observed in species, except *O. rufipogon* under the existing environmental conditions, was may be due to the stomatal closure (Panda *et al*., 2008; Gauthami *et al*., 2014). The present study also reported a weak positive correlation between the photosynthetic rate and stomatal conductance. Siddique *et al*. (1999) reported that a strong relationship between net photosynthetic rate and stomatal conductance is an indication of the reduction in net photosynthetic rate, mostly due to stomatal closure, whereas a weak relationship indicates that the net photosynthetic rate is regulated by non-stomatal factors. An increase in sub stomatal CO_2 concentration (C int) suggests the predominance of non-stomatal limitation to photosynthesis, whereas a decrease in C int indicates the stomatal limitations dominated for the photosynthetic inhibition (Panda *et al*., 2008). Results of the present

Table 4: Physiological characters of five wild rice species.

Variable	*O. eichingeri* (Mean ± SE)	*O. granulata* (Mean ± SE)	*O. nivara* (Mean ± SE)	*O. rhizomatis* (Mean ± SE)	*O. rufipogon* (Mean ± SE)	p value	LSD*
PAR	1777.0 ± 29.9	1638.0 ± 48.0	1612 ± 101	1749.4 ± 88.3	1372.6 ± 34.0	0.1938	391.27
Evap	1.618 ± 0.187	0.774 ± 0.0926	0.722 ± 0.0442	0.778 ± 0.0248	2.06 ± 0.121	< 0.0001	0.3623
GS	98.8 ± 13.8	65.2 ± 8.87	52.0 ± 3.66	52.0 ± 2.24	228.2 ± 19.0	< 0.0001	47.829
LT	34.74 ± 0.201	33.2 ± 0.0447	33.82 ± 0.24	34.86 ± 0.271	33.12 ± 0.0735	0.0134	0.9557
PN	0.28 ± 1.12	2.68 ± 1.29	0.96 ± 1.55	2.0 ± 1.39	5.86 ± 1.48	0.0622	4.6716
C Int	398.4 ± 25.6	324.4 ± 27.9	349.6 ± 47.1	336.2 ± 32.5	342.2 ± 15.1	0.6212	120.71

One way ANOVA was used to compare the mean values. PAR - photosynthetically active radiation (µmol $m^{-2}s^{-1}$); (p=0.1938) Evap - leaf evaporation rate (mmol $m^{-2}s^{-1}$); (p<0.0001), GS - stomatal conductance (mmol $m^{-2}s^{-1}$); (p<0.0001), LT - leaf temperature (°C); (p=0.0134), PN net photosynthetic rate (µmol $m^{-2}s^{-1}$); (p=0.0622), C Int - sub-stomatal CO_2 concentration (µmol mol^{-1}). (p=0.6212). *LSD (least significant difference) at p = 0.05.

study indicated that the sub stomatal CO_2 concentration was not significantly different ($p = 0.6212$) among species (Table 4).

The vascular tissues are the most important structural components in plant tissues, which are responsible for the transport of assimilates, minerals and water (Hose *et al*., 2001; Cholewa & Griffith, 2004).

The present study showed that the vein density was not significantly different ($p = 0.0612$) among species (Table 5). However, the vascular bundle cell size and density are strongly correlated with the transpiration and photosynthetic rate of the species, which has a large culm (He & Zhang, 2003). Among the five wild rice species, *O. rufipogon* (Table 4) recorded the highest rate ($p < 0.0001$) of transpiration and stomatal conductance ($p < 0.0001$). Meanwhile, *O. rufipogon* showed a higher net photosynthesis rate among the tested species. *O. rufipogon* naturally grows and survives in environments such as deep-water habitats where water is not a limiting factor. Therefore, is an ideal species for super rice breeding when water is available at sufficient levels (Liu *et al*., 2015). As found in *O. rufipogon,.* cultivars or species with a large culm has shown a higher apoplastic transport ability (Gong *et al*., 2006), which might help transfer water and nutrients more rapidly and efficiently thus, contributing to higher grain Bulliform cells are large, thin-walled and highly vacuolated cells that play a vital role in controlling leaf rolling in response to drought, salinity and high temperature (Itoh *et al*., 2005). The efflux of water from bulliform cells induce adaxial leaf curling (Liu *et al*., 2016). Expansion of the adaxial epidermal cells while an increase in bulliform cells, is closely related to abaxial rolling of the leaf. Furthermore, abaxial leaf rolling and their functions are bidirectional (Li *et al*., 2010). The present study showed that *O. eichingeri* and *O. rufipogon* had the highest number of bulliform cells ($p=0.05$) per cluster (Figure 3) whereas *O. granulata* showed the lowest number. The largest bulliform cell cluster and the highest width of middle bulliform cells were found in *O. rhizomatis* (Table 5). Moreover, the wider bulliform cell cluster and higher distance between the two clusters in the *O. rufipogon* indicated a systematic modification in morphology and anatomy involved in the development of rice in terms of drought resistance.

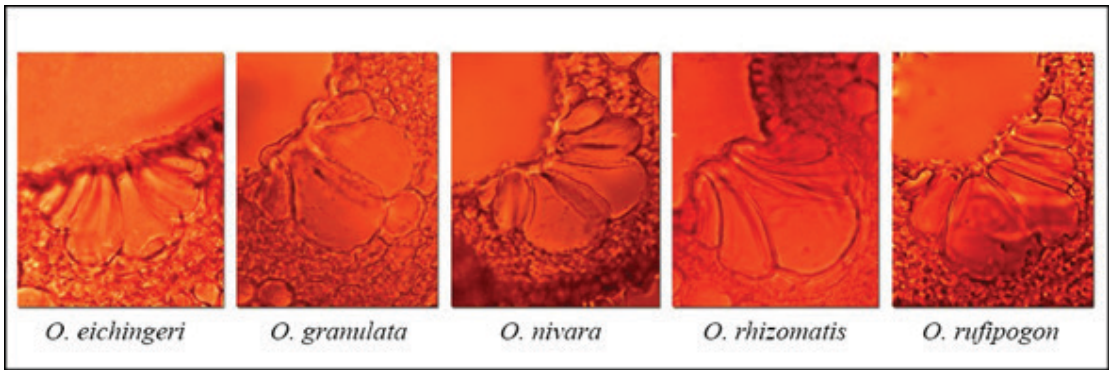


Figure 3: Light micrographs illustrating representative bulliform cell cluster of the five wild rice species (magnification 10×40)

Table 5: Anatomical characteristics of bulliform cell clusters in the leaves of five wild rice species.

Variable	*O. eichingeri* (Mean±SE)	*O. granulata* (Mean±SE)	*O. nivara* (Mean ± SE)	*O. rhizomatis* (Mean ± SE)	*O. rufipogon* (Mean ± SE)	LSD*
A	6.667 ± 0333	4.333 ± 0.333	5.667 ± 0.882	5.333 ± 0.333	6.333 ± 0.333	1.5753
B	177.62 ± 8.52	159.8 ± 17.6	113.8 ± 23.4	212.2 ± 13.0	246.7 ± 43.0	84.826
C	34.11 ± 1.98	48.03 ± 1.43	47.73 ± 5.63	84.16 ± 7.07	53.39 ± 5.74	17.107
D	34.15 ± 4.04	33.26 ± 0.438	46.3 ± 11.1	39.44 ± 1.65	61.4 ± 13.7	27.664
E	4999 ± 591	5201 ± 386	4838 ± 1125	11696 ± 838	7831 ± 2133	3910.4
F	12.667 ± 0.333	11.00 ± 0.000	11.33 ± 0.333	11.00 ± 0.000	12.33 ± 0.333	0.6875
G	27.67 ± 1.45	63.33 ± 3.28	20.00 ± 1.000	32.67 ± 1.76	40.67 ± 1.45	5.5269
H	8.667 ± 0.667	8.667 ± 0.333	113.33 ± 6.98	23.33 ± 0.882	184.33 ± 6.98	13.2

Further, the number of bulliform cells per cluster had a positive significant correlation indicating that a relatively higher cell number or cluster area in a species might play an important role in the adaptation to dry conditions (Giuliani *et al*., 2013). Stomata are microscopic apparatus in leaf epidermis enabling exchange of air and mainly contribute to the photosynthetic efficiency. The highest and the lowest stomatal density were recorded in *O. granulata* and *O. nivara*, respectively (Table 5). The stomatal density and stomatal size are the anatomical traits that contribute to gas diffusion (Giuliani *et al*., 2013). Leaf gas exchange was controlled by different stomatal traits such as stomata number, density and size (Panda *et al*., 2008). Stomatal density was influenced by the number of stomata per row, although on the abaxial surface, a greater number of rows across the leaf have also contributed to the stomatal density. The highest trichome density was observed in *O. rufipogon* (Table 5) indicating its enhancing antibiosis and antixenosis properties thus, reducing insect landing and feeding on leaf surface (Tian *et al*., 2012). This character could be exploited by breeders in the selection of superior genotypes in terms of phenotypic performance.

Specific habitat information for field collectors

Information on habitat preference, geographical distribution and life history traits are the most important facts that drive efficient sampling of wild genetic resources in their natural habitats. Some of these habitats are threatened due to various human activities (Sandamal *et al*., 2018a). Therefore, immediate actions are needed to conserve these valuable rice genetic resources (Abhayagunasekara *et al*., 2018).

O. nivara was mainly confined to the low country dry and intermediate zones. It was not found in the wet zone or upcountry dry/wet regions (Liyanage & Senanayake, 2010). *O. nivara* is distributed extensively in the dry zone and approximately more than 2 ha area in certain natural habitats can be seen. Swampy areas, at the edges of ponds and lakes, and beside streams are the major natural habitats of *O. nivara*. It generally begins seedling emergence with monsoon rain and grows in shallow water. However, continuous water logging condition is not required throughout the life cycle (life cycle observations). It is an annual plant; flowering occurs from January to May and peak mature panicles were recorded from April to May (Ratnasekera *et al*., 2019).

Natural populations of *O. rufipogon* were distributed mainly in the coastal belt from Puttalam to Matara in the intermediate and wet zones (Liyanage & Senanayake, 2010), thus differ in *O. nivara* habitats. The typical natural habitats of *O. rufipogon* were stream banks, marshy lands, swamps, and deep-water lake edges (Sandamal *et al*., 2018a). It grows in water 10 cm – 5 m deep. Perennial *O. rufipogon* is photoperiod sensitive plant with a bimodal flowering pattern and peak mature panicles were observed in April and October, separately (Ratnasekera *et al*., 2019). The reproductive stage of *O. rufipogon* occurred over a longer period than in *O. nivara*.

O. eichingeri is mainly distributed in evergreen forests and dry, mixed evergreen forests located in the intermediate and dry zones. The habitat included forest margins, disturbed and undisturbed forests, and stream banks under shaded or open conditions with well-draining soil condition (Liyanage, 2002). *O. eichingeri* shows a year-round flowering pattern. It was found in association with other wild rice species such as *O. nivara*, *O. granulata* and *O. rhizomatis* (*field observations*).

O. granulata is distributed in the intermediate zone of Sri Lanka. It grows in shady or partially shady habitats often in sloping upland. Moreover, it is well established in the degraded primary or secondary forest area. *O. granulata* shows a year-round flowering pattern.

O. rhizomatis grows naturally in the intermediate and dry zones of Sri Lanka. It grows in periodically flooded areas in the open or under partial shade in primary and secondary forests in dry and intermediate zones. Plants can be seen during the period of late December to May. Mature panicles were observed during the February and March and seed shattering happened in late March.

CONCLUSION

This study has dissected the morphological, leaf anatomical and physiological traits of wild relatives of cultivated rice in Sri Lanka (*O. nivara, O. rufipogon, O. eichingeri, O. rhizomatis and O. granulata*) and reports a significant morpho-physiological and anatomical diversity of the traits. Qualitative parameters such as the panicle type, awning, stigma colour, lemma and palea pubescence, seed coat colour, blade pubescence, and ligule shape showed vast differences among the five species and are useful and promising characters in field identification. Quantitative traits such as flag leaf length, flag leaf width, culm length, culm diameter, panicle length, 100-grain weight, and plant height are distinctive parameters among five species that could be used for further confirmation of species. The plant physiological characters such as net photosynthetic rate, transpiration rate, stomatal conductance and sub-stomatal CO_2

concentration are ideal to be considered in super rice breeding. Anatomical traits such as bulliform cells per cluster, stomatal density, and trichome density could be useful in addressing biotic and abiotic stresses in rice breeding. The leaf morphology, physiology and anatomy are not always inter-related. Hence, recombining those traits could open up new avenues to re-engineer new leaf types of genus *Oryza*. Increasing desirable traits are also possible using re-engineered trait from appropriate wild species.

Physiological functions and anatomical features of the five wild rice species vary from each other indicating the potential of using such functional traits in rice breeding programmes. There were significant correlations between several functional and structural traits, and physiological traits such as transpiration and photosynthesis. The findings of the present study help clear the way for field identification, conservation of the existing rice gene pool as well as provide useful information on important traits of the five rice genotypes for further utilisation.

Conflict of Interest

All authors declared that there is no conflict of interest involved in this work.

Acknowledgement

Financial assistance provided by the National Research Council of Sri Lanka (Grant number NRC 15-108) is gratefully acknowledged. We thank Tharindra Jayarathne and W.M.R.H. Wickramasinghe for their technical assistance with regards to photography and design colour images on the manuscript.

REFERENCES

Abhayagunasekara A.V.C., Bandaranayake P.C.G., Samarasinghe W.L.G. & Pushpakumara D.K.N.G. (2018). Diversity of wild rice species of Sri Lanka: some productive traits. *Proceedings of the 7th Young Scientist Forum Symposium,* Sri Lanka, pp. 7–10.

Alvarez J.M., Rocha J.F. & Machado S.R. (2008). Bulliform cells in Loudetiopsis chrysothrix (Nees) Conert and Tristachya leiostachya Nees (Poaceae): structure in relation to function. *Brazilian Archives of Biology and Technology* **51**: 113–119.
DOI: http://dx. doi.org/10.1590/S1516-89132008000100014

Ammiraju J.S. et al. (11 authors) (2010). Spatio-temporal patterns of genome evolution in allotetraploid species of the genus *Oryza*. *The Plant Journal* **63**(3): 430–442.
DOI: https://doi.org/10.1111/j.1365-313X.2010.04251.x

Banaticla-Hilario M.C.N. (2012). An ecogeographic analysis of *Oryza* series Sativae in Asia and the Pacific. *PhD thesis*, Wageningen University, The Netherlands.

Banaticla-hilario M.C.N., Sosef M.S., Mcnally K.L., Hamilton N.R.S. & Van Den Berg R.G. (2013). Ecogeographic variation in the morphology of two Asian wild rice species, *Oryza nivara* and *Oryza rufipogon*. *International Journal of Plant Sciences* **174**: 896–909.
DOI: https://doi.org/10.1086/670370

Cholewa E. & Griffith M. (2004). The unusual vascular structure of the corm of Eriophorum vaginatum: implications for efficient retranslocation of nutrients. *Journal of Experimental Botany* **55**(397): 731–741.
DOI: https://doi.org/10.1093/jxb/erh054

Chuanren D., Bochu W., Pingqing W., Daohong W. & Shaoxi C. (2004). Relationship between the minute structure and the lodging resistance of rice stems. *Colloids and Surfaces B: Biointerfaces* **35**: 155–158.
DOI: https://doi.org/10.1016/j.colsurfb.2004.03.005

Crawford A.J., McLachlan D.H., Hetherington A.M. & Franklin K.A. (2012). High temperature exposure increases plant cooling capacity. *Current Biology* **22**(10): 396–397.
DOI: https://doi.org/10.1016/j.cub.2012.03.044

Duan S., Lu B., Li Z., Tong J., Kong J., Yao W., Li S. & Zhu Y. (2007). Phylogenetic analysis of AA-genome *Oryza* species (Poaceae) based on chloroplast, mitochondrial, and nuclear DNA sequences. *Biochemical genetics* **45**: 113–129.
DOI: https://doi.org/10.1007/s10528-006-9062-x

Gauthami P., Subrahmanyam D., Padma V., Kiran T.V., Rao Y.V., Rao P. & Voleti S.R. (2014). Variation in leaf photosynthetic response of rice genotypes to post-anthesis water deficit. *Indian Journal of Plant Physiology* **19**(2): 127–137.
DOI: https://doi.org/10.1007/s40502-014-0086-7

Giuliani R., Koteyeva N., Voznesenskaya E., Evans M.A., Cousins A.B. & Edwards G.E. (2013). Coordination of leaf photosynthesis, transpiration, and structural traits in rice and wild relatives (Genus *Oryza*). *Plant Physiology* **162**: 1632–1651.
DOI: https://doi.org/10.1104/pp.113.217497

Gong H., Randall D. & Flowers T. (2006). Silicon deposition in the root reduces sodium uptake in rice (*Oryza sativa* L.) seedlings by reducing bypass flow. *Plant, Cell and Environment* **29**: 1970–1979.
DOI: https://doi.org/10.1111/j.1365-3040.2006.01572.x

Gupta N.K., Meena S.K., Gupta S. & Khandelwal S.K. (2002). Gas exchange, membrane permeability, and ion uptake in two species of Indian jujube differing in salt tolerance. *Photosynthetica* **40**(4): 535–539.
DOI: https://doi.org/10.1023/A:1024343817290

Hayat S., Hayat Q., Alyemeni M.N., Wani A.S., Pichtel J. & Ahmad A. (2012). Role of proline under changing environments: a review. *Plant Signaling and Behavior* **7**(11): 1456–1466.
DOI: https://doi.org/10.4161/psb.21949

He W.M. & Zhang X.S. (2003). Responses of an evergreen shrub Sabina vulgaris to soil water and nutrient shortages

in the semi-arid Mu Us Sandland in China. *Journal of arid environments* **53**: 307–316.
DOI: https://doi.org/10.1006/jare.2002.1051

Hose E., Clarkson D., Steudle E., Schreiber L. & Hartung W. (2001). The exodermis: a variable apoplastic barrier. *Journal of Experimental Botany* **52**: 2245–2264.
DOI: https://doi.org/10.1093/jexbot/52.365.2245

Huckelhoven R. & Kogel K.H. (1998). Tissue-specific superoxide generation at interaction sites in resistant and susceptible near-isogenic barley lines attacked by the powdery mildew fungus (*Erysiphe graminis* f. sp. *hordei*). Molecular Plant-Microbe Interactions **11**(4): 292–300.
DOI: https://doi.org/10.1094/MPMI.1998.11.4.292

Ikeda R. & Vaughan D.A. (1991). The distribution of resistance genes to the brown planthopper in rice germplasm. *Rice Genetics Newsletter* **8**: 1–3.

Itoh J.I., Nonomura K.I., Ikeda K., Yamaki S., Inukai Y., Yamagishi H., Kitano H. & Nagato Y. (2005). Rice plant development: from zygote to spikelet. *Plant and Cell Physiology* **46**: 23–47.
DOI: https://doi.org/10.1093/pcp/pci501

Jackson M.T. (1995). Protecting the heritage of rice biodiversity. *Geo Journal* **35**: 267–274.
DOI: https://doi.org/10.1007/BF00989134

Kadioglu A. & Terzi R. (2007). A dehydration avoidance mechanism: leaf rolling. *The Botanical Review* **73**: 290–302.
DOI: https://doi.org/10.1663/0006-8101(2007)73[290:ADAMLR]2.0.CO;2

Khan M.H., Dar Z.A. & Dar S.A. (2015). Breeding strategies for improving rice yield - a review. *Agricultural Sciences* **6**(5): 467.

Khush G.S. (1997). Origin, dispersal, cultivation and variation of rice. *Plant Molecular Biology* **35**: 25–34.
DOI: https://doi.org/10.1023/A:1005810616885

Khush G.S. (2001). Green revolution: the way forward. *Nature Reviews Genetics* **2**: 815–822.
DOI: https://doi.org/10.1038/35093585

Li L., Shi Z.Y., Li L., Shen G.Z., Wang X.Q., An L.S. & Zhang J.L. (2010). Overexpression of ACL1 (abaxially curled leaf 1) increased bulliform cells and induced abaxial curling of leaf blades in rice. *Molecular Plant* **3**: 807–817.
DOI: https://doi.org/10.1093/mp/ssq022

Lim J.S., Manan Z.A., Hashim H. and Alwi S.R.W. (2013). Towards an integrated, resource-efficient rice mill complex. *Resources, Conservation and Recycling* **75**: 41–51.
DOI: https://doi.org/10.1016/j.resconrec.2013.04.001

Liu R., Zheng X.M., Zhou L., Zhou H.F. & Ge S. (2015). Population genetic structure of *Oryza rufipogon* and *Oryza nivara*: implications for the origin of *O. nivara*. *Molecular Ecology* **24**: 5211–5228.
DOI: https://doi.org/10.1111/mec.13375

Liu X., Li M., Liu K., Tang D., Sun M., Li Y., Shen Y., Du G. & Cheng, Z. (2016). Semi-Rolled Leaf2 modulates rice leaf rolling by regulating abaxial side cell differentiation. *Journal of Experimental Botany* **67**(8): 2139–2150.
DOI: https://doi.org/10.1093/jxb/erw029

Liyanage A.S.U. (2002). *Eco-Geographic Survey of Crop Wild Relatives.* Agriculture Press, Gannoruwa, Peradeniya.

Liyanage A.S.U. & Senanayake G. (2010). *The Atlas of Selected Crop Wild Relatives in Sri Lanka*. Department of Agriculture, Peradeniya, Sri Lanka.

Liyanage A.S.U., Hemachandra P.V., Edirisinghe D.K., Senevirathna S.K. & Takahashi J. (2002). Surveying and mapping of wild species of *Oryza* in Sri Lanka. *Japanese Journal of Tropical Agriculture* **46**: 14–22.

Lu B.R., Zheng K., Qian H. & Zhuang J. (2002). Genetic differentiation of wild relatives of rice as assessed by RFLP analysis. *TAG Theoretical and Applied Genetics* **106**: 101–106.
DOI: https://doi.org/10.1007/s00122-002-1013-2

Madurangi S.A.P., Ratnasekera D., Hemachandra P.V. & Senanayake S.G.J.N. (2012). Evaluation of brownplanthopper *Nilaparvata lugens* (Stal) resistance in *Oryza nivara* wild rice accessions found in Sri Lanka. *Proceedings of International Forestry & Environment Symposium* **15**:172–175.

Mathobo R., Marais D. & Steyn J.M. (2017). The effect of drought stress on yield, leaf gaseous exchange and chlorophyll fluorescence of dry beans (*Phaseolus vulgaris* L.). *Agricultural Water Management* **180**: 118–125.
DOI: https://doi.org/10.1016/j.agwat.2016.11.005

Micol J.L. & Hake S. (2003). The development of plant leaves. *Plant Physiology* **131**(2): 389–394.
DOI: https://doi.org/10.1104/pp.015347

Panda D., Sharma S.G. & Sarkar R.K. (2008). Chlorophyll fluorescence parameters, CO_2 photosynthetic rate and regeneration capacity as a result of complete submergence and subsequent re-emergence in rice (*Oryza sativa* L.). *Aquatic Botany* **88**(2): 127–133.
DOI: https://doi.org/10.1016/j.aquabot.2007.08.012

Rajkumar G., Weerasena O. & Fernando K. (2015). A study of genetic diversity in *Oryza rhizomatis* DA Vaughan accessions using AFLP markers and morphological traits. *Tropical Plant Research* **2**(1): 10–16.

Rathore M., Singh R., Kumar B. & Chauhan B. (2016). Characterization of functional trait diversity among Indian cultivated and weedy rice populations. *Scientific Reports* **6**(1): 1–9.
DOI: https://doi.org/10.1038/srep24176

Ratnasekera D., Tennakoon A., Sandamal S., Wijerathna P., Wickramasinghe W.M.R.H., Iroshan A. & Jayaweera W.M.C.S. (2019). Divergence pattern of life-history traits in *Oryza nivara* and *Oryza rufipogon* natural populations in Sri Lanka. *Association for Tropical Biology and Conservation - Asia Pacific Conference*, Sri Lanka, pp. 106-107.

Ren F., Lu B.R., Li S., Huang J. & Zhu Y. (2003). A comparative study of genetic relationships among the AA-genome *Oryza* species using RAPD and SSR markers. *Theoretical and Applied Genetics* **108**: 113–120.
DOI: https://doi.org/10.1007/s00122-003-1414-x

Sandamal S., Tennakoon A., Ratnasekera D., Amarasekera D.A.B.N. & Marambe B. (2018a). Eco-geographic variation of common wild rice-*Oryza rufipogon* Griff. in Sri Lanka. *Plantae Scientia* **1**(2): 36-43.

DOI: https://doi.org/10.32439/ps.v1i02.36–43

Sandamal S., Tennakoon A., Meng Q.L., Marambe B., Ratnasekera D., Melo A. & Ge S. (2018b). Population genetics and evolutionary history of the wild rice species *Oryza rufipogon* and *O. nivara* in Sri Lanka. *Ecology and Evolution* **8**(23): 12056–12065.
DOI: https://doi.org/10.1002/ece3.4665

Sarla N., Bobba S. & Siddiq E. (2003). ISSR and SSR markers based on AG and GA repeats delineate geographically diverse *Oryza nivara* accessions and reveal rare alleles. *Current Science* **84**: 683–690.

Sattler R. & Hall B.K. (1994). Homology, homeosis, and process morphology in plants. Homology: *The Hierarchical Basis of Comparative Biology Copyright* 423–475.

Siddique M.R.B., Hamid A. & Islam M.S. (1999). Drought stress effects on photosynthetic rate and leaf gas exchange of wheat. *Botanical Bulletin of Academia Sinica* **40**: 141–145.

Tan B.L. & Norhaizan M.E. (2020). Rice demands: A brief description. In: Rice By-products: *Phytochemicals and Food Products Application*, pp. 7–11. Springer, USA.
DOI: https://doi.org/10.1007/978-3-030-46153-9_2

Tian D., Tooker J., Peiffer M., Chung S.H. & Felton G.W. (2012). Role of trichomes in defense against herbivores: comparison of herbivore response to woolly and hairless trichome mutants in tomato (*Solanum lycopersicum*). *Planta* **236**: 1053–1066.
DOI: https://doi.org/10.1007/s00425-012-1651-9

Vaughan D.A. (1989). The genus Oryza L.: current status of taxonomy. *IRRI Research Paper Series* **138**: 1–21.

Zhu T., Xu, P.Z., Liu, J.P., Peng, S., Mo X.C. & Gao L.Z. (2014). Phylogenetic relationships and genome divergence among the AA-genome species of the genus *Oryza* as revealed by 53 nuclear genes and 16 intergenic regions. *Molecular Phylogenetics and Evolution* **70**: 348–361.
DOI: https://doi.org/10.1016/j.ympev.2013.10.008

Supplementary Table 1: Simple linear correlation coefficient between pairs of all the traits

	SH	LL	FLL	FLW	CL	CD	PH	PL	100GW	GL	GW	NCC	DBC	CBW	SD	TD	PAR	Evap	GS	LT
LL	0.887*																			
FLL	0.855*	0.966*																		
FLW	-0.713*	-0.876*	-0.834*																	
CL	0.986*	0.861*	0.831*	-0.680*																
CD	0.731*	0.717*	0.831*	-0.689*	0.705*															
PH	0.975*	0.85*	0.853*	-0.650*	0.988*	0.755*														
PL	0.622*	0.531*	0.698*	-0.353	0.623*	0.778*	0.719*													
100GW	0.752*	0.929*	0.827*	-0.811*	0.745*	0.461	0.696*	0.244												
GL	0.873*	0.957*	0.901*	-0.746*	0.875*	0.580*	0.851*	0.468	0.944*											
GW	0.611*	0.836*	0.731*	-0.707*	0.597*	0.341	0.549*	0.116	0.952*	0.877*										
NCC	0.253	0.238	0.351	-0.420	0.247	0.613*	0.293	0.497	0.068	0.057	-0.088									
DBC	0.340	-0.017	0.030	0.118	0.332	0.357	0.343	0.302	-0.224	-0.019	-0.286	-0.055								
CBW	0.66*	0.559	0.563	-0.448	0.683*	0.514*	0.715*	0.529*	0.486	0.556*	0.409	0.525*	-0.050							
SD	-0.153	-0.386	-0.557*	0.408	-0.084	-0.662*	-0.179	-0.586*	-0.170	-0.209	-0.172	-0.479	0.136	-0.124						
TD	0.979*	0.940*	0.905*	-0.805*	0.967*	0.762*	0.952*	0.588*	0.826*	0.904*	0.677	0.287	0.244	0.616*	-0.224					
PAR	-0.651*	-0.543*	-0.477	0.431	-0.663*	-0.295	-0.597*	-0.212	-0.569*	-0.626*	-0.498	-0.183	0.008	-0.570*	-0.162	-0.603*				
Evap	0.471	0.250	0.288	-0.444	0.449	0.569*	0.464	0.377	0.025	0.078	-0.142	0.606*	0.461	0.442	-0.077	0.448	-0.219			
GS	0.723*	0.457	0.440	-0.508	0.716*	0.574*	0.712*	0.445	0.276	0.363	0.099	0.476	0.484	0.640*	0.081	0.673*	-0.450	0.914*		
LT	-0.505	-0.455	-0.255	0.431	-0.511	0.015	-0.385	0.195	-0.622*	-0.552*	-0.570	0.120	0.046	-0.260	-0.436	-0.494	0.696*	0.017	-0.263	
PN	0.516*	0.288	0.296	-0.164	0.506	0.242	0.502	0.441	0.123	0.290	-0.036	-0.113	0.510	0.007	0.125	0.470	-0.306	0.380	0.487	-0.170

*Significant at probability = 0.05; SH: seedling height; LL: ligule length; FLL: flag leaf length; FLW: flag leaf width; CL: culm length; CD: culm diameter; PH: plant height; PL: panicle length; 100GW: 100 grain weight; GL: grain length; GW: grain width; NCC: number of cells per cluster; DBC: distance between two cluster; CBW: cluster base width; SD: stomatal density; TD: trichome density; PAR: photosynthetically active radiation; Evap: transpiration rate from the leaf; GS: stomatal conductance; LT: leaf temperature; PN: net photosynthetic rate.

J.Natn.Sci.Foundation Sri Lanka 2021 49 (3): 383 - 391
DOI: http://dx.doi.org/10.4038/jnsfsr.v49i3.10011

RESEARCH ARTICLE

Improved protocol for efficient regeneration of coconut (*Cocos nucifera* L.) anther derived embryos

D Bandupriya[1*] and P Waidyarathne[2]
[1] *Department of Plant Sciences, Faculty of Science, University of Colombo, Colombo 03.*
[2] *Plant Physiology Division, Coconut Research Institute, Lunuwila.*

Submitted: 04 June 20; Revised:11 March 21; Accepted: 26 March 21

Abstract: The occurrence of severe shoot necrosis and other constraints such as low frequency of embryo induction and poor regeneration into plants, restrict the use of coconut androgenesis in practice. Fine-tuning of the protocol by addressing the above constraints was carried out with the intention of scaling-up haploid plant production. Out of the carbon types, sucrose and maltose, when added in concentrations of 90.0 gL^{-1} and 120.0 gL^{-1} showed significantly higher ($p < 0.05$) embryo (44.0 % and 36.0 %, respectively) production. Out of the concentrations used in the study, 20.0 µM 6-benzylaminopurine (BAP) showed significantly ($p = 0.001$) higher shoot generation (47.6 %) as well as significantly ($p = 0.006$) longer shoot production (31.7 %) during the study period. The effect of $CaCl_2$ on the reduction of shoot necrosis was also tested. $CaCl_2$ showed a significant ($p = 0.001$) effect on reducing the shoot necrosis. The lowest occurrence of shoot necrosis (25.0 %) was observed in 4.0 mM $CaCl_2$ treatment. Continuous sub-culturing of shoots with initial signs of shoot necrosis to elevated $CaCl_2$ levels until the rooting stage facilitated the recovery. Transfer of the shoots frequently into a fresh medium was not beneficial for the suppression of necrosis. Shoots maintained in the medium enriched with 4.0 mM $CaCl_2$ were transferred for acclimatisation, and this is the first report of transferring haploid coconut plants to acclimatisation conditions. The rooted shoots produced through the optimised protocol were acclimatised successfully. The prevention of shoot loss due to shoot necrosis will be beneficial for further refinement of the coconut anther culture protocol.

Keywords: Anther culture, calcium chloride, carbon source, regeneration, shoot necrosis.

INTRODUCTION

Coconut plays a vital role in the economy of tropical countries such as the Philippines, Indonesia, India, and Sri Lanka. Genetic improvement of coconut for high yield and other desirable traits is a priority research area for uplifting the coconut industry. Due to the long life span and high heterozygosity, coconut breeding through conventional methods is a long, difficult and expensive process (Nguyen *et al.,* 2015). Moreover, the production of true hybrids is hampered by high heterozygosity of coconut palm. At present, coconut breeding is done either by mass selection or crossing between varieties that have high variation within a population (Batugal *et al*., 2009). Thus, the resultant progenies are not true hybrids. Alternative approaches to produce homozygous pure lines in coconut are highly desirable in order to improve coconut plantations. The production of double haploids (DH) is the fastest route to initiate homozygosity in plants and has been experimented with a large number of crop species (Dunwell, 1985; Abdollahi & Rashidi, 2018; Bhatia *et al*., 2018). DHs are produced by doubling the chromosomes of haploid plants, which can occur spontaneously or by chemical treatments, resulting in individuals with two identical copies of each chromosome (Dunwell, 1985).

* Corresponding author (dbandupriya@pts.cmb.ac.lk https://orcid.org/:0000-0001-6257-1472)

In higher plants, doublehaploids can be introduced either through androgenesis (anther or microspore culture) or gynogenesis (ovary or megaspore culture). The availability of a few mega spores and the difficulty of fine-dissection of gametes hinder the use of mega spores for haploid plant production (Rajcan *et al.*, 2011).

The development of an effective protocol for DHs and its further application in breeding programmes is the only practical alternative for enhancing the coconut breeding strategy for the production of true hybrids. This could reduce the time required for the development of homozygous populations. Anther culture has been reported by Perera *et al*. (2008). Some of the critical factors that are required to induce microspore embryogenesis such as culture medium, pollen developmental stage, stress pre-treatment, and anther density have been discussed (Nguyen *et al*., 2015; Bandupriya *et al*., 2016). The problems associated with shoot necrosis and shoot death during in vitro culture limit further growth of plantlets up to acclimatisation stage. Moreover, the low frequency of microspore-derived embryo induction and poor plant regeneration restrict the use of anther culture technique in further developments. Thus, the use of boosted $CaCl_2$ concentrations in the regeneration medium and frequent sub-culturing was tested for overcoming shoot necrosis. Further, attempts were made to improve the anther culture protocol by studying the effect of different carbon sources on androgenic induction and of 6-benzylaminopurine (BAP) on plant regeneration efficiency.

METHODOLOGY

Plant material and explant preparation

Rachilla was collected from inflorescences of the variety 'Sri Lanka Tall' at three weeks before the splitting (3WBS) stage as described by Perera *et al*. (2008), from an adult coconut palm growing at Bandirippuwa Estate, Lunuwila, Sri Lanka. At this stage, anthers contain pollen grains at thelate uni-nucleate stage (Perera *et al.*, 2008). The middle portion of each rachillae (containing male flowers) were wrapped in aluminium foil and given a heat shock at 38 ºC for 6 d. Pre-treated anthers were excised from the male flowers and pooled anthers were surface sterilised using 2.0 % (v/v) commercial bleach (Clorox®) solution with a few drops of liquid detergent for 10 min, followed by four rinses with sterilised distilled water under aseptic conditions.

Effect of carbon source on androgenesis induction

Culture initiation and regeneration were based on the methods described by Perera *et al*. (2009) with modifications. Modified Eeuwens Y_3 medium (Eeuwens, 1976) was used in all steps until plants were transferred to the soil. A medium consisted of 100 µM 2, 4-dichlorophenoxyacetic acid (2,4-D) and 100 µM Naphthaleneacetic acid (NAA) was used as the androgenesis induction medium. The effect of the type and concentration of the carbon source on androgenesis induction was studied by culturing the pre-treated anthers into solidified media supplemented with sucrose, maltose and glucose at concentrations of 40, 90, 120 and 150 gL^{-1}. After adjusting the pH to 5.8, activated charcoal (Heycarb, Sri Lanka) at a concentration of 0.1 % (w/v) and phytagel 0.25 % (w/v) were added to the medium and autoclaved at 121ºC for 20 min. Fifteen anthers were cultured (abaxial side up) per Petri plate (90 × 18 mm) each containing 40.0 mL of culture medium. Five Petri plates were used for each treatment. The Petri plates were incubated in the dark at 27 ± 1 ºC until embryos emerged. The number of anthers that produced embryos was counted and recorded after 08 months from culture initiation.

Effect of BAP on regeneration

The embryos were sub-cultured into somatic embryo induction medium with reduced 2,4-D (70.0 µM) solidified with 0.25 % (w/v) phytagel, followed by maturation medium devoid of any growth regulators and solidified with 0.30 % (w/v) phytagel. Mature embryos were collected, bulked and cultured on the germination medium supplemented with different concentrations of BAP (5.0, 10.0, 20.0, and 25.0 µM) for further proliferation and shoot initiation. Three sub cultures were added into the same fresh medium until shoots emerged. Well-developed germinating embryos were then transferred to regeneration medium supplemented with 0.45 µM gibberellic acid (GA_3). Cultures were maintained for 6 wks in each media mentioned above before being transferred to the next medium. Finally, continuous sub-culturing was done at 6-week intervals (unless otherwise stated) into fresh GA_3 containing media until shoots developed. All culture media contained 0.1 % (w/v) activated charcoal. The cultures were maintained in the dark at 27 ± 1 ºC until the embryos germinated. The germinated embryos (with shoot sprouts) were then exposed to 16 h photoperiod (PAR; 25 µmolm-$2s^{-1}$). The number of embryos converted into shoots was counted. Shoots longer than 1.5 cm was counted in each treatment after 08 months from the first culture of embryos into BAP containing medium. The experiment was repeated three times.

Rooting and acclimatisation of anther cultured plants

Regenerated shoots were transferred to a medium

containing 0.5 µM indole acetic acid (IAA) to induce rooting. Once the roots developed, the plantlets were transferred to a liquid medium supplemented with the same concentration of IAA. Plants were maintained in a medium supplemented with high $CaCl_2$ (4.0 mM) until they were transferred to the soil medium for acclimatisation. Plants with 3–4 well-developed leaves and a healthy root system (Figure 3c) were carefully removed from the liquid medium and each plant was transferred to a propagator containing a potting mixture of sand, soil, and coir dust (1:1:1).

Reduction in shoot necrosis

Two methods were tested for the reduction of shoot necrosis at shoot multiplication stage. To determine the effect of $CaCl_2$ on shoot necrosis, different levels of $CaCl_2$ (2.0, 3.0, and 4.0 mM) were incorporated into germination medium and the same levels were maintained until plants were transferred to soil. Twenty germinating embryos were used for each treatment and the number of plants showing necrosis in each treatment was recorded.

Sub-culturing the shoots into a fresh shoot multiplication medium at 3-wk intervals instead of 6 wks was also attempted. A new set of anther-derived shoots was used for this experiment. Eight shoots were used for each treatment and the experiment was carried out twice.

Experimental design and data analysis

The experiment was designed as a two-factor factorial laid on a completely randomised design (CRD) to determine the effect of carbon source on androgenesis induction. Three sugar types and four concentrations were considered as factors. The experiment was repeated three times. Percentage of embryo production data were analysed using two-way ANOVA after confirming the normality of the data with Anderson Darling normality test (AD = 0.622, p = 0.093). Post-hoc evaluations were done with Tukey's test to find the best sugar type with the correct concentration combination.

The experiment to determine the effect of four BAP concentrations on shoot regeneration was designed as a simple CRD experiment with three replicates. The percentage of embryos converted into shoots and the percentage of embryos that produced shoots longer than 1.5 cm were tested for normality with Anderson Darling test, and one-way ANOVA was used for data analysis.

Elucidating the effect of $CaCl_2$ on the reduction of shoot necrosis was done using three different $CaCl_2$ concentrations in a simple CRD experiment. Twenty germinating embryos were used in each treatment.

Binary logistic models were used to compare the probability of shoot necrosis (as it is a binary response) based on the 03 $CaCl_2$ concentrations as a categorical predictor and to compare the effect of the number of subcultures on necrosis.

RESULTS AND DISCUSSION

Androgenesis was successfully induced in cultured anthers of coconut (Figure 1a – d). Shoots emerged either through a germination point in the embryo or by splitting the haustorial tissue. Single or multiple shoots were successfully developed into complete plantlets.

Effect of type of sugar and concentration on embryo production

Androgenesis efficiency was determined based on the percentage number of embryos produced in cultured anthers under different sugar treatments. The results revealed that the concentration and the type of sugar and their interaction indicate the effect of sugar type on the level of embryo production, is dependent on the sugar concentration.

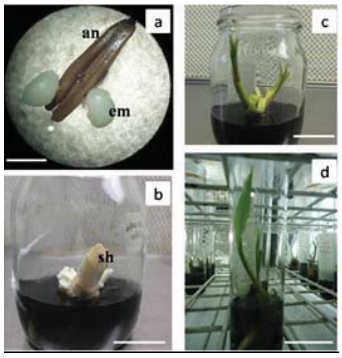


Figure 1: Plant regeneration through coconut anther culture. a - embryos (em) derived from anthers (an) after three months of culture initiation (Bar 2 mm); b - germinating embryo, the emerging shoot (Sh) through a germination point in the embryo (Bar 2.0 cm); c - multiple shoots developed from a single embryo (Bar 2.5 cm); d - complete plantlet (Bar 4.0 cm)

Embryo production was promoted in all four sucrose concentrations. Post hoc evaluation of the interaction effect revealed that the 90 gL^{-1} sucrose concentration performed superior to the other sugar treatments (Figure 2) and recorded the highest percentage of embryo production (44.0 %, Figure 2). There was a reduction in embryo production with the increase of sucrose concentration in the medium. The incorporation of maltose instead of sucrose showed a reduction in embryo production when maltose was added either as 40.0 gL^{-1} or 90.0 gL^{-1} concentration. Maltose added at a concentration of 120.0 gL^{-1} produced significantly higher embryo production equal to 90.0 gL^{-1} sucrose. The effect of higher concentrations (especially 150.0 gL^{-1}) on embryo production was found to be unfavourable and showed a reduced embryo production in both sucrose and maltose. Interestingly at 120.0 gL^{-1} concentration, both sucrose and maltose showed a similar production of embryos. In general, glucose at all concentrations showed the least production of embryos (Figure 2).

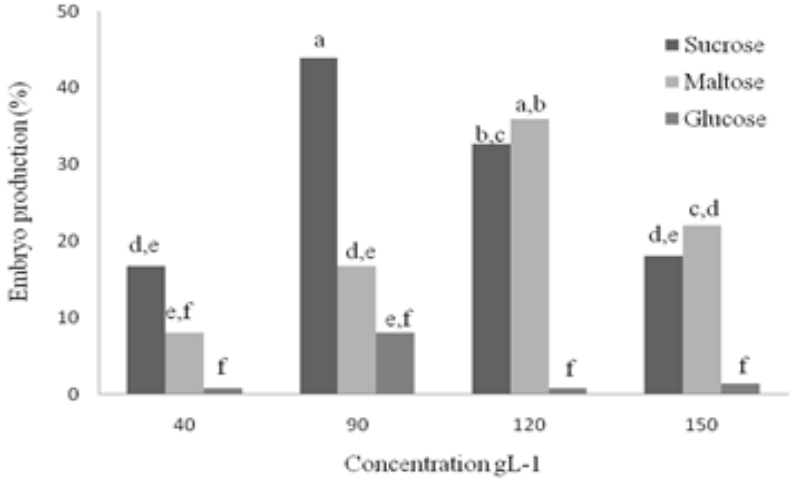


Figure 2: Effect of different sugar types and concentrations on androgenic responses of coconut anthers in androgenesis induction medium. Percentages (mean) with different letters of each parameter are significantly different ($p < 0.05$).

Effect of BAP on shoot regeneration

Results revealed that the different concentrations of BAP act significantly on embryos to generate shoots ($F = 16.5$, $p = 0.001$) and to produce healthy shoots (longer than 1.5 cm) eight months after embryos are transferred to germination medium ($F = 9.22$, $p = 0.006$).

Out of the concentrations used in the study, 20.0 µM BAP showed significantly high shoot generation as well as significantly longer shoots during the study period (Table 1). The percentage of embryos converted to shoots in the 20.0 µM BAP concentration was threefold compared to the control medium supplemented with 5.0 µM. Further increase of BAP concentration reduced the conversion of embryos into shoots (Table 1).

Table 1: Effect of different BAP concentrations in the germination medium on plant regeneration in anther culture of coconut

Concentration of BAP (µM)	Percentage embryos converted to produce shoots*	Percentage embryos with shoots longer than 1.5 cm*
5.0	15.74 ^{B}	07.87 ^{B}
10.0	19.91 ^{B}	15.74 ^{B}
20.0	47.62 ^{A}	31.75 ^{A}
25.0	27.78^{B}	19.91 ^{AB}

^{A,B} Means with the same letters along the columns are not significantly different at $p < 0.05$ at 95 % confidence level

Effect of $CaCl_2$ on shoot necrosis

Shoots derived through androgenesis and raised in regeneration medium [supplemented with normal $CaCl_2$ (2.0 mM)] were affected by shoot necrosis and eventually died (Figure 3a). The symptoms started at either in leaves or immature stem in almost all the cultures. This serious problem made it difficult to raise plantlets up to acclimatisation stage. In order to elucidate the effect of $CaCl_2$ on the reduction of shoot necrosis, shoots derived from coconut anthers were cultured in the germination medium supplemented with elevated $CaCl_2$ concentrations. The germination medium used in this particular experiment comprised the best BAP (20.0 µM) concentration, which was determined in a previous experiment. It was found that $CaCl_2$ can cause a significant effect on shoot necrosis (G-square = 13.63, p = 0.001). The results of goodness-of-fit tests (Deviance, Pearson & Hosmer-Lemeshow) are all greater than the significance level of 0.05 (Chi-Square = 0.64, p = 0.72), which indicates that the use of binary logistic models is appropriate.

Table 2: The effect of $CaCl_2$ concentration on the reduction of shoot necrosis

$CaCl_2$ Concentration (mM)	Percentage necrosis Mean ± SE
2.0	84.21 ± 8.59^A
3.0	66.7 ± 12.6^A
4.0	25.00 ± 11.2^B

^{A,B} Means with the same letters along the columns are not significantly different at $p < 0.05$ at 95 % confidence level

According to the results presented in Table 2, plantlets treated with 2.0 mM $CaCl_2$ concentration showed significantly high necrosis than with 4.0 mM treatment. Three millimolar (3.0 mM) $CaCl_2$ and 4.0 mM $CaCl_2$ concentrations also showed statistical significance for having different levels of necrosis in respective cultures. However, there was no significantly different occurrence of necrosis between the $CaCl_2$ concentrations 2.0 mM and 3.0 mM. The highest level of shoot necrosis (84.21 %) was observed in the medium containing 2.0 mM $CaCl_2$, which is the concentration present in the normal Y_3 medium used in routine culturing.

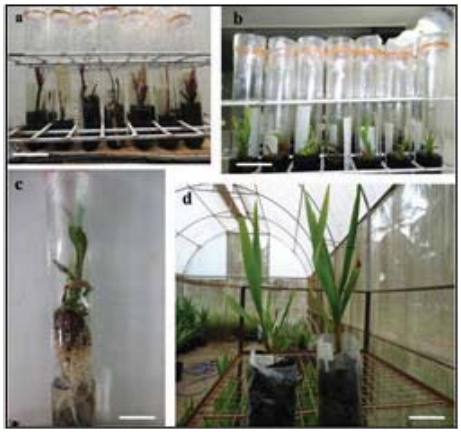


Figure 3. The effect of $CaCl_2$ concentration on the reduction of shoot necrosis. a – dying shoots in the regeneration medium (supplemented with 2.0 mM $CaCl_2$) due to shoot necrosis (Bar 4.0 cm); b - healthy shoots in regeneration medium supplemented with 4.0 mM $CaCl_2$ (Bar 3.0 cm); c - complete plantlet in the rooting medium (Bar 2.0 cm); d - plants at acclimatisation stage (Bar 6.0 cm)

Continuous sub-culturing of the shoots with initial signs of shoot necrosis into media supplemented with 4.0 mM $CaCl_2$ until they developed a good root system has facilitated the recovery. Shoots maintained in the medium enriched with 4.0 mM $CaCl_2$ were transferred for acclimatisation (Figure 3c, 3d). This is the first report of transferring coconut plants developed through androgenesis into acclimatisation conditions.

Effect of number of subcultures on necrosis

Subculture of necrotic shoots into shoot multiplication medium at three-week intervals instead of six weeks did not significantly alleviate the problem of necrosis. It was revealed (Table 3) that there is no significant relationship between the number of subcultures and the occurrence of necrosis (Chi sq. = 0, p = 1.00). The odds ratio 1 indicated that both levels have a similar chance of occurrence of necrosis.

Table 3: Effect of sub-culturing on shoot necrosis reduction.

Subculture interval (weeks)	% shoot necrosis ± SD
3	56.25 ± 06.25
6	56.25 ± 18.75

Carbohydrate that acts as the source of carbon and energy during plant regeneration is a common, important component in coconut tissue culture media (Nguyen *et al*., 2015). Although sucrose is reported as the common sugar type in many of the plant tissue culture media (Yaseen *et al*., 2013), other sugars such as maltose, glucose and some tri-saccharides and pentoses have the potential to metabolise during androgenesis (Yaseen *et al*., 2013). According to the results obtained in the present study, the carbohydrate source is one of the major components that support the conversion of anthers into embryos. Both, the type of carbohydrate and carbohydrate concentration, affect the results obtained for the tested parameters. Out of the carbohydrates tested, sucrose and maltose were superior to glucose. A sucrose concentration of 90.0 gL^{-1} and maltose concentration of 120.0 gL^{-1} showed significantly higher ($p < 0.05$) embryo (44.0 % and 36.0 %, respectively) production. Similar results of using high concentrations of sucrose have been reported elsewhere showing significantly higher embryo production in maize (Buter, 1997). Sucrose is an easily metabolised sugar, which shows a variety of effects on plant cell and tissue culture (Vitova *et al.,* 2002). It is reported that sucrose can control the expression of pathogenesis-related genes in plants (Herbers *et al*., 1996). Moreover, it has been reported that high concentrations of carbohydrates improve embryogenesis by creating an osmotic stress (Agarwal *et al.,* 2004). In addition under osmotic stress conditions, polyamine synthesis in plant cells increases causing favourable conditions for embryogenesis (Litz, 1986). However, sucrose levels higher than 90.0 gL^{-1} showed an inhibitory effect on coconut anthers. Higher levels of sucrose (150.0 gL^{-1}) adversely affected the production of embryos (Figure 2). Similar observations on the reduction of embryo production upon elevated sucrose levels have been reported for species such as barley (Marsolais & Kasha, 1985), rye (Guo & Pulli, 2000), and *Cucumis sativus* (Ashok & Murthy, 2004).

Maltose has been superior to sucrose as a carbohydrate source for androgenesis in several species including cereals. Culturing of barley microspores in media supplemented with sucrose, glucose, or fructose was found to be deleterious, whereas maltose acted favourably on embryo production (Scott & Lyne, 1994). The capacity of barley microspores to differentiate and induce green plantlets has been enhanced by both maltose and malt extracts (Finnie *et al*., 1989). The effect has been determined in relation to the osmotic regulation of microspores during the induction phase (Sunderland & Dunwell, 1977). However, in the present study maltose effect did not surpass the same observed in sucrose but showed similar results with sucrose at a comparatively higher concentration of maltose.

The effect of different concentration regimes of BAP on the conversion and further proliferation of embryos into shoots was investigated. BAP is considered as a chemically stable cytokinine in plant tissues and it is the commonly preferred cytokinine by plant tissue culturists (Klems *et al*., 2000). Initial work on coconut androgenesis has shown that conversion of embryos into plantlets is possible in the presence of 5.0 µM BAP (Perera *et al*., 2008; 2009). However, extremely low percentage (7.0 %) of shoot conversion has been reported. In a recent study, Perera *et al.* (2020) reported that BAP concentration plays a significant role in converting anther derived embryos into shoots. Maximum embryo sprouting has been reported in the media supplemented with 25.0 and 35.0 µm BAP with a record of 50.0 % and 53.0 % conversions, respectively. However, the greatest shoot development recorded in the study conducted by Perera *et al.* (2020) was less than 30.0 %, in the medium supplemented with 35.0 µm BAP. Nevertheless, in the present study, nearly 50.0 % of the embryos were converted into plantlets when 20.0 µM BAP was used. The use of modified Y_3 medium formulated specifically for coconut in vitro culture instead of Murashige and Skoog medium is one of the major differences between these two studies. Moreover, different culture incubation durations in BAP incorporated media were maintained in these two studies.

Hormones usually tend to show the maximum shooting response at its optimum concentration. Farahani *et al.* (2008) reported that the shoot multiplication of *Musa acuminata* was affected by the concentration of BAP. Later in 2015, Ferdous *et al.* revealed the maximum single shoot formation and longest shoot formation in M. acuminata at 0.5 mg/L BAP. Similarly, determination of the precise concentration of BAP for maximum shoot regeneration has been reported in several other studies in different crop plants. Gubi *et al.* (2004), Kadota and Niimi (2003), Klems *et al.* (2000) and Jafari *et al*. (2011) reported that overexposure of cultures to higher concentrations of BAP might lead to hyperhydric shoots, which was not observed in the present study. However, Katoda and Niimi (2003) reported that the occurrence of hyperhydric cultures is less in BAP supplemented culture media when compared to media incorporated with synthetic cytokinines such as *N*-(2-chloro-4-pyridyl)-N9-phenylurea (CPPU) and 1-phenyl-3- (1,2,3-thiadiazol-5-yl) urea (TDZ). Shoots consisting of single, double or multiple (Figure 1c) shoots were produced in BAP

supplemented media, which is in accordance with the previous reports of coconut androgenesis (Perera *et al*., 2009). Ploidy analysis of coconut anther-derived plants has been performed previously on the current protocol by flow cytometry analysis and recorded a high double haploid yield (Perera *et al*., 2008). Thus, ploidy analysis studies were not executed for the current study since similar conditions were used as in the previous study.

Shoot necrosis is one of the obstacles associated with the androgenesis of coconut. The survival of shoots regenerated through androgenesis was difficult due to high shoot necrosis. An increase in calcium concentration in the regeneration medium from 2.0 mM to 4.0 mM has recorded higher recovery of anther-derived shoots affected by shoot necrosis, and reduced shoot necrosis from 84.21 % to 25.00 %. Similar results have been reported in *vitro* for several other perennial crops such as grapes (Surakshitha *et al*., 2019), oak (Vieitez *et al*., 1989), banana (Martin *et al*., 2007) and *Trichosantes dioica* (Kishore *et al*., 2015). As discussed in previous reports, the occurrence of necrosis in coconut shoots developed through androgenesis may be associated with the calcium deficiency. Calcium is a major nutrient required for plant growth and it is responsible for the growth and differentiation of plant cells, formation of the cell wall, maintain membrane permeability and (Hepler, 2005; Stael *et al*., 2012). Thus, calcium deficiency in plant tissues could disturb metabolic activities in developing tissues, and as such, metabolic imbalances could be visualised as growth abnormalities like shoot necrosis (Surakshitha *et al*., 2019). According to Hirschi (2004), upward movement of calcium ion in the xylem sap is basically due to an efficient transpiration system. Therefore, it is suggested that conditions existing in the culture vessels that limit an efficient transpiration stream may be another way by which the mobility of calcium ions is limited in the in vitro plantlets, causing deficiency symptoms such as shoot necrosis. Since two times of the normal Ca^{+} concentration of the Y_3 medium was sufficient to reduce shoot necrosis, the effect may have caused low or no effect on shoot necrosis in this particular situation. The recovery or low necrosis in plantlets in high Ca^{+2} ion medium may be due to enhancing the mitotic process, possibly by regulating other hormonal signalling functions as described by Hepler and Wayne (1985). Bairu et al. (2009) reported that elevated BAP levels increased the occurrence of shoot necrosis in *Harpagophytum procumbens.*

However, mitigation of shoot necrosis was possible in the present study even at a higher BAP (20.0 µM) concentration, when a Ca^{+2} rich medium was used. Plantlets recovered in the Ca^{+2} rich medium were successfully acclimatised and subjected to greenhouse conditions (Figure 3d). Androgeneis is a highly genotype dependent activity (Bhatia *et al.*, 2017). Since the modified protocol described above was developed for a Tall coconut variety, which is a cross pollinating variety, the successful application of adrogenesis in this variety will enable successful coconut breeding. When androgenesis protocols are being developed for other coconut varieties, the above mentioned facts could be considered to develop better protocols.

CONCLUSIONS

In conclusion, the present study demonstrated that manipulating different stages of androgenesis process in vitro could enhance the plantlet production up to a considerable level. Successful embryo production was achieved in sucrose and maltose when applied at 90.0 gL^{-1} or 120.0 gL^{-1}, respectively. Twenty micromolar (20 µM) BAP showed the best shoot regeneration in anther derived embryos. The increase of $CaCl_2$ concentration significantly affected alleviation of the problem of necrosis. Transfer of shoots at very early stages and continuous subculture of shoots into high $CaCl_2$ containing (4.0 mM) regeneration medium effectively reduced the occurrence of necrosis.

Conflicts of interest:

The authors have declared that there is no conflict of interest.

Acknowledgments

Financial assistance for the research was provided by the Coconut Research Institute, Sri Lanka through consolidate funds. Authors acknowledge Mrs. T.R. Gunathilake, Mr. E.S. Santha, Mrs. A. Keerthisinghe, and Ms. Nissansala Maduwanthie for the assistance given for culture maintenance. The authors also acknowledge Dr L.K. Weerakoon for reading the manuscript critically.

REFERENCES

Abdollahi M.R. & Rashidi S. (2018). Production and conversion of haploid embryos in chickpea (*Cicer arietinum* L.) anther cultures using high 2,4-D and silver nitrate containing media. *Plant Cell Tissue and Organ* Culture **133**: 39–49. DOI: https://doi.org/10.1007/s11240-017-1359-4

Agarwal S., Kanwar K. & Sharma D.R (2004). Factors affecting secondary somatic embryogenesis and embryo maturation

in *Morus alba* L. *Scientia Horticulturae* **102**: 359–368. DOI: https://doi.org/10.1016/j.scienta.2004.04.002

Ashok K.H.G. & Murthy H.N. (2004). Effects of sugars and amino acids on androgenesis of *Cucumis sativus* L. *Plant Cell Tissue and Organ Culture* **78**: 201–208. DOI: https://doi.org/10.1023/B:TICU.0000025637.56693.68

Bairu M.W., Jain N., Stirk W.A, Dolezal K. & Staden J.V. (2009). Solving the problem of shoot-tip necrosis in *Harpagophytum procumbens* by changing the cytokinin types, calcium and boron concentrations in the medium. *South African Journal of Botany* **75**: 122–127. DOI: https://doi.org/10.1016/j.sajb.2008.08.006

Bandupriya H.D.D., Femando S.C. & Vidhanaarachchi V.R.M. (2016). Micropropagation and androgenesis in coconut: an assessment of Sri Lankan implication. *Cocos* **22**: 31–47. DOI:https://doi.org/10.4038/cocos.v22i1.5810

Batugal P., Bourdeix R. & Baudouin L. (2009). Coconut breeding. In: Breeding Plantation Tree Crops: *Tropical Species* (eds. S.M. Jain & P.M. Priyadarshan), pp. 327–375. Springer, New York, USA. DOI:https://doi.org/10.1007/978-0-387-71201-7_10

Bhatia R., Dey S.S., Parkash C., Sharma K., Sood S. & Kumar R. (2018). Modification of important factors for efficient microspore embryogenesis and doubled haploid production in field grown white cabbage (*Brassica oleracea* var. capitata L.) genotypes in India. *Scientia Horticulturae* **233**: 178–187. DOI: https://doi.org/10.1016/j.scienta.2018.01.017

Bhatia R., Dey S.S., Sood S., Sharma K., Parkash C. & Kumar R. (2017). Efficient microspore embryogenesis in cauliflower (*Brassica oleracea* var. botrytis L.) for development of plants with different ploidy level and their use in breeding programme. *Scientia Horticulturae* **216**: 83–92. DOI: https://doi.org/10.1016/j.scienta.2016.12.020

Buter B. (1997). In *vitro* haploid production in maize. In: In *vitro Haploid Production in Higher Plants* (eds. S. Mohan Jain, S.K. Sopory & R.E. Veilleux), pp 37–71. Kluwer Academic Publishers, Dordrecht, The Netherlands. DOI: https://doi.org/10.1007/978-94-017-1862-2_2

Dunwell J. M. (1985). Anther and ovary culture. In: *Cereal Tissue and Cell Culture* (eds. S.W.J. Bright & M.G.K. Jones), pp. 1–44. Matinus Nijhoff/ Dr. W. Junk Publishers, Dortrecht, The Netherlands. DOI: https://doi.org/10.1007/978-94-009-5133-4_1

Eeuwens C.J. (1976). Mineral requirements for growth & callus initiation of tissue explants excised from mature coconut palm (*Cocos nucifera* L.) cultured in *vitro*. *Physiologia Plantarum* **36**: 23–28. DOI: https://doi.org/10.1111/j.1399-3054.1976.tb05022.x

Farahani F., Aminpoor H., Sheidai M., Noormohammadi Z. & Mazinani M.H. (2008). An improved system for in vitro propagation of banana (*Musa acuminate* L.) cultivars. *Asian Journal of Plant Sciences* **7**: 116–118. DOI:https://doi.org/10.3923/ajps.2008.116.118

Ferdous M.H., Billah A.A.M., Mehraj H., Taufiqu T. & Uddin A.F.M.J. (2015). BAP and IBA pulsing for in vitro multiplication of banana cultivars through shoot-tip culture. *Journal of Biosciences and Agricultural Research* **3**: 87–95. DOI:https://doi.org/10.18801/jbar.030215.35

Finnie S.J., Powell W. & Dyer A.F. (1989). The effect of carbohydrate composition and concentration on anther culture response in barley (*Hordeum vulgare* L.). *Plant Breeding* **103**: 110–118. DOI: https://doi.org/10.1111/j.1439-0523.1989.tb00358.x

Gubi J., Lajchova Z., Farago J. & Jurekova Z. (2004). Effect of growth regulators on shoot induction and plant regeneration in tomato (*Lycopersicon esculentum* Mill.). Biologia **59**: 405–408.

Guo Y.D. & Pulli S. (2000). Isolated microspore culture and plant regeneration in rye (*Secale cereale* L.). Plant *Cell Reports* **19**: 875–880. DOI: https://doi.org/10.1007/s002990000194

Hepler P.K. (2005). Calcium: a central regulator of plant growth and development. *Plant Cell* **17**: 2142–2155. DOI: DOI: https://doi.org/10.1105/tpc.105.032508

Hepler P.K. & Wayne R.O. (1985). Calcium and plant development. *Annual Review of Plant Physiology* **36**: 397–439. DOI:https://doi.org/10.1146/annurev.pp.36.060185.002145

Herbers K., Meuwly P., Frommer W.B., Métraux J.P. & Sonnewald U. (1996). Systemic acquired resistance mediated by the ectopic expression of invertase: possible hexose sensing in the secretory pathway. *Plant Cell* **8**: 793–798. DOI: https://doi.org/10.1105/tpc.8.5.793

Hirschi K.D. (2004). The calcium conundrum. Both versatile nutrient and specific signal. *Plant Physiology* **136**: 2438–2442. DOI: https://doi.org/10.1105/tpc.8.5.793

Jafari N., Othman R.F. & Khalid K. (2011). Effect of benzylaminopurine (BAP) pulsing on in vitro shoot multiplication of *Musa acuminata* (banana) cv. Berangan. *African Journal of Biotechnology* **10**: 2446–2450.

Kadota M. & Niimi Y. (2003). Effect of cytokinin types and their concentration on shoot proliferation and hyperhydricity in in *vitro* pear cultivar shoots. *Plant Cell Tissue and Organ Culture* **72**: 261–265. DOI: https://doi.org/10.1023/A:1022378511659

Kishore K., Patnaik S. & Shukla A. K. (2015). Optimization of method to alleviate in *vitro* shoot tip necrosis in Trichosantes dioica Roxb. *Indian Journal of Biotechnology* **14**: 107–111.

Klems M.J., Balla J., Machackova I., Eder J. & Prochazka S. (2000). The uptake and metabolism of H-3-benzylaminopurine in tobacco (*Nicotiana tabacum* L.) and cucumber (*Cucumis sativus* L.) explants. *Plant Growth Regulation* **31**: 135–142. DOI: https://doi.org/10.1023/A:1006374611228

Litz R.E. (1986). Effect of osmotic stress on embryogenesis in Carica suspension culture. *Journal of the American Society for Horticultural Science* **111**: 969–972.

Marsolais A.A. & Kasha K.J. (1985). The role of sucrose and auxin in a barley anther culture medium. *Canadian Journal*

of Botany **63**: 2209–2212.
DOI:https://doi.org/10.1139/b85-313

Martin K.P., Zang C.L., Slater A. & Madassery J. (2007). Control of shoot necrosis and plant death during micropropagation of banana and plantain (*Musa* spp.). *Plant Cell Tissue and Organ Culture* **88**: 51–59.
DOI: https://doi.org/10.1007/s11240-006-9177-0

Nguyen Q.T., Bandupriya H.D.D., Villalobo A.L., Sisunandar S., Foale M. & Adkins S.W. (2015). Tissue culture and associated biotechnological interventions for the improvement of coconut (*Cocos nucifera* L.): a review. Planta **242**: 1059–1076.
DOI: https://doi.org/10.1007/s00425-015-2362-9

Perera P.I.P., Hocher V., Verdeil J-L., Bandupriya H.D.D., Yakandawala D.M.Y. & Weerakoon L.K. (2008). Androgenic potential of coconut (*Cocos nucifera* L.). *Plant Cell Tissue and Organ Culture* **92**: 293–302.
DOI: https://doi.org/10.1007/s11240-008-9337-5

Perera P.I.P., Motha K.H. & Vidhanaarchchi V.R.M. (2020). Morphological and histological analysis of anther-derived embryos of coconut (*Cocos nucifera* L.). *Plant Cell Tissue and Organ Culture* **140**: 685–689.
DOI: https://doi.org/10.1007/s11240-019-01762-9

Perera P.I.P., Yakandawala D.M.Y., Hocher V., Verdeil J-L. & Weerakoon L.K. (2009). Effect of growth regulators on microspore embryogenesis in coconut anthers. *Plant Cell Tissue and Organ Culture* **96**: 171–180.
DOI: https://doi.org/10.1007/s11240-008-9473-y

Rajcan I., Boersma J.G. & Shaw E.J. (2011). Plant genetic techniques: plant breeder's toolbox. In: *Comprehensive Biotechnology*, second edition (ed. M.M. Young), pp. 133–147. Academic Press, USA.
DOI:// https doi.org/10.1016/B978-0-08-088504-9.00252-X

Scott P. & Lyne R.L. (1994). The effect of different carbohydrate sources upon the initiation of embryogenesis from barley microspores. *Plant Cell Tissue and Organ Culture* **36**: 129–133.
DOI: https://doi.org/10.1007/BF00048323

Stael S., Wurzinger B., Mair A., Mehlmer N., Vothknecht U.C. & Teige M. (2012). Plant organellar calcium signalling: an emerging field. *Journal of Experimental Botany* **63**: 1525–1542.
DOI: https://doi.org/10.1093/jxb/err394

Sunderland N. & Dunwell J.M. (1977). Anther and pollen culture In: *Plant Tissue and Cell Culture* (ed. H.E. Street), pp. 223–265, Botany Monographs volume 11. Blackwell Scientific Publications. Oxford, London.

Surakshitha N.C., Soorianathasundaram K., Ganga M. & Raveendran M. (2019). Alleviating shoot tip necrosis during *in vitro* propagation of grape cv. Red Globe. *Scientia Horticulturae* **248**: 118–125.
DOI: https://doi.org/10.1016/j.scienta.2019.01.013

Vieitez A.M., Sanchez C. & San-Jose C. (1989). Prevention of shoot tip necrosis in shoot cultures of chestnut and oak. *Scientia Horticultura*e **41**: 101–109.
DOI: https://doi.org/10.1016/0304-4238(89)90059-9

Vitova L., Stodulkova E., Bartonickova A. & Lipavska H. (2002). Mannitol utilisation by celery (Apium graveolens) plants grown under different conditions *in vitro*. *Plant Science* **163**: 907–916.
DOI: https://doi.org/10.1016/S0168-9452(02)00240-6

Yaseen M., Ahmad T., Sablok G., Standardi A. & Hafiz I.A. (2013). Role of carbon sources for in vitro plant growth and development. *Molecular Biology Reports* **40**: 2837–2849.
DOI: https://doi.org/10.1007/s11033-012-2299-z

J.Natn.Sci.Foundation Sri Lanka 2021 49 (3): 393 - 402
DOI: http://dx.doi.org/10.4038/jnsfsr.v49i3.10305

RESEARCH ARTICLE

In vitro culture of *Nymphaea nouchali* seeds; a conservation approach for a vulnerable species

PIP Perera[1], DPSTG Attanayaka[1], K Yakandawala[1], DMD Yakandawala[2], CN Fernando[1], UPE Arachchi[1], WPDS Gunawardhana[1] and WDCD Weerasinghe[1]

[1] *Faculty of Agriculture and Plantation Management, Wayamba University of Sri Lanka, Makandura, Gonawila.*
[2]*Department of Botany, Faculty of Science, University of Peradeniya, Peradeniya.*

Submitted: 25 November 2020; Revised:11 March 2021; Accepted: 01 April 2021

Abstract: *Nymphaea nouchali* var. *nouchali* is a non-viviparous species with a slow natural propagation by rhizomes or seeds. The plant is threatened in its natural habitats due to several reasons and is included in the vulnerable category of the National Red List of Sri Lanka. *In vitro* contamination-free culture method was developed to initiate mass propagation of the species. Results were validated through molecular and microscopic studies. Bacterial growth occurred in the seeds disinfected *via* the standard method using Clorox™. The mature seeds scarified with 75 % H_2SO_4 for 60 seconds gave contamination-free cultures with optimum seed germination. Scanning Electron micrographs of mature seeds showed the rows containing trichomes running between the two poles of seeds and the sclereids between the rows of trichomes to be the potential habitats for bacteria. Light micrographs showed the thick seed coat that causes a physical dormancy. Sulphuric acid treatment was effective in degrading the trichomes completely and the seed coat partially. The highest seed germination (65.5 %) was obtained with seeds cultured/ treated with 75 % H_2SO_4 on the solidified MS medium. The basal stem of the well-grown seedlings *in vitro* gave rise to mini rhizomes. Molecular analysis showed the close genetic relatedness within and among the isolated plant populations where the seeds were collected. The *in vitro* protocol developed in this study can be used for propagation of seedlings of this vulnerable species for maintaining the biodiversity by population enhancement through restoration and introduction into new habitats.

Keywords: Biodiversity, *in vitro* seed culture, *Nymphaea nouchali*, seed dormancy, seed morphology, sulphuric acid scarification.

INTRODUCTION

The genus *Nymphaea* L., (Nymphaeaceae) commonly known as water lilies are aquatic plants distributed in tropics and temperate regions. The genus harbours around 40–50 species and they are phenotypically diverse, mostly due to varying hydrological and edaphic conditions in their surroundings (Polina & Alexy, 2007). These species can be either annuals or perennials with their rhizomes anchored in mud and grown in open water bodies. The water lilies are one of the most eye catching groups of aquatic plants with year-round flowering (Guruge *et al*., 2016). Thus, they have attracted the attention of the botanists, horticulturists and plant enthusiasts.

Three *Nymphaea* species have been recorded in Sri Lanka viz., *N. pubescens* Willd., *N. rubra* Roxb. ex Andrews and *N. nouchali* Burm. f. (synonym *N. stellata* Willd.) (Dassanayake, 1996; Guruge *et al*., 2016). Recent taxonomic studies have revealed the occurrence of two varieties of *Nymphaea nouchali* Burm. f. in Sri Lanka, as *N. nouchali* var. *nouchali* and *N. nouchali* var. *versicolour* (Sims) (Guruge *et al*., 2017). *Nymphaea nouchali* var. *nouchali* (hereafter mention as *Nymphaea nouchali*) is native to southern and eastern parts of Asia (Dassanayake, 1996).

* Corresponding author (prasanthi@wyb.ac.lk https://orcid.org/:0000-0003-3302-2778)

Nymphaea nouchali has economical, pharmaceutical, social and cultural values. All parts of the plant (seeds, tender leaves, flowers, flower peduncles, and rhizomes) are rich in starch and are edible (Raja *et al*., 2010). Because of the medicinal properties including antibacterial, antioxidant, antimicrobial and anti-diabetic activities (Raja *et al*., 2010; Dash, 2013; Lim, 2014) they are widely used in Ayurvedic medicine (Jayaweera, 1982). The social importance of the plant is evident by selecting the blue-flowered variety of *N. nouchali* as the National Flower of Sri Lanka. The plant is populated in aquatic landscapes.

Nymphaea nouchali is a non-viviparous plant having a comparatively low propagation rate vegetatively *via* rhizomes and sexually by seeds (Yakandawala & Yakandawala, 2011). It is grown in the lowland waterbodies of wet and dry zones of Sri Lanka (Dassanayake, 1996; Guruge *et al*., 2016). The plant is now threatened in its natural habitats due to anthropogenic activities, drying of waterbodies and the competition posed by the viviparous *Nymphaea* × *erangae for* habitats, resulting in whipping off some populations completely (Yakandawala & Yakandawala, 2011; Yakandawala *et al*., 2017). In addition, the compatibility of *Nymphaea* × *erangae* for hybridization has also posed an additional threat to the genetic purity of native *N. nouchali* (Yakandawala & Yakandawala, 2011).

Evidence for natural hybridization among *Nymphaea* species has been reported elsewhere in the world, where different karyotypes ranging from 2n = 28 to 84 has been observed, generating uncertainty in species identification (Raja *et al*., 2010; Nierbauer *et al*., 2014). This creates problems for conservation of the species. Hybridization of natives with alien species has been reported to support the evolution of invasiveness and is identified as a major threat to the extinction of native species, where the island flora is thought to be more vulnerable (Veitch & Clout, 2002; Ellstrand & Schierenbeck, 2006; Yakandawala & Yakandawala, 2011). Due to these reasons the native *N. nouchali* has been recognised under the vulnerable category in the National Red List (MOE, 2012), highlighting the importance of conserving the pure populations.

Mass propagation of *N. nouchali* through tissue culture is a promising method to overcome the limitations associated with conserving this vulnerable species. *In vitro* cultures of *Nymphaea* have been reported in different species; *Nymphaea* 'Daubeniana', a highly viviparous species, using leaves (Jenks *et al*., 1990), *Nymphaea* hybrid 'James Brydon', using rhizome tips (Lakshmanan, 1994), and *N. alba,* using seeds (Sumlu *et al*., 2010) in different countries. Fernando *et al.* (2016) reported heavy contamination rates occurred in the leaf explants of *Nymphaea* at different maturity stages. Our attempts made for inducing somatic embryogenesis in the leaf explants of *N. nouchali* were not successful in demonstrating the recalcitrance of this non-viviparous species.

Among the *in vitro* culture methods, seeds have a direct application for the conservation of endemic or threatened plant species. Thus, in the present study emphasis was given to establish the contamination free culture protocol with the aim of mass propagating this vulnerable species for conservation purpose. Molecular and microscopic evidence were used for validating the results obtained in this research.

METHODOLOGY

Plant materials

Mature pods of *N. nouchali* were collected from the plant populations existing in natural habitats of marshy lands in Sri Lanka, after careful observation for the unique morphological characters in the petals, stamens, stigmatic heads, and leaf lamina of the species, as described by Yakandawala & Yakandawala (2011) and Guruge *et al*. (2017). The seeds obtained from the burst opened pods were soaked for three days to facilitate sedimentation of seeds by fermenting the outer pulp. After removing the debris, the seeds were isolated and washed thoroughly with soap water, followed by tap water for 30 min.

Basal culture medium

Murashige and Skoog medium (1962), supplemented with 20.0 g/L sucrose and 100.0 mg/L Myo-Inositol (w/v), was used as the basal culture medium unless otherwise mentioned. The pH was adjusted to 5.8. Solidified media with agar (6.0 g/L; w/v) were used unless otherwise mentioned. Media were sterilised by autoclaving at 121 ˚C, 15 Pa for 15 min (HVP 50, Hirayama, Saitama, Japan). Petri plates (90 × 10 mm) containing 25 mL of culture medium were used for inoculating the seeds. The cultures were maintained at 28 ℃ in light supplied with CFL (6500 K) with a photoperiod of 16/8 h.

The seeds were sterilised under the laminar flow cabinet and treated with different chemicals mentioned in the following experiments.

Effect of concentrations and duration of Clorox™

Seeds treated with 70% ethanol for 1 min were subjected to ten sterilisation protocols with five Clorox™ concentrations; 10, 30, 50, 70 or 100 % (v/v), each with two exposure times, 10- or 20-min. Experiment was repeated three times. A total of 375 seeds were used for each treatment.

Effect of concentrations and duration of H_2SO_4

After disinfecting with 70 % ethanol for 1 min and 10 % Clorox™ for 10 min, seeds were treated with five concentrations of H_2SO_4, 10, 25, 50, 75, and 100 % (v/v) for five exposure times 15, 30, 60, 90, 180 s. Untreated seeds were used as the control. Experiment was repeated three times and a total of 450 seeds were used for each treatment.

Sixty microliters of Tween were used as the surfactant in the Clorox™ solution. After treating with each disinfectant, the seeds were washed thoroughly with distilled water three times by agitating each for 2 min.

Optimisation of culture conditions for seed germination

Germination of the seeds treated with acid was further tested by culturing them onto three different media. After disinfecting the seeds with 70 % ethanol for 1 min followed by 10 % Clorox™ for 10 min as mentioned above, the seeds were treated with 25, 50 and 75 % (v/v) H_2SO_4 for 1 min. MS basal medium supplemented with 0.5 mg/L 2, 4-D and 2 mg/L BAP solidified with 6.0 g/L agar, the same medium in liquid form and the Albert's solution (2.22 g/L), were tested for seed germination. The responses were compared with the untreated seeds. The experiment was repeated three times and a total of 360 seeds were used for each treatment.

Data analysis

Factorial experiment with completely randomised design (CRD) was used. Observations were made using a stereo microscope in weekly intervals. The data were analysed using SAS statistical package (SAS, 1999). Chi-square or maximum likelihood ANOVA was conducted using the Proc CatMod procedures of PC-SAS to analyse counted data. Treatment means were compared using SE, 95 % confidence intervals or orthogonal contrast coefficients where appropriate (Compton, 1994).

Microscopic analysis

Freshly isolated, acid treated, contaminated and germinated seeds were fixed separately in FAA (50 % ethanol + 1 0% formaldehyde + glacial acetic acid, 18:1:1) for 72 h and dehydrated with a graded ethanol series, 30, 50 and 70 % (v/v) for 2 h and stored in 70 % ethanol, until morphological and histological observations were made. Microscopic analysis was done for a sample of 10 seeds from each category.

Morphological observations were made using a scanning electron microscope (SEM; Carl Zeiss EVO LS 10, Germany). For the histological analysis, the samples were further dehydrated with 90, 95 and 100 % ethanol and embedded in paraffin wax. Sections (4–5 µm) were obtained using a rotary microtome (Microteck, Germany) and stained with 5 % Napthol Blue Black (Sigma-Aldrich, Lyon, France) for 5 min at 60 ^{0}C. Slides were observed under the light microscope (Optika B-500, Ponteranica, Italy) and photographed (Optika B-500, Ponteranica, Italy).

Genetic Diversity analysis

Samples collected from three locations; Chilaw, Puttalam, and Kurunegala, were analysed using Random Amplified Polymorphic DNA (RAPD) markers to assess the genetic variation of, within, and among isolated populations of *N. nouchali*. The RAPD technique was selected to assess the three populations due to the characteristics of the test; low-cost, rapidity, simplicity, quantity of DNA required, use of universal primers that work in any genome, high potential to detect polymorphism (Goulart *et al*., 2005), and non-availability of genetic information of the species. The standard CTAB method was used with several modifications to extract the Genomic DNA from leaves (Doyle & Doyle, 1990; Priya *et al*., 2017). Ten primers from the series of OPK (2, 4, 5, 7, 8, 10, 11, 12, 14, 16, 18, and 20) were initially tested and five were selected (4, 8, 10, 14, 18) based on the polymorphism among the samples.

Five primers (Table 1) were used, and each primer was repeated three times. PCR Amplification was carried out in volumes of 20 µL containing 50–100 ng of genomic DNA, 200 µM each of dNTPs, 2.5 µM $MgCl_2$, 1X buffer and 2.5 units of GoTaq Polymerase (Promega Technologies, USA) and 16.5 ng/µL of random 10-mer primer (Operon Technologies, California) with a Simpli AmpTM Thermal Cycler (Applied Biosystems, USA) with an initial denaturation step at 94 °C for 1 min, 45 cycles at 94 °C for 1 min, 36 °C for 1 min and 72 °C for 2 min. A final extension step at 72 °C for 5 min was also included. The RAPD products were subjected to gel electrophoresis on 1.2 % agarose gel with a ladder of 1 kb DNA (BIORIN GmbH) and photographed using gel documentation apparatus (QUANTUM CX5, GmbH). RAPD assays for each primer were repeated three times and bands were scored as presence (1) or absence (0) for each primer across all genotypes. The estimated genetic distances between populations and a dendrogram were obtained using dominant diploid data from the program POPGENE, version 1.32, according to the method as per Nei (1972, 1978) based on unweighted pair group method with arithmetic mean (UPGMA) algorithm (Yeh, 1999).

RESULTS AND DISCUSSION

Being an aquatic plant, contamination is the most critical challenge in any *in vitro* culture method. The explants collected from *N. nouchali* grown in marshy habitats naturally contain endogenous contaminants apart from heavy population of exogenous ones. Fernando *et al.* (2016) reported that the main contaminants in different leaf stages of *Nymphaea* are different; fungi in mature leaf stages and bacteria in immature leaf stages limit establishing *in vitro* cultures. Seeds of *N. nouchali* are enclosed in the pods; however, mature pods collected for culture initiation are partially dehisced at the time of explant collection, exposing the seeds to the heavily contaminated water in the marshy lands. Furthermore, the seed extraction protocol *via* fermentation may also cause the accumulation of heavy bacterial populations on the seed coat. Therefore, establishment of contamination free seed cultures is an essential step to be optimised, aiming at mass propagation technique for the species.

Effect of sterilisation agent for initiating contamination free seed cultures

Effect of Clorox™

In the factorial experiment, an interaction effect was not observed in tested factors of Clorox™ concentration and the exposed time duration. Therefore, they were considered as major factors. Clorox™ concentration (χ^2 =130.36; $p < 0.0001$) and the duration ($\chi^2 = 4.7$; $p < 0.0302$) were effective in controlling the contamination (Figure 1). The lowest contamination rate (39.2 %) was observed in seeds sterilised with 100 % Clorox™. It indicated the less effectiveness of the concentrated Clorox™ in disinfection.

Tested concentrations affected significantly on the germination of both contaminated and non-contaminated seeds ($\chi^2 = 62.01$; $p < 0.0001$ and $\chi^2 = 56.05$; $p < 0.0001$, respectively). However, the seed germination was significantly lower in 100 % Clorox™ (13.9 %; $\chi^2 = 12.15$; $p < 0.0005$) compared to 50 % Clorox™. Longer duration (20 min) for Clorox™ was effective in seed sterilisation (47.7%; $\chi^2 = 4.7$; $p < 0.0302$) but negatively affected on seed germination (2.03 %; $\chi^2 = 13.02$; $p < 0.0003$), compared to the 10 min duration.

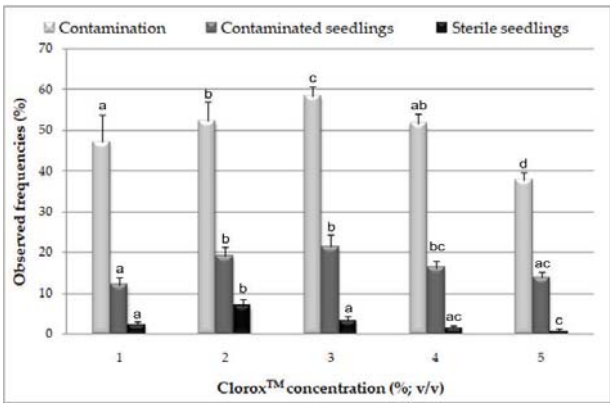


Figure 1: Effect of Clorox™ concentration on sterilization of *Nymphaea nouchali* var. nouchali seeds. Means with different letters are significantly different at $p < 0.05$ at 95 % confidence level; n = 750

Results indicated that 70 % ethanol and Clorox™ at any concentration is not sufficient for disinfection of seeds, leaving some bacteria in the hidden habitats of the seed surface. After culturing the seeds into the nutrient medium, remaining microbes were multiplied and subsequently contaminated the cultured seeds. Results further revealed that the germination frequency is higher in contaminated seeds than non-contaminated ones. The bacteria available on the surface of contaminated seeds may act on degradation of the hard seed coat to remove the physical dormancy, thus triggering their germination. The dormancy imposed by the physical barrier of the seed coat limits the gas exchange and passage of water (Penfield, 2017), limiting the germination. Thick seed coat in *N. nouchali* may contribute to the

morphophysical dormancy (Dalziell *et al*., 2018) that can be removed by bacterial degradation under natural conditions. Seed dormancy has been reported in genus *Nymphaea* (Smits *et al*., 1995; Dalziell *et al*., 2018). Sumlu *et al*. (2010) reported that seed germination was observed only in the scarified seeds of *N. alba*, with hard seed coat by cutting it mechanically. In the present study a few seedlings obtained from non-contaminated seeds may be due to the removal of dormancy through the scarification occurred in the seed preparation process. Therefore, attempts were made in this study to test the hypothesis by scarifying the seeds with a strong acid.

Effect of H_2SO_4

Sulphuric acid was effective in establishing the contamination free cultures. Interaction effect was not observed among the tested factors of acid concentration and the exposed time duration to the acid. Therefore, they were considered as major factors. The H_2SO_4 concentrations tested were effective in disinfection (χ^2 = 365.10; $p < 0.0001$), germination (χ^2 = 68.57; $p < 0.0001$) and gave rise to the non-contaminated seedlings (χ^2 = 125.03; $p < 0.0001$) (Figure 2a). Among the tested concentrations, a complete disinfection was observed in seeds treated with 50–100 % acid. The concentrations of 10 and 25 % acid, reduced the contamination rate over the control of untreated.

The greatest germination (6.3 %) was observed in the seeds treated with 75 % acid that gave 100 % disinfection. Among the seedlings derived from the seeds treated with 10 % and 25 % acid, both contaminated and non-contaminated seedlings were observed. Although 100 % H_2SO_4 was effective in complete disinfection of the seeds, none of them germinated indicating the loss of seed viability in the concentrated acid. Observations revealed that treatment with 75 % acid was effective on both purposes, i.e. complete eradication of contaminants and to remove the physical seed dormancy through scarification.

The duration that the seeds were exposed to the acid was also affected on disinfection (χ^2 = 148.17; $p < 0.0001$), germination (χ^2 = 12.66; $p < 0.0131$) and giving rise to the non-contaminated seedlings (χ^2 = 10.89; $p < 0.0279$) (Figure 2b). The lowest contamination rate, the greatest germination rate and the greatest frequency of the non-contaminated seedlings among the germinated were observed in the seeds treated for 60 and 180 seconds. Based on this observation, acid treatment for 60 seconds duration was selected as the optimum.

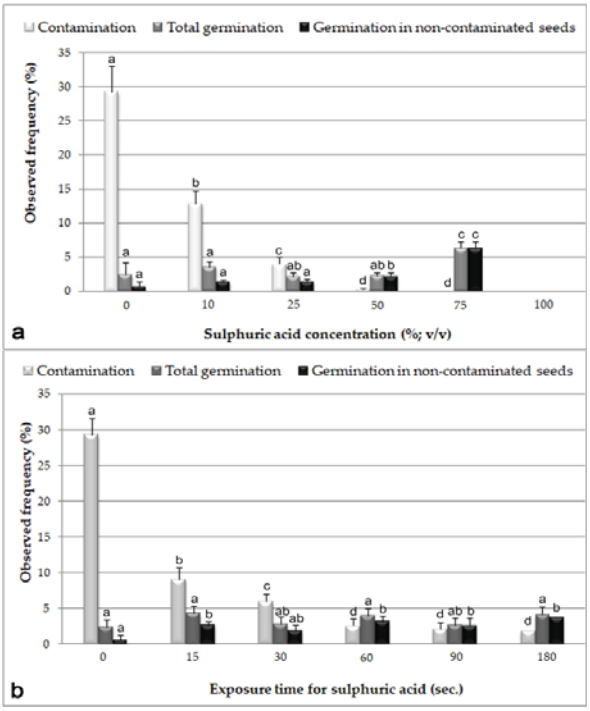


Figure 2: Effect of H_2SO_4 on establishing the contamination free cultures and germination of Nymphea nouchali var. nouchali seeds (a) acid concentration; control, without acid treatment n = 450; other treatment n = 2250; (b) duration n = 1800; means with different letters are significantly different at $p < 0.05$ at 95% confidence level

Sulphuric acid has been used in sterilisation protocols very rarely in other crops; 1 N acid for 1 minute in *Citrus sinensis* buds (Niedz *et al*., 2002) and undiluted acid for 1 minute in *Phaseolus vulgar* seeds (Malik & Saxena, 1991). However, this is the first report of using it in sterilisation of the seeds of genus *Nymphaea*.

Higher germination frequency resulted with elevated concentrations of acid revealed that the scarification is effective with high concentrations, indicating the loss of morphological or physical dormancy. Use of H_2SO_4 for scarification has also been reported in other aquatic seeds viz., *Astragalus vulnerariae*, where 40 % H_2SO_4 was used for 15 minutes (Dilaver *et al*., 2017). Zero germination observed in the seeds treated with 100 % acid indicates the loss of viability by exposing the embryo to the concentrated acid through the cracks made by extensively degraded seed coat. The higher germination rate observed in the seeds treated with 75 % acid revealed that the scarification made at this concentration, be the optimum level for weakening the seed coat to facilitate permeability, and thereby to eliminate the physical dormancy. Weakened or cracked seed coat acts as the site for water entry as reported in

other crops viz., *Stylosanthes humilis* (Chaves *et al*., 2017) and *Aspalathus linearis* (Kelly & Van Staden, 1985). The lower contamination rates observed in 50 % and 25 % acids revealed that the concentrations were detrimental to the bacteria though the trichomes were not degraded.

Optimisation of culture conditions for seed germination and seedling growth

The factorial experiment conducted for further testing the effect of H_2SO_4 concentrations (25, 50 and 75 %) on seed germination in different media revealed the occurrence of interaction ($\chi^2 = 608.41$; $p < 0.0001$). The greatest germination (65.5 %) was obtained by culturing the seeds treated with 75 % H_2SO_4 onto the solidified MS medium (Figure 3).

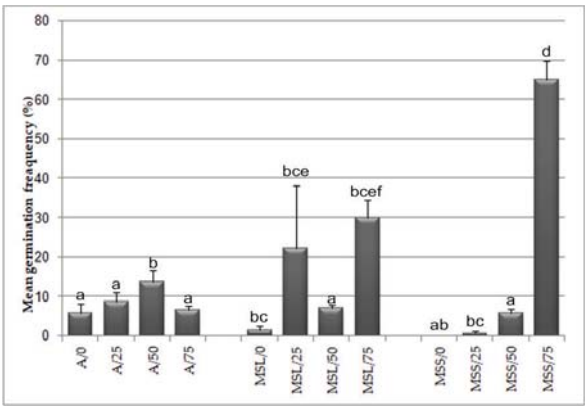


Figure 3: Effect of culture medium and sulphuric acid concentration on seed germination of *Nymphaea nouchali* var. *nouchali*. Seeds treated with different concentrations of sulphuric acid (0, 25, 50, 75 %) were cultured on to the Albert solution (A) and MS liquid (MSL) and solidified (MSS) media. Means with different letters are significantly different at $p < 0.05$ at 95 % confidence level. n= 360

The plants derived from the acid treated seeds showed a normal growth containing a well-grown shoot and the roots (Figure 4a, c) indicating that acid scarification had no adverse effect. The basal stem of the well-grown seedlings gave rise to mini rhizomes. Bacteria was the contaminant in all cultures. Milky jelly-appearance of the contamination suggested the presence of bacterial contamination (Figure 4b).

The results revealed that the physical state of the culture medium and the medium composition has a significant effect on seed germination. Although *N. nouchali* is an aquatic plant, solidified medium favoured for seed germination. Reason could be the proper gas exchange in the scarified seeds caused for better germination. The composition of the medium, i.e. MS basal medium supplemented with 0.5 mg/L 2, 4-D and 2.0 mg/L BAP, was effective in giving rise to well-grown complete seedlings, containing both shoots and roots with a higher germination rate (≈ 66 %). Seed is a highly available explant material in *N. nouchali* where each pod contains thousands of seeds. Therefore, mass scale plant production is possible through the protocol developed in the study. *In vitro*-derived contamination free seedlings can be further used for producing clones through micropropagation. The mini rhizomes formed in the well-grown seedlings would also serve as an additional source of explant for *in vitro* propagation of this species.

Microscopic analysis

The scanning electron micrographs (SEM) clearly showed the trichomes and sclereids available on the seed coat in fresh seeds. Rows of long trichomes parallel to the long axis of the seed (Figure 4d), and the sclereids arranged in regular transverse rows (Figure 4e) in between the rows are the characteristic features of the *N. nauchali* seeds. This observation is comparable with some species of Mexican *Nymphaea* including *N. ampla*, *N. elegans* and *N. gracilis*, etc. (Bonilla-Barbosa *et al*. 2000), as it provides information on the diversity of the genus *Nymphaea*. The seed coat morphology also demonstrated the reason for the occurrence of heavy contamination in the cultured seeds *in vitro*. The trichomes and the grooved sclereids facilitate safe harbouring of bacteria during sterilisation process causing heavy bacterial contamination after culture initiation. Concentrated H_2SO_4 and 75 % acid degraded those trichomes completely (Figure 4f and 4g) and caused total eradication of microbes from the cultures as shown in Figure 2a, thus eliminating the occurrence of contamination. The 50–10 % acid was not strong enough to degrade the trichomes. Although 50 % acid did not degrade the trichomes, it may be adequately strong to kill the microbes completely, whereas the other two concentrations (10 % and 25 %) were not as effective as such. *In vitro* seed propagation was reported only in some *Nymphaea* species; *N. alba* (Sumlu *et al*., 2010; Latowski *et al*., 2014) and *N. lotus var. thermalis* (Blidar *et al*., 2019), which have no trichomes on the seed coats.

Evidence also provided through light micrographs to demonstrate the effect of acid treatment in seed culture. The thick seed coat in fresh seeds (Figure 4h) may cause physical seed dormancy. In the seeds treated

with concentrated acid, the cracks were observed in the seed coat (Figure 4i and 4j). It may allow moving the strong acid (100 %) into the embryo, leading to loss of seed viability, where none of the seeds were germinated as mentioned above. In the seeds treated with 75 % acid, cracks were not visible, thus maintained the seed viability. Degradation of the seed coat occurred to different degrees, with 75–50 % acid concentration causing germination of seeds by removing physical dormancy. In the untreated seeds, heavy bacterial action degraded the seed coat (Figure 4k) to the same degree of different acid treatments, allowing seed germination.

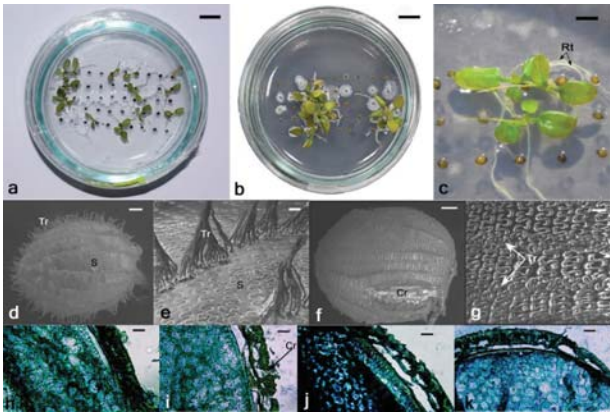


Figure 4: Morphological and microscopic aspects of seed culture of Nymphea nouchali var. nouchali. (a-c) Morphological aspect seed germination: (a) non-contaminated (bar = 1 cm) and (b) contaminated seeds (bar = 1 cm); (c) healthy seedlings derived from acid treated seeds (bar = 1.4 mm). (d-g) Scanning electron micrographs; (d) a fresh seed (bar = 100 µm); (e) close view of the seed coat. Note the trichomes (Tr) arranged in lines and sclereids (s) in between those lines (bar = 20 µm). (f) A seed treated with 100 % acid. Note the cracked seed coat (bar = 100 µm) and (g) close view of the seed coat after treating with 100 % acid. Note the complete degradation of the trichomes. (h-k) Histological aspect the testa (T) of the seeds: (h) a fresh seed (bar = 200 µm); (i) a seed treated with 100 % (bar = 200 µm); (j) seed treated with 75 % acid (bar = 200 µm); (k) a contaminated seed (bar = 200 µm)

Genetic diversity of isolated plant populations

Although the study showed that seed culturing is a promising technique for *in vitro* propagation of *N. nouchali*, unawareness of the level of genetic purity of the seedling populations is a barrier for implementing the technique for conservation of true-to-type species. RAPD technique has been used for similar genetic studies for several species such as *Ensete ventricosum* (Birmeta *et al*., 2004), *Cedrus atlantica* (Renau-Morata *et al*., 2005; Mendonça *et al*, 2014) and *Hevea brasilensis* (Nakkanong *et al*., 2008; Liyanage *et al*., 2014).

A total of 56 bands were generated with an average frequency of 11.2 bands per primer (Figure 5a). The genetic distance matrix obtained for individuals of three populations showed that the genetic distance among nine genotypes collected from three locations ranged from low to moderate values of 0.0364 to 0.5596 with an average of 0.3404 (Table 2).

This indicates a closer genetic relatedness between the populations. As estimated from the distant matrix, the lowest within-population genetic variation with an average of 0.26 was found in the Chilaw population compared to the 0.30 and 0.31 of the Kurunegala and Puttalam populations, respectively. UPGMA clustering grouped the nine genotypes into two major clusters (A and B; Figure 5b). However, the results indicated high genetic relatedness within the plant populations.

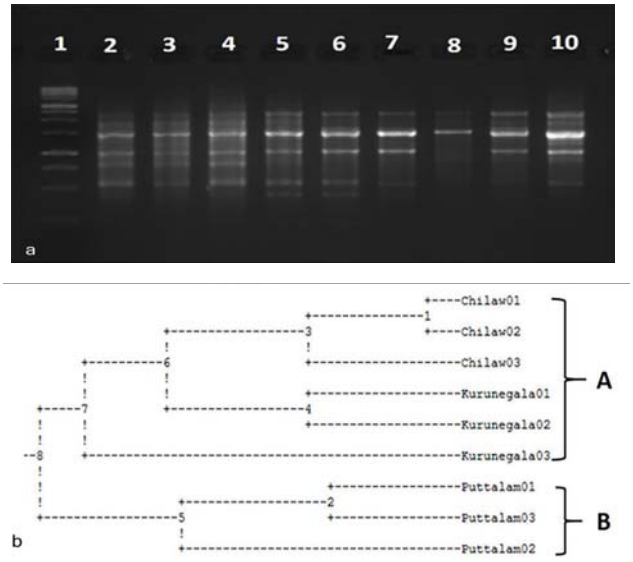


Figure 5: Assessment of genetic diversity in isolated *Nymphaea nouchali var. nouchali* plant populations using randomly amplified polymorphic DNA (RAPD). (a) RAPD profile generated for primer OPK 04. Lane 1 - 1kb ladder; Lane 2, 3, 4 - genotypes from Chilaw; Lane 5, 6, 7 - Genotypes from Kurunegala; lane 8, 9, 10 - Genotypes from Puttalam. (b) UPGMA dendogram showing the relationship among nine genotypes. Note the two clusters generated A and B

Table 1. Banding pattern details of the 10-mer random primers

Primer	Sequence	Total number of bands observed	Number of polymorphic bands	Polymorphism %
OPK 04	5- CCGCCCAAAC-3	10	6	60.00
OPK 08	5- GAACACTGGG-3	9	4	44.44
OPK 10	5- GTGCAACGTG-3	14	9	64.29
OPK 14	5- CCCGCTACAC-3	11	4	36.36
OPK 18	5- CCTAGTCGAG-3	12	7	58.33

Table 2. Genetic similarity/distances among the nine genotypes of *Nymphaea nouchali*

Genotype	1	2	3	4	5	6	7	8	9
Chilaw-01 (1)	***	0.96	0.86	0.80	0.70	0.66	0.75	0.63	0.73
Chilaw-02 (2)	0.04	***	0.86	0.80	0.66	0.66	0.71	0.59	0.73
Chilaw-03 (3)	0.15	0.15	***	0.84	0.70	0.70	0.71	0.59	0.70
Kurunegala-01 (4)	0.22	0.22	0.18	***	0.86	0.71	0.73	0.57	0.64
Kurunegala-02 (5)	0.36	0.41	0.36	0.15	***	0.71	0.73	0.64	0.68
Kurunegala-03 (6)	0.41	0.41	0.36	0.34	0.34	***	0.59	0.61	0.61
Puttalam-01 (7)	0.29	0.34	0.34	0.31	0.31	0.53	***	0.73	0.88
Puttalam-02 (8)	0.47	0.53	0.53	0.56	0.44	0.50	0.31	***	0.79
Puttalam-03 (9)	0.31	0.31	0.36	0.44	0.39	0.50	0.13	0.24	***

Nei's genetic identity (above diagonal) and genetic distance (below diagonal) are shown

The study revealed that a stable level of genetic uniformity existed within and among the three *N. nouchali* populations, suggesting the non-occurrence of gene contamination in these locations by undesirable alien alleles introduced by outcrossing. The close genetic relationship of these populations also suggests that the heterogeneity expected of the *in vitro* seedlings from these populations to be low, hence showing the possibility of using the technique for mass propagation from seeds collected from the isolated habitats.

CONCLUSIONS

Seed morphology is identified as the cause for heavy culture contamination and dormancy of the seeds. The higher germination rate observed in the seeds treated with 75 % acid for 60 seconds revealed that the treatment is optimal for elimination of the microbes by degrading the trichomes. Furthermore, it was effective in eliminating the physical dormancy by partially degrading the seed coat. The protocol developed in this study can be used for mass propagation of *N. nouchali*. As future prospectus, the explants collected from the *in vitro* seedlings can be used for developing clonal propagation techniques through somatic embryogenesis or micropropagation of the species, which lacks any specialised organs for vegetative reproduction in mass scale. The output of the study would help in habitat restoration and establishing new populations enabling this vulnerable species to conserve for future generations. Further, the mass propagation of seed-derived plants could also be used to supply the demand for ornamental plant industry. This is the first report of using H_2SO_4 as sterilisation agent of the genus *Nymphaea* that enabled successful propagation through seeds *in vitro*.

Acknowledgments

Authors gratefully acknowledge the financial support given by the National Research Council, Sri Lanka (Grant Number 15-013) and Wayamba University of Sri Lanka through the grants SRHDC/RP/04/15/01 and SRHDC/RP/04/18/01. The authors thank Mr. J.D.J.S. Kularatne and Prof. R. Abeynayake for the assistance in statistical analysis and Ms. A. Fernando for the technical support.

Conflicts of Interest

The authors declare no conflict of interest.

REFERENCES

Birmeta G., Nybom H. & Bekele E. (2004). Distinction between wild and cultivated enset (*Ensete ventricosum*) gene pools in Ethiopia using RAPD markers. *Heriditas* **140**(2): 139–148. DOI: https://doi.org/10.1111/j.1601-5223.2004.01792.x

Blidar C.F., Tripon I.M. & Ilea C. (2019). In vitro conservation of genetic resources of *Nymphaea lotus var. thermalis* (DC.) Tuzs., an endangered plant species. *Romanian Biotechnological Letters* **24**(3): 448–457. DOI: https://doi.org/10.25083/rbl/24.3/448.457

Bonilla-Barbosa J., Novelo A., Orozco Y.H. & Márquez-Guzmán J. (2000). Comparative seed morphology of Mexican *Nymphaea species*. *Aquatic Botany* **68**(3): 189–204. DOI: https://doi.org/10.1016/S0304-3770(00)00125-X

Chaves I.S., Silva N.C.Q. & Ribeiro D.M. (2017). Effect of the seed coat on dormancy and germination in *Stylosanthes humilis* H.B.K. seeds. *Journal of Seed Science* **39**(2): 114–122. DOI: https://doi.org/10.1590/2317-1545v39n2167773

Compton M.E. (1994). Statistical methods suitable for the analysis of plant tissue culture data. *Plant Cell Tissue and Organ Culture* **37**: 217–242. DOI: https://doi.org/10.1007/BF00042336

Dalziell E.L., Baskin C.C., Baskin J.M., Young R.E., Dixon K.W. & Merritt D.J. (2018). Morphophysiological dormancy in the basal angiosperm order Nymphaeales. Annals of Botany 123(1): 95–106. DOI: https://doi.org/10.1093/aob/mcy142

Dash B.K., Kumer S.M., Alam K., Hossain K., Islam R., Banu N.A., Rahman S. & Jamal A.H.M. (2013). Antibacterial activity of *Nymphaea nouchali* (Burm. f) Flower. *Annals of Clinical Microbiology and Antimicrobials* **12**(1): 27. DOI: https://doi.org/10.1186/1476-0711-12-27

Dassanayake M.D. (1996). Nymphaeaceae. In: *A Revised Handbook to the Flora of Ceylon* (eds. M.D. Dassanayake, W.D. Clayton). Oxford & IBH Publ. Co. Pvt., Ltd., New Delhi, India.

Dilaver Z., Mirzapour M. & Kendir H. (2017). Breaking seed dormancy and micropropagation of perennial vulneraria milkvetch (*Astragalus vulnerariae* DC.). *Acta Scientiarum Polonorum-hortorum Cultus* **16**(4): 79–88. DOI: https://doi.org/10.24326/asphc.2017.4.9

Doyle J.J. & Doyle J.L. (1990). Isolation of plant DNA from fresh tissue. *Focus* **12**: 13–15.

Ellstrand N.C. & Schierenbeck K.A. (2006). Hybridization as a stimulus for the evolution of invasiveness in plants. *Euphytica* **148**: 35–46. DOI: https://doi.org/10.1073/pnas.97.13.7043

Fernando C.N., Yakandawala K., Attanayaka D.P.S.T.G. & Perera P.I.P. (2016). Optimization of the sterilization protocol and the leaf stage for minimizing the contamination risk in cultured violet flowered *Nymphaea leaves*. *Transactions of Persatuan Genetik Malaysia* **3**: 103–106.

Goulart M.F., Ribeiro S.P. & Lovato M.B. (2005). Genetic, morphological and spatial characterization of two populations of *Mabea fistulifera* Mart. (Euphorbiaceae), in different successional stages. *Brazilian Archives of Biology And Technology* **48**(2): 275–284. DOI: https://doi.org/10.1590/S1516-89132005000200015

Guruge D.P.G.S.K., Yakandawala D. & Yakandawala K. (2016). Confirming the identity of newly recorded *Nymphaea rubra roxb.* ex andrews discerning from *Nymphaea pubescens* willd. using morphometrics and molecular sequence analyses. *Bangladesh Journal of Plant Taxonomy* **23**(2): 107–117. DOI: https://doi.org/10.3329/bjpt.v23i2.30819

Guruge S., Yakandawala D. & Yakandawala K. (2017). A taxonomic synopsis of *Nymphaea nouchali* Burm. f. and infraspecific taxa. *Journal of the National Science Foundation of Sri Lanka* **45**(3): 307–318. DOI: https://doi.org/10.4038/jnsfsr.v45i3.8194

Jayaweera D.M.A. (1982). *Medicinal Plants Used in Ceylon.* National Science Council of Sri Lanka Publications, Colombo

Jenks M.A., Kane M.E., Marousky F., McConnel D. & Sheehan T. (1990). In vitro establishment and epiphyllous plantlet regeneration of *Nympheae* 'doubeniana'. *HortScience* **25**(12): 1664. DOI: https://doi.org/10.21273/HORTSCI.25.12.1664

Jenks M.A., Kane M.E. & Mc Connell D.B. (2000). Shoot organogenesis from petiole explants in the aquatic plant *Nymphoides indica*. *Plant Cell Tissue and Organ Culture* **63**: 1–8. DOI: https://doi.org/10.1023/A:1006471027372

Kelly K.M. & Van Staden J. (1985). Effect of acid scarification on seed coat structure, germination and seedling vigour of Aspalathus linearis. *Journal of Plant Physiology* **121**(1): 37–45. DOI: https://doi.org/10.1016/S0176-1617(85)80089-4

Lakshmanan P. (1994) In vitro establishment and multiplication of *Nymphaea* hybrid 'James Brydon'. *Plant Cell Tissue and Organ Culture* **36**: 145–148. DOI: https://doi.org/10.1007/BF00048326

Latowski K., Toma C., Dąbrowska M. & Zviedre E. (2014). Taxonomic features of fruits and seeds of *Nymphaea* and Nuphar taxa of the Southern Baltic region. *Limnological Review* **14**(2): 83–91. DOI: https://doi.org/10.2478/limre-2014-0009

Lim T.K. (2014). *Nymphaea nouchali. In: Edible Medicinal and Non-Medicinal Plants*, volume 8, Flowers, pp. 519–525. Springer Science + Business Media, Dordretcht New York, London, UK.

Liyanage K.K., Sumanasinghe V.A., Attanayaka D.P.S.T.G. & Baddewithana B.W.A.N. (2014). Identification of recommended *Hevea braislensis* (Willd.ex.A. Juss) Mull. Arg. clones grown in Sri Lanka by RAPD analysis. *Tropical Agricultural Research* **25**(2): 188–200. DOI: https://doi.org/10.4038/tar.v25i2.8141

Malik K.A. & Saxena P.K. (1991). Regeneration in *Phaseolus vulgar* L. promotive role of N6-benzylaminopurine in cultures from juvenile leaves. *Planta* **184**(1): 48–150. DOI: https://doi.org/10.1007/BF00208248

Mendonça E.G., de Souza A.M., Vieira, F. de A., Estopa R.A., Reis C.A.F. & de Carvalho D. (2014). Using Random Amplified Polymorphic DNA to assess genetic diversity and structure of natural *Calophyllum brasiliense* (Clusiaceae) populations in riparian forests. *International Journal of Forestry Research* **2014**: 1–8.
DOI: https://doi.org/10.1155/2014/305286

MOE (2012). The National Red List 2012 of Sri Lanka; Conservation Status of the Fauna and Flora, 8th edition, pp. 476. Ministry of Environment, Colombo, Sri Lanka.

Murashige T. & Skoog F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiolgia Plantarum* **15**: 473–497.
DOI: https://doi.org/10.1111/j.1399-3054.1962.tb08052.x

Nakkanong K., Nualsri C. & Sdoodee S. (2008). Analysis of genetic diversity in early introduced clones of rubber tree (*Hevea brasiliensis*) using RAPD and microsatellite markers. Songklanakarin *Journal of Science and Technology* **30**(5): 553–560.

Nei M. (1972). Genetic distance between populations. *The American Naturalist* **106**(949): 283–292.
DOI: https://doi.org/www.jstor.org/stable/2459777

Nei M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**(3): 583–590.

Niedz R.P., Hyndman S.E., Wynn E.T. & Bausher M.G. (2002). Normalizing sweet orange [C. *sinensis* (L.) Osbeck] somatic embryogenesis with semi-permeable membranes. *In vitro Cellular and Developmental Biology-Plant* **38**(6): 552–557.
DOI: https://doi.org/10.1079/IVP2002331

Nierbauer K.U., Kanz B. & Zizka G. (2014). The widespread naturalisation of *Nymphaea hybrids* is masking the decline of wild-type Nymphaea alba in Hesse, Germany. Flora **209**(2): 122–130.
DOI: https://doi.org/10.1016/j.flora.2013.12.005

Penfield S. (2017). Seed dormancy and germination. *Current Biology* **27**(17): 874–878.
DOI: https://doi.org/10.1016/j.cub.2017.05.050

Polina A.V. & Alexy B.S. (2007). Morphological variation of Nymphaea (*Nymphaeaceae*) in European Russia. *Nordic Journal of Botany* **25**(5–6): 329–338.
DOI: https://doi.org/10.1111/j.0107-055X.2007.00140.x

Priya P., Shukla M.R., Alan J. & Saxena P.K. (2017). Iron supplementation promotes in vitro shoot induction and multiplication of Baptisia australis. *Plant Cell Tissue and Organ Culture* **129**(1): 145–152.
DOI: https://doi.org/10.1007/s11240-016-1165-4

Raja M.K.M.M., Sethiya N.K. & Mishra S.H. (2010). A comprehensive review on Nymphaea stellata: A traditionally used bitter. *Journal of Advanced Pharmaceutical Technology and Research* **1**(3): 311–319.
DOI: https://doi.org/10.4103/0110-5558.72424

Renau-Morata B., Nebauer S.G., Sales E., Allainguillaume J., Caligari P. & Segura J. (2005). Genetic diversity and structure of natural and managed populations of *Cedrus atlantica* (Pinaceae) assessed using Random Amplified Polymorphic DNA. *American Journal of Botany* **92**(5): 875–884.
DOI: https://doi.org/10.3732/ajb.92.5.875

SAS Institute Inc. SAS/STAT user's guide, version 7-1. SAS Institute Inc. Cary, North Carolina, 1999.

Smits A.J.M., Schmitz G.H.W., Van Der Velde G. & Voesenek L.A.C.J. (1995). Influence of ethanol and ethylene on the seed germination of three nymphaeid water plants. *Freshwater Biology* **34**(1): 39–46.
DOI: https://doi.org/10.1111/j.1365-2427.1995.tb00421.x

Sumlu S., Atar H.H. & Khawar K.M. (2010). Breaking seed dormancy of water lily (*Nymphaea alba* L.) under *in vitro* condition. *Biotechnology & Biotechnological Equipment* **24**(1): 1582–1586.
DOI: https://doi.org/10.2478/V10133-010-0009-3

Veitch C.R. & Clout M.N. (2002). Turning the tide: the eradication of invasive species. *Proceedings of the International Conference on the Eradication of Island Invasives.* IUCN Species Specialist Group. Gland, Switzerland and Cambridge, UK.

Yakandawala D., Guruge S. & Yakandawala K. (2017). The identity of the violet flowered water lily (*Nymphaeaceae*) and its hybrid origin in the wetland ecosystems of Sri Lanka. *Journal of the National Science Foundation of Sri Lanka* **45**(4): 381–392.
DOI: https://doi.org/10.4038/jnsfsr.v45i4.8232

Yakandawala D. & Yakandawala K. (2011). Hybridization between native and invasive alien plants: an overlooked threat to the biodiversity of Sri Lanka. *Ceylon Journal of Science* (Bio Sci) **40**(1): 13–23.
DOI: https://doi.org/10.4038/cjsbs.v40i1.3403

Yeh F.C., Yang R.C. & Boyle T. (1999). POPGENE Version 1.31: Microsoft Windows-Based Freeware for Population Genetic Analysis. Univ. of Alberta Press, Edmonton, Canada.

J.Natn.Sci.Foundation Sri Lanka 2021 49 (3): 403 - 417
DOI: http://dx.doi.org/10.4038/jnsfsr.v49i3.10146

RESEARCH ARTICLE

An integrated corpus-based text mining approach used to process military technical information for facilitating EFL troopers' linguistic comprehension: US anti-tank missile systems field manual as an example

LC Chen[1,2], KH Chang[3, 4*] and SC Yang[2]
[1] *Department of Foreign Languages, R.O.C. Military Academy, Kaohsiung 830, Taiwan.*
[2] *Institute of Education, National Sun Yat-Sen University, Kaohsiung 804, Taiwan.*
[3] *Department of Management Sciences, R.O.C. Military Academy, Kaohsiung 830, Taiwan.*
[4] *Institute of Innovation and Circular Economy, Asia University, Taichung 413, Taiwan.*

Submitted: 02 September 2020; Revised: 28 December 2020; Accepted: 26 March 2021

Abstract: Military knowledge is an uncommon research field and is often classified as confidential information. Furthermore, when US military knowledge is adopted by English as a foreign language (EFL) countries, properly interpreting military texts brings about challenges. Taking Asian militaries as examples of EFL countries, not every trooper has sufficient English proficiency and capability to read and comprehend complicated military knowledge databases. In addition, under limited training time and lack of suitable reference materials, it is difficult to popularise and improve the efficiency of the courses that study US field manuals (FMs), which are important books that introduce US military combat tactics and strategies, military operation procedures, weapon systems, and others. Nevertheless, in many EFL countries, English learning is integrated into the education system to promote internationalisation and enhance global competitiveness. Thus, the English proficiency of nationals in most EFL countries is not negligible. Based on these considerations, this paper discusses the integration of the corpus software and cooperation of linguists and military experts to conduct syntax analysis and taxonomy of military terminology to enable EFL troopers with non-excellent English proficiency to understand the intricate US military domain knowledge and develop the military corpus as an auxiliary language training material. The US Army FMs of anti-tank missile systems are adopted as an empirical example to illustrate the proposed approach. Analytical findings will become critical reference indicators for defence language institutes (DLI) of EFL militaries in developing military English training materials and for processing military information.

Keywords: Anti-tank missile systems, Asian militaries, corpus software, military corpus, military terminology, syntax analysis.

INTRODUCTION

Corpus-based analytical approaches are considered as big data analysis; its sources of big data (Christ *et al.,* 2019) are natural languages (NLs) that are compiled to become corpus from enormous texts and discourses (Koops & Lohmann, 2015; Brindle, 2016; Beller & Bender, 2017; Chen *et al*., 2020). Statistics-based algorithm corpus analysis studies are indispensable in today's digital era. Due to the developments in computer technology, large amounts of texts which include articles, news reports and discourses, among others can be stored electronically in computers (Ferguson, 2001; Daskalovska, 2015; Coats, 2019). Thus, corpus programs directly process text information, via computers for further analysis (Chen & Chang, 2019; Smalheiser *et al.,* 2019). In terms of military information, it has sophisticated domain knowledge and scientific techniques, moreover, it is difficult to collect and is often categorised as confidential data (Trembach, 2019). Hence, this kind of analysis is pertinent to countries where English is a foreign language (EFL) and weaponries and military operation techniques are mainly adopted from the US military. A chunk of the military domain knowledge might appear ambiguous to troopers who use English as a foreign language (EFL) troopers. Seeking proper information processing toolkits is necessary for improving the efficiency of EFL military information processing. To conquer the problems resulting from language difficulies, computer-assisted language learning (CALL)

* Corresponding author (evenken2002@yahoo.com.tw https://orcid.org/:0000-0002-9630-7386)

has been identified as a key tool for improving foreign languages acquisition since the invention of computers.

CALL is an educational research term that discusses the interrelationship between language teaching, learning, and digital technologies. The core of CALL is to adopt various computer-based technological resources to enhance the efficacy of language teaching and learning. With modern technological advancement, CALL applications have also evolved from single unit systems to multi-function systems that are more complex and much closer to simulating real-life situations (Gonen, 2019). For example, Hsieh *et al.* (2017) embedded a social media app, LINE, into a foreign language course. The results of their study showed that a social communication program would be an affordable, ubiquitous, and easy to use pedagogical assistant tool for stimulating students' learning motivation and improving students' language learning efficacy. Harvey-Scholes (2018) used the computer program, N-gram method, to detect native Spanish speakers' English writing errors for improving EFL students' writing proficiency. Cheng and Tsai (2019) assimilated head-mounted displays used for creating virtual realties into field trip education. They found that this pedagogical method that embedded novel technology can effectively enhance students' learning motivation and reduce exam anxiety. CALL applications have been gradually utilised by language teachers (or instructors) as digitalised coworkers in their language classrooms. Nowadays, CALL is more learner-centered, with the availability of learning materials for providing autonomous and continuous learning approaches (Lai *et al.,* 2016). The aforementioned studies mainly discuss the CALL approaches used in language courses for enhancing EFL learners' learning achievements. When languages are used for specific purposes, the problems to be solved may be slightly different. US military information consists of linguistic and domain knowledge, both English and military; thus, it is considered as an example of English for Specific Purposes (ESP) case.

ESP mainly discusses English linguistic knowledge that is applied in specific domains and pedagogical approaches or self-learning methods for making EFL learners acquire disciplinary literacy. Namely, ESP curricula and learning tools' developments lean towards the domains' needs. Disciplinary literacy, as Zygouris-Coe (2012) defined it, is mainly centered on the learning domain knowledge and putting the domain knowledge into an actual environment, where the language use is English. It emphasises a middle-ground between English acquisition and its application to domain knowledge. Moreover, in order to enable ESP learners to use English for gaining expertise in domains, concept-embedding words and lexical bundles (LBs) inevitably need to be extracted and researched. Shanahan and Shanahan (2017), proposed a pedagogical strategy for learning ESP. They proposed that learners should focus on vocabulary for general purposes at the initial stage. When the learners attain certain comprehensive levels of vocabulary and grammar, they will begin to acquire the requisite vocabulary to handle schools' academic needs. Finally, the learners would acquire the specific linguistic knowledge for a discipline and expertise in a domain. Based on the theories postulated by Zygouris-Coe (2012) and Shanahan and Shanahan (2017), it is obvious that 'vocabularies' are the essential elements in ESP research. Furthermore, retrieving concept-embedding words and LBs such as terminology, technical words, and domain-specific phrases are also essential tasks. In ESP research cases, many pedagogical approaches and knowledge processing approaches are applied in surmounting language barriers that are particularly caused by EFL (Flowerdew, 2000; Zygouris-Coe, 2012). However, there are no absolute advantages for English native speakers in gaining domain knowledge (Zygouris-Coe, 2012; Shanahan & Shanahan, 2017; Viswanathan *et al.,* 2020). For instance, Derbentseva *et al.* (2007) structured concept maps to illustrate intricate domain knowledge for improving students' subject learning proficiency in Canada, where English is the first language. In addition, in EFL environments, corpus-based approaches are popular because its results align more with domain experts and linguists' expectations. Li *et al.* (2019) reviewed the literature that used corpus-based methods to analyse tourism information and pointed out the intimate correlation between corpus analysis and big data analysis. Li (2016) used Wordsmith tool, the most popular corpus program, to extract the word list and keyword list from the corpus of JRC-Acquis (EN) in order to identify vague terms that are used in legal documents. Munoz (2015), in addition to use the corpus program to retrieve the keyword list from the agricultural corpus, also conducted a taxonomy of keywords so that the data output will obtain greater benefits in ESP courses' development and ESP learning. Other research activities that employed corpus-based approaches to probe linguistic patterns in texts of different professional disciplines such as medicine (Li *et al.* 2019; Siefridt *et al.,* 2020), engineering (Liu & Han, 2015; Nekrasova-Beker, 2019), linguistic and language education (Henry & Roseberry, 2001; Green, 2019; Kim & Nam, 2019), and others, have also significantly resolved complex ESP cases.

Recently, statistics-based algorithm corpus programs have been advanced gradually by modern computer technologies, but the limitations of corpus-

based studies still keep emerging, especially in cross-disciplinary researches (Cho & Yoon, 2013; Sholokhov *et al.,* 2020; Siefridt *et al.,* 2020). One of the limitations might be identifying the right person for analysing the results of corpus programs. This person is the proper interpreter for deciphering ESP cases in corpus-based approaches. Furthermore, without the verification assessments by experts or satisfaction feedback from those in the specific field, it is difficult to prove the actual benefits of the results of some corpus-based studies. In order to make military corpus analysis results satisfy military domain usages, this paper integrates a corpus-based CALL software and synergism of domain and linguistic experts to process the corpus of US Army anti-tank weapon systems FMs. The proposed method can be separated into two phases: (1) machine processing that is implemented by AntConc 3.5.8 (Anthony, 2019), a corpus software, (2) manual annotations including syntax analysis and conducting military terminology in second language (TISL) taxonomy by linguistic and military experts; and consists of a total of 7 steps. The results illustrate how the proposed approach generates domain-oriented results for EFL troopers as auxiliary language learning materials in acquiring US Army domain knowledge.

METHODOLOGY

For military domain, military knowledge embraces professional and uncommon genre types, terminologies and scientific knowledge that cause civilians or even military personnel some difficulty in receiving information. Thus, the proposed approach integrates machine processing and military experts' annotation to process military corpus for retrieving domain knowledge, inducting genre types, and conducting military TISL taxonomy.

The proposed approach can be divided into two phases, machine process, and manual annotation, and covers seven steps (Figure 1). Steps 1 to 3 belong to Phase I - machine process; AntConc 3.5.8 (Anthony, 2019) is the primary analytical corpus software to process the target corpus. Steps 4 to 6 belong to Phase II - manual annotation. In step 5, linguists and domain experts conduct syntax analysis by clustering a function word list. In step (6), linguists and domain experts conduct the military TISL taxonomy based on checking the word list and keyword list, checking LBs of tokens, and checking concordance lines of abbreviations and acronyms. Step 7 clusters the results in step 5 and 6 for military ESP training courses. Detailed descriptions and illustrations of each step is introduced as follows:

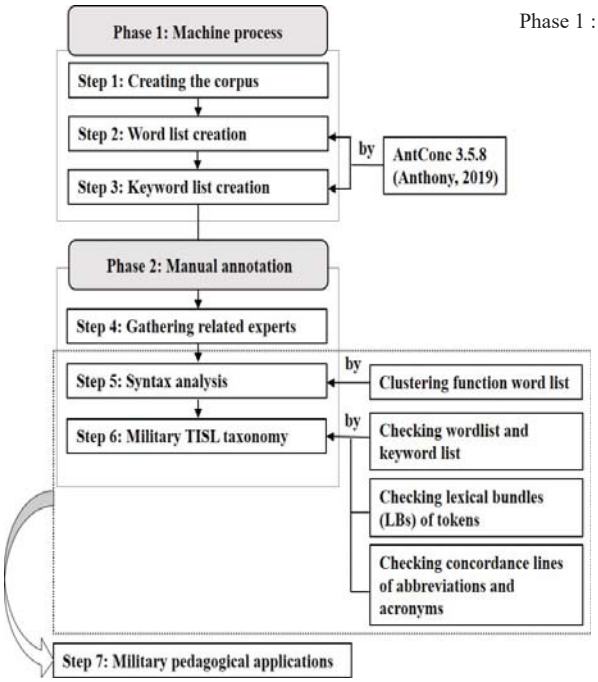


Figure 1: The procedures of the proposed approach

Step 1: Creating the corpus

Although Antconc, Wmatrix, and Sketch Engine have similar functions in corpus analysis, nevertheless, considering the budgets and the necessity of internet, we would prefer Antconc because of its competitive advantages, such as affordable (i.e. free to access), ubiquitous (i.e. can be used anywhere), and easy (i.e. concise operative interfaces). Moreover, it can be operated without installation or an internet connection. Hence, this paper adopts AntConc 3.5.8 (Anthony, 2019), as the primary corpus software to process the target corpus.

Once textual data are collected, data conversion (transforming data from *.docx*, or *.pdf* into *.txt*) is inevitable because AntConc 3.5.8 (Anthony, 2019) only accepts particular data formations (i.e. *.txt, UTF-8*).

Step 2: Word list creation

Once all input data is prepared for analysis, users will choose the 'Word List' section and click the 'Start' button to generate the word list of the input corpus, and record it for further analysis.

Step 3: Keyword list creation

Normally, the algorithm of keyword list generator based on a log likelihood test to compare two corpus data, input corpus data and reference corpus data. The software utilises statistical algorithm to find words with high frequency in the input corpus but with low frequency in the benchmark corpus to compute their keyness values, to identify keywords. Keywords are considered significant features of the input corpus.

The selection of the benchmark corpus needs to base on genre types, namely, those two corpora have to have different genres (i.e. specific purposes vs. general purposes). Hence, the biggest and the most adopted general purposes genre type corpora include the corpus of contemporary American English (COCA) and the British National Corpus (BNC). Those corpora provide free access, and are ideal benchmark corpus data (e.g., Li, 2016).

After the word list is created and the benchmark corpus, input users will select the 'Keyword List' section and click the 'Start' button on the corpus software to generate the keyword list of the input corpus and record it for further analysis.

Step 4: Gathering related experts

Experts with linguistic analysis and military expertise are gathered to conduct the following procedures. All results from the corpus software, such as word list, keyword list, lexical bundles (LBs) of tokens and keywords, and so on, need to be analysed by experts in this step. Linguistic experts are expected to interpret genres of texts, while domain experts are expected to retrieve domain knowledge.

Step 5: Syntax analysis

In this step, the gathered experts, based on the word list that created in step 2, cluster a high-frequency function word list. Function words may seem literally meaningless; nevertheless, those are critical elements for structuring sentences, paragraphs, and even articles. Thus, high-frequency function words are critical clues for implementing syntax analysis.

Step 6: Military TISL taxonomy

In order to make extracted military TISL more meaningful for EFL troopers, in this step, the gathered experts will check the word list and keyword list (results in step 2 & 3) to extract military TISL, check LBs of tokens to avoid missing critical phrase-style terminologies, and check concordance lines of abbreviations and acronyms to retrieve the complete LBs and hidden meanings of abbreviations and acronyms. Eventually, the gathered experts will re-categorise military TISL.

Step 7. Military pedagogical applications

Results in step 5 and step 6 will become important in ESP training materials for EFL troopers before they enter the actual weaponry training.

RESULTS AND DISCUSSION

Overview of the military corpus data

In this study, the compiled military corpus data includes FM 3-23.25 Light anti-Armor Weapons (LAW) (US Department of the Army Washington, DC., 2001), FM 3-22.34 TOW Weapon System (US Department of the Army Washington, DC., 2003), FM 3-22.37 JAVELIN - Close Combat Missile System, Medium (US Department of the Army Washington, DC., 2008). Those weapon systems have been widely adopted by many countries and their FMs embrace the most complete technical information of the US Army current anti-tank weapon systems.

In the aforementioned military technical FMs, terms that are used for detailed weaponry specification, operating procedures, and tactical usages are introduced. Even if EFL speakers who have high proficiency in English research FMs, the interpretation of results may still cause information distortion because they lack military background knowledge. To verify the proposed approach in analysing military texts, the researchers adopted the compiled military corpus as an empirical example for importing to the proposed two-phase approach. The corpus contained three technical books (i.e. FMs), 5,346 word types and 108,605 tokens.

Table 1: Lexical characteristics of the target corpus

Components of US Army Anti-Tank missile systems FMs		Word types	Tokens	TTR
Missile types	FM serial number			
Javelin Missile	FM 3-22.37	3,294	46,838	7.03 %
TOW Missile	FM 3-22.34	3,430	38,541	8.90 %
Light Anti-armor Weapon (LAW)	FM 3-23.25	2,518	23,226	10.84 %
Whole corpus		5,346	108,605	4.92 %

Its type/token ratio (TTR) is 4.92 % (Table 1). FMs were segmented into each chapter (as sub-corpus), as it allowed the researchers to easily identify concordance plots and the etymology of words in the manuals. FMs' figures, references and tables were eliminated. The elements of analysis were tokens, clusters, and concordances. In addition, this paper chose COCA as the benchmark corpus. COCA is the largest (contains 9,412,521 words) and genre-equivalent corpus of contemporary American English. It contains diverse texts which include discourses, fictions, newspapers, magazines, academic papers, and so on. Thus, using COCA as the reference corpus would be an ideal way to retrieve keywords from the target corpus.

Resulting data of machine processing in Phase I

In Phase I, AntConc 3.5.8 (Anthony, 2019) analysed the target corpus and generated the word list and keyword list. The raw data results are described as follows.

(1) Generating word list

The word list is the data resulting from step 2. The corpus program uses its statistic-based algorithm to integrate and count tokens' frequency and to rank tokens. The word list indicated 5,346 words which ranked in frequency from high to low (see Table 2). High frequency words can be considered as the core elements of the target corpus. Moreover, low frequency words can also be considered as unique features of the target corpus.

Table 2: An illustrative example of generating word list (partial data)

Rank	Frequency	Tokens	Rank	Frequency	Tokens	Rank	Frequency	Tokens
1	10988	the	35	400	s	68	205	indicator
2	3267	and	36	393	by	69	204	during
3	2667	to	37	390	range	70	202	use
4	2348	of	38	377	if	71	195	d
5	2249	a	39	376	e	72	191	one
6	1530	is	40	368	targets	73	188	right
7	1298	in	41	339	vehicle	74	183	battery
8	1018	on	42	330	can	75	183	time
9	1007	figure	43	328	should	76	175	two
10	889	for	44	317	round	77	171	left
11	865	gunner	45	312	must	78	168	vehicles
12	863	target	46	308	b	79	167	these
13	814	or	47	291	unit	80	166	trigger
14	718	be	48	285	weapon	81	164	each
15	653	are	49	284	table	82	164	rear
16	643	with	50	272	used	83	160	forward
17	633	at	51	257	light	84	159	enemy
18	615	missile	52	253	launcher	85	157	squad
19	581	from	53	253	meters	86	156	engagement
20	565	it	54	250	launch	87	152	its
21	544	position	55	243	switch	88	151	IR
22	517	training	56	235	sight	89	150	antiarmor
23	514	TOW	57	230	system	90	150	no
24	509	when	58	226	area	91	147	up
25	504	as	59	222	has	92	145	c
26	492	an	60	219	gunnery	93	145	cover
27	471	CLU	61	219	seeker	94	143	day
28	457	m	62	219	will	95	142	weapons
29	456	that	63	217	section	96	141	end
30	451	fire	64	214	may	97	141	NVS
31	430	Javelin	65	210	he	98	141	other
32	409	firing	66	210	trainer	99	139	using
33	404	not	67	208	all	100	138	place
34	404	this						

(2) Generating Keyword list

The keyword list is the data output from step 3. The mechanism of generating keyword list is that the corpus software calculates 'keyness' of words by its algorithm, likelihood test, to find words that frequently appear in the target corpus but infrequently appear in the benchmark corpus. The keyword list, in this case (log-likelihood test (4-term), $p < 0.05$ (+ Bonferroni), covered 1,185 words and showed more specific words of the target corpus. In addition, it allowed us to filter function words or more generally-use words (Table 3).

Table 3: An illustrative example of generating keyword list (partial data)

Rank	Freq.	Keyness	Keywords	Rank	Freq.	Keyness	Keywords
1	865	7646	gunner	51	82	679.91	MGS
2	863	5838.75	target	52	312	642.31	must
3	1007	5675.24	figure	53	130	634.9	procedures
4	615	5024.43	missile	54	80	633.3	qualification
5	10988	4802.45	the	55	89	632.97	infrared
6	514	4289.22	TOW	56	226	630.88	area
7	471	4216.11	CLU	57	70	626.34	FTT
8	430	3761.36	JAVELIN	58	70	626.34	NFOV
9	409	3029.92	firing	59	98	621.61	armor
10	544	2694.15	position	60	128	617.97	mode
11	517	2652.34	training	61	68	608.45	WFOV
12	368	2519.37	targets	62	70	595.31	loader
13	253	2240.87	launcher	63	142	586.15	weapons
14	451	1977.8	fire	64	97	585.58	consists
15	219	1937.1	gunnery	65	115	581.38	combat
16	390	1914.39	range	66	63	563.7	TFTT
17	339	1913.54	vehicle	67	78	558.86	connector
18	219	1897.59	seeker	68	85	555.15	brightness
19	285	1794.99	weapon	69	125	551.9	display
20	253	1737.53	meters	70	115	541.1	engage
21	291	1582.35	unit	71	109	536.13	soldier
22	317	1520.29	round	72	61	535.59	TGT
23	250	1498.35	launch	73	70	534.48	warhead
24	205	1498.01	indicator	74	63	533.49	misfire
25	210	1428.04	trainer	75	59	527.91	firer
26	243	1416.54	switch	76	75	498.65	tactical
27	150	1342.27	antiarmor	77	97	498.05	tasks
28	235	1287.18	sight	78	76	493.15	ammunition
29	141	1261.72	NVS	79	71	487.13	stationary
30	151	1228.28	IR	80	136	486.88	conditions
31	133	1190.13	FOV	81	109	484.89	command
32	376	1108.08	e	82	108	483.59	positions
33	183	1058.92	battery	83	87	481.38	adjust
34	166	1018.3	trigger	84	91	474.08	tables
35	157	980.12	squad	85	99	466.5	objects
36	308	962.79	b	86	129	464.91	leader
37	156	944.72	engagement	87	230	464.56	system
38	105	939.55	nightsight	88	272	459.41	used
39	217	899.51	section	89	145	456.39	cover
40	284	892.83	table	90	58	455.1	platoon

Continued in page 409

Continued from page 408

41	164	891.84	rear	91	55	453.04	traversing
42	99	885.85	daysight	92	64	449.02	TM
43	168	867.89	vehicles	93	160	448.09	forward
44	159	843.26	enemy	94	103	444.47	sector
45	90	805.31	backblast	95	115	442.44	shoulder
46	90	772.32	gunners	96	105	441.03	assembly
47	84	751.62	BCU	97	57	439.84	tracer
48	257	717.93	light	98	62	437.07	tracker
49	122	698.71	tube	99	53	435.56	BST
50	114	688.35	indicators	100	48	429.48	handgrip

Resulting data of experts' annotations in Phase II

(1) Gathering related experts

For many EFL countries' military, US military FMs are highly complex and critical because they involve a foreign language and military domain knowledge. Even if the corpus program is able to categorise and to process the target corpus, the contribution of raw data results to EFL troopers remains low. Thus, the researchers gathered related experts including linguists, military experts, and experts in performance evaluations (see Table 4) and appointed an assessment team to operate the analytical program in Phase I and to optimise the data results in Phase II.

Table 4: Introduction of linguistic and domain experts

Researchers	Academic position	Academic specialties	Military specialty	Military rank	Years in active duty
Expert 1	Professor	Fuzzy Logic, Soft Computing, Applied Mathematics, Military Operations Research	Artillery	Colonel	23 years
Expert 2	Assistant Professor	Performance Evaluation, Military Operations Research	Artillery	Lieutenant Colonel	19 years
Expert 3	Lecturer	Corpus Linguistics, Computer-assisted language learning (CALL), Natural Language Processing (NLP)	Infantry	Major	9 years
Expert 4	Weaponry Instructor	Management Science, also specialised in LAW, Javelin missile, TOW missile systems	Infantry	Major	9 years

(2) Syntax analysis

According to the word list from step 2, all tokens were ranked by its frequency (refer to example data on Table 2). High frequency words were considered as important indicators for identifying core linguistic patterns in the target corpus. The word list featured many high-frequency function words. Those words may seem unrelated to the domain and easily be neglected; nevertheless, those are important elements of grammatical structures such as auxiliary verbs, prepositions, conjunctions, pronouns, and so on. Grammatical structures may confuse EFL troopers in their attempt at understanding FMs. Thus, the experts retrieved function words from the range of the top 500 high frequency words in the corpus of US Army antitank weapons FMs. The words were categorised into eight groups based on their grammatical functions (Table 5), then outlined the following linguistic evidences to conduct syntax analysis, for giving EFL military personnel important linguistic insights before they involve researches or training courses in US anti-tank missile systems.

Group 1. *"To-infinitive"* and *"for"* represent purposes and reasons:

When "to-infinitive" clauses are placed after nouns or noun phrases, they indicate what the things refer to or intend to do. Using the target corpus as examples, the researchers discovered that "to-infinitive" sentence construction rules were highly used for indicating purposes of activities and terminologies:

1-1 *Raise or lower your knees* ***to adjust*** *for elevation on the target.*

1-2 *... supporting fires* ***to allow*** *screening units* ***to break*** *contact and withdraw.*

Table 5: Clustering function word list

Classification	Rank in word list	Frequency	Function words
Group 1	3	2667	to
	10	889	for
Group 2	4	2348	of
Group 3	7	1298	in
	391	46	under
Group 4	9	1007	figure
	49	284	table
Group 5	24	509	when
	25	504	as
	69	204	during
	146	107	before
	131	115	after
	240	67	until
Group 6	38	377	if
Group 7	43	328	should
	45	312	must
	64	214	may
Group 8	176	93	such

1-3 … *with a carry bag, which provides space* ***to carry*** *a CLU, a technical manual* …

1-4 … *improving the kill mechanisms of our missiles to defeat the improved armor.*

1-5 …*is used to develop the skills required to engage targets under field conditions.*

When "for" is placed in front of nouns or noun phrases, it represents the purposes of objects, actions, and so on. In this case, "for + nouns (NPs)" were used to explain equipment's functions or purposes of important operating procedures:

1-6 *Track gate adjustment* ***for a bunker involves the perceived size of …***

1-7 … *do MGS self-test* ***for battery****.*

1-8 *The gunner should use the NFOV* ***for classification and recognition****.*

1-9 *Inspect the open end of the round* ***for dirt and foreign material****.*

1-10 *The trainers must know the appropriate combat techniques* ***for employing these weapons.***

Field manuals can be considered as a type of equipment user guide. They teach readers how to operate systems or component parts and explain the purposes of operating procedures. Thus, "to-infinitive" and "for" are important grammatical rules for developing specific genre.

Group 2. *"nouns (noun phrases) + of + nouns (noun phrases)"* composed of terminologies or indicates relationships of nouns:

When "of" is placed between noun phrases, the combinations show its relationships of possession, belonging, or connection. It is a kind of a strong supplementary narrative usage to show the relationship between a noun and another. In this case, the researchers noted that many terminologies were connected by "of" and developed to LBs for domain usages. See follows:

2-1 *Ensure all the standard* ***principles of camouflage*** *are followed.*

2-2 … *the Javelin may not be* ***the weapon of choice*** *in the urban environment* …

2-3 *See Table 4-2 for* ***frequency of events*** *as required by DA Pam 350-38 STRAC.*

2-4 *... not to fire until given* ***the command of execution****.*

2-5 … *leader selects a primary position and* ***sector of fire*** *for each weapon.*

Group 3. Using *"in, under"* to describe conditions and situations:

"In, under" are words used to describe some procedures or activities that happen in certain conditions or situations. According to the word list, the researchers found that details in US Army field manuals explain some conditions that may be used to initiate some procedures or some specific functions, as follows:

3-1 … *FTT allows gunner training to be conducted* ***in a field environment*** …

3-2 *It can be employed* ***in all weather conditions*** *as long as the …*

3-3 *... Javelin's 2,000-meter range allows flexibility* ***in choosing ambush positions****.*

3-4 … *be used at any time of day* ***under any weather conditions****.*

3-5 *... and firing the TOW* ***under NBC and limited visibility conditions*** …

Group 4. *"Figure and Table"* as illustration and data explanation:

The high frequency words such as '*figure* and *table*' highlights the importance of "illustration and data explanation" in field manuals. Texts are combined with assistant materials (e.g., photos, graphics, sketches, tables, charts, etc.) in order to show procedures, introduce equipment, and analyse the capabilities of weapon systems in a more precise and detailed manner. Those features hint that, "illustration and data explanation" is critical to making readers understand the abstractive domain knowledge, hence, avoiding misuse of weapon systems.

4-1 ***Figure*** *A-1 shows the probability of survival for …*

4-2 *The eyepiece (****Figure*** *1-14) allows the gunner to see the CLU …*

4-3 ***Figure*** *5-2 Javelin command launch unit.*

4-4 *See* ***Table*** *3-1, a notional training schedule.*

4-5 ***Table*** *6-1. Armored vehicle kills.*

Group 5. Words for describing operating sequences and timing of uses of weapon systems:

Words such as "when, as, during, before, after, until" are adopted to tell users "when or under what kinds of circumstances, someone will or should do something". In the typical manual genre, those words are not only used to express tenses but also used to express the important operating sequences of weapons.

5-1 *WARNING:* ***When*** *firing the M136 AT4, do not place …*

5-2 ***As*** *Javelin gunners destroy their targets, leaders should …*

5-3 ***During*** *combat or field training, TOW crews will ...*

5-4 … *activating the seeker* ***before*** *assuming a firing position.*

5-5 *... soft targets can normally continue to fight* ***after*** *being attacked by light anti-armor weapons.*

Group 6. Conditional clauses:

FMs use many conditional clauses to give users scenarios, possible situations, the next steps to take or consequences of actions taken. The common sentences structures identified are: "If something happened, someone should do …" Examples are as follows:

6-1 ***If*** *a misfire occurs in combat ...*

6-2 ***If*** *facilities and equipment are not available to …*

6-3 ***If*** *in a firing position, moves the round ...*

6-4 ***If*** *possible, they should construct reinforced position …*

6-5 ***If*** *the gunner is not engaging a target …*

Group 7. Giving suggestions, indicating importance and anticipating scenarios:

The researchers found in FMs that "should, must, may" indicated three different levels of authors' intentions. When describing tactic usages, FMs used "should" to give suggestions to readers. This allows for flexibility and does not put constraint on readers' tactical approaches. When referring to safety procedures, or safety concerns of weapon systems, FMs used "must" to highlight something that is necessary and nonnegotiable. When

FMs used the word "may", they gave readers scenarios to foresee situations that may happen. Those messages remind readers of early preparation to avoid occurring surprising and emergency incidents.

7-1 *Each position* ***should*** *allow flank fire and have cover and …*

7-2 *The Infantry* ***should*** *be able to cover dismounted AAs to …*

7-3 *However, trainers and leaders* ***must*** *adopt new safety procedures to ensure …*

7-4 *To fire the AT4, the firer* ***must*** *apply firm and steady forward pressure to …*

7-5 *The launcher electronics* ***may*** *also be damaged.*

Group 8. Using *"such as"* to give lists of items or examples. See as follows:

e.g. 8-1 *Backlighting occurs when an IR source,* ***such as*** *a tank's exhaust, emits IR …*

e.g. 8-2 *… can also be used against soft targets,* ***such as*** *bunkers, field fortifications, automobiles, and …*

e.g. 8-3 *is heat produced by a slow (****such as*** *a bonfire) or very quick (****such as*** *…*

e.g. 8-4 *… one time on a prearranged signal* ***such as*** *a command, whistle, booby trap, mine, or …*

e.g. 8-5 *… an object in the target scene,* ***such as*** *a far tree line.*

(3) Military TISL taxonomy

(i) Checking and refining wordlist and keyword list

Word list and keyword list are an important analysis results from corpus software. However, words such as function words, meaningless words, and some characters existed abundantly on those lists. Thus, the first filtering process is to eliminate those kinds of words for making word list and keyword list more domain oriented.

(ii) Checking LBs of tokens

In this case, the researchers found that some terminologies may exist in the form of phrases. Thus, it is necessary to check LBs of tokens. The researchers based on the setting of cluster size (min.2 and max.5) and term position (both on the left and right) to check each tokens on the keyword list and recheck potential military-oriented words on the wordlist. For example, a the word, "top" is ranked No. 585 in the keyword list may be irrelevant to the military domain if the focus is only on the surface explanation of the word. Nevertheless, when searching for the word in the clusters/n-gram, the results showed "top attack", "top attack mode", "top indicator(s)" on the list (Figure 2). This is confirmed by the military experts that the term belongs to one of the most important terms in the Javelin missile-system. To avoid missing critical information, checking LBs of each potential token is crucial.

(iii) Checking concordance lines of abbreviations and acronyms

In the word list and keyword list, the researchers found that FMs adopted many acronyms to form terminologies. In addition, the researchers also classified "abbreviations" and "acronyms" as different groups to highlight their importance. Thus, understanding the LBs of acronyms is crucial, otherwise it will be hard to comprehend the terminological meanings. According to US military FMs, "abbreviations" and "acronyms" have been explained in detail, but retrieving information directly from those FMs seems to be inefficient and lack integration of knowledge. AntConc 3.5.8 (Anthony, 2019) is an appropriate platform for providing concordance evidence to extract LBs of acronyms and abbreviations, and military domain knowledge (see Figure 3).

(iv) Re-categorising military TISL

Keywords and high frequency words represent identity, core knowledge, critical information and specific terminologies of the target corpus. Referred to Munoz's (2015) research, the experts gathered based on their specialties (1) to eliminate function words, meaningless words and unrelated letters on the word list and keyword list; (2) checking LBs of tokens to avoid missing critical phrase-style terms; (3) checking concordance lines of abbreviations and acronyms to extract definition of those; and (4) classify terminologies into seven groups in order to illustrate the whole frame of military TISL in US Army FMs of antitank weapon systems based on terminologies' functions, meanings, usages, and characteristics. The categorisation (Figure 4) can be defined as: Group 1. Weapon systems; Group 2. Critical component parts and accessories; Group 3. Procedures, actions, and operations; Group 4. People; Group 5. Measurements; Group 6. Abbreviations; Group 7. Acronyms. The groups were created based on aforementioned criteria. The compartmentalisation of military TISL in this case, facilitates the efficiency of understanding the military professional terminologies.

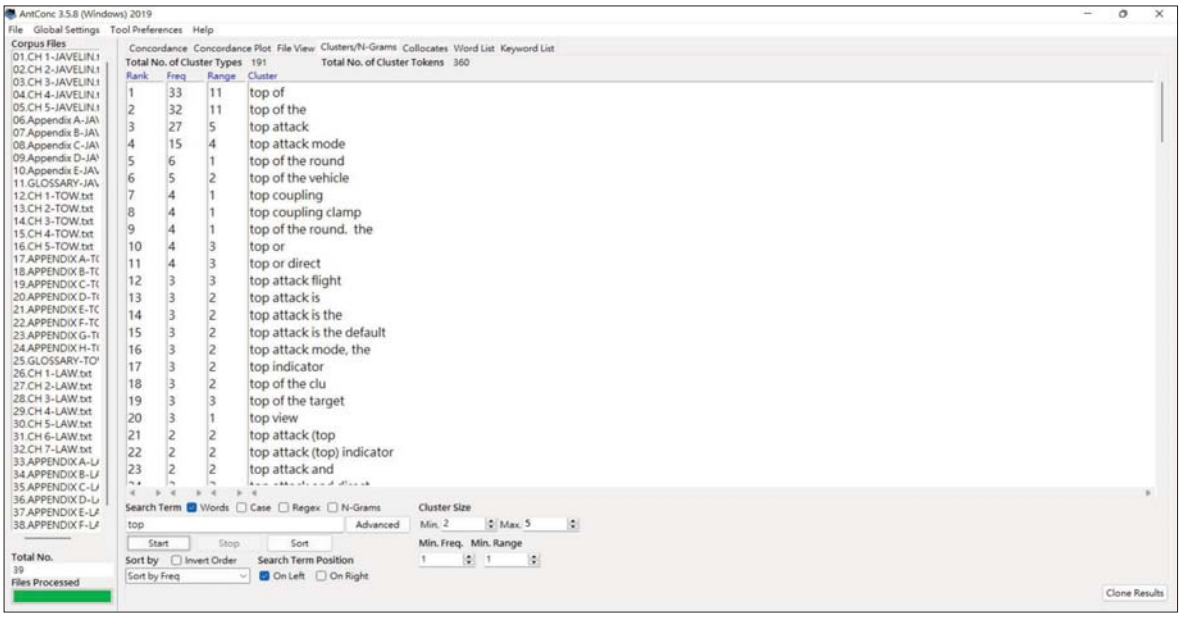


Figure 2: An illustrative example of extracting LBs of a token from AntConc 3.5.8

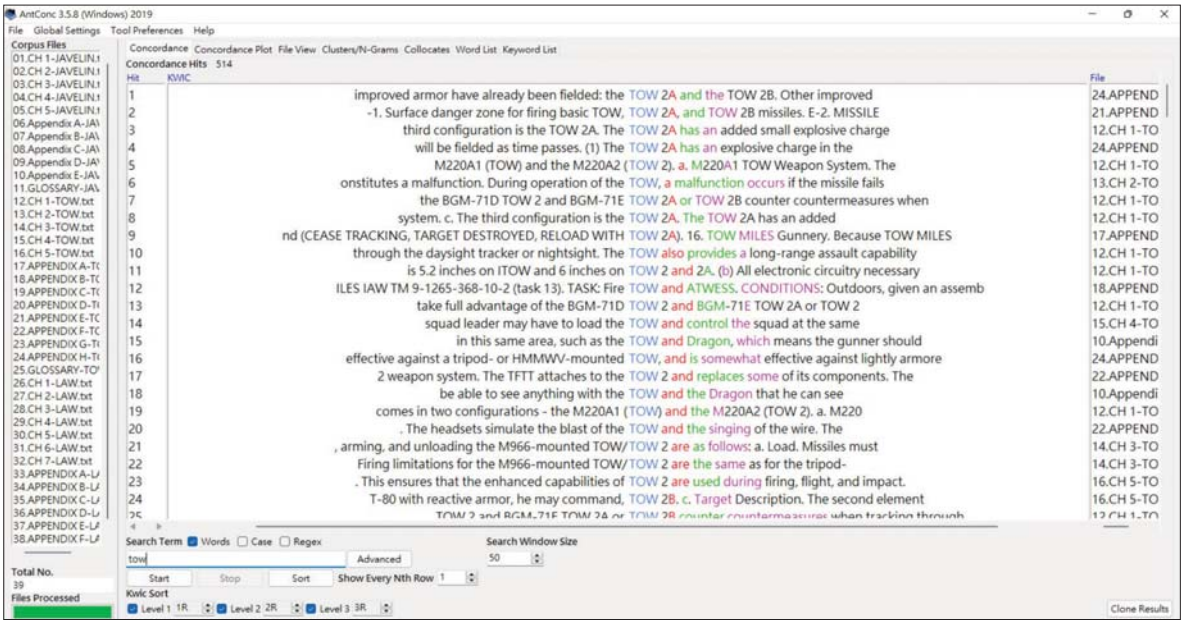


Figure 3: An illustrative example of extracting LBs of an acronym from AntConc 3.5.8

Group 1. The categorisation "Weapon systems" indicated terms which referred to weapons (1-1, 1-2, 1-3) and ammunitions (1-4, 1-5):

1-1 *The **TOW** is mainly an antitank weapon used for …*

1-2 *The **Javelin** is a fire-and-forget, shoulder-fired …*

1-3 ***LAW** is a lightweight, self-contained, anti-armor weapon …*

1-4 *The encased **missile** is a priority item and should be …*

1-5 *The Javelin **round** is bulky and restricts movement in heavily …*

Group 2. The categorisation, "Critical component parts and accessories", represented important parts and accessory equipment, for example, digital controllers (2-1), battery (2-2), critical parts of missiles (2-3), situation indicators (2-4), observation equipment (2-5):

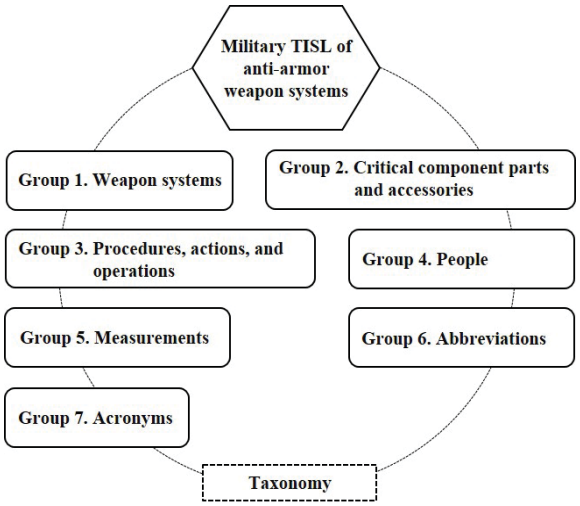


Figure 4: Architecture diagram of Military TISL of anti-armor weapon systems

2-1 … *system consisting of a* ***command launch unit (CLU)*** *and a round.*

2-2 … *or the rechargeable* ***BB390A/U battery****, and is equipped with a connector that …*

2-3 *After* ***seeker*** *activation, the BCU indicator flashes to indicate …*

2-4 *The* ***BCU indicator*** *is located on the bottom left side …*

2-5 … ***CLU*** *houses the* ***day sight****,* ***night vision sight (NVS)****,* ***controls****, and* ***indicators****.*

Group 3. The categorisation, "Procedures, actions, and operations", included verbs and nominalisations related to operating weapon systems (3-1, 3-2, 3-3, 3-4, 3-5):

3-1 *Breath control is as important when* ***firing*** *a light anti-armor weapon …*

3-2 *... the gunner squeezes the fire trigger to* ***launch*** *the missile.*

3-3 *When* ***aiming*** *the AT4, remember to* ***aim*** *by placing …*

3-4 *... turn the system on or off and* ***adjust*** *the brightness of the eyepiece display.*

3-5 *The gunner strives to* ***engage*** *enemy vehicles in the 1,000-to 2,000-meter range.*

Group 4. The categorisation, "People", indicated to operators (4-1, 4-2), friendly (4-2, 4-5), and enemy (4-3, 4-4):

4-1 *When* ***the gunner*** *fires the Javelin, it ignites a launch …*

4-2 *The TOW* ***gunner*** *or* ***squad leader*** *can use the night sight method or the …*

4-3 … *identifying targets (to include friend or* ***foe****), prioritizing target's …*

4-4… *warhead is capable of defeating any known* ***enemy armor****.*

4-5 *The* ***commander*** *chooses a method of assessing the trainers …*

Group 5. The categorisation, "Measurements", embraced units of measurements (5-1, 5-2, 5-5), ranges of weapons (5-3), and tools of measurements (5-4):

5-1 *Russian 40* ***mm*** *Antitank Grenade Launcher, RPG-7V (Figure H-6).*

5-2 *Armor penetration: 450 to 2,000* ***mm****.*

5-3 *The Javelin's* ***2,000-meter range*** *allows flexibility in choosing ambush positions.*

5-4 *Most armies use* ***laser range finders*** *and target designators.*

5-5 … *estimate it as a fast-moving vehicle (10* ***mph*** *or faster).*

Group 6. The categorisation, "Abbreviations", showed the urgency and efficiency of military messages while communicating:

6-1 … ***focus adjust (FOC ADJ)****,* ***sight select (SGT SEL)****, and* ***filter select (FLTR SEL)*** *switches.*

6-2 … *course must be conducted* ***in accordance with (IAW)*** *the Javelin POI established by the US IS.*

6-3 … *gate* ***adjust*** *(GATE* ***ADJ****),* ***contrast and brightness (CTRS and BRT)****, and* ***attack select (ATTK SEL)*** *…*

6-4 … ***sight select (SGT SEL)****, and* ***filter select (FLTR SEL)*** *switches.*

6-5 *The gunner pushes the* ***attack select (ATTK SEL)*** *switch on the right handgrip to …*

Group 7. Finally, the categorisation, "Acronyms", showed combinations of words developed into terminologies, those may represent weapons (7-1), equipment (7-2, 7-3), and tactical terms (7-4, 7-5):

7-1 ***TOW**: tube-launched, optically tracked, wire-guided.*

7-2 *The day field-of-view (**FOV**) can be used for surveillance and target …*

7-3 ***BCU**: battery coolant unit.*

7-4 ***AAs**: avenues of approaches.*

7-5 ***OPFOR**: opposing force.*

To sum this section, the contributions of the proposed approach can be summarised as follows: (1) results of syntax analysis provide EFL troopers with syntax patterns that high frequently used in the target corpus for facilitating their military information reading and translating efficiency, (2) results of military TISL taxonomy enhance EFL troopers' TISL acquisition efficiency, and extract technical information in detail. The results presented in this paper indicate insights into the types of syntaxes and TISL used in US Army FMs of anti-tank weapon systems. The analysis made by expert assessment team enabled the results of a corpus-based approach based on linguistic and domain aspects. The findings reveal important pedagogic implications in military training courses at EFL military training facilities where US army anti-tank weapon systems are adopted by them.

CONCLUSION

Syntax analysis and vocabulary taxonomy are also critical for improving the accuracy and efficiency of corpus analysis and NLP. Military technical information is an uncommon scientific field; if the military simply seek linguists or information engineers' assistance in processing military information, the analytical results might be distorted especially some information seems insignificant but embedded deeply in domain knowledge.

Language is an important channel to communicate and to acquire information, but it evolves in different domains. Information processing programs based on certain algorithms may not handle complicated NLs' linguistic rules nor generate high precision resulting data to satisfy each domain. Corpus programs or NLP techniques are ideal toolkits, but machines are not always 100 % accurate, thus proper manual annotation is inevitable. The researchers integrated a corpus software, linguists, and domain experts' specialties to process information, and to make resulting data more meaningful and more applicable to military training purposes.

This paper highlights the value of compiling a narrow-angled specialised corpus to conduct syntax analysis and domain-oriented TISL taxonomy especially customised to address the needs in specific areas with the collaboration of linguists and domain experts, rather than conducting general linguistic analysis. More specifically, this paper suggests that when conducting corpus-based approaches in processing ESP cases, researchers should recruit domain experts to process the linguistic evidence of the specific corpus. In processing the corpus of US Army anti-tank weapon systems FMs, the proposed approach can consolidate and analyse the idiomatic syntaxes from the perspective of a linguist by clustering function words from the wordlist of the target corpus, and categorise military TISL from the perspective of military experts by cross checking wordlist and keyword list, checking LBs to retrieve terminological phrases, and checking concordance lines to retrieve complete terms of acronyms and abbreviations.

The proposed approach highlights the values of combination of a corpus-based approach and related experts' cooperation in data processing. The significant features can be summarised as follows: (1) the results of syntax analysis and military TISL taxonomy are more in accordance with EFL troopers' needs in learning military knowledge in English, (2) the proposed approach can integrate large amounts of domain texts and be smoothly utilised by linguist and military experts to conduct in-depth analysis and decipher during information processing, (3) the proposed approach adopts AntConc 3.5.8 (Anthony, 2019), free costs, open access, and with user-friendly operating platforms, to reduce the costs and enhance the efficiency of texts information processing; this approach especially suitable for military that has low defense budget to develop training materials.

In the future, the linguistic analytical results can become valuable reference data and criteria of TISL taxonomy and identification for improving the efficacy of ESP courses developments, and for enhancing the accuracy of corpus analysis to rapidly fetch key information.

Acknowledgement

The authors would like to thank the Ministry of Science and Technology, Taiwan, for financially supporting this research under Contract No. MOST HO-2410-H-145-001 and MOST 109-2410-H-145-002.

REFERENCES

Anthony L. (2019). AntConc (Version 3.5.8): Corpus Software. Available at *https://www.laurenceanthony.net/software/antconc/*, Accessed 30 April 2020.

Beller S. & Bender A. (2017). Theory, the final frontier? A corpus-based analysis of the role of theory in psychological articles. *Frontiers in Psychology* **8**: 951.
DOI: https://doi.org/10.3389/fpsyg.2017.00951

Brindle A. (2016). A corpus analysis of discursive constructions of the Sunflower Student Movement in the English-language Taiwanese press. *Discourse and Society* **27**(1): 3–19.
DOI: https://doi.org/10.1177/0957926515605957

Chen L.C. & Chang K.H. (2019). Creating a novel type of information database of military terminology: an example of US Army medical and casualty evacuation. *Basic and Clinical Pharmacology and Toxicology* **125**(S9): 8.
DOI: https://doi.org/10.1111/bcpt.13358

Chen L.C., Chang K.H. & Chung H.Y. (2020). A novel statistic-based corpus machine processing approach to refine a big textual data: an ESP case of COVID-19 news reports. *Applied Sciences-Basel* **10**(16): 5505.
DOI: https://doi.org/10.3390/app10165505

Cheng K.H. & Tsai C.C. (2019). A case study of immersive virtual field trips in an elementary classroom: Students' learning experience and teacher-student interaction behaviors. *Computers and Education* **140**: 103600.
DOI: https://doi.org/10.1016/j.compedu.2019.103600

Cho H. & Yoon H. (2013). A corpus-assisted comparative genre analysis of corporate earnings calls between Korean and native-English speakers. *English for Specific Purposes* **32**(3): 170–185.
DOI: https://doi.org/10.1016/j.esp.2013.03.001

Christ A., Penthin M. & Kroner S. (2019). Big data and digital aesthetic, arts, and cultural education: hot spots of current quantitative research. *Social Science Computer Review* **39**(5):821–843.
DOI: https://doi.org/10.1177/0894439319888455

Coats S. (2019). Articulation rate in American English in a corpus of youtube videos. *Language and Speech* **63**(4): 799–831.
DOI: https://doi.org/10.1177/0023830919894720

Daskalovska N. (2015). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning* **28**(2): 130–144.
DOI: https://doi.org/10.1080/09588221.2013.803982

Derbentseva N., Safayeni F. & Canas A. J. (2007). Concept maps: Experiments on dynamic thinking. *Journal of Research in Science Teaching* **44**(3): 448–465.
DOI: https://doi.org/10.1002/tea.20153

Ferguson G. (2001). If you pop over there: a corpus-based study of conditionals in medical discourse. *English for Specific Purposes* **20**(1): 61–82.
DOI: https://doi.org/10.1016/S0889-4

Flowerdew L. (2020). The Academic literacies approach to scholarly writing: a view through the lens of the ESP/Genre approach. *Studies in Higher Education* **45**(3): 579–591.
DOI: https://doi.org/10.1080/03075079.2019.1576165

Gonen S.I.K. (2019). A qualitative study on a situated experience of technology integration: reflections from pre-service teachers and students. *Computer Assisted Language Learning* **32**(3): 163–189.
DOI: https://doi.org/10.1080/09588221.2018.1552974

Green C. (2019). Enriching the academic wordlist and secondary vocabulary lists with lexicogrammar: toward a pattern grammar of academic vocabulary. *System* **87**: 102158.
DOI: https://doi.org/10.1016/j.system.2019.102158

Harvey-Scholes C. (2018). Computer-assisted detection of 90% of EFL student errors. *Computer Assisted Language Learning* **31**(1-2): 144–156.
DOI: https://doi.org/10.1080/09588221.2017.1392322

Henry A. & Roseberry R.L. (2001). A narrow-angled corpus analysis of moves and strategies of the genre: 'Letter of application'. *English for Specific Purposes* **20**(2): 153–167.
DOI: https://doi.org/10.1016/S0889-4906(99)00037-X

Hsieh J.S.C., Wu W.C.V. & Marek M.W. (2017). Using the flipped classroom to enhance EFL learning. *Computer Assisted Language Learning* **30**(1-2): 1–21.
DOI: https://doi.org/10.1080/09588221.2015.1111910

Kim J.E. & Nam H. (2019). How do textual features of L2 argumentative essays differ across proficiency levels? A multidimensional cross-sectional study. *Reading and Writing* **32**(9): 2251–2279.
DOI: https://doi.org/10.1007/s11145-019-09947-6

Koops C. & Lohmann A. (2015). A quantitative approach to the grammaticalization of discourse markers evidence from their sequencing behavior. *International Journal of Corpus Linguistics* **20**(2): 232–259.
DOI: https://doi.org/10.1075/ijcl.20.2.04koo

Lai C., Yeung Y. & Hu J.J. (2016). University student and teacher perceptions of teacher roles in promoting autonomous language learning with technology outside the classroom. *Computer Assisted Language Learning* **29**(4): 703–723.
DOI: https://doi.org/10.1080/09588221.2015.1016441

Li L.Q., Zhao J., Hou L., Zhai Y.K., Shi J.M. & Cui F.F. (2019). An attention-based deep learning model for clinical named entity recognition of Chinese electronic medical records. *BMC Medical Informatics and Decision Making* **19**(1): Article Number 235.
DOI: https://doi.org/10.1186/s12911-019-0933-6

Li Q., Li S.B., Zhan S., Hu J. & Hu J.J. (2019). A review of text corpus-based tourism big data mining. *Applied Sciences-Basel* **9**(16): Article Number 3300.
DOI: https://doi.org/10.3390/app9163300

Li S.L. (2016). A corpus-based study of vague language in legislative texts: strategic use of vague terms. *English for Specific Purposes* **45**: 98–109.
DOI: https://doi.org/10.1016/j.esp.2016.10.001

Liu J. & Han L.N. (2015). A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes* **39**: 1–11.
DOI: https://doi.org/10.1016/j.esp.2015.03.001

Munoz V.L. (2015). The vocabulary of agriculture semi-popularization articles in English: A corpus-based study. *English for Specific Purposes* **39**: 26–44.
DOI: https://doi.org/10.1016/j.esp.2015.04.001

Nekrasova-Beker T.M. (2019). Discipline-specific use of language patterns in engineering: A comparison of published pedagogical materials. *Journal of English for Academic Purposes* **41**: 100774.

DOI: https://doi.org/10.1016/j.jeap.2019.100774

Shanahan T. & Shanahan C. (2017). Disciplinary literacy: Just the FAQs. *Educational Leadership* **74**(5): 18–22.

Sholokhovnn A., Kinnunen T., Vestman V. & Lee K. A. (2020). Voice biometrics security: Extrapolating false alarm rate via hierarchical Bayesian modeling of speaker verification scores. *Computer Speech and Language* **60**: 101024. DOI: https://doi.org/10.1016/j.csl.2019.101024

Siefridt C., Grosjean J., Lefebvre T., Rollin L., Darmoni S. & Schuers, M. (2020). Evaluation of automatic annotation by a multi-terminological concepts extractor within a corpus of data from family medicine consultations. *International Journal of Medical Informatics* **133**: 104009. DOI: https://doi.org/10.1016/j.ijmedinf.2019.104009

Smalheiser N.R Luo, M.Q.Addepalli S. & Cui X. K. (2019). A manual corpus of annotated main findings of clinical case reports. *Database-The Journal of Biological Databases and Curation* **2019**: 143. DOI: https://doi.org/10.1093/database/bay143

Trembach S. (2019). From information to knowledge to wisdom: the cold war battle for information superiority and its implications for thriving in the age of data smog. *LIBRI* **69**(1): 1–12. DOI: https://doi.org/10.1515/libri-2018-0008

US Department of the Army (2001). *Light Antiarmor Weapons*: FM 3-23.25. US Department of the Army, Washington DC, USA.

US Department of the Army (2003). TOW *weapon system*: FM 3-22.34. US Department of the Army, Washington DC, USA.

US Department of the Army (2008). *Javelin - close combat missile system, medium*: FM 3-22.37. US Department of the Army, Washington DC, USA.

Viswanathan U.M., Sultana A.S. & Shankar S. (2020). Linguistic resources to facilitate science education *Current Science* **118**(2): 271–273. DOI: https://doi.org/10.18520/cs/v118/i2/271–273

Zygouris-Coe V.I. (2012). Disciplinary literacy and the common core state standards. *Topics in Language Disorders* **32**(1): 35–50. DOI: https://doi.org/10.1097/TLD.0b013e31824561a

J.Natn.Sci.Foundation Sri Lanka 2021 49 (3): 419 - 427
DOI: http://dx.doi.org/10.4038/jnsfsr.v49i3.9974

RESEARCH ARTICLE

Paraoxonase 1 phenotype distribution in a cohort of healthy Sri Lankan population

S Marasinghe[1*] and R Sivakanesan[2]
[1]*Postgraduate Institute of Science, University of Peradeniya, Peradeniya*
[2]*Department of Biochemistry, Faculty of Medicine, University of Peradeniya, Peradeniya.*

Submitted: 15 May 2020; Revised: 15 May 2021; Accepted: 21 May 2021

Abstract: Human serum paraoxonase 1(PON1) is an enzyme synthesised mainly by the liver. Paraoxonase 1 Q192R polymorphism involves amino acid substitution of glutamine (Q isoform) to arginine (R isoform) at position 192. Paraoxonase 1 activity and polymorphism is associated with many disease conditions. In the Sri Lankan context, data is lacking on the distribution of Paraoxonase 1 phenotypes. The aim of this study was to assess the phenotypic distribution of Paraoxonase 1 in a cohort of healthy Sri Lankan individuals, using the dual substrate method. Serum samples of 77 apparently healthy individuals were used for the study. Paraoxonase activity was measured by a kinetic method using paraoxon as the substrate at 412 nm and 25° C. Salt stimulated paraoxonase 1 activity was measured in the presence of 1M NaCl. Arylesterase activity was measured by a kinetic method using phenylacetate as the substrate at 270 nm and 25° C. The ratio of salt stimulated paraoxonase activity to arylesterase activity was used to assign phenotypes. Basal and salt stimulated paraoxonase activities were bimodally distributed. Arylesterase activity was unimodally distributed. Salt stimulated paraoxonase activity/ arylesterase activity ratio was trimodally distributed. The three modes corresponded to QQ (lower activity), QR (intermediate activity) and RR (higher activity) phenotypes. The percentage distribution of QQ, QR and RR phenotypes were 36 %, 51 % and 13 %, respectively. This study has set the baseline data on phenotypic distribution of paraoxonase1 in a cohort of healthy Sri Lankan individuals.

Keywords: Arylesterase, distribution, paraoxonase, phenotypes, polymorphism.

INTRODUCTION

Human serum paraoxonase 1(PON1; EC3.1.8.1) is a glycoprotein with a molecular mass of 43kDa. It is synthesised mainly in the liver and contains 354 amino acids. In blood, it is associated with Apo-A component of high density lipoprotein (HDL) (Gaidukov & Thawfik, 2005). PON1 is a calcium-dependent esterase which shows both paraoxonase and arylesterase activity (Mackness *et al*., 2001). It has the ability to hydrolyse organophosphates (Mackness *et al*., 1998) and biologically active lipoperoxides (Hashemi *et al*., 2011), rendering a protective effect against organophosphate toxicity and formation of atherosclerotic plaques. Recent studies have revealed its ability to hydrolyse homocystein thiolactone, a risk factor for cardiovascular diseases (Zafiropoulos *et al*., 2010). Cardio protective characteristic of PON1 is supported by several studies conducted *in vitro* (Mackness *et al*., 1991, González *et al*., 2019; Grzegorzewska *et al*., 2021) and *in vivo* (Shih *et al*., 1998).

Serum PON 1 activity significantly differs among individuals with a 40-fold variation (Mackness &

* Corresponding author (sanjila.marasinghe@gmail.com https://orcid.org/:0002-6156-2870)

Mackness, 2015). These differences are modulated by hereditary and acquired factors (Draganov & La Du, 2004). The PON1 activity is influenced by diet, drugs, alcohol consumption, smoking and environmental toxins. Serum PON1 activity was reported to be decreased by intake of high fat diet (Sutherland *et al*., 1999). The intake of antioxidant polyphenols such as quercetin and glabridin was associated with increased PON1 activity (Aviram *et al*., 1999). Few studies reported the ability of statins and fibrates to increase PON1 activity (Balogh *et al*., 2001; Paragh *et al*., 2003). However, contrasting results were observed in some studies, emphasising the inability of statins and fibrates to influence PON1 activity (Durrington *et al*., 1998). Consumption of 40 g/day of alcohol was associated with increased PON1 activity and mass in mg/L (van der Gaag *et al*., 1999). A study has revealed the inhibitory effect of smoking on PON1 activity (Nishio & Watanabe, 1997). In another study, serum PON 1 concentration and activity were decreased due to smoking (James *et al*., 2000). The PON1 activity was inhibited by organophosphates (Sözmen *et al*., 2002).

The PON1 gene is located on the long arm of chromosome 7q21-22. The 192 R/Q polymorphism of human serum PON1 has two isoforms, namely, PON1 Q and PON1 R (Mackness & Mackness, 2015). The PON1 Q contains a glutamine at position 192 and it shows low activity towards paraoxon hydrolysis. PON1 R contains an arginine at position 192 (Deakin & James, 2004). It shows a six-fold higher activity towards paraoxon hydrolysis compared to Q isoform. But the arylesterase activity is similar in both isozymes (Gaidukov *et al*., 2006). The R allozyme shows a greater degree of stimulation of its paraoxon-hydrolyzing activity in 1 M NaCl than the Q allozyme (Eckerson *et al*., 1983a). This qualitative property can be used to identify two different allozymes based on their response to salts. Eckerson *et al* (1983b) proposed that heterozygous PON1 phenotype (QR) can be obviously differentiated from both homozygous phenotypes (QQ and RR) on the basis of its ratio of salt stimulated paraoxonase activity to arylesterase activity. The ratio of salt-stimulated PON1 activity to arylesterase activity (P/A ratio) was trimodally distributed (Eckerson *et al*., 1983b). The trimodal distribution corresponds to QQ, QR and RR phenotypes. The least frequent value between two modes is called the antimode. The population was divided at the two antimodes to segregate the three phenotypes (QQ, QR and RR).

PON1 activity is found to be reduced in many disease conditions such as cardiovascular diseases (González *et al*., 2019; Murillo-González *et al*., 2020; Grzegorzewska *et al*., 2021), thyroid disorders (Azizi *et al*., 2003), cancers (El-Lebedy *et al*. 2014), diabetes mellitus (Savu *et al*., 2014), chronic renal failure (Prakash *et al*., 2010), chronic liver damage (Ferre *et al*., 2002) and psychiatric disorders (Moreira *et al*., 2019). The comparison of PON1 activity and/ polymorphism between healthy individuals and patients require baseline data of healthy individuals. In the Sri Lankan context, data of PON1 activity and distribution of PON1 phenotypes among healthy individuals are lacking. The aim of this study was to evaluate the PON1 activity and phenotype distribution in a cohort of healthy Sri Lankan individuals using the dual substrate method.

MATERIALS AND METHODS

The study protocol and procedures were approved by the Ethics Review Committee, Postgraduate Institute of Science, University of Peradeniya on 10th November 2013. Informed consent was obtained from all participants before commencing the study. The study was conducted in the Department of Biochemistry, Faculty of Medicine, University of Peradeniya.

Study subjects

Seventy seven apparently healthy individuals without any history of hypertension, smoking, cardiovascular diseases, diabetes, renal or hepatic diseases and cancer were enrolled for the study. The study group comprised 43 females and 34 males with a mean age of 57.82 ± 16.38 years. The number of subjects recruited was based on previous studies conducted as described below. Organophosphate intoxication on human serum paraoxonase was studied in 28 organophosphate poisoning patients with 66 control subjects by Sözmen *et al*. (2002). Species and substrate specific stimulation of human plasma paraoxonase 1 by high chloride concentration was conducted in 15 male healthy volunteers by Bełtowski *et al*. (2002). The study of Paraoxonase -1 gene polymorphism in a healthy population of Khorramabad, Iran involved 64 healthy individuals (Chehari *et al*., 2014).

Sample collection

Blood was collected by venipuncture with minimal stasis from individuals who have been resting for at least 20 minutes prior to collection of blood. Precautions were taken to avoid haemolysis of blood. Blood was collected with only moderate suction. The needle was always removed from the syringe before emptying it slowly into

the collecting tube. The samples were not collected in EDTA containing vacutainers as paraoxonase activity is irreversibly inactivated by the chelation of Ca. Blood was allowed to clot, and the serum was separated by centrifuging at 4000 rpm for 10 min. The serum samples were stored frozen at -20 ˚C and assayed within two weeks of collection. The samples were completely thawed before the assay.

Paraoxonase activity assay

Serum paraoxonase activity was estimated according to the protocol described by Eckerson *et al*. (1983a). The rate of hydrolysis of paraoxon (diethyl-p-nitrophenyl phosphate) was measured by monitoring the increase in absorbance at 412 nm at 25˚ C for 75 s in 15 s intervals. The basal assay mixture included 1.0 mM paraoxon and 1.0 mM $CaCl_2$ in 0.05 M glycine buffer at a pH of 10.5. One unit of paraoxonase activity was defined as 1 µmol of p-nitrophenol formed per minute and activity was expressed as U/L of serum. The liberated amount of p-nitrophenol was calculated using molar extinction coefficient of p-nitrophenol (18290 $M^{-1}cm^{-1}$). Salt stimulated PON1 activity was measured in the presence of 1 M NaCl. A blank without serum was used to correct any spontaneous hydrolysis of the substance.

Arylesterase activity assay

Arylesterase activity was measured using phenylacetate as the substrate (Josse *et al*., 1999). The change in absorbance was measured at 270 nm for 75 s in 15 s intervals. The amount of phenol formed was calculated using molar extinction coefficient (1310 $M^{-1}cm^{-1}$) for phenol. One unit of arylesterase activity was defined as 1 µmol phenol generated/min at a pH of 8.0 and at 25˚ C. The activity is expressed as kU/L. Paraoxon, phenylacetate, calcium chloride, and sodium choloride were purchased through local suppliers from Sigma-Aldrich (St. Louis, Missouri, United States).

PON1-Q192R phenotyping

PON1-Q192R phenotype distributions were determined using both paraoxonase and arylesterase activity according to the method of Eckerson *et al*. (1983b). Paraoxon is identified as a distinguishing substrate with a polymorphic distribution of activity, and phenylacetate is a non-distinguishing substrate for two allozymes. The distribution normality of variables was assessed using Kolmogorov-Smirnov test. Univariate density estimation was performed using R statistical software (R 3.4.2). Each variable was fitted a mixture of normal distributions with model-based clustering.

Statistical analysis

Statistical analysis was performed using MINITAB16 software (MINITAB inc., State College, PA, USA). A p value < 0.05 was considered as statistically significant. All values were presented as mean ± SD. The range of each variable was calculated by dividing the lowest observed value from the highest observed value. The significance of differences between groups was assessed by using independent Students' t-test, Mann Whitney test and ANOVA. The significance of association between variables was evaluated by using chi-square test.

RESULTS AND DISCUSSION

The cohort comprised 55.8 % females and 44.2 % males. Mean age of the participants was 57.82 ± 16.38 years. There was no significant difference in age between males and females. The mean basal paraoxonase activity was 205.27 ± 115.00 U/L, ranging from 34.54 U/L to 628.30 U/L (Figure 1). The mean salt stimulated PON1 activity was 320.4 ± 218.9 U/L, ranging from 50.3 U/L to 1233.6 U/L (Figure 2). The mean arylesterase activity was 159.53 ± 37.11 kU/L, ranging from 56.2 kU/L to 253.9 kU/L (Figure 4). The range for basal PON1 activity, salt stimulated activity, and arylest erase activity among the population was 593.8, 118.3, and 179.7 respectively .The basal paraoxonase (Figure 1) and salt stimulated paraoxonase (Figure 2) activities were bimodally distributed, whereas arylesterase activity (Figure 4) was unimodally distributed. Degree of stimulation of PON1 activity by 1 M NaCl was normally distributed (Figure 3).

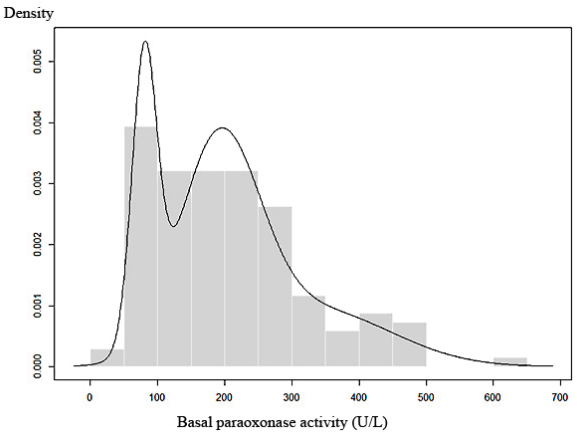


Figure 1: Estimated density curve of basal paraoxonase activity in healthy individuals

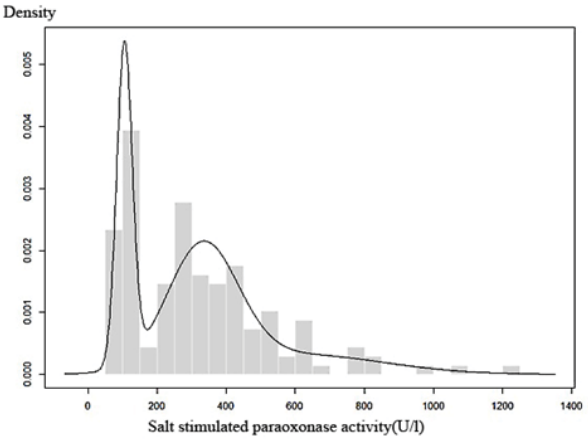


Figure 2: Estimated density curve for salt stimulated paraoxonase activity in healthy individuals

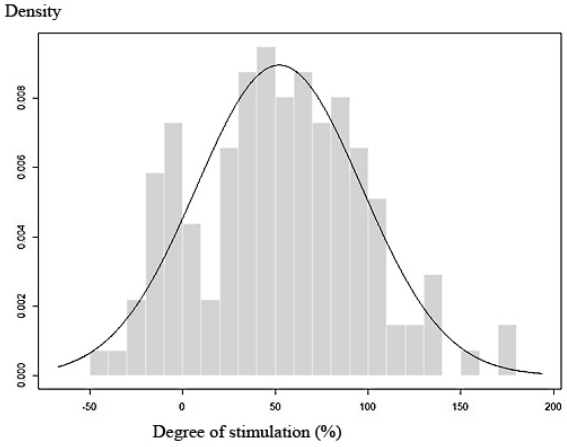


Figure 3: Estimated density curve for degree of stimulation of paraoxonase activity by 1M NaCl in healthy individuals

The addition of NaCl caused approximately 1.52 ± 0.44 fold increment in PON1 activity in the range of 0.55 to 2.72. However, the addition of NaCl also caused an inhibition of PON1 activity in twenty-three individuals, ranging from 0.63% to 45.4%.

Eckerson *et al*. (1983a) also reported that the degree of salt stimulation separated individuals into two very discrete classes; one stimulated -23% to 35% (the non-salt responsive type), and the other stimulated 60 % and greater (the salt-responsive type).

Frequency distribution histograms for percent stimulation of paraoxonase activity by 1.0 M NaCl was obtained by Eckerson *et al*. (1983a), and the present study included both non-salt responsive and salt-responsive types, which implies that both attributes to salt response should be considered for phenotype categorisation.

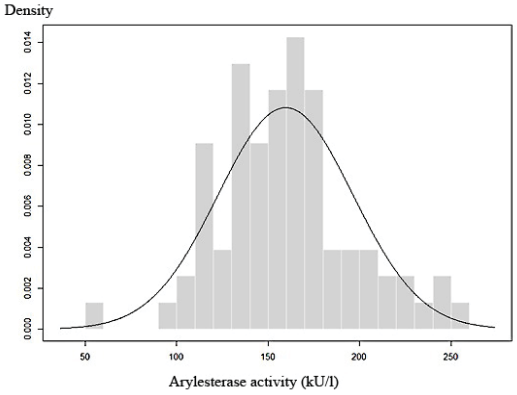


Figure 4: Estimated density curve of arylesterase activity in healthy individuals

We observed the relationship between age and salt responsiveness of an individual. The number of individuals whose PON1 activity was inhibited by the addition of NaCl was higher in the younger age group (18–54 years) than in the older age group (54–90 years). The number of individuals whose PON1 activity was stimulated by the addition of NaCl was higher in the older age group (54–90 years) than in the younger age group (18–54 years) (p = 0.005) (Table 1).

Table 1: Number of individuals whose PON1 activity was inhibited / stimulated by addition of 1M NaCl according to age

Age group (years)	Number of individuals whose PON1 activity was inhibited by NaCl	Number of individuals whose PON1 activity was stimulated by NaCl
18–54	17	48
54–90	6	66

There were no significant differences in basal and salt stimulated PON1 activity and P/A ratio between males and females. Arylesterase activity was significantly higher in females than in males (Table 2).

Table 2: Paraoxonase and arylesterase activities of healthy Sri Lankan individuals according to gender

Parameters	Female (n=87)	Male (n=50)	p value
Basal paraoxonase activity (U/L)	212.3 ± 116.7	193.1 ± 112.1	0.344
Salt stimulated PON1 activity (U/L)	331.2 ± 226.7	301.6 ± 205.4	0.436
Arylesterase activity (kU/L)*	168.2 ± 39.9	148.5± 30.3	0.029
P/A ratio	2.07 ± 1.18	1.79 ± 1.23	0.059

p value < 0.05. * Significantly different between males and females

The ratio of salt stimulated paraoxonase activity to arylesterase activity disclosed a trimodal distribution in the studied cohort (Figure 5). The three modes corresponded to QQ, QR and RR phenotypes. Subjects with salt stimulated PON1 activity to arylesterase activity ratio < 1.5 were classified as the QQ group (homozygous low activity) (n = 28), those with ratios between 1.5 and 3.5 as the QR group (heterozygous moderate activity) (n = 39), and those with ratios > 3.5 (n = 10) as the RR group (homozygous high activity). The percentage distribution of QQ, QR and RR phenotypes were 36 %, 51 % and 13 %, respectively.

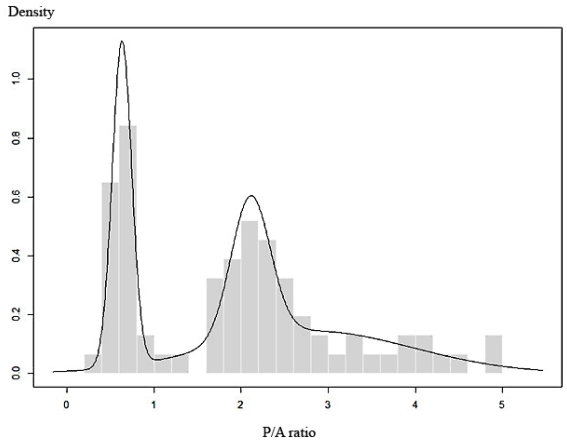


Figure 5: Estimated density curve of P/A ratio in healthy individuals,

Means of basal PON1 activity ($p < 0.001$), salt stimulated PON1 activity ($p < 0.001$), and P/A ratio ($p < 0.001$) were significantly different among the 3 phenotypes (Table 3). The above three parameters were in the order of QQ < QR < RR phenotypes. Mean arylesterase activity was significantly higher in QQ phenotype compared to QR phenotype ($p = 0.034$). There were no significant differences in mean arylesterase activity between QR and RR phenotypes and QQ and RR phenotypes. Mean basal PON1 activity of RR phenotype was 3-fold higher than the mean basal PON1 activity of QQ and mean salt stimulated PON1 activity of RR phenotype was 5 fold higher than the mean salt stimulated PON1 activity of QQ.

Table 4 depicts the differences in mean PON1 and arylesterase activities between males and females, according to PON1 phenotypes. Arylesterase activity was significantly higher in females than in males in QQ phenotype. P/A ratio was significantly higher in males compared to females in QQ phenotype. Basal PON 1 activity, salt stimulated PON1 activity and arylesterase activity was significantly higher in females compared to males in QR phenotype. Significant differences were not observed in all four parameters between males and females in RR phenotype.

Table 3: Paraoxonase and arylesterase activities in healthy individuals according to phenotype

Parameter	Phenotype		
	QQ (n=28)	QR (n=39)	RR (n=10)
Basal PON1 activity (U/L)[1]	115.3 ± 76.2	218.0 ± 78.0	382.3 ± 113.3
Salt stimulated PON1 activity (U/L)[2]	126.6 ± 56.8	344.4 ± 105.3	670.2 ± 231.8
Arylesterase activity (kU/L)	173.7 ± 36.6[a]	150.1 ± 34.9[b]	156.8 ±38.1[a,b]
P/A ratio[3]	0.67 ± 0.19	2.28 ± 0.43	4.20 ± 0.48

[1,2,3] Significantly different among phenotypes $p < 0.001$; [a, b] different superscript in the row for Arylesterase activity indicate significantly different, $p = 0.034$

PON1 is an antioxidant enzyme with paraoxonase, esterase and lactonase activity. PON1 is involved in metabolism of many drugs containing lactone or cyclic carbonate moieties. However, natural substrate of PON1 inside human body is still unclear (Draganov & La Du, 2004). Numerous studies have revealed associations between PON1 activity and various disease conditions (Aldonza *et al*., 2017; Ertürk *et al*., 2017; Wei *et al*., 2017; Passaro *et al*., 2018; Moreira *et al*., 2019; Matsumoto *et al*., 2020; Murillo-González *et al*., 2020). Determination of PON1 activity and phenotypic distribution in healthy individuals may facilitate the unveiling of any possible relationships between PON1 activity and/or phenotypic distribution and particular disease conditions. Identification of relationships between PON1 activity/polymorphism and disease conditions may improve the existing knowledge on pathophysiology of disease conditions. The ability of paraoxonase in destroying biologically active lipids in mildly oxidised LDL was first documented by Watson *et al*. (1955). The inhibitory effect of PON on HDL oxidation was established based on the reduced HDL peroxide and aldehyde formation (Aviram *et al*., 1998). Since then, numerous studies have been carried out on the cardio-protective role of PON. A recent review concluded that the true physiological substrates for PON are still not known (Taler-Verčič *et al*., 2020). As such, further studies on the enzymes' molecular mechanism may help to identify possible natural substrates for PON1. Modulation of PON1 activity by pharmacological, environmental and behavioural interventions may exert a beneficial therapeutic approach to hinder the detrimental effects of certain diseases, which exhibit an association with serum levels of PON1 within the body.

Table 4: Paraoxonase and aryesterase activities according to phenotypes and gender

Parameter	Phenotype	Male	Female	p value
Basal PON 1 activity (U/L)	QQ	104.9 ± 46.6	129.2± 104.4	0.926
	QR[1]	182.5 ± 42.5	235.8 ± 86.0	0.030
	RR	347.2 ± 94.0	417.4 ± 130.2	0.531
Salt stimulated PON1 activity (U/L)	QQ	122.4 ± 51.4	132.1 ± 65.1	0.871
	QR[2]	296.0 ± 52.5	368.7 ± 117.0	0.031
	RR	597.1 ± 139.6	743 ± 297	0.676
Arylesterase activity (kU/L)	QQ[3]	161.44 ± 33.78	190 ± 35	0.039
	QR[4]	134.02 ± 21.06	158.12 ± 37.91	< 0.001
	RR	144.9 ± 24.2	168.7 ± 48.3	0.403
P/A ratio	QQ[5]	0.72 ± 0.21	0.62 ± 0.16	0.048
	QR	2.22 ± 0.3	2.31 ± 0.48	0.743
	RR	4.10 ± 0.51	4.31 ± 0.48	0.531

Values are represented as mean ± SD. [1,2,3,4,5] Significantly different between males and females

In this study, salt stimulated PON1 and arylesterase activities were measured to determine the phenotypic distribution of PON1 in a cohort of healthy Sri Lankan individuals. Based on the salt-stimulated PON1 to arylesterase activity ratio, it was possible to distinguish the three paraoxonase phenotypes (QQ, QR and RR). The percentage distribution of QQ, QR and RR phenotypes in our cohort was 36 %, 51 % and 13 % respectively. The findings were compared with the results from a Croatian cohort, where 39 %, 48 % and 13 % individuals belonged to QQ, QR and RR phenotypes, respectively (Juretiæ *et al*., 2001). Close values were observed in another study conducted in Iran, in which frequencies of QQ, QR and RR phenotypes were reported to be 48.1 %, 41.3 % and 10.6 %, respectively (Sepahvand *et al*., 2007). The genotype frequencies for paraoxonase 1-Q192R were reported as 47 % (QQ), 41 % (QR) and 12 % (RR) in a healthy population of Khorramabad, Iran by Chehari *et al*. (2014), which was almost similar to that of reported by Sepahvand *et al*. (2007). Contrasting results were observed in a Thai population with 14.4 %, 51.9 % and 33.7 % of individuals belonging to QQ, QR and R, respectively (Porntadavity *et al*., 2009).

Mean PON1 activity of the studied Sri Lankan cohort was found to be 205.27 ± 115.00 U/L ranging from 34.54 U/L to 628.30 U/L. Basal PON1 activity exhibited close resemblance to values reported in a Croatian population (251 ± 143 U/L) [Juretiæ *et al*., 2001], with an interquartile range of 236 and a Thai populations 239.7 ± 83.9 U/L [Porntadavity *et al*., 2009]. Mean basal paraoxonase activity in the present study was considerably higher compared to those observed in many other populations including Iran [81.8 ± 57.0 U/L, ranging from 19.2 to 290 U/L) (Sepahvand *et al*., 2007) and 98.79 ± 68.79 U/L (Naderi *et al*., 2011)], Czech Republic [125.2 ± 69.4 U/L (Novak *et al*., 2010)], Bulgaria [128.79 ± 15.1 U/L (Doneva-Basheva *et al*., 2013)], Hungary [188 ± 55 U/L (Paragh *et al*., 2002)] and Turkey [178 ± 79 U/L (Dirican *et al*., 2004)]. A study conducted by Elkiran *et al*. (2007) in Turkey revealed a higher value (395.8 ± 116.6 U/L) than our finding for basal PON1 activity.

We observed significantly higher arylesterase activities in females than in males. However, opposite results were reported in a Thai study (Female: 137.0 ± 25.7 vs. Male: 160.0 ± 37.9) (Porntadavity *et al*., 2009), and no differences were reported in an Iranian study (Female: 79.8 ± 11.9 vs. Male: 81.6 ± 17) (Sepahvand *et al*., 2007). The significant differences among the three phenotypes in basal PON1 activity, salt stimulated PON1 activity and P/A ratio in our study was similar to the observation made on the three phenotypes in Thai study (Porntadavity *et al*., 2009). This should be reworded as 'The range observed in basal PON1 activity and salt stimulated PON1 activity in our population was higher than that of

Thai population with basal PON1 activity of 384.5 U/L and salt stimulated PON1 1022.2 U/L' (Porntadavity *et al*., 2009), whereas range in arylesterase activity was similar in both populations. Salt responsiveness of PON1 in Thai population (2.3 fold) (Porntadavity *et al*., 2009) was moderately higher than our population (1.52 ± 0.44). In our study, we found significantly higher arylesterase activity in females compared to males in the QQ phenotype. The P/A ratio was significantly higher in males compared to females in the QQ phenotype. Mean arylesterase activity was significantly higher in males (176.2 ± 41.3) than in females (146.4 ± 25.0) in the QQ phenotype in the Thai population. We observed significantly higher basal PON1 activity, salt stimulated PON1 activity and arylesterase activity in females compared to males in the QR phenotype. However, significantly higher mean salt stimulated PON1 activity (560.5 ± 157.5 Vs 469.4 ± 113.8) and mean arylesterase activity (171.8 ± 41.1Vs 136.4 ± 28.4) were observed in males than females in QR phenotype in the Thai population. Hence, there are contrasting differences in PON and arylesterase activities among the phenotypes in different populations.

The ability of PON1 to hydrolyse organophosphates varies depending on the PON1 phenotype. Previous studies have revealed that the QR and RR phenotypes are significantly more resistant to OP toxicity compared to the QQ phenotype (Sepahvand *et al*., 2007). Hence, we can suggest that 64 % of our population exhibits a resistance against OP toxicity. Sri Lanka is an agricultural country which uses considerable amounts of OP substances as pesticides and herbicides. Therefore, people may have genetically adapted to resist toxic effects of OP substances. Studies assessing the correlation between PON1 phenotype and atherosclerosis risk have revealed that QQ phenotype is more cardio protective compared to other two phenotypes (Sepahvand *et al*., 2007). According to the results of our study we can hypothesise that majority of our population doesn't have this protective effect and may be susceptible to develop atherosclerosis. A study is in progress with acute coronary syndrome to verify our hypothesis.

CONCLUSION

In this study, basal and salt stimulated paraoxonase activities were bimodally distributed. Arylesterase activity was unimodally distributed. Salt stimulated paraoxonase activity/ arylesterase activity ratio was trimodally distributed. The three modes corresponded to QQ (lower activity), QR (intermediate activity) and RR (higher activity) phenotypes. The percentage distribution of QQ, QR and RR phenotypes were 36 %, 51 % and 13 %, respectively. This study has set the baseline data on phenotypic distribution of paraoxonase1 in a cohort of healthy Sri Lankan individuals. The findings may contribute in future studies involving assessment of PON1 activity and/or polymorphism in various disease conditions.

REFERENCES

Aldonza M.B.D., Son Y.S., Sung H.J., Ahn J.M., Choi Y.J., Kim Y.I., Cho S. & Cho J.Y. (2017). Paraoxonase-1 (PON1) induces metastatic potential and apoptosis escape via its antioxidative function in lung cancer cells. *Oncotarget* **8**(26): 42817.
DOI: https://doi.org/10.18632/oncotarget.17069

Aviram M., Rosenblat M., Bisgaier C.L., Newton R.S., Primo-Parmo S.L. & La Du B.N. (1998). Paraoxonase inhibits high-density lipoprotein oxidation and preserves its functions. A possible peroxidative role for paraoxonase. *Journal of Clinical Investigation* **101**(8): 1581–1590.
DOI: https://doi.org/10.1172/JCI1649

Aviram M., Rosenblat M., Billecke S., Erogul J., Sorenson R., Bisgaier C.L., Newton R.S. & La Du B. (1999). Human serum paraoxonase (PON1) is inactivated by oxidised low density lipoprotein and preserved by antioxidants. *Free Radicals Free Radical Biology and Medicine* **26**: 892–904.
DOI: https://doi.org/10.1016/s0891-5849(98)00272-x

Azizi F., Raiszadeh F., Solati M., Etemadi A., Rahmani M. & Arabi M. (2003). Serum paraoxonase 1 activity is decreased in thyroid dysfunction. *Journal of Endocrinological Investigations* **26**(8): 703–709.
DOI: https://doi.org/10.1007/BF03347350

Balogh Z., Seres I., Harangi M., Kovacs P., Kakuk G. & Paragh G. (2001). Gemfibrozil increases paraoxonase activity in type 2 diabetic patients: a new hypothesis of the beneficial action of fibrates? *Diabetes and Metabolism* **27**: 604–610.

Bełtowski J., Wójcicka G. & Marciniak A. (2002). Species- and substrate-specific stimulation of human plasma paraoxonase 1 (PON1) activity by high chloride concentration. *Acta Biochimica Polonica* **49**(4): 927–936.
DOI: https://doi.org/10.18388/abp.2002_3752

Chehari K., Sepahvand F., Ghobadi S., Ismaili A. & Alavy E. (2014). Study of Paraoxonase -1 Gene Polymorphism in a Healthy Population of Khorramabad, Iran. *Journal of Applied Biotechnology Reports* **1**(2): 85–88.

Deakin S.P. & James R.W. (2004). Genetic and environmental factors modulating serum concentrations and activities of the antioxidant enzyme paraoxonase-1. Clinical Science 107(5):435-47.

DOI: https://doi.org/10.1042/CS20040187

Dirican M., Akça R., Sarandöl E. & Dilek K. (2004). Serum paraoxonase activity in uremic predialysis and hemodialysis patients. *Journal of Nephrology* **17**: 813–818.

Doneva-Basheva K., Anastasov A., Postadzhyan A., Kamenova Z. & Vlaykova T. (2013). Serum paraoxonase and arylesterase activity of PON1 in acute coronary syndrome. *Trakia Journal of Sciences* **1**: 39–49.

Draganov D.I. & La Du B.N. (2004). Pharmacogenetics of paraoxonases: a brief review. *Naunyn-Schmiedeberg's Archives of Pharmacology* **369**: 78–88.
DOI: https://doi.org/10.1007/s00210-003-0833-1

Durrington P.N., Mackness M.I., Bhatnagar D., Julier K., Prais H., Arrol S., Morgan J. & Wood G.N. (1998) Effects of two different fibric acid derivatives on lipoproteins, cholesteryl ester transfer, fibrinogen, plasminogen activator inhibitor and paraoxonase activity in type IIb hyperlipoproteinaemia. *Atherosclerosis* **138**: 217–225.
DOI: https://doi.org/10.1016/s0021-9150(98)00003-3

Eckerson H.W., Romson J., Wyte C. & La Du B.N. (1983a). The human serum paraoxonase polymorphism: Identification of phenotypes by their response to salts. *American Journal of Human Genetics* **35**: 214–227.

Eckerson H.W., Wyte C.M. & La Du B.N. (1983b). The human serum paraoxonase/ arylesterase polymorphism. *American Journal of Human Genetics* **35**: 1126–1138.

Elkiran E.T., Mar N., Aygen B., Gursu F., Karaoglu A. & Koca S. (2007). Serum paraoxonase and arylesterase activities in patients with lung cancer in a Turkish population. *BMC Cancer* **7**: 48.
DOI: https://doi.org/10.1186/1471-2407-7-48

El-Lebedy D., Kafoury M., Abd-El Haleem D., Ibrahim A., Awadallah E. & Ashmawy I. (2014). Paraoxonase-1 gene Q192R and L55M polymorphisms and risk of cardiovascular disease in Egyptian patients with type 2 diabetes mellitus. *Journal of Diabetes and Metabolic Disorders* **13**: 125–131.
DOI: https://doi.org/10.1186/s40200-014-0125-y

Ertürk C., Altay M.A., Bilge A. & Celik H. (2017). Is there a relationship between serum ox-LDL, oxidative stress, and PON1 in knee osteoarthritis? *Clinical Rheumatology* **36**: 2775–2780.
DOI: https://doi.org/10.1007/s10067-017-3732-4

Ferre N., Camps J., Prats E., Vilella E., Paul A., Figuera L. & Joven J. (2002). Serum paraoxonase activity: a new additional test for the improved evaluation of chronic liver damage. *Clinical Chemistry* **48** (2): 261–268.
DOI: https://doi.org/10.1093/clinchem/48.2.261

Gaidukov L., Rosenblat M., Aviram M., Tawfik D.S. (2006). The 192R/Q polymorphs of serum paraoxonase PON1 differ in HDL binding, lipolactonase stimulation and cholesterol efflux. *Journal of Lipid Research* **47**: 2492–2502.
DOI: https://doi.org/10.1194/jlr.M600297-JLR200

Gaidukov L. & Thawfik D.S. (2005). High affinity, stability, and lactonase activity of serum paraoxonase PON1 anchored on HDL with ApoA-I. *Biochemistry* **44** (35): 11843–11854.
DOI: https://doi.org/10.1021/bi050862i

González F., Ponce-RuÍz, N., Rojas-GarcÍa A.E., Bernal-Hernández Y.Y., Mackness M., Ponce-Gallegos J., Cardoso-Saldaña G., Jorge-Galarza E., Torres-Tamayo M. & Medina-Díaz I.M. (2019). PON1 concentration and high-density lipoprotein characteristics as cardiovascular biomarkers. Archives of Medical Sciences. *Atherosclerotic Diseases* **4**: e47–e54.
DOI: https://doi.org/10.5114/amsad.2019.84447

Grzegorzewska A.E., Adamska P., Iwańczyk-Skalska E., Ostromecka K., Niepolski L., MarcinkowskiW., Mostowska A., Warchoł W., Żaba C. & Jagodziński P.P. (2021). Paraoxonase 1 concerning dyslipidaemia, cardiovascular diseases, and mortality in haemodialysis patients. *Scientific Reports* **11**: 6773.
DOI: https://doi.org/10.1038/s41598-021-86231-0

Hashemi M., Kordi-Tamandani D. M., Sharifi N., Moazeni-Roodi A., Kaykhaei M., Narouie B., Torkmanzehi A. (2011). Serum paraoxonase and arylesterse activities in metabolic syndrome in Zahedan, Southeast Iran. *European Journal of Endocrinology* **164**: 219–222.
DOI: https://doi.org/10.1530/EJE-10-0881

James R.W., Leviev I. & Righetti A. (2000). Smoking is associated with reduced serum paraoxonase activity and concentration in coronary artery disease patients. *Circulation* **101**(19): 2252–2257.
DOI: https://doi.org/10.1161/01.cir.101.19.2252

Josse D., Xie W., Renault F., Rochu D., Schopfer L. M., Masson P. & Lockridge O. (1999). Identification of residues essential for human paraoxonase (PON1) arylesterase/ organophosphatase activities. *Biochemistry* **38**: 2816.
DOI: https://doi.org/10.1021/bi982281h

Juretiæ D., Tadijanoviæ M., Rekiæ B., Simeon-Rudolf V., Reiner E. & Barièiæ M. (2001). Serum Paraoxonase Activities in Hemodialyzed Uremic Patients: Cohort Study. *Clinical Sciences* **42**(2): 146–150.

Mackness M.I., Arrol S. & Durrington P.N. (1991). Paraoxonase prevents accumulation of lipoperoxides in low-density lipoprotein. *FEBS Letters* **286**(1–2): 152–154.
DOI: https://doi.org/10.1016/0014-5793(91)80962-3

Mackness B., Davies G.K., Turkie W., Lee E., Roberts D.H., Hill E., Roberts C., Durrington P.N. & Mackness M.I. (2001). Paraoxonase status in coronary heart disease: are activity and concentration more important than genotype? Arteriosclerosis *Thrombosis and Vascular Biology* **21**: 1451–1457.
DOI: https://doi.org/10.1161/hq0901.094247

Mackness M. & Mackness B. (2015). Human paraoxonase-1 (PON1): gene structure and expression, promiscuous activities and multiple physiological roles. *Gene* **567**(1): 12–21.
DOI: https://doi.org/10.1016/j.gene.2015.04.088

Mackness B., Mackness M. & Durrington P.N. (1998). Human serum paraoxonase. *General Pharmacology* **31** (3): 329–336.
DOI: https://doi.org/10.1016/S0306-3623(98)00028-7

Matsumoto A., Maes M., Supasitthumrong T., Maes A., Michelin A., De Oliveira Semeão L., De Lima Pedrão J.V., Moreira E.G., Kanchanatawan B. & Barbosa D. (2020). Deficit

schizophrenia and its features are associated with PON1 Q192R genotypes and lowered paraoxonase 1 (PON1) enzymatic activity: effects on bacterial translocation. *CNS Spectrums* **26**(4): 1–10.
DOI: https://doi.org/10.1017/S1092852920001388

Moreira E. G., Correia D. G., Bonifácio K. L., Moraes J. B. D., Cavicchioli F. L., Nunes C. S., Nunes S. O. V., Vargas H.O., Barbosa, D.S. & Maes M. (2019). Lowered PON1 activities are strongly associated with depression and bipolar disorder, recurrence of (hypo) mania and depression, increased disability and lowered quality of life. *The World Journal of Biological Psychiatry* **20**(5): 368–380.
DOI: https://doi.org/10.1080/15622975.2017.1322219

Murillo-González F.E. et al. (11 authors) (2020). PON1 lactonase activity and its association with cardiovascular disease. *Clinica Chimica Acta* **500**: 47–53.
DOI: https://doi.org/10.1016/j.cca.2019.09.016

Naderi M., Hashemi M., Komijani-Bozchaloei F., Moazeni-Roodi A. & Momenimoghaddam M. (2011). Serum paraoxonase and arylesterase activities in patients with pulmonary tuberculosis. *Pathophysiology* **18**: 117–120.
DOI: https://doi.org/10.1016/j.pathophys.2010.05.002

Nishio E. & Watanabe Y. (1997). Cigarette smoke extract inhibits plasma paraoxonase activity by modification of the enzyme's free thiols. *Biochemical and Biophysical Research Communications* **236**: 289–293.
DOI: https://doi.org/10.1006/bbrc.1997.6961

Novak F., Vavrova L., Kodydkova J., Hynkova M., Zak A. & Novakova O. (2010). Decreased paraoxonase activity in critically ill patients with sepsis. *Clinical and Experimental Medicine* **10**(1): 21–25.
DOI: https://doi.org/10.1007/s10238-009-0059-8

Paragh G., Balla P., Katona E., Seres I., Égerházi A. & Degrell I. (2002). Serum paraoxonase activity changes in patients with Alzheimer's disease and vascular dementia. *Europian Archives of Psychiatry and Clinical Neuroscience* **252**: 63–67.
DOI: https://doi.org/10.1007/s004060200013

Paragh G., Seres I., Harangi M., Balogh Z., Illyes L., Boda J.Z.S. & Kovacs P. (2003). The effect of micronised fenofibrate on paraoxonase activity in patients with coronary heart disease. *Diabetes and Metabolism* **29**: 613–618.
DOI: https://doi.org/10.1016/s1262-3636(07)70077-0

Passaro A., Vigna G. B., Romani A., Sanz J. M., Cavicchio C., Bonaccorsi G., Valacchi G. & Cervellati C. (2018). Distribution of paraoxonase-1 (PON-1) and lipoprotein phospholipase A2 (Lp-PLA2) across lipoprotein subclasses in subjects with type 2 diabetes. *Oxidative Medicine and Cellular Longevity* 2018: Article ID 1752940.
DOI: https://doi.org/10.1155/2018/1752940

Porntadavity S., Tantrarongroj S., Pidetcha P., Dansethakul P. & Suwannathon L. (2009). Paraoxonase1 phenotype distribution in Thais. *Journal of Medical Association of Thailand* **92**(3): 405–412.

Prakash M., Phani N.M., Kavya R. & Supriya M. (2010). Paraoxonase: Its antiatherogenic role in chronic renal failure. *Indian Journal of Nephrology* **20**(1): 9–14.
DOI: https://doi.org/10.4103/0971-4065.62088

Savu O., Serafinceanu C., Grajdeanu I. V., Iosif L., Gaman L. & Stoian I. (2014). Paraoxonase lactonase activity, inflammation and antioxidant status in plasma of patients with type 1 diabetes mellitus. *Journal of International Medical Research* **42**(2): 523–529.
DOI: https://doi.org/10.1177/0300060513516287

Sepahvand F., Shafiei M., Ghaffari S. M., Rahimi-Moghaddam P. & Mahmoudian M. (2007). Paraoxonase phenotype distribution in a healthy Iranian population. *Basic and Clinical Pharmacology and Toxicology* **101**: 104–107.
DOI: https://doi.org/10.1111/j.1742-7843.2007.00080.x

Shih D.M. et al. (12 authors) (1998). Mice lacking serum paraoxonase are susceptible to organophosphate toxicity and atherosclerosis. *Nature* **394**: 284–287.
DOI: https://doi.org/10.1038/28406

Sözmen E.Y., Mackness B., Soömen B., Durrington P., Girgin F.K., Aslan L. & Mackness M. (2002). Effect of organophosphate intoxication on human serum paraoxonase. *Human and Experimental Toxicology* **21**: 247–252.
DOI: https://doi.org/10.1191/0960327102ht244oa

Sutherland W.H., Walker R.J., de Jong S. A., van Rij A.M., Phillip V. & Walker H.L. (1999). Reduced postprandial serum paraoxonase activity after a meal rich in used cooking fat. *Arteriosclerosis, Thrombosis, and Vascular Biology* **19**: 1340–1347.
DOI: https://doi.org/10.1161/01.ATV.19.5.1340

Taler-Verčič A., Marko Goličnik M. & Aljoša Bavec A. (2020). The structure and function of paraoxonase-1 and its comparison to paraoxonase-2 and -3. *Molecules* **25**(24): 5980.
DOI: https://doi.org/10.3390/molecules25245980

van der Gaag M.S., van Tol A., Scheek L.M., Richard W., James RW., Urgert R., Schaafsma G. & Hendriks H.F.J. (1999). Daily moderate alcohol consumption increases serum paraoxonase activity; a diet-controlled, randomized intervention study in middle-aged men. *Atherosclerosis* **147**: 405–410
DOI: https://doi.org/10.1016/s0021-9150(99)00243-9

Watson A.D., Berliner J.A, Hama S.Y., La Du B.N., Faull K.F., Fogelman A.M. & Navab M. (1995). Protective effect of high density lipoprotein associated paraoxonase. Inhibition of the biological activity of minimally oxidized low density lipoprotein. *Journal of Clinical Investigation* **96**(6): 2882–2891.
DOI: https://doi.org/10.1172/JCI118359

Wei L.K. et al. (11 authors) (2017). Polymorphisms of MTHFR, eNOS, ACE, AGT, ApoE, PON1, PDE4D, and ischemic stroke: meta-analysis. J*ournal of Stroke and Cerebrovascular Diseases* **26**(11): 2482–2493.
DOI: https://doi.org/10 10.1016/j.jstrokecerebrovasdis.2017.05.048

Zafiropoulos A., Linardakis M., Jansen E.H.J.M., Tsatsakis A., Kafatos A. & Tzanakakis G. N. (2010). Paraoxonase 1 R/Q alleles are associated with differential accumulation of saturated versus 20:5n3 fatty acid in human adipose tissue. *Journal of Lipid Research* **51**: 1991–2000.
DOI: https://doi.org/10.1194/jlr.P004960

J.Natn.Sci.Foundation Sri Lanka 2021 49 (3): 429 - 439
DOI: http://dx.doi.org/10.4038/jnsfsr.v49i3.9939

RESEARCH ARTICLE

Radii problems and some other properties of certain classes of analytic functions with boundary rotation

A Saliu* **and KI Noor**
Department of Mathematics, COMSATS University Islamabad, Park Road, Tarlai Kalan, Islamabad 45550, Pakistan.

Submitted: 28 September 2020; Revised: 07 January 2021;Accepted 28 February 2021

Abstract: In this work, some generalized classes of convex, close-to-convex, quasi-convex and strongly convex functions $f(z)$ in the frame of a unit disc U are introduced. The first task is to discuss some geometric properties of a class of functions $f(z)$ for which $f(U)$ has a boundary rotation of at most $m\pi$, $m \geq 2$. Furthermore, the smallest disc of the linear combinations of functions belonging to the classes $\mathcal{V}_m[\mathcal{A},\mathcal{B}]$, $\mathcal{CCV}_{m_1m_2}[\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D}]$ and $\mathcal{QCV}_{m_1m_2}[\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D}]$. Condition for univalence, covering theorem and distortion result of a more generalized strongly close-to-convex functions are also obtained. Meanwhile, some remarkable cases and the consequences of our investigation are also highlighted.

Keywords: Analytic function, Univalent function, Strongly close-to-convex function, Subordination.

INTRODUCTION

Theory on geometric functions is the part of the complex analysis, which considers the geometric characterization of analytic functions and established around the turn of the twentieth century. Regardless of the well-known coefficient problem, the Bieberbach conjecture that was explained by Louis de Branges in 1984 recommends different methodologies and bearings of concentrates in the field of Univalent function theory. As a result, several authors paid their attention to establishing certain subclasses of analytic functions defined by differential subordination, such as the classes of starlike, convex, close-to-convex and quasi-convex functions (Duren, 2001; Goodman, 1983).

One natural extension of these classes is the class $\mathcal{V}_m$ ($m \geq 2$) of functions with bounded boundary rotation. Löwner (1917)started work related to this topic but Paatero (1931) systematically developed its properties. This direction of research has experienced many advancements and contributions, see (Brannan,1969; Noor & Thomas, 1980; Noor, 1981; 1988; 2002; 2009; Noor & Malik, 2010; Saliu, 2019; Afis & Noor, 2020; Saliu & Noor, 2020). Following the aforementioned works and taking into account the open unit disc U, the aim of the present study is to introduce the novel generalizations of convex, close-to-convex, quasi-convex and strongly convex functions. Then the smallest disc of the linear combinations of functions belonging to these classes was found. Moreover, the geometric features of a class of functions for which the image domain has boundary rotation under certain conditions was investigated. A more general concept of strongly close-to-convex functions is also initiated. In this direction, conditions for the univalence, covering theorem and distortion results for this novel class are obtained.

* Corresponding author (saliugsu@gmail.com; https://orcid.org/0000-0002-0647-3824)

METHODOLOGY

Some mathematical preliminaries and definitions that will play a key role in proving the main results are presented here.

Let $\mathcal{H}$ denotes the class of functions $f(z)$ in the open unit disk U with the series form

$$f(z) = z + a_2 z^2 + a_3 z^3 + \dots. \quad ...(1)$$

If $f(z)$ and $g(z)$ are analytic functions in U, then $f(z)$ is subordinate to $g(z)$ (written as $f(z) \prec g(z)$) if there exist a Schwarz function $w(z)$ such that $f(z) = g(w(z)),\ z \in \mathcal{U}$.

Janowski (1973) introduced the class $\mathcal{P}[\mathcal{A},\mathcal{B}],\ -1 \le \mathcal{B} < \mathcal{A} \le 1$ of functions $p(z)$ satisfying the inequality

$$\left|p(z) - \frac{1-\mathcal{A}\mathcal{B}r^2}{1-\mathcal{B}^2r^2}\right| \le \frac{(\mathcal{A}-\mathcal{B})r}{1-\mathcal{B}^2r^2}, \quad 0 \le r < 1. \quad ...(2)$$

Also, the class $\mathcal{P}^{\lambda}[\mathcal{A},\mathcal{B}]$ was introduced by Raina *et al*. (2020) and defined it as

$$\left|p^{\frac{1}{\lambda}}(z) - \frac{1+\mathcal{A}\mathcal{B}r^2}{1-\mathcal{B}^2r^2}\right| \le \frac{(\mathcal{A}-\mathcal{B})r}{1-\mathcal{B}^2r^2}, \quad 0 \le r < 1,\ 0 < \lambda \le 1. \quad ...(3)$$

As a special case, $\mathcal{P}^{\lambda}[\mathcal{A},\mathcal{B}]$ reduces to $\mathcal{P}[\mathcal{A},\mathcal{B}]$ for $\lambda = 1$. Similar to the class $\mathcal{P}_m,\ m \ge 2$ introduced by Pinchuk (1971), Noor (1991a) defined the class $\mathcal{P}_m[\mathcal{A},\mathcal{B}]$ and used it to find the radius of convexity and starlikeness for the classes $\mathcal{V}_m[\mathcal{A},\mathcal{B}]$ and $\mathcal{R}_m[\mathcal{A},\mathcal{B}]$ respectively. For more details regarding these classes, refer to (Noor, 1988, 1991b, 1994; 2009; Noor & Malik, 2010).

We define the following classes of functions as follows:

Deftnition 2.1. *Let* $f(z)$ be of the form (1) and $m_1, m_2 \ge 2,\ -1 \le \mathcal{B} < \mathcal{A} \le 1,\ -1 \le \mathcal{C} < \mathcal{D} \le 1$. Then we say $f \in \mathcal{CCV}_{m_1 m_2}[\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D}]$ if there exists a function $g \in \mathcal{V}_{m_1}[\mathcal{A},\mathcal{B}]$ such that

$$\frac{f'(z)}{g'(z)} \in \mathcal{P}_{m_2}[\mathcal{C},\mathcal{D}]. \quad ...(4)$$

Deftnition 2.2. *Let* $f(z)$ be of the form (1) and $m_1, m_2 \ge 2,\ -1 \le \mathcal{B} < \mathcal{A} \le 1,\ -1 \le \mathcal{C} < \mathcal{D} \le 1$. Then we say $f \in \mathcal{QCV}_{m_1 m_2}[\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D}]$ if there exists a function $g \in \mathcal{V}_{m_1}[\mathcal{A},\mathcal{B}]$ such that

$$\frac{(zf'(z))'}{g'(z)} \in \mathcal{P}_{m_2}[\mathcal{C},\mathcal{D}]. \quad ...(5)$$

Deftnition 2.3. *Let* $f(z)$ be of the form (1) and $m \ge 2$, $-1 \le \mathcal{B} < \mathcal{A} \le 1,\ -1 \le \mathcal{C} < \mathcal{D} \le 1$. Then we say $f \in \mathcal{K}^{\lambda}{}_m[\mathcal{A},\mathcal{B}]$ if there exists a function $g \in \mathcal{V}_m[\mathcal{A},\mathcal{B}]$ such that

$$\frac{f'(z)}{g'(z)} \in \mathcal{P}^{\lambda}[\mathcal{A},\mathcal{B}], \quad 0 < \lambda \le 1. \quad ...(6)$$

By specifying certain values of the underlying parameters in the above definitions, we obtain the following sub- classes, which have been studied by many researchers:

(i) $\mathcal{CCV}_{22}[\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D}]$, is the well-known subclass of close-to-convex function introduced by Silvia(1983).

(ii) For $\mathcal{C} = \mathcal{A},\ \mathcal{D} = \mathcal{B}$ and $m_1 = m_2 = m$, we obtain the class $\mathcal{T}_m[\mathcal{A},\mathcal{B}]$ that was studied by Noor (1991b).

(iii) $\mathcal{CCV}_{22}[1,-1,1,-1] \equiv \mathcal{CCV}$, is the usual well-known class of close-to-convex functions introduced and examined by Kaplan(1952).

(iv) For $\mathcal{C} = \mathcal{A},\ \mathcal{D} = \mathcal{B}$ and $m_1 = m_2 = 2$, we get the class $\mathcal{QCV}[\mathcal{A},\mathcal{B}]$ that was explored in (Aliintaş & Kiliç, 2018).

(v) $\mathcal{QCV}[1,-1,1,-1]$ is the well-known class of quasi convex functions that was first introduced and examined in (Noor,1980).

(vi) For $\lambda = 1,\ m = 2,\ \mathcal{K}^{\lambda}{}_m[\mathcal{A},\mathcal{B}]$ becomes the class of functions considered by Raina *et al*. (2020).

(vii) $\mathcal{K}^{\lambda}{}_m[1,-1] = \mathcal{K}^{\lambda}{}_m$, is the class explored by Noor (2002).

To establish the main results, the following results are required.

Lemma 2.1. (Stump, 1971) If $|u-a| \le d$ *and* $|v-a| \le d$, where *a* and *d* are real with $a > d \ge 0$, and

$$w = u\frac{1}{1+\lambda e^{i\alpha}} + v\frac{1}{1+\lambda^{-1}e^{-i\alpha}}, \quad ...(7)$$

where α is real, λ>0 and $\alpha \in [0,\pi)$,then

$$\mathrm{Re}\, w \ge a - d\sec\frac{\alpha}{2}. \quad ...(8)$$

The following lemma exhibits some practice related

to our consequences.

Lemma 2.2. *Let* $p \in \mathcal{P}_m[\mathcal{A},\mathcal{B}]$. Then

$$\frac{1}{2\pi}\int_0^{2\pi}|p(z)|^2 d\theta \leq \begin{cases} 1+\frac{m^2(\mathcal{B}-\mathcal{A})^2}{4\mathcal{B}^2}\left(\frac{\mathcal{B}^2r^2}{1-\mathcal{B}^2r^2}\right), & \mathcal{B}\neq 0 \\ 1+\frac{m^2\mathcal{A}^2r^2}{4}, & \mathcal{B}=0 \end{cases} \quad ...(9)$$

$$\int_0^{2\pi}|p(z)|^2 d\theta \leq \begin{cases} r+\frac{m^2(\mathcal{B}-\mathcal{A})^2}{8\mathcal{B}^3}\left[\log\left(\frac{1+\mathcal{B}r}{1-\mathcal{B}r}\right)-2\mathcal{B}r\right], & \mathcal{B}\neq 0 \\ r+\frac{m^2\mathcal{A}^2r^3}{12}, & \mathcal{B}=0 \end{cases} \quad ...(10)$$

Proof. Since $p \in \mathcal{P}_m[\mathcal{A},\mathcal{B}]$, then

$$p(z)=\frac{m+2}{4}p_1(z)-\frac{m-2}{4}p_2(z),$$

where

$$p_i(z) \prec \frac{1+\mathcal{A}z}{1+\mathcal{B}z}, \quad i=1,2 \quad ...(11)$$

and by subordination property,

$$|p(z)| \leq \max_{|z|\leq r}\left|\frac{1+\mathcal{A}z}{1+\mathcal{B}z}\right|.$$

Let $p_i(z)=1+\sum_{n=1}^{\infty}d_{i,n}z^n$ ...(12)

and we can write

$$\frac{1+\mathcal{A}z}{1+\mathcal{B}z}=1+\sum_{n=1}^{\infty}\vartheta_n z^n, \quad \text{where}$$

$$\vartheta_n=(-1)^n\mathcal{B}^{n-1}(\mathcal{B}-\mathcal{A}), \ \ n=1,2,3,\ldots. \quad ...(13)$$

Thus, by Perseval's identity and subordination property, we get

$$\begin{aligned}\frac{1}{2\pi}\int_0^{2\pi}|p(z)|^2 d\theta &\leq 1+\sum_{n=1}^{\infty}\left|\left(\frac{m+2}{4}\right)d_{1,n}-\left(\frac{m-2}{4}\right)d_{2,n}\right|^2 r^{2n} \\ &\leq 1+\sum_{n=1}^{\infty}\left[\left(\frac{m+2}{4}\right)|d_{1,n}|+\left(\frac{m-2}{4}\right)|d_{2,n}|\right]^2 r^{2n} \\ &\leq 1+\sum_{n=1}^{\infty}\frac{m^2}{4}|\vartheta|^2 r^{2n} \\ &= 1+\frac{m^2}{4}\frac{(\mathcal{B}-\mathcal{A})^2}{\mathcal{B}^2}\sum_{n=1}^{\infty}\mathcal{B}^{2n}r^{2n} \\ &= 1+\frac{m^2(\mathcal{B}-\mathcal{A})^2}{4\mathcal{B}^2}\left(\frac{\mathcal{B}^2r^2}{1-\mathcal{B}^2r^2}\right).\end{aligned} \quad ...(13)$$

Using (14), we obtain

$$\begin{aligned}\frac{1}{2\pi}\int_0^r\int_0^{2\pi}|p(z)|^2 d\theta d\rho &\leq r+\frac{m^2(\mathcal{B}-\mathcal{A})^2}{4\mathcal{B}^2}\int_0^r\left(\frac{1}{1-\mathcal{B}^2\rho^2}-1\right)d\rho \\ &= r+\frac{m^2(\mathcal{B}-\mathcal{A})^2}{8\mathcal{B}^3}\left[\log\left(\frac{1+\mathcal{B}r}{1-\mathcal{B}r}\right)-2\mathcal{B}r\right].\end{aligned}$$

If $B = 0$,

$$\begin{aligned}\frac{1}{2\pi}\int_0^{2\pi}|p(z)|^2 d\theta &\leq 1+\left[\left(\frac{m+2}{4}\right)\mathcal{A}+\left(\frac{m-2}{4}\right)\mathcal{A}\right]^2 r^2 \\ &= 1+\frac{m^2}{4}\mathcal{A}^2r^2\end{aligned}$$

and

$$\frac{1}{2\pi}\int_0^r\int_0^{2\pi}|p(z)|^2 d\theta d\rho \leq r+\frac{m^2}{12}\mathcal{A}^2r^3.$$

Remark 2.1. As a consequence of Lemma 2.2, we have that

(i) for m = 2, the result coincide with the one obtained by Cho & Kumar (2019),

(ii) for $p \in \mathcal{P}_m[1,-1]$,

$$\frac{1}{2\pi}\int_0^r|p(z)|^2 d\theta \leq \frac{1+(m^2-1)r^2}{1-r^2} \quad ...(15)$$

and

$$\frac{1}{2\pi}\int_0^r\int_0^{2\pi}|p(z)|^2 d\theta d\rho \leq \frac{m^2}{2}\log\left(\frac{1+r}{1-r}\right)+(1-m^2)r. \quad ...(16)$$

Inequality (15) coincides with the one given by Noor (1981), whereas inequality (16) shows an improvement in the result given in (Noor & Noor, 1992) for the case $\rho = 0$ therein.

Lemma 2.3. (Raina *et al*., 2020) *Let* $p \in \mathcal{P}^{\lambda}[\mathcal{A},\mathcal{B}]$.

Then

$$\left(\frac{1-\mathcal{A}r}{1-\mathcal{B}r}\right)^{\lambda} \leq |p(z)| \leq \left(\frac{1+\mathcal{A}r}{1+\mathcal{B}r}\right)^{\lambda} \quad \mathcal{B}\neq 0, \quad ...(17)$$

$$(1-\mathcal{A}r)^{\lambda} \leq |p(z)| \leq (1+\mathcal{A}r)^{\lambda} \quad \mathcal{B}=0. \quad ...(18)$$

Lemma 2.4. *Let* $f \in \mathcal{V}_m[\mathcal{A},\mathcal{B}]$, where $m \geq 2$ and $-1 \leq \mathcal{B} < \mathcal{A} \leq 1$. Then

$$|\arg f'(z)| \leq \begin{cases} \frac{m}{2}\frac{\mathcal{A}-\mathcal{B}}{\mathcal{B}}\arcsin(\mathcal{B}r), & \text{if } \mathcal{B}\neq 0 \\ \frac{m}{2}\mathcal{A}r, & \mathcal{B}=0 \end{cases} \quad ...(19)$$

Proof. Using the representation for $f \in \mathcal{V}_m[\mathcal{A},\mathcal{B}]$ in (Noor, 1994), we write

$$f'(z) = \frac{(f_1'(z))^{\frac{m+2}{4}}}{(f_2'(z))^{\frac{m-2}{4}}}, \quad ...(20)$$

for some $f_1(z),\ f_2(z) \in \mathcal{V}_2[\mathcal{A},\mathcal{B}]$. It is known in (Silvia, 1983) that if $f \in \mathcal{V}_2[\mathcal{A},\mathcal{B}]$, then

$$|\arg f'(z)| \leq \begin{cases} \frac{\mathcal{A}-\mathcal{B}}{\mathcal{B}}\arcsin(\mathcal{B}r), & \text{if } \mathcal{B}\neq 0 \\ \mathcal{A}r, & \mathcal{B}=0. \end{cases} \quad ...(21)$$

Using this result and (20), the proof is completed.

Lemma 2.5. (Noor, 1994) *Let* $f \in \mathcal{V}_m[\mathcal{A},\mathcal{B}]$. Then

$$\frac{(1-\mathcal{B}r)^{q_1}}{(1+\mathcal{B}r)^{q_2}} \leq |f'(z)| \leq \frac{(1+\mathcal{B}r)^{q_1}}{(1-\mathcal{B}r)^{q_2}}, \quad \mathcal{B}\neq 0,$$
$$e^{-\frac{\mathcal{A}mr}{2}} \leq |f'(z)| \leq e^{\frac{\mathcal{A}mr}{2}}, \quad \mathcal{B}=0,$$

where

$$q_1 = \left(\frac{\mathcal{A}-\mathcal{B}}{\mathcal{B}}\right)\left(\frac{m+2}{4}\right) \quad \text{and} \quad q_2 = \left(\frac{\mathcal{A}-\mathcal{B}}{\mathcal{B}}\right)\left(\frac{m-2}{4}\right). \quad ...(22)$$

Throughout this work, unless otherwise stated, we assume $-1 \leq \mathcal{B} < \mathcal{A} \leq 1,\ -1 \leq \mathcal{D} < \mathcal{C} \leq 1$ and $m, m_1, m_2 \geq 2$.

RESULTS AND DISCUSSION

In this section, we shall proceed with the statements and proofs of our main results.

Theorem 3.1. *Let* $A(r,f)$ be the area bounded by the image curve $|z| = r,\ r \in (0,1)$ under the function $f \in \mathcal{V}_m[\mathcal{A},\mathcal{B}]$. Then

$$L(r,f) \leq \begin{cases} 2\left[\frac{\pi A(r,f)}{r}\left(r + \frac{m^2(\mathcal{B}-\mathcal{A})^2}{8\mathcal{B}^3}\left(\log\left(\frac{1+\mathcal{B}r}{1-\mathcal{B}r}\right) - 2\mathcal{B}r\right)\right)\right]^{\frac{1}{2}}, & \mathcal{B}\neq 0 \\ 2\left[\frac{\pi A(r,f)}{r}\left(r + \frac{m^2\mathcal{A}^2r^3}{12}\right)\right]^{\frac{1}{2}}, & \mathcal{B}=0. \end{cases}$$

Proof. Since $f \in \mathcal{V}_m[\mathcal{A},\mathcal{B}]$, then

$$\frac{(zf'(z))'}{f'(z)} = p(z), \quad \text{where} \quad p \in \mathcal{P}_m[\mathcal{A},\mathcal{B}]. \quad ...(23)$$

By the definition of arc length, we have

$$\begin{aligned} L(r,f) &= \int_0^{2\pi} |zf'(z)|d\theta \\ &= \int_0^r \int_0^{2\pi} |(zf'(z))'|d\theta d\rho \\ &= \int_0^r \int_0^{2\pi} |f'(z)|\left|\frac{(zf'(z))'}{f'(z)}\right| d\theta d\rho \\ &\leq \left(\int_0^r \int_0^{2\pi} |f'(z)|^2 d\theta d\rho\right)^{\frac{1}{2}} \times \left(\int_0^r \int_0^{2\pi} |p(z)|^2 d\theta d\rho\right)^{\frac{1}{2}} \\ &\leq \left(2\pi\sum_{n=1}^{\infty} n|a_n|^2 r^{2n-1}\right)^{\frac{1}{2}} \left(\int_0^r \int_0^{2\pi} |p(z)|^2 d\theta d\rho\right)^{\frac{1}{2}}. \end{aligned} \quad ...(24)$$

On applying Lemma 2.2, and using the fact that

$$A(r,f) = \pi\sum_{n=1}^{\infty} n|a_n|^2 r^{2n}, \quad a_1 = 1,$$

we obtain the required result.

As a consequence of the Theorem 3.1, we have Corollary 3.1. *Let* $f \in \mathcal{V}_m$. Then

$$L(r,f) \leq 2\left[\frac{\pi A(r,f)}{r}\left(\frac{m^2}{2}\log\left(\frac{1+r}{1-r}\right) + (1-m^2)\right)\right]^{\frac{1}{2}}$$

or equivalently as

$$L(r,f) \leq m\left[\pi A(r,f)\log\left(\frac{1}{1-r}\right)\right]^{\frac{1}{2}} \quad \text{as} \quad r \to 1$$

and for $f \in \mathcal{V}_2$, we have

$$L(r,f) = O\left[\pi A(r,f)\log\left(\frac{1}{1-r}\right)\right]^{\frac{1}{2}} \quad \text{as} \quad r \to 1,$$

which was the result obtained by Nunokawa (1969) and O is a constant.

Theorem 3.2. If $f \in \mathcal{V}_m[\mathcal{A},\mathcal{B}]$, then so is $F'(z) = (-f'(z)\cdot f'(-z))^{\frac{1}{2}}$.

Proof. Suppose $f \in \mathcal{V}_m[\mathcal{A},\mathcal{B}]$. Then in (Noor, 1991a), we have the representation

$$f'(z) = \exp\left\{\frac{\mathcal{A}-\mathcal{B}}{2\mathcal{B}}\int_{-\pi}^{\pi}\log\left(1+\mathcal{B}ze^{it}\right)d\mu(t)\right\}, \quad \mathcal{B}\neq 0 \quad ...(25)$$

and

$$f'(z)=\exp\left\{\frac{\mathcal{A}}{2}\int_{-\pi}^{\pi}ze^{it}d\mu(t)\right\},\quad \mathcal{B}=0, \qquad ...(26)$$

where μ is a real valued of bounded variation on $[-\pi, \pi]$ satisfying the conditions

$$\int_{-\pi}^{\pi}d\mu(t)=2 \quad \text{and} \quad \int_{-\pi}^{\pi}|d\mu(t)|\le m.$$

There fore for $\mathcal{B}\neq 0$,

$$F'(z)=-\exp\left\{\frac{\mathcal{A}-\mathcal{B}}{4\mathcal{B}}\int_{-\pi}^{\pi}\log\left(1+\mathcal{B}ze^{it}\right)d\mu(t)\right\}$$

$$\times\exp\left\{\frac{\mathcal{A}-\mathcal{B}}{4\mathcal{B}}\int_{-\pi}^{\pi}\log\left(1-\mathcal{B}ze^{it}\right)d\mu(t)\right\}.$$

Thus,

$$\frac{(zF'(z))'}{F'(z)}=1+\frac{\mathcal{A}-\mathcal{B}}{4}\int_{-\pi}^{\pi}\frac{ze^{-it}}{1+\mathcal{B}ze^{it}}d\mu(t)-\frac{\mathcal{A}-\mathcal{B}}{4}\int_{-\pi}^{\pi}\frac{ze^{-it}}{1-\mathcal{B}ze^{it}}d\mu(t)$$

$$=\frac{1}{2}\left\{\frac{1}{2}\int_{-\pi}^{\pi}\frac{1+\mathcal{A}ze^{-it}}{1+\mathcal{B}ze^{it}}d\mu(t)+\frac{1}{2}\int_{-\pi}^{\pi}\frac{1-\mathcal{A}ze^{-it}}{1-\mathcal{B}ze^{it}}d\mu(t)\right\}$$

$$=\frac{1}{2}\left(p(z)+p(-z)\right),\quad \text{where}\quad p(z)\in\mathcal{P}_m[\mathcal{A},\mathcal{B}].$$

In view of the convexity property of the class $\mathcal{P}_m[\mathcal{A},\mathcal{B}]$, it follows that $F\in\mathcal{V}_m[\mathcal{A},\mathcal{B}]$, which is the required result. The case $\mathcal{B}=0$ also follows the same arguments.

Theorem 3.3. *Let* $f, g\in\mathcal{V}_m[\mathcal{A},\mathcal{B}]$ and suppose that for any $\mathcal{A}$ and $\mathcal{B}$ satisfying the condition $3\mathcal{A}\mathcal{B}+\mathcal{A}^2+\mathcal{B}^2<0$, $F(z)=\beta f(z)+(1-\beta)g(z)\in\mathcal{V}_m[1,-1]$ $(\beta\in\mathbb{C}-\{0,1\})$ if $|z|<r_0$, wherer r_0 is the smallest positive root satisfying the equation:

$$\begin{cases}\left(1-\mathcal{A}\mathcal{B}r^2\right)\cos\left(\frac{\gamma}{2}+\frac{m}{2}\left(\frac{\mathcal{A}-\mathcal{B}}{\mathcal{B}}\right)\arcsin(\mathcal{B}r)\right)-(\mathcal{A}-\mathcal{B})r=0, & \mathcal{B}\neq 0\\ \cos(\frac{\gamma}{2}+\frac{\mathcal{A}mr}{2})-\mathcal{A}r, & \mathcal{B}=0,\end{cases} \qquad ...(27)$$

where $0\le\gamma=\arg\left(\frac{\beta}{1-\beta}\right)<\pi$.

Proof. Since $f, g\in\mathcal{V}_m[\mathcal{A},\mathcal{B}]$, then there exist $p_1, p_2, h_1, h_2\in\mathcal{P}[\mathcal{A},\mathcal{B}]$ such that

$$\frac{(zf'(z))'}{f(z)}=\frac{m+2}{4}p_1(z)-\frac{m-2}{4}p_2(z) \qquad ...(28)$$

$$\frac{(zf'(z))'}{f(z)}=\frac{m+2}{4}p_1(z)-\frac{m-2}{4}p_2(z) \qquad ...(29)$$

By direct calculation,

$$\frac{(zF'(z))'}{F(z)}=\frac{(zf'(z))'}{f(z)}\left[1+\left(\frac{\beta}{1-\beta}\frac{f'(z)}{g\prime(z)}\right)^{-1}\right]^{-1}+\frac{(zg'(z))'}{g(z)}\left[1+\frac{\beta}{1-\beta}\frac{f'(z)}{g\prime(z)}\right]^{-1} \qquad ...(30)$$

$$=\frac{m+2}{4}\left\{\left[1+\frac{\beta}{1-\beta}\frac{f'(z)}{g\prime(z)}\right]^{-1}h_1(z)+\left[1+\left(\frac{\beta}{1-\beta}\frac{f'(z)}{g\prime(z)}\right)^{-1}\right]^{-1}p_1(z)\right\}$$

$$-\frac{m-2}{4}\left\{\left[1+\frac{\beta}{1-\beta}\frac{f'(z)}{g\prime(z)}\right]^{-1}h_2(z)+\left[1+\left(\frac{\beta}{1-\beta}\frac{f'(z)}{g\prime(z)}\right)^{-1}\right]^{-1}p_2(z)\right\}$$

$$=\frac{m+2}{4}w_1(z)-\frac{m-2}{4}w_2(z), \qquad ...(31)$$

where we have used (28) and (29), and

$$w_i=\frac{u_i}{1+\lambda e^{i\alpha}}+\frac{v_i}{1+\lambda^{-1}e^{-i\alpha}},\quad i=1,2, \qquad ...(32)$$

with $u_i=h_i$, $v_i=p_i$ $(i=1,2)$ and

$$\lambda=\left|\frac{\beta}{1-\beta}\frac{f'(z)}{g'(z)}\right|,\quad \alpha=\arg\left(\frac{\beta}{1-\beta}\frac{f'(z)}{g'(z)}\right).$$

In view of Lemma 2.4, we have

$$|\alpha|\le\gamma+m\frac{\mathcal{A}-\mathcal{B}}{\mathcal{B}}\arcsin(\mathcal{B}r)$$

and

$$\sec\frac{\alpha}{2}\le\sec\left(\frac{\gamma}{2}+\frac{m}{2}\frac{\mathcal{A}-\mathcal{B}}{\mathcal{B}}\arcsin(\mathcal{B}r)\right).$$

Applying Lemma 2.1 to (32), we obtain

$$\mathrm{Re}w_i\ge a-d\sec\frac{\alpha}{2},$$

where

$$a=\frac{1-\mathcal{AB}r^2}{1-\mathcal{B}^2r^2},\quad d=\frac{(\mathcal{A}-\mathcal{B})\,r}{1-\mathcal{B}^2r^2} \qquad ...(33)$$

and $a>b$ holds for any $\mathcal{A}$ and $\mathcal{B}$ satisfying the condition $3\mathcal{AB}+\mathcal{A}^2+\mathcal{B}^2<0$. Thus, the right side of (33) is positive if

$$0<\left(1-\mathcal{AB}r^2\right)\cos\left(\frac{\gamma}{2}+\frac{m}{2}\frac{(\mathcal{A}-\mathcal{B})}{\mathcal{B}}\arcsin\mathcal{B}r\right)-(\mathcal{A}-\mathcal{B})\,r.$$

Let

$$\mathcal{T}(r)=\left(1-\mathcal{AB}r^2\right)\cos\left(\frac{\gamma}{2}+\frac{m}{2}\frac{(\mathcal{A}-\mathcal{B})}{\mathcal{B}}\arcsin\mathcal{B}r\right)-(\mathcal{A}-\mathcal{B})\,r.$$

Then

$$\mathcal{T}(0)=\cos\frac{\gamma}{2}>0\quad\text{and}$$

$$\mathcal{T}\left(\frac{1}{\mathcal{B}}\sin\left(\frac{\mathcal{B}(\pi-\gamma)}{m\,(\mathcal{A}-\mathcal{B})}\right)\right)=-\frac{(\mathcal{A}-\mathcal{B})}{\mathcal{B}}\sin\left(\frac{\mathcal{B}(\pi-\gamma)}{m\,(\mathcal{A}-\mathcal{B})}\right)<0.$$

This means that $\mathcal{T}(r)$ has a solution in the interval $\left[0,\frac{1}{\mathcal{B}}\sin\left(\frac{\mathcal{B}(\pi-\gamma)}{m(\mathcal{A}-\mathcal{B})}\right)\right]$. Hence from (32), we conclude that $F\in\mathcal{V}_m[1,-1]$.

For $\mathcal{A}=1,\ \mathcal{B}=-1,\ m=2$ in Theorem 3.3, we have the following:

Corollary 3.2. (Stump, 1971) Let $f,g\in\mathcal{V}_2[1,-1]$. Then $F(z)=\beta f(z)+(1-\beta)g(z)$ is convex in the disc $|z|<r_c$, and r_c is the smallest positive root of the equation:

$$(1+r^2)\cos\left(\frac{\gamma}{2}+2\arcsin r\right)-2r=0, \qquad ...(34)$$

where $0\le\gamma=\arg\left(\frac{\beta}{1-\beta}\right)<\pi$.

Theorem 3.4. Let $f,g\in\mathcal{CCV}_{m_1m_2}[\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D}]$ and $0\le\gamma=\arg\left(\frac{\beta}{1-\beta}\right)<\pi$ and suppose for any C and D satisfying the condition $3\mathcal{CD}+\mathcal{C}^2+\mathcal{D}^2<0$, $F(z)=\beta f(z)+(1-\beta)g(z)\in\mathcal{CCV}_{m_1m_2}[1,-1,1,-1]$ in the disk $|z|<\min\{r_1,r_2\}$, where r_1 with $m=m_1$ and r_2 is the least roots of the equation:

$$\begin{cases}\left(1-\mathcal{CD}r^2\right)\cos\left(\frac{\gamma}{2}+\frac{m_2}{2}\left(\frac{\mathcal{C}-\mathcal{D}}{\mathcal{D}}\right)\arcsin(\mathcal{D}r)\right)-(\mathcal{C}-\mathcal{D})r=0, & \mathcal{D}\neq0\\ \cos(\frac{\gamma}{2}+\frac{\mathcal{C}m_2r}{2})-\mathcal{C}r, & \mathcal{D}=0.\end{cases} \qquad ...(35)$$

Proof. Let $f,g\in\mathcal{CCV}_{m_1m_2}[\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D}]$. Then there exist $h(z),q(z)\in\mathcal{V}_{m_1}[\mathcal{A},\mathcal{B}]$ such that

$$\frac{f'(z)}{h'(z)},\ \frac{g'(z)}{q'(z)}\in\mathcal{P}_{m_2}[\mathcal{C},\mathcal{D}]. \qquad ...(36)$$

Suppose $\mathcal{D}\neq0$ and let $G(z)=\beta h(z)+(1-\beta)q(z)$. Then by Theorem 3.3, $G\in\mathcal{V}_{m_1}[1,-1]$ in the disk $|z|<r_1$.

Therefore by simple calculation, it follows that

$$\frac{F'(z)}{G'(z)}=\frac{f'(z)}{h'(z)}\left[1+\frac{\beta}{1-\beta}\frac{h'(z)}{q'(z)}\right]^{-1}+\frac{g'(z)}{q'(z)}\left[1+\left(\frac{\beta}{1-\beta}\frac{h'(z)}{q'(z)}\right)^{-1}\right]^{-1}. \qquad ...(37)$$

From (36), it implies there exist $h_3,h_4,p_3,p_4\in\mathcal{P}[\mathcal{C},\mathcal{D}]$ such that (37) becomes

$$\begin{aligned}\frac{F'(z)}{G'(z)}=&\frac{m_2+2}{4}\left\{\left[1+\frac{\beta}{1-\beta}\frac{h'(z)}{q\prime(z)}\right]^{-1}h_3(z)+\left[1+\left(\frac{\beta}{1-\beta}\frac{h'(z)}{q\prime(z)}\right)^{-1}\right]^{-1}p_3(z)\right\}\\&-\frac{m_2-2}{4}\left\{\left[1+\frac{\beta}{1-\beta}\frac{h'(z)}{q\prime(z)}\right]^{-1}h_4(z)+\left[1+\left(\frac{\beta}{1-\beta}\frac{h'(z)}{q\prime(z)}\right)^{-1}\right]^{-1}p_4(z)\right\}\\=&\frac{m+2}{4}w_1(z)-\frac{m-2}{4}w_2(z),\end{aligned}$$

where

$$w_i=\frac{u_i}{1+\lambda e^{i\alpha}}+\frac{v_i}{1+\lambda^{-1}e^{-i\alpha}},\ \ i=3,4,$$

with $u_i=h_i,\ v_i=p_i\ (i=3,4)$ and

$$\lambda=\left|\frac{\beta}{1-\beta}\frac{h'(z)}{q'(z)}\right|,\quad \alpha=\arg\left(\frac{\beta}{1-\beta}\frac{h'(z)}{q'(z)}\right).$$

r_c

Using Lemma 2.1 with

$$a=\frac{1-\mathcal{CD}r^2}{1-\mathcal{D}^2r^2},\quad d=\frac{(\mathcal{C}-\mathcal{D})\,r}{1-\mathcal{D}^2r^2} \qquad ...(38)$$

and proceeding in the same manner as in Theorem 3.3, we obtain the required result.

Corollary 3.3. *Let* $f, g \in \mathcal{CCV}_2[1,-1]$. Then $F(z) = \beta f(z) + (1-\beta)g(z)$ is close-to-convex in the disc$|z| < r_c$, and r_c is the smallest positive root of (34).

Theorem 3.5. *Let* $f, g \in \mathcal{QCV}_{m_1 m_2}[\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}]$ and $0 \le \gamma = \arg\left(\frac{\beta}{1-\beta}\right) < \pi$ and suppose for any $\mathcal{C}$ and $\mathcal{D}$ satisfying the condition $3\mathcal{CD} + \mathcal{C}^2 + \mathcal{D}^2 < 0$,

$F(z) = \beta f(z) + (1-\beta)g(z) \in \mathcal{QCV}_{m_1 m_2}[1,-1,1,-1]$ in the disk $|z| < \min\{r_1, r_2\}$, where r_1 is defined in Theorem 3.3 with $m = m_1$ *and* r_2 is the least roots of (35).

Proof. The proof follows the same setting as in Theorem 3.4.

Corollary 3.4. *Let* $f, g \in \mathcal{QCV}_2[1,-1]$. Then $F(z) = \beta f(z) + (1-\beta)g(z)$ is quasi convex in the disc $|z| < r_c$, and r_c is the smallest positive root of (34).

Using Kaplan(1952)method, we prove Theorem 3.6 and the rest of the proofs associated with corresponding theorems in this presentation follow the techniques of Noor (1983).

Theorem 3.6. *Let* $f \in K_m^{\lambda}[\mathcal{A}, \mathcal{B}]$. Then for $z = re^{1\theta}$ $\theta_1 < \theta_2$,

$$\int_{\theta_1}^{\theta_2} \mathrm{Re}\,\frac{(zf'(z))'}{f'(z)} d\theta > -\pi\left(\lambda + \frac{\mathcal{A}-\mathcal{B}}{1-\mathcal{B}}\left(\frac{m}{2}-1\right)\right).$$

Proof $\frac{f'(z)}{g'(z)} \in \mathcal{P}^{\lambda}[\mathcal{A}, \mathcal{B}] \Rightarrow zf'(z) = zg'(z)p^{\lambda}(z)$ for some $g \in \mathcal{V}_m[\mathcal{A}, \mathcal{B}]$ and $p \in \mathcal{P}[\mathcal{A}, \mathcal{B}]$. Let $p(z) = \arg zf'(z)$ and $q(z) = \arg zg'(z)$. Then by definition

$$|p(z) - q(z)| < \frac{\lambda\pi}{2} \qquad ...(39)$$

Let $P(r,\theta) = p(re^{i\theta}) + \theta$ and $Q(r,\theta) = q(re^{i\theta}) + \theta$ be defined for all $r \in [0,1)$ and $\theta \in [0, 2\pi]$. Then

$$|P(r,\theta_2) - P(r,\theta_1)| \le |P(r,\theta_2) - Q(r,\theta_2)| + |Q(r,\theta_2) - Q(r,\theta_1)| + |Q(r,\theta_1) - P(r,\theta_1)|. \qquad ...(40)$$

It is easy to see that for $g \in \mathcal{V}_m[\mathcal{A}, \mathcal{B}]$,

$$g'(z) = (g_1')^{\frac{\mathcal{A}-\mathcal{B}}{1-\mathcal{B}}}, \quad g_1 \in \mathcal{V}_m. \qquad ...(41)$$

Since every function $g_1 \in \mathcal{V}_m$ is a close-to-convex function of order $\frac{m}{2} - 1$ (Brannan, 1969), then

$$\int_{\theta_1}^{\theta_2} \mathrm{Re}\,\frac{(zg'(z))'}{g'(z)} d\theta > -\pi\left(\frac{\mathcal{A}-\mathcal{B}}{1-\mathcal{B}}\right)\left(\frac{m}{2}-1\right),$$

i.e.,

$$|Q(r,\theta_2) - Q(r,\theta_1)| < \pi\left(\frac{\mathcal{A}-\mathcal{B}}{1-\mathcal{B}}\right)\left(\frac{m}{2}-1\right). \qquad ...(42)$$

In view of (39), (40) and (42), we obtain

$$|Q(r,\theta_2) - Q(r,\theta_1)| < \lambda + \pi\left(\frac{\mathcal{A}-\mathcal{B}}{1-\mathcal{B}}\right)\left(\frac{m}{2}-1\right). \qquad ...(43)$$

Hence, for $\theta_1 < \theta_2$ and $z = re^{i\theta}$, we obtain the result.

Remark 3.1.

(i) From Theorem 3.6, it is obvious that $f \in K_m^{\lambda}[\mathcal{A}, \mathcal{B}]$ is a close-to-convex function and hence univalent for $2 \le m \le 2\left[1 + \left(\frac{1-\mathcal{B}}{\mathcal{A}-\mathcal{B}}\right)(1-\lambda)\right]$.

(ii) Goodman (1971) defines the class $\mathcal{K}(\beta)$ of normalized analytic functions f which are close-to-convex functions of order $\beta \ge 0$ as follows: A function $f(z) = z + \sum_{n=2}^{\infty} a_n z^n$ belongs to $\mathcal{K}(\beta)$ if f is analytic in $\mathcal{U}$, $f'(z) \ne 0$ and for $z = re^{i\theta}$, $0 \le \theta_1 < \theta_2 \le 2\pi$, $\beta \ge 0$,

$$\int_{\theta_1}^{\theta_2} \mathrm{Re}\,\frac{(zf'(z))'}{f'(z)} d\theta > -\pi\beta.$$

We note that

$$K_m^{\lambda}[\mathcal{A}, \mathcal{B}] \subset \mathcal{K}\left(\lambda + \frac{\mathcal{A}-\mathcal{B}}{1-\mathcal{B}}\left(\frac{m}{2}-1\right)\right)$$

and the functions in $K_m^{\lambda}[\mathcal{A}, \mathcal{B}]$ need not to be finitely-valent for

$$m > 2\left[1 + \left(\frac{1-\mathcal{B}}{\mathcal{A}-\mathcal{B}}\right)(1-\lambda)\right].$$

From Theorem 3.6, we have the following covering theorem.

Theorem 3.7. *Let* $2 \le m \le 2\left[1 + \left(\frac{1-\mathcal{B}}{\mathcal{A}-\mathcal{B}}\right)(1-\lambda)\right]$. Then the range of U under functions $f \in K_m^{\lambda}[\mathcal{A}, \mathcal{B}]$ contains the disc

$$|z| < \frac{2}{\lambda(\mathcal{A}-\mathcal{B}) + m\left(\frac{\mathcal{A}-\mathcal{B}}{1-\mathcal{B}}\right) + 4}.$$

Proof. Since $f \in K_m^{\lambda}[\mathcal{A},\mathcal{B}]$, then from (41), we have for $z \in \mathcal{U}$,

$$f'(z) = (g_1'(z))^{1-\alpha} p^{\lambda}(z), \quad f \in \mathcal{V}_m,\ p \in \mathcal{P}[\mathcal{A},\mathcal{B}],\ \alpha = \frac{1-\mathcal{A}}{1-\mathcal{B}}. \qquad ...(44)$$

Let $f(z)$ be given by (1), $g_1(z) = z + b_2 z^2 + b_3 z^3 + \ldots$ and $p(z) = 1 + c_1 z + c_2 z^2 + \ldots$. Then by (44), we obtain

$$1 + 2a_2 z + 3a_3 z^2 + \cdots = \left(1 + 2b_2 z + 3b_3 z^2 + \ldots\right)^{1-\alpha} \times \left(1 + c_1 z + c_2 z^2 + \ldots\right)^{\lambda}$$

and equating coefficiients of z, we obtain $2a_2 = \lambda c_1 + 2\alpha b_2$. Therefore,

$$|a_2| \le \frac{1}{2}\left(\lambda|c| + 2(1-\alpha)|b_2|\right) \qquad ...(45)$$

$$\le \frac{1}{2}\left[\lambda(\mathcal{A}-\mathcal{B}) + m\left(\frac{\mathcal{A}-\mathcal{B}}{1-\mathcal{B}}\right)\right], \qquad ...(46)$$

where we have used the bounds for functions in $\mathcal{P}[\mathcal{A},\mathcal{B}]$ (Saliu, 2019) and in $\mathcal{V}_m$ (Lehto, 1952).

The Koebe one-quarter theorem states that each omitted value w of the univalent function $f(z)$ of the form (1) satisfies

$$|w| > \frac{1}{2+|a_2|}. \qquad ...(47)$$

In view of (45) and (47), we have the required result. To establish the next theorem, we need the following hypergeometric functions defined as follow:

If $\mathrm{Re}(c) > Re(b) > 0$, then

$${}_2F_1(a,b;c;z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-zt)^{-a}dt$$

and

$${}_1F_1(b,c;z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}e^{tz}dt.$$

Theorem 3.8. *Let* $f(z) \in K_m^{\lambda}[\mathcal{A},\mathcal{B}]$. *Then for* $\mathcal{B} \le 0$,

$$\Phi_2(\mathcal{A},\mathcal{B},m,r) \le |f(z)| \le \Phi_1(\mathcal{A},\mathcal{B},m,r),$$

where

$$\Phi_1(\mathcal{A},\mathcal{B},m,r) = \begin{cases} \frac{1}{b\mathcal{B}}\sum_{k=0}^{\infty}\binom{\lambda}{k}\left(\frac{\mathcal{A}}{\mathcal{B}}\right)^{\lambda-k}\left(\frac{\mathcal{B}-\mathcal{A}}{\mathcal{B}}\right)^{k} 2^{\gamma_k+1}\left[{}_2F_1(a,b;c;-1) - r_1^{1-q_2}\,{}_2F_1(a,b;c;-r_1)\right], \mathcal{B} \ne 0, \\ \sum_{k=0}^{\infty}\binom{\lambda}{k}\frac{\mathcal{A}^k r^{k+1}}{k}\,{}_1F_1\left(k+1,k+2;\frac{\mathcal{A}m}{2}\right), \quad \mathcal{B}=0, \end{cases}$$

$$\Phi_2(\mathcal{A},\mathcal{B},m,r) = \begin{cases} \frac{1}{b\mathcal{B}}\sum_{k=0}^{\infty}\binom{\lambda}{k}\left(\frac{\mathcal{A}}{\mathcal{B}}\right)^{\lambda-k}\left(\frac{\mathcal{B}-\mathcal{A}}{\mathcal{B}}\right)^{k} 2^{\gamma_k+1}\left[r_2^{1-q_2}\,{}_2F_1(a,b;c;-r_2) - {}_2F_1(a,b;c;-1)\right], \mathcal{B} \ne 0, \\ \sum_{k=0}^{\infty}(-1)^k\binom{\lambda}{k}\frac{\mathcal{A}^k r^{k+1}}{k}\,{}_1F_1\left(k+1,k+2;\frac{-\mathcal{A}m}{2}\right), \quad \mathcal{B}=0, \end{cases}$$

q_2 is being given by (27)

$$\gamma_k = \left(\frac{\mathcal{A}-\mathcal{B}}{\mathcal{B}} - k\right),\ a = \gamma + 2,\ b = 1 - q_2,\ r_1 = \frac{1-\mathcal{B}r}{1+\mathcal{B}r},$$

$$r_2 = r_1^{-1} \text{and}\, c = b + 1. \qquad ...(49)$$

The inequality (48) cannot be improved because the function

$$f(z) = \int_0^z \frac{(1+\mathcal{B}x\xi)^{q_1}}{(1-\mathcal{B}x\xi)^{q_2}}\left(\frac{1+\mathcal{A}x\xi}{1+\mathcal{B}x\xi}\right)^{\lambda} d\xi, \quad |x| = 1, \qquad ...(50)$$

attains the equality.

Proof. Let $z = re^{i\theta}$. Then for $f \in \mathcal{K}_m^{\lambda}[\mathcal{A}, \mathcal{B}]$,

$$|f(z)| \leq \int_0^r |g(\rho e^{i\theta})||p(\rho e^{i\theta})|d\rho, \quad p \in \mathcal{P}^{\lambda}[\mathcal{A}, \mathcal{B}],\ g \in \mathcal{V}_m[\mathcal{A}, \mathcal{B}]. \qquad ...(51)$$

In view of Lemma 2.3, Lemma 2.5 and binomial expansion we obtain

$$\begin{aligned}|f(z)| &\leq \int_0^r \frac{(1+\mathcal{B}\rho)^{q_1}}{(1-\mathcal{B}\rho)^{q_2}} \left(\frac{1+\mathcal{A}\rho}{1+\mathcal{B}\rho}\right)^{\lambda} d\rho \\ &= \sum_{k=0}^{\infty} \binom{\lambda}{k} \left(\frac{\mathcal{A}}{\mathcal{B}}\right)^{\lambda-k} \left(\frac{\mathcal{B}-\mathcal{A}}{\mathcal{B}}\right)^{k} \int_0^r \frac{(1+\mathcal{B}\rho)^{q_1}}{(1-\mathcal{B}\rho)^{q_2}} \frac{d\rho}{(1+\mathcal{B}\rho)^{k}} \\ &= \sum_{k=0}^{\infty} \binom{\lambda}{k} \left(\frac{\mathcal{A}}{\mathcal{B}}\right)^{\lambda-k} \left(\frac{\mathcal{B}-\mathcal{A}}{\mathcal{B}}\right)^{k} I, \end{aligned} \qquad ...(52)$$

where

$$I = \int_0^r \left(\frac{1+\mathcal{B}\rho}{1-\mathcal{B}\rho}\right)^{q_2} (1+\mathcal{B}\rho)^{\gamma} d\rho. \qquad ...(53)$$

Using the transformation $u = \frac{1+\mathcal{B}\rho}{1-\mathcal{B}\rho}$, we have that

$$\begin{aligned} I &= -\frac{2^{\gamma+1}}{\mathcal{B}} \int_1^{r_1} u^{-q_2}(1+u)^{-(\gamma+2)} du \\ &= \frac{2^{\gamma+1}}{\mathcal{B}} \left[I_1 - I_2\right], \end{aligned} \qquad ...(54)$$

where

$$I_1 = \int_0^1 u^{-q_2}(1+u)^{-(\gamma+2)} du \qquad ...(55)$$

$$= \frac{1}{b}\, {}_2F_1(a, b; c; -1), \qquad ...(56)$$

with

$$a = \gamma + 2,$$
$$b = 1 - q_2 > 0 \ \text{ since } \ \mathcal{B} \leq 0,$$
$$c = 2 - q_2 = b + 1,$$

and

$$I_2 = \int_0^{r_1} u^{-q_2}(1+u)^{-(\gamma+2)} du.$$

Let $u = r_1 v$. Then

$$\begin{aligned} I_2 &= r_1^{1-q_2} \int_0^{r_1} v^{-q_2}(1+v)^{-(\gamma+2)} dv \\ &= \frac{r_1^{1-q_2}}{b}\, {}_2F_1(a, b; c; -r_1). \end{aligned} \qquad ...(57)$$

Using (57), (55) in (54) and (54) in (52), we have

$$|f(z)| \leq \frac{1}{b\mathcal{B}} \sum_{k=0}^{\infty} \binom{\lambda}{k} \left(\frac{\mathcal{A}}{\mathcal{B}}\right)^{\lambda-k} \left(\frac{\mathcal{B}-\mathcal{A}}{\mathcal{B}}\right)^{k} 2^{\gamma+1} \Big[{}_2F_1(a, b; c; -1) - r_1^{1-q_2}\, {}_2F_1(a, b; c; -r_1)\Big]. \qquad ...(58)$$

For the case $B = 0$, we have from (51), Lemma 2.3 and Lemma 2.5 that

$$\begin{aligned} |f(z)| &\leq \int_0^r e^{\frac{\mathcal{A}m\rho}{2}} (1+\mathcal{A}\rho)^{\lambda} d\rho \\ &= \sum_{k=0}^{\infty} \binom{\lambda}{k} \mathcal{A}^k \int_0^r e^{\frac{\mathcal{A}m\rho}{2}} \rho^k d\rho \\ &= \sum_{k=0}^{\infty} \binom{\lambda}{k} \mathcal{A}^k r^{k+1} \int_0^1 u^k e^{\frac{\mathcal{A}mr}{2}u} du, \quad \text{by putting} \quad \rho = ur \\ &= \sum_{k=0}^{\infty} \binom{\lambda}{k} \frac{\mathcal{A}^k r^{k+1}}{k}\, {}_1F_1\left(k+1, k+2; \frac{-\mathcal{A}mr}{2}\right). \end{aligned} \qquad ...(59)$$

To prove the lower bound of (48), we consider a point z_0 ($|z_0| = r < 1$) such that $|f(z)| \geq |f(z_0)|$ ($\forall\, z\colon |z| = r$). Let C be an arc in U which is mapped by the function $w = f(z)$ onto a line segment L connecting origin to the point $f(z_0)$ and lying completely in the image of U under f. Thus, by Lemma 2.3 and Lemma 2.5, we get

$$\begin{aligned} |f(z)| \geq |f(z_0)| &= \int_L |dw| \\ &= \int_C |f'(z)||dz| \geq \int_0^r \frac{(1-\mathcal{B}\rho)^{q_1}}{(1+\mathcal{B}\rho)^{q_2}} \left(\frac{1-\mathcal{A}\rho}{1-\mathcal{B}\rho}\right)^{\lambda} d\rho. \end{aligned} \qquad ...(60)$$

Employing the same techniques used for finding the upper bound in (3.30), we obtain the lower bound. This completes the proof.

For λ = 1, in Theorem 3.8, we obtain the following corollary:

Corollary 3.5. *Let* $f(z) \in \mathcal{K}_m^1[\mathcal{A},\mathcal{B}]$. *Then for* $\mathcal{B} \leq 0$, $\Psi_2(\mathcal{A},\mathcal{B},m,r) \leq |f(z)| \leq \Psi_1(\mathcal{A},\mathcal{B},m,r)$,

where

$$\Psi_1(\mathcal{A},\mathcal{B},m,r) = \begin{cases} \frac{2^{\gamma_1+1}}{b\mathcal{B}^2}\Big\{\Big[2\mathcal{A}\,{}_2F_1(\gamma_1+3,b;c;-1)+(\mathcal{B}-\mathcal{A})\,{}_2F_1(\gamma_1+2,b;c;-1)\Big] \\ -r_1^{1-q_2}\Big[2\mathcal{A}_2F_1(\gamma_1+3,b;c;-r_1)+(\mathcal{B}-\mathcal{A})\,{}_2F_1(\gamma_1+2,b;c;-r_1)\Big]\Big\}, \quad \mathcal{B}\neq 0, \\ \frac{2}{\mathcal{A}m^2}\Big\{e^{\frac{\mathcal{A}mr}{2}}\,[(1+\mathcal{A}r)m-2]+2-m\Big\}, \quad \mathcal{B}=0 \end{cases}$$

and

$$\Psi_2(\mathcal{A},\mathcal{B},m,r) = \begin{cases} \frac{2^{\gamma_1+1}}{b\mathcal{B}^2}\Big\{r_2^{1-q_2}\Big[2\mathcal{A}_2F_1(\gamma_1+3,b;c;-r_2)+(\mathcal{B}-\mathcal{A})\,{}_2F_1(\gamma_1+2,b;c;-r_2)\Big] \\ -\Big[2\mathcal{A}\,{}_2F_1(\gamma_1+3,b;c;-1)+(\mathcal{B}-\mathcal{A})\,{}_2F_1(\gamma_1+2,b;c;-1)\Big]\Big\}, \quad \mathcal{B}\neq 0, \\ \frac{2}{\mathcal{A}m^2}\Big\{e^{-\frac{\mathcal{A}mr}{2}}\,[2-(1-\mathcal{A}r)m]+m-2\Big\}, \quad \mathcal{B}=0. \end{cases}$$

The inequality is sharp for the function defined by (50) with λ=1.

Proof. For $\mathcal{B}\neq 0$, and $\lambda = 1$, (52) becomes

$$\begin{aligned} |f(z)| &\leq \int_0^r \frac{(1+\mathcal{B}\rho)^{q_1}}{(1-\mathcal{B}\rho)^{q_2}}\left(\frac{1+\mathcal{A}\rho}{1+\mathcal{B}\rho}\right)d\rho \\ &= \frac{\mathcal{A}}{\mathcal{B}}\int_0^r \frac{(1+\mathcal{B}\rho)^{q_1}}{(1-\mathcal{B}\rho)^{q_2}}d\rho + \\ &\qquad \left(1-\frac{\mathcal{A}}{\mathcal{B}}\right)\int_0^r \frac{(1+\mathcal{B}\rho)^{q_1}}{(1-\mathcal{B}\rho)^{q_2}}(1+\mathcal{B}\rho)^{-1}d\rho. \end{aligned} \qquad ...(61)$$

Evaluating the two integrals in (61) the same way we evaluated the integral (53),we easily obtain the upper bound. Applying the technique used to obtain (60) for λ = 1 and implementing the procedures to prove the upper bound in Corollary 3.5, then the lower bound is obvious.

In the case B = 0 and λ = 1, then by Lemma 2.3 and Lemma 2.5, (52) reduces to

$$|f(z)| \leq \int_0^r e^{\frac{\mathcal{A}m\rho}{2}}\,(1+\mathcal{A}\rho)\,d\rho, \qquad ...(62)$$

which gives the upper bound in this case, using integration by parts. In view of the method used to obtain (3.38),the lower bound is proved for $\mathcal{B}=0$ and $\lambda=1$.

For $\mathcal{B}=-1$ and $\mathcal{A}=1$ in Theorem 3.8, our investigation reduces to the one obtained by Noor (2002) and contained in the following corollary:

Corollary 3.6. *Let* $f \in \mathcal{K}_m^{\lambda}$. Then

$$\frac{1}{m+2\lambda}\left[\left(\frac{1-r}{1+r}\right)^{\frac{m}{2}+\lambda}-1\right] \leq |f(z)| \leq \frac{1}{m+2\lambda}\left[\left(\frac{1+r}{1-r}\right)^{\frac{m}{2}+\lambda}-1\right]. \qquad ...(63)$$

This result is sharp.

CONCLUSION

In this study, generalized close-to-convex, quasi-convex and strongly close-to-convex classes of functions were introduced. In addition, geometric properties of said functions, including the arc length result, radii problems, the necessary condition for univalence, covering theorem and distortion results were investigated. Many consequences that showed the validity of the present investigations were also highlighted.

REFERENCES

Afis S. & Noor K.I. (2020). On subclasses of functions with boundary and radius rotations associated with crescent domains. *Bulletin of the Korean Mathematical Society* **57**(6): 1529–1539.
DOI: https://doi.org/10.4134/BKMS.b200039.

Aliinta¸s O. & Kili¸c O¨.O¨.(2018). Coefficient estimates for a class containing quasi-convex functions. *Turkish Journal of Mathematics* **42**(5): 2819–2825.
DOI: http://doi:10.3906/mat-1805-90.

Brannan D. (1969). On functions of bounded boundary rotation I. *Proceedings of the Edinburgh Mathematical Society* **16**(4): 339–347.
DOI: https://doi.org/10.1017/S001309150001302X.

Cho N. & Kumar V. (2019). Arc length for the Janowski classes. *Analele Stiintifice ale Universitatii Al I Cuza din Iasi-Matematica* **Tomul LXV**(f. 1).

Duren P.L. (2001). *Univalent functions*, volume 259. Springer Science & Business Media.

Goodman A.W. (1971).*Close-to-Convex Functions of Higher Order*,volume 18. Amer Mathematical Soc. 201 Charles ST, Providence, RI 02940-2213.

Goodman A.W. (1983). *Univalent functions*. Publ. Tempa.

Janowski W. (1973). Some extremal problems for certain families of analytic functions I. *Annales Polonici Mathematici* **28**(3): 297–326.
DOI: https://doi.org/10.4064/ap-28-3-297-326.

Kaplan W. (1952). Close-to-convex schlicht functions. *The Michigan Mathematical Journal* **1**(2): 169–185.
DOI: https://doi.org/10.1307/mmj/1028988895.

Lehto O. (1952). On the distortion of conformal mappings with bounded boundary rotation. *Academiae Scientiarum Fennicae. Annales. Mathematica. Acad. Sci. Fennica, Helsinki* **124**.

Löwner K. (1917). Untersuchungen u¨ber die Verzerrung bei konformen Abbildungen des Einheitskreises $|z|<1$, die durch Funktionen mit nicht verschwindender Ableitung geliefert werden. *Verh.S¨achs. Ges. Wiss. Leipzig* **69**:89–106.

Noor K.I. & Malik B. (2010).On linear combinations of certain analytic functions. *J. Math. Anal* **1**(1):35.NoorK.I.& Thomas D. (1980). Quasi-convex univalent functions. *International Journal of Mathematics and Mathematical Sciences* **3**(2): 255–266.

Noor K.I. (1981). On analytic functions related with functions of bounded boundary rotation. *Rikkyo Daigaku sugaku zasshi* **30**(2): 113–118.

Noor K.I. (1983). On a generalization of close-to-convexity. *International Journal of Mathematics and Mathematical Sciences* **6**(2): 327–333.
DOI: https://doi.org/10.1155/S0161171283000289.

Noor K.I. (1988). Some radius of convexity problems for analytic functions of bounded boundary rotation. *Punjab* **21**: 71–81.

Noor K.I. (1991a). On radii of convexity and starlikeness of some classes of analytic functions. *International Journal of Mathematics and Mathematical Sciences* **14**(4): 741–746.
DOI: https://doi.org/10.1155/S016117129100100X.

Noor K.I. (1991b). On some integral operators for certain families of analytic function. *Tamkang J. Math* **22**: 113–117.

Noor K.I. (1994). On functions of bounded boundary rotation of complex order. *Soochow J. Math* **20**:101–111.Noor K.I. (2002). On a class related with strongly close *to* convex functions. *Mathematica (Cluj)* **44**: 67.

Noor K.I. (2009). Some properties of analytic functions with bounded radius rotation. *Complex Variables and Elliptic Equations* **54**(9): 865–877.

Noor K.I. & Noor M.A. (1992). Higher order close-to-convex functions. In: *Math. Japonica*. Citeseer.

Nunokawa M. (1969). On Bazilevi˘c and convex functions. *Transactions of the American Mathematical Society* **143**: 337–341.
DOI: https://doi.org/10.2307/1995252.

PaateroK.V.(1931).*U¨ber die Konforme Abbildung von Gebieten deren R¨ander von beschr¨ankter Drehung sind: Akad. Abh*. Ph.D. thesis.

Pinchuk B. (1971). Functions of bounded boundary rotation. *Israel Journal of Mathematics* **10**(1): 6–16.
DOI: https://doi.org/10.1007/BF02771515.

Raina R.K., Sharma P. & Sokol J. (2020). A class of strongly close-to-convex functions. *Boletim da Sociedade Paranaense de Matem´atica* **38**(6): 9–24.
DOI: https://doi.org/10.5269/bspm.v38i6.38464.

Saliu A. (2019). On generalized k-uniformly close-to-convex functions of Janowski type. *International Journal of Analysis and Applications* **17**(6): 958–973.
DOI: https://doi.org/10.28924/2291-8639.

Saliu A. & Noor K.I. (2020). On Janowski close-to-convex functions associated with conic regions. *International Journal of Analysis and Applications* **18**(4): 614–623.
DOI: 10.28924/2291-8639-18-2020-614.

Silvia E. (1983). Subclasses of close-to-convex functions. *International Journal of Mathematics and Mathematical Sciences* **6**(3): 449–458.

StumpR.K.(1971).Linear combinations of univalent functions with complex coefficients. *Canadian Journal of Mathematics* **23**(4): 712–717.
DOI: https://doi.org/10.4153/CJM-1971-080-6.

J.Natn.Sci.Foundation Sri Lanka 2021 49 (3): 441 - 450
DOI: http://dx.doi.org/10.4038/jnsfsr.v49i3.9979

RESEARCH ARTICLE

Construct validity and reliability of the Sinhala version of the Chalder fatigue questionnaire in a cohort following dengue infection in Sri Lanka

ND Perera[1*], SP Prathapan[3], DA Wass[3] and A Wijewickrama[1]
Ministry of Health, No 231, De Saram Place, Colombo 10.
National Institute of Infections Diseases, Mandavila Road, Angoda.
Faculty of Medicine, University of Sri Jayawardenapura, Gangodawila, Nugegoda.

Submitted:17 May 2020; Revised: 30 April 2021; Accepted: 09July 2021

Abstract: The objective of the study was to culturally adapt, translate and assess the validity and reliability of the 11-item Chalder fatigue questionnaire (CFQ) among adults (18 – 60 years) confirmed with dengue infection admitted to a tertiary care hospital in Colombo District, Sri Lanka. Modified Delphi technique was used in cultural adaptation and assessing face, content and consensual validity. A descriptive cross-sectional validation study was conducted among 110 patients. CFQ was administered one month after having dengue fever for assessing post-infectious fatigue (PIF). CFQ-Sinhala version (CFQ-S) was assessed for construct validity and reliability. Construct validity of CFQ-S was described with hypothesised scale structure and with confirmatory factor analysis. The culturally adapted CFQ-S confirmed the original two-factor structure among adults after one month of having dengue infection. CFQ-S demonstrated satisfactory internal consistency of ≥ 0.7. Cronbach's alpha coefficient was 0.85 for the overall scale. The test-retest reliability was assessed by calculating the intra-class correlation coefficient between the two assessments and reported as 0.89 on the overall scale. The study confirmed satisfactory levels of validity and the reliability of the CFQ-S, a valid tool to screen for PIF. The two-factor model described by the original author was confirmed as the best fitting model by triangulation of results.

Keywords: Chalder fatigue scale, construct validity, dengue fever, post-infectious fatigue, reliability

INTRODUCTION

The global incidence of dengue has amplified 30-fold over the past fifty years. It is endemic in many tropical and sub-tropical countries and the reporting of the first outbreak is also on the rise making it a universal concern. Bhatt *et al.* (2012) have projected the global incidence of dengue to be approximately 390 million cases per year, almost three times higher than the estimate of the World Health Organization (WHO). In Sri Lanka, the total confirmed as dengue infected for the year 2018 was 32989, of which 21.7 % (n = 7174) was reported from Colombo District (Ministry of Health, 2019).

The post-infectious period following a dengue infection is a relatively less studied research area. Post-infectious fatigue (PIF) after dengue infection and other viral infections have been observed in several studies. Fatigue is a subjective sensation of tiredness, lack of energy and exhaustion. When fatigue becomes chronic and accompanied by a disability, it is considered as an illness. Many physicians and researchers have looked into an entity named chronic fatigue syndrome (Gelder *et al.*, 2009). Fukuda *et al.* (1994) proposed a conceptual framework to define and study fatigue. Fatigue lasting more than or equal to one month is termed prolonged fatigue (Fukuda *et al.*, 1994).

A review of fatigue measuring scales by Hjollund (2007) reveals that there are 252 different approaches to assess fatigue. There are ad-hoc approaches, assessing by single questions, using multi-system scales, or by 'fatigue specific' scales. The most commonly used fatigue

* Corresponding author (nadeekayasanth@gmail.com https://orcid.org/:0003-2725-3086)

precise scales are the fatigue severity scale, fatigue questionnaire/Chalder fatigue scale, multi-dimensional fatigue inventory, Piper fatigue scale, fatigue impact scale, etc. The Chalder fatigue questionnaire, (CFQ) fatigue impact scale and the Piper fatigue scale have been used to assess fatigue after infection (Hjollund *et al.*, 2007).

Several studies have assessed PIF following an infection of dengue. One study carried out in Singapore assessing PIF using the fatigue questionnaire (FQ)/ Chalder fatigue questionnaire (CFQ) two months following hospitalisation has described this clinical entity. CFQ, a tool validated in several settings assesses the physical and mental dimensions of fatigue. Feeling of the presence of exhaustion and lack of energy has been measured in the physical fatigue section and subjective feeling of being psychologically exhausted with consensus on concentration, recall and speech has encompassed in the mental fatigue section (Chalder *et al.*, 1993; Seet *et al.*, 2007; Cella & Chalder, 2010).

Chalder fatigue questionnaire (CFQ)

CFQ was originally developed as a 14-item scale to measure the severity of fatigue, and special care was taken to develop the tool as a generic measure. Symptoms that are directly related to fatigue are included and those symptoms only associated with chronic fatigue syndrome are excluded from the tool. The rating of items is on a four-point Likert scale as in the general health questionnaire (GHQ). Two methods for scoring have been described as a bimodal system (GHQ method) and a four-point system (four-point Likert score) (Chalder*et al.*,1993).

Initially, 275 newly registered patients at a general practice setting completed the CFQ. Another 100 consecutive attendees completed the CFQ and the fatigue section of the revised clinical interview schedule (CIS-R). Exploratory factor analysis was conducted in both sets of samples by principal component analysis (PCA) and a two-factor model; physical fatigue dimension and mental fatigue dimension has emerged. Considering the results, an 11- item fatigue scale was developed and it has shown better validity and reliability over the 14- item scale. The internal consistency reliability was calculated for the revised version and the Cronbach's alpha for the 11-item version was 0.89 with the physical fatigue and the mental fatigue sub-scales having Cronbach's alpha values of 0.845 and 0.821, respectively. Criterion validity had been assessed in the sample of 100 attendees by computing a two by two table and a receiver operating characteristic (ROC) curve. The cut off value was decided at 3/4 with a sensitivity of 75.5 % and specificity of 74.5 %. These results should be interpreted with caution, since the CIS-R is not a 100 % criterion measure and considering the small sample size of 100 (Chalder *et al.*,1993). There is evidence for post-infectious fatigue/persistent fatigue following a dengue infection from international as well as local studies. The CFQ has been translated and cross-culturally adapted in different settings such as Brazil and China and in Hong-Kong among diverse study populations (Cho *et al.*,2007; Won & Fielding, 2010; Fong *et al.*, 2015; Jing *et al.*, 2016;). The study conducted in Brazil tested psychometric measures by Cho *et al.* (2007) among primary care attendees. A pilot study (n = 204) and a proper validation study (n = 304) had been conducted. Study participants were assessed with the CFQ and the fatigue section of the CIS-R. The Brazilian version of the fatigue scale was shown to reproduce the two dimensional factor structure following PCA. The Cronbach's alpha was 0.88 confirming satisfactory reliability (Cho *et al.*, 2007).

Wong and Fielding (2010) reported findings from their study on the construct validity of a Chinese version (Cantonese version) of the CFQ. The study participants (n = 201) were assessed by the Chalder fatigue scale, short form health survey (SF-12) and hospital anxiety and depression scale (HADS). Confirmatory factor analysis (CFA) was tested for one factor, two factor and a three factor model. A two-factor correlated model showed model fit, which was quite similar to the original English version. Good internal consistency was demonstrated with a Cronbach's alpha of 0.863 (Wong & Fielding, 2010).

Considering the local studies, Ball *et al.* (2011) had used CFQ in assessing fatigue among the general population and all participants were assessed with the CFQ Bradford somatic inventory (BSI) and the short form 36 Health Survey questionnaire. There were a total of 37 items and a confirmatory factor analysis was conducted via M-plus. In their study, they have included 13 items of fatigue as one sub-scale with the other two scales, which is not exactly similar to the CFQ scale used in the current study, which has only 11 items (Ball *et al.*, 2011). Although there were several local studies assessing the prevalence of fatigue among dengue patients the researchers were unable to gather evidence of a proper validation study of a tool measuring fatigue among patients with dengue infection or any other infectious disease. Therefore, it is considered as a timely requirement to evaluate the validity and reliability of a suitable tool to assess post-infectious fatigue among patients suffering from a dengue infection.

METHODOLOGY

The current study adopted a systematic process in selecting a suitable tool to assess PIF, cultural adaptation, translation to Sinhala language and to evaluate judgemental validity, construct validity and reliability.

Operationalisation of the concept of PIF was done with an extensive literature survey and listing down of all the available definitions and finalised with a group of experts in Medicine, Psychiatry, Neurology, Community Medicine and Immunology. The definition of post-infectious fatigue, following dengue infection was operationalised as 'a subjective feeling of tiredness, lack of energy and exhaustion lasting for at least one-month duration following dengue infection' (Fukuda *et al.*, 1994; Seet *et al.*, 2007; Gelder *et al.*,2009).

A tool was selected based on certain criteria; tools originally developed in English language, after the year 1980 and with an acceptable level of validity and reliability. After reviewing nearly twenty tools, five tools were selected based on the operationalised definition and the context of the study. Fatigue severity scale -FSS (Krupp *et al.*, 1989), Piper fatigue scale-revised -PFS (Piper *et al.*, 1989), Chalder fatigue questionnaire -CFQ /Fatigue questionnaire - FQ (Chalder *et al.*,1993), multidimensional fatigue inventory -MFI (Smets *et al.*, 1995) and fatigue section of SF-36 (Ware & Sherbourne, 1992) were selected for further review. A Modified Delphi technique was carried out in selecting the most appropriate tool for the current research. All the experts unanimously selected the CFQ as the best tool to assess the post-infectious fatigue in the first round. The underlying reasons for the selection of the CFQ were; it follows the operationalised definition; it is a simple yet a multidimensional tool; easy to administer with an acceptable level of validity and reliability; it has been used to assess post-infectious fatigue following dengue infection in Singapore and in Sri Lanka previously in an unpublished study (Seet *et al.*, 2007). Although it has been developed to assess fatigue severity in general practice settings, it has also been used in various situations to assess fatigue (Cella & Chalder, 2010).

Permission was obtained from the author to use the CFQ following cultural adaptation. Further discussions were conducted with the author about the suitability of the tool to assess post-infectious fatigue among dengue patients. The tool has been originally developed as a self-administered tool, considering the different educational level of the participants in the current context, CFQ was used as an interviewer-administered tool with the permission of the author.

Several techniques have been discussed in the literature on the translation of technical instruments. In this research, the forward and backward translation method was selected (Tsang *et al.*, 2017; World Health Organization, 2017).

Cultural adaptation of the CFQ was done using a Modified Delphi technique with a team of experts (n=8) from the fields of clinical medicine, psychiatry, community medicine, neurology and psychology. This iterative procedure was conducted in two rounds. During the first round, the panel was provided with a concept note explaining the objectives, detailed description on CFQ and information regarding the research. In the first round, the following were assessed based on a five-point Likert scale; relevance in assessing fatigue among adult patients with dengue in Sri Lanka and the appropriateness of the words used in the local context. They were further asked to indicate their suggestions on how an item should be modified if they assign a score less than or equal to three for an item. Further, expert opinion was obtained on the scoring method for the tool and the cut-off guideline. The mean scores ranged from 3 to 5 for an item. If an item received an average mark of less than three by more than 50 % of the expert panel, it was considered as the cut off to remove an item from the questionnaire. However, no items scored less than three and therefore no item was removed from the tool. The expert panel did not suggest any additional items to be added and as a result, the item structure remained unchanged.

Item number one, four, six and eleven had received equal to or less than an average mark of four, therefore these items were highlighted in the second iteration. The principal investigator (PI) discussed with the experts regarding the suggested amendments, and modifications were done to the above items. The older version and the modified version were presented with the average marks in the concept note of the second iteration. The expert panel was revisited and opinion was obtained regarding the cultural acceptability of those items again. After the second iteration, all the item scores were summarised. An average mark of more than four was received for all the items in the tool and the tool was finalised.

The culturally adapted and translated CFQ-S was pre-tested among ten patients, aged 18 – 60 years, admitted to Colombo South Teaching Hospital (CSTH), who were diagnosed with dengue infection. They were interviewed one-month post-infection. Validity refers to how accurately a study instrument measures the intended variable (Abramson & Abramson, 1999; Friss & Sellers, 2014). 'Fatigue' is a subjective and abstract phenomenon,

which does not have a concrete gold standard (Chalder *et al.*, 1993; Dittner *et al.*, 2004).

During the development stage of the original tool, Chalder *et al* (1993) had considered the fatigue section of the revised clinical interview schedule (RCIS) as a gold standard and reported receiver operating curve (ROC) statistics. They have taken the optimal cut off as ¾ with a sensitivity of 75.5 % and a specificity of 74.5 % , which the results should be interpreted with this limitation. In the current study, the convergent validity or the criterion validity of the Chalder fatigue scale (S) was not considered due to the unavailability of the Sinhala validated version of CIS−R among adults in the local setting.

Therefore, a triangulation approach was used; with the use of several complementary validation methods which would provide the most accurate approximate assessment. Hence, judgmental validity and construct validity were assessed in the study (Abramson & Abramson, 1999).

Face validity of the CFQs was assessed, concerning subjective qualities, such as whether it assesses the level of fatigue among dengue patients. The responses by the study participants in the pre-test were referred in assessing the face validity.

There is evidence from the literature that the CFQ-English version is a valid scale to assess the severity of fatigue. During the translation process, measures were taken to ensure the semantic and conceptual equivalence of the translated version, so that the content validity will be agreeable in the translated version.

Further, the appraisal by the multi-disciplinary team of experts (n = 8) approved the content and consensual validity of the CFQ. Each item in the scale was assessed for the following; relevance in assessing PIF among adult dengue patients, appropriateness of the wording used and acceptability in the local context in assessing PIF among adult dengue patients.

Procedure in appraising construct validity

A descriptive cross-sectional validation study was conducted in a tertiary care hospital in Colombo, Sri Lanka, from April to June 2018. Adults, who were resident in Colombo District for the past six months, of 18-60 years of age and clinically diagnosed and confirmed as dengue infected by a consultant physician and/or by the presence of NS I antigen and/or dengue specific IgM in their serum, comprised the study population. Those who were diagnosed with a mental illness, pregnant mothers, those who could not comprehend well and patients who were unable to respond to an interviewer-administered questionnaire in Sinhala were excluded. For the current study, considering a subject to variable (STV) ratio of 10:1, the sample size was calculated as 110 with an added 20 % to account for loss to follow up the total sample size accounted for 138.

The non-probability consecutive sampling method was used in recruiting the patients from the medical units in CSTH and they were followed up to one month post-infection at the hospital. The socio-demographic details were collected when recruiting patients on the date of discharge from the hospital. Culturally adapted translated and validated CFQ was used for assessment of fatigue at one month post-infection via an interviewer-administered questionnaire. Ethical approval was taken from the Ethics Review Committee of the University of Kelaniya and CSTH. Permission was taken from the Director/CSTH to conduct the study. Data were collected by the nursing graduates after a training on objectives of the study and data collection techniques.

Preparatory data analysis

Scoring was done according to the instructions given with the Chalder fatigue questionniare. According to the scoring system, the higher scores indicated a greater level of fatigue (Chalder *et al.*, 1993). Before carrying out data analysis, the dataset was evaluated for appropriateness and compliance with the assumptions required by the analytical techniques in CFA.

The CFQ is based on continuous scores and had two sub-scales of fatigue, namely, physical fatigue and mental fatigue as described by the author (Chalder *et al.*, 1993). The scores were recorded on a four-point Likert scale. The values for each item varied from zero to three and the aggregate scores varied from zero to 33. The univariate standardised skewness values in all eleven items ranged from -0.765 to 4.045 and the univariate kurtosis values ranged from -2.010 to 4.690. In the sample, three and two items out of 11 items showed high skewness and kurtosis values, respectively.

The Kolmogorov-Smirnov test and the Shapiro-Wilk test are the other tests used to determine the normality of the dataset. In the sample, both tests were significant ($p < 0.05$). The results of all three techniques showed that the data were not normally distributed. Therefore, the Robust Maximum Likelihood (RML) estimation was used in conducting confirmatory factor analysis with LISREL software.

Adequacy of the sample size was assessed by Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy , which reported a value of 0.847. Bartlett's test of sphericity is used to test the null hypothesis that variables in the sample correlation matrix are not correlated and showed a chi-square value of 517.021(df = 55, p value < 0.001). The correlation matrix was observed in the data to evaluate inter-item correlations. Most of the correlations (68%) were more than 0.3, which is indicative of a satisfactory inter-item correlation for factor analysis. Almost all item combinations showed variance inflation factor (VIF) values less than 3, suggestive of absence of multicollinearity.

Data processing and statistical analysis

Univariate analysis was conducted by SPSS-21 version to analyse the descriptive data of the sample. Descriptive statistics are presented as frequency distributions.

Construct validity of the CFQ was assessed using two methods; assessment of the hypothesised scale structure and performing confirmatory factor analysis. To assess the hypothesised scale structure, a multi-trait scaling analysis was conducted using the SPSS-21 software using two methods. This procedure is centred on an analysis of the item scale correlations. Therefore, correlations of each item with the sub-scales were assessed. Depending on the item-scale correlations, item-convergent and item-discriminant validity were evaluated. In the first method, confirmation of the item convergent validity is defined as a correlation of 0.40 or greater between an item and its own sub-scale. Confirmation of the item discriminant validity is established by comparing the degree of the correlation with an item with its particular sub-scale in comparison with other sub-scales. Item discriminant validity was supported by two criteria. First, the highest correlation in a row is the correlation between the item with its own sub-scale, discriminating the other sub-scale. Second, by checking whether each item is correlating significantly with its own sub-scale; 'whether the correlation between an item and its hypothesised scale is more than two standard errors higher, than its correlation with the other scales'. The cut off value for detecting the significance level was calculated by multiplying the standard error of each item by 1.96 and subtracting the resulting value from the correlation score of its own sub-scale. Each item was considered as a scaling success if the particular cut off value is higher than the correlation value of the other sub-scale (Hays *et al.*, 1998).

In the second method, the Average Variance Extracted (AVE) and the Composite Reliability (CR), was assessed to confirm convergent validity. The recommended minimum of AVE to have a satisfactory convergent validity was 0.5, and the CR value should be > 0.7. To assess the discriminant validity, the AVE values were compared with the squared inter-construct correlation and to fulfill satisfactory discriminant validity, the AVE of each domain should be higher than the squared inter-construct correlation (Renko *et al.*, 2001).

There is evidence that the CFQ is composed of two sub-scales; physical fatigue and mental fatigue. (Chalder *et al.*, Cho *et al.*, 2007; Cella & Chalder, 2010; 1993; Won & Fielding, 2010). A confirmatory factor analysis was conducted to assess whether the hypothesised scale structure can be reproduced in the study sample via the Linear Structural Relations (LISREL) 8.8 software. In assessing the overall goodness of fit in model evaluation, several model fit indices were looked into, since each index would provide different information regarding the assessment model. They are described under three categories of model fit indices; absolute fit indices, relative fit indices, parsimony fit indices. It is recommended that at least one index from each category should be within the expected level to decide on the acceptability of the model. The following fit indices were assessed and the desired level for model fit is presented within parenthesis; Absolute fit indices (Satorra Bentler scaled chi-square test ($p > 0.05$), root mean square of approximation (RMSEA- < 0.08), goodness of fit index (GFI- > 0.90) and adjusted goodness of fit index (AGFI- > 0.90). The relative fit indices were; comparative fit index (CFI - > 0.95) and non-normed fit index (NNFI - > 0.95). parsimony fit indices were; (Parsimony Goodness of Fit Index (PGFI - >0.5) and parsimonious normed fit index (PNFI - >0.5) (Brown, 2006).

Assessment of CFA

CFA was assessed in two phases;

i. In the first phase, two-factor model was assessed: The first seven items were loaded on to a subscale named 'physical fatigue (PF)' and the last four items were loaded on to a subscale named "mental fatigue (MF)", which has been evaluated by the original author.

ii. In the second phase, the modifications suggested by the LISREL software to improve the model fit were

considered. Several modifications were done. Two error covariance was added between the two subscales and a path was drawn from mental fatigue item 3 (MF3) to physical fatigue (PF) sub-scale.

Assessment of reliability

The reliability of the CFQ was assessed, since it is an essential technique in predicting both random and systematic error in any measurement tool (Streiner *et al.*, 2015). The internal consistency and test-retest methods were used to measure reliability by SPSS version 21. Test re-test reliability of the tool measuring post-infectious fatigue was evaluated by administering the same tool to a sub-sample of 20 patients selected randomly, with an interval of seven days. Internal consistency was evaluated by computing the Cronbach's alpha of the post-infectious fatigue assessment tool. According to Nunnally's criterion, internal consistency estimates of a magnitude of 0.7 or greater was considered acceptable (Abramson & Abramson, 1999). For test-retest reliability, a correlation coefficient (Spearman's *r*) of 0.70 or greater was considered acceptable (Litwin, 1995).

RESULTS AND DISCUSSION

Selection, cultural adaptation and translation of the Chalder fatigue questionnaire

Operationalisation and selection of an appropriate tool to measure post-infectious fatigue was a major challenge at the planning stage, which was achieved with a systematic process as described in the methods section. Abramson and Abramson (1999) describe that by operationalising how we measure the outcome of interest should be expressed with objectively apparent details, should be easy to understand and unambiguous.

The possibility and the implications of developing a new tool vs. validating an existing tool were assessed extensively. At the planning stage, there were no Sri Lankan studies that had assessed PIF following dengue infection. There were only two local studies that had assessed fatigue among twins (general population) and Navy officers (Ball *et al.*, 2011; Hanwella *et al.*, 2014). A survey among Sri Lankan physicians revealed that 77 % were reported as post-viral fatigue (Kularatne, 2005). There was textbook evidence of the presence of fatigue following dengue infection. Yet those descriptions were not comprehensive (Kumar & Clark, 2005; Walker *et al.*, 2014). Apart from these findings little was known about the topic. On the other hand, 'fatigue' as a generic term has been debated and assessed at many stages by different authorities over fifty years. Therefore, considering all these factors, it was finally decided that using an already established generic tool to assess post-infectious fatigue would have more scientific advantages than developing a new tool.

Modified Delphi technique was used in deciding the most appropriate tool, in the cultural adaptation process and in assessing face, content and consensual validity of the CFQ. Following cultural adaptation, the item number did not alter after two iterations. Item number one, four, six and eleven were modified. The CFQ was translated into Sinhala language giving due consideration to safeguard semantic equivalence and theoretical equivalence.

This objective was accomplished by involving translators with technical expertise as well as language expertise. During forward translation, the translators were provided with a guide to the translation process. Further, a team comprising the translators, supervisors and the PI discussed the suitability of each translated item. In contrast, to evade bias, the back translators were not given any explanation regarding the tool and kept blind to the original English version.

Pre-testing of the tool was conducted among a similar age group (18–60 years) and considering the similar eligibility criteria used in the study, they were not included as study participants.

Validity of the Sinhala version of the CFQ

Face, content and consensual validity were confirmed by another panel of experts in the fields of Community Medicine, Psychiatry, Neurology, Immunology, Psychology and Clinical Medicine, through a Modified Delphi process. Exploratory factor analysis was not considered in the study based on several factors; the items of the Chalder Fatigue Scale did not change after the cultural adaptation process and after evaluating judgmental validity, only the wordings of items one, four, six and eleven were changed to improve how the meaning of the items was delivered to the participants and there was prior evidence from the literature on its factor structure. In conducting the validation study to appraise construct validity, a total of 140 patients were recruited and 20 patients were lost to follow up, ending with a final sample of 120 (response rate – 85.7 %). The mean age was 29.6 years with a standard deviation of 10.1 years. The socio-demographic details are presented in Table 1.

Table 1: Characteristics of the study sample

Characteristic	Frequency (n=120)	Percentage (%)
Age of the patient		
18 - 25 years	54	45.0
26 – 35 years	35	29.2
36 – 45 years	19	15.8
46 - 55 years	12	10.0
Gender		
Male	73	60.8
Female	47	39.2
Status of dengue infection		
Dengue fever	52	42.3
Dengue haemorrhagic fever	68	57.7

Multi-trait scaling analysis

Item convergent and discriminant validity were tested for a two factor model using two methods.

Method I

The first seven items were included in the 'physical fatigue' domain and the last four items were included in the "mental fatigue" domain. Item convergent validity was established in the CFQ for the two-factor structure as each item correlates with its sub-scale with a correlation of > 0.4. The item to physical fatigue domain correlations varied from 0.778 to 0.473. The items of mental fatigue domain correlations varied from 0.557 to 0.363. In the mental fatigue sub-scale, only one item was having a correlation of 0.363, and which approximates with 0.4. The item discriminant validity was supported since each item correlates more strongly with its own sub-scale than with the other sub-scale. Further, each item was assessed for item scaling; the correlation between an item and its own sub-scale was significantly higher (> 1.96 standard errors) than the correlation with the other sub-scales. All the 11 items showed success in item scaling. When calculating item-scale correlations, the own item was excluded from the scale total to adjust for inflation of the correlation (Hays *et al.*, 1998). The results are further described in Table 2.

Table 2: Multi trait correlation matrix for the Chalder fatigue questionnaire in two-factor model

Item	Physical fatigue sub-scale score	Mental fatigue sub-scale score	Standard Error	Cut off value (-1.96SE)	Scaling success
1. Do you feel tired? (PF1)	0.657	0.173	0.059	0.541	success
2. Do you need to rest more? (PF2)	0.763	0.361	0.055	0.655	success
3. Do you feel sleepy or drowsy? (PF3)	0.554	0.398	0.058	0.44	success
4. Do you have problems in starting activities? (PF4)	0.473	0.286	0.049	0.377	success
5. Do you lack energy? (PF5)	0.732	0.345	0.054	0.626	success
6. Do you lack muscle strength? (PF6)	0.631	0.172	0.053	0.527	success
7. Do you feel weak? (PF7)	0.778	0.231	0.054	0.672	success
8. Do you have difficulties in concentrating? (MF1)	0.374	0.520	0.049	0.424	success
9. Do you make slips of the tongue when speaking? (MF2)	0.091	0.363	0.033	0.299	success
10. Do you find it more difficult to find the correct word? (MF3)	0.143	0.441	0.031	0.38	success
11. Do you have problems in remembering things? (MF4)	0.362	0.557	0.055	0.449	success

Method II

Factor loadings were explored following varimax rotation and the factors were loaded in to two domains; first seven items in to one domain and the latter four into another domain. The average variance extracted (AVE) and the composite reliability (CR) values were calculated, and presented in Table 3.

Table 3 : Average variance extracted (AVE) and the composite reliability (CR) for the two factor model

Domain	AVE(≥ 0.5)	CR (≥ 0.7)
Physical Fatigue domain	0.6	0.9
Mental Fatigue domain	0.5	0.8

The AVE for both domains were ≥ 0.5 and the CR was ≥ 0.7, and confirms satisfactory convergent validity (Rienko *et al.*, 2001). This model was further used to evaluate discriminant validity. The average variance extracted (AVE) in both constructs were compared with the squared inter-construct correlation. The Spearman correlation coefficient between the physical fatigue and the mental fatigue domains were 0.387. The squared inter-construct correlation was calculated as 0.1497. It was evident that the AVE in both domains were more than the squared inter-construct correlation, and confirms discriminant validity (Rienko *et al.*, 2001).

Confirmatory factor analysis

The two-factor model was tested and this model showed acceptable model fit in the fit indices representing all three categories. Next, the modifications suggested by the LISREL software were tested as described in the methods section. The results are presented in Table 4.

As described in Table 4, the Satorra-Bentler scaled chi-square test, RMSEA, GFI, AGFI, CFI, NNFI, PGFI and PNFI indices showed satisfactory levels for model fit. The two-factor structure consisting of physical and mental sub-scales was the most accepted factor structure of CFQ in literature, which was confirmed in this study as well (Chalder *et al.*, 1993; Cho *et al.*, 2007; Cella & Chalder, 2010; Won & Fielding, 2010).

There were suggestions from LISREL software regarding methods to further improve model fit. These methods of improving model fit included adding error covariance between the items and adding error covariance between an item and a factor. There was evidence from the literature on testing on suggested modifications by the software (Won & Fielding, 2010). The fit indices were presented for a modified two-factor model. Considering the model fit indices of the modified models, the values were slightly better for the two-factor model in chi-square statistic, RMSEA and the parsimony fit indices. The values were similar for GFI, CFI and NNFI. Considering the above all modifications, the two-factor model added with the suggested error covariance and the path change showed the highest satisfactory indices. though with the suggested modifications the model showed improved model fit, the changes in model fit indices did not vary much. Brown (2006) argues against adding correlated error terms to an already fitting model to improve model fit. Since these suggestions by the software solely depends on the data, it may affect the generalisability of the findings (Brown, 2006). Therefore, by considering all the results and expert opinion from the consultant psychiatrist and the statistician, the original two-factor model was selected as the best-fitted model in the current study. The standardised parameter estimates for the original two-factor model is presented in figure 1.

Table 4: Results of confirmatory factor analysis

	Absolute model fit indices				Relative fit indices		Parsimony fit indices	
Model	Chi-Square	RMSEA	GFI	AGFI	CFI	NNFI	PGFI	PNFI
Two factor-original	X^2=57.76							
	df = 43,p = 0.07	0.054	0.9	0.85	0.98	0.98	0.59	0.73
Modified two factor [a]	X^2=44.86							
	df = 41,p = 0.31	0.028	0.92	0.87	1.0	0.99	0.57	0.71
Modified two factor [b]	X^2=43.41							
	df = 40,p = 0.33	0.027	0.92	0.88	1.0	0.99	0.56	0.69

X^2 = Satorra-Bentler scaled chi square test (desired value p > 0.05), RMSEA = root mean square error of approximation (desired value < 0.08), GFI = Goodness of fit index (desired value > 0.9), AGFI = adjusted goodness-of-fit index (desired value > 0.9) , CFI = comparative fit index (desired value > 0.95), NNFI = non-normed fit index (desired value > 0.95), PGFI = parsimony goodness of fit index (desired value > 0.5), PNFI = parsimonious normed fit index (desired value > 0.5)

[a] set error covariance between mf1 & pf4,

[b] set error covariance between mf1 & pf4 and pf2 & pf1

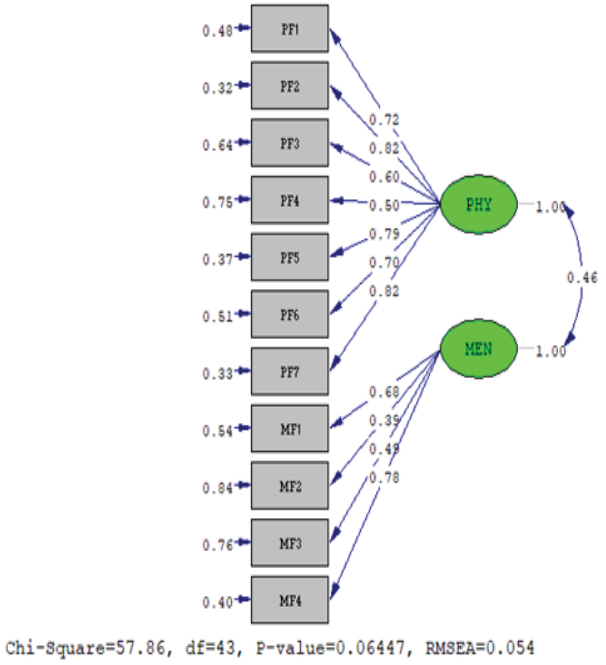


Figure 1: Standardised parameter estimates for the factor structure of the original two-factor model

Reliability

Validity and reliability are two complementing characteristics that improve the quality of data, bridging the phenomena of interest and the actual measurements Considering the internal consistency, the Cronbach's alpha coefficient was 0.85 for the overall scale. Cronbach's alpha of 0.874 and 0.673 were reported for the physical fatigue and mental fatigue sub-scale, respectively. It is considered that the Cronbach's alpha coefficient of > 0.7 as having satisfactory internal consistency (Abramson & Abramson, 1999). It is discussed that the scales with a lesser number of items (less than ten) might get a Cronbach's alpha value up to 0.5 because the Cronbach's alpha is very sensitive to the number of items in a scale (Pallant, 2013). Hence the Cronbach's alpha of 0.673 for the mental fatigue sub-scale which contains four items was justifiable. During the development of the initial tool, the original author reported an overall Cronbach's alpha of 0.89 for the revised 11 item scale. For physical fatigue sub-scale and mental fatigue sub-scale, the Cronbach's alpha were 0.84 and 0.82, respectively (Chalder *et al.*, 1993). The test-retest reliability was assessed by calculating the intra-class correlation coefficient between the two assessments. The questionnaire was administered by the PI with an interval of one week. The correlation coefficients were more than 0.7, indicating good test-retest reliability and all the coefficients were statistically significant ($p < 0.001$).

CONCLUSION

In summary, the CFQ (S) was culturally adapted and validated ensuring scientific guidelines, as it was a relatively less investigated area locally and globally. The CFQ (S) had shown satisfactory validity and reliability among Sinhala conversant, 18 to 60 year old people at the local setting. Therefore, this study confirmed that the culturally adapted, translated and validated Sinhala version of the Chalder fatigue questionnaire is a valid tool to screen for post-infectious fatigue among adults who had dengue infection, amid those who were conversant with the Sinhala language.

Acknowledgements

Authors are grateful to the University of Sri Jayewardenepura for the financial support (ASP/01/RE/MED/2018/73) and Professor Trudie Chalder for granting permission to use the Chalder fatigue questionnaire with cultural adaptation.

Competing interests: The authors declare that there are no competing interests.

REFERENCES

Abramson J.H. & Abramson Z.H. (1999). *Survey Methods in Community Medicine.* Churchill Livingston; London,UK.

Ball H.A., Siribaddana S.H., Sumathipala A., Kovas Y., Glozier N., Rijsdijk F., McGuffin P. & Hotopf M. (2011). Genetic and environmental contributions to the overlap between psychological, fatigue and somatic symptoms: a twin study in Sri Lanka. *Twin Research and Human Genetics* **14**(1): 53–63.
DOI: https://doi.org/10.1375/twin.14.1.53

Bhatt S., Gething P., Brady O., Messina J., Farlow A. & Moyes C. (2012). The global distribution and burden of dengue. *Nature* **496** (7446): 504–507.
DOI: https://doi.org/10.1038/nature 12060

Brown T.A. (2006). *Confirmatory Factor Analysis for Applied Research,* Guilford Publications, Inc.72, Spring Street, New York, USA.

Cella M. & Chalder T. (2010). Measuring fatigue in clinical and community settings. *Journal of Psychosomatic Research* **69**(1): 17–22.
DOI: https://doi.org/10.1016/j.jpsychores.2009.10.007

Chalder T., Berelowitz G., Pawlikowska T., Watts L. & Wessely S.W.D. (1993). Development of fatigue scale. *Journal of Psychosomatic Research* **37**:147– 53.

Cho H.J., Costa E., Menezes P.R., Chalder T., Bhugra D. & Wessely S. (2007). Cross-cultural validation of the Chalder Fatigue Questionnaire in Brazilian primary care. *Journal of Psychosomatic Research* **62**(3): 301–304.
DOI: https://doi.org/10.1016/j.jpsychores.2006.10.018

Dittner A. J., Wessely S. C. & Brown R. G. (2004). The assessment of fatigue: A practical guide for clinicians and researchers. *Journal of Psychosomatic Research* **56**(2):157–170.
DOI: https://doi.org/10.1016/S0022-3999(03)00371-4

Fong T.C., Chan J.S., Chan C.L., Ho R.T., Ziea E.T., Wong V.C., Ng B.F. & Ng S.M. (2015). Psychometric properties of the Chalder Fatigue Scale revisited: an exploratory structural equation modelling approach. *Quality of Life Research* **24**(9) : 2273-2278.
DOI: https://doi.org/10.1007/s11136-015-0944-4.

Friis R.H. & Sellers T.A. (2014). *Epidemiology for Public Health Practice*, 5th Edition, Jones & Bartlett Publishers. Wall Street Burlington,USA.

Fukuda K., Straus S.E., Hickie I., Sharpe M.C., Dobbins J.G. & Komaroff A. (1994). The chronic fatigue syndrome: a comprehensive approach to its definition and study. *Annals of Internal Medicine* **121**(12): 953–959.
DOI:https://doiorg/10.7326/0003-4819-121-12-1994 12150-00009.

Gelder M.G., Andreasen N.C., Lopez-Ibor Jr J.J., & Geddes J.R. (eds.) (2009). *New Oxford Textbook of Psychiatry*,2nd edition, Oxford University Press UK.

Hanwella R., Jayasekera N.E. & De Silva V.A. (2014). Fatigue symptoms in Sri Lanka Navy personnel deployed in combat areas. *Ceylon Medical Journal* **59**(2): 39–44.
DOI: http://dx.doi.org/10.4038/cmj.v59i2.7062

Hays R.D., Hayashi T., Carson S. & Ware J.E. (1998). *User's Guide for the Multitrait Analysis Program* (*MAP*). The Rand Corporation, Santa Monika, USA.

Hjollund N.H., Andersen J.H. & Bech P. (2007). Assessment of fatigue in chronic disease: a bibliographic study of fatigue measurement scales. *Health and Quality of life Outcomes* **5**(1): 12.
DOI: https://doi.org/10.1186/1477-7525-5-12

Howard K.I. & Forehand G.A. (1962). A method for correcting item-total correlations for the effect of relevant item inclusion. *Educational and Psychological Measurement* **22**(4): 731–735.
DOI: https://doi.org/10.1177/001316446202200407

Jing M.J., Lin W.Q., Wang Q., Wang J.J., Tang J., Jiang E.S., Lei Y.X. & Wang P.X. (2016). Reliability and construct validity of two versions of the Chalder fatigue scale among the general population in mainland China. *International Journal of Environmental Research and Public Health* **13**(1):147
DOI: https://doi.org/10.3390/ijerph13010147

Krupp L.B., LaRocca N.G., Muir-Nash J. & Steinberg A.D. (1989). The fatigue severity scale: application to patients with multiple sclerosis and systemic lupus erythematosus. *Archives of Neurology* **46**(10) : 1121–1123.
DOI:https:/doiorg/10.1001 archneur.1989.00520460115022

Kularatne S.A. (2005). Survey on the management of dengue infection in Sri Lanka: opinions of physicians and pediatricians. *Southeast Asian Journal of Tropical Medicine and Public Health.* **36**(5): 1198.
DOI: https://doi.org/10.1186/s13104-015-1085-0.

Kumar P. & Clark M. (2005). *Clinical Medicine*, 6th edition, Saunders Elsevier, London, UK.

Litwin M.S. (1995). *How to Measure Survey Reliability and Validity*. Sage Publications, USA.

Ministry of Health. (2019). *Epidemiology of Dengue*; National Dengue Control Unit: Ministry of Health, Colombo.

Pallant J. (2013). *SPSS Survival Manual*, McGraw-Hill Education, UK.

Piper B.F., Lindsey A.M., Dodd M.J., Ferketich S., Paul S.M. & Weller S. (1989). In: *Key Aspects of Comfort: Management of Pain, Fatigue, and Nauses* (eds. S.G. Funk, E.M. Tornquist, M.T. Champagne, L.A. Copp & R.A. wiese), PP.199-240. Springer Publishing Company, New York, USA.

Renko H.Y. & Autio E. (2001). Social capital, knowledge acquisition and knowledge exploitation in young technology-based firms. *Strategic Management Journal* **22**(6-7): 587–613.
DOI: https://doi.org/10.1002/smj.183

Seet R.C., Quek A.M. & Lim E.C. (2007). Post-infectious fatigue syndrome in dengue infection. *Journal of Clinical Virology* **38**(1):1–6.
DOI:https://doi.org/ 10.1016/j.jcv.2006.10.011.

Smets E.M., Garssen B., Bonke B.D. & De Haes J.C. (1995) The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *Journal of Psychosomatic Research* **39**(3): 315–25.
DOI:https://doi.org/ 10.1016/0022-3999(94)00125-o.

Streiner D.L., Norman G.R. & Cairney J. (2015). *Health Measurement Scales: a Practical Guide to their Development and Use*. Oxford University Press, USA.

Tsang S., Royse C.F. & Terkawi A.S. (2017). Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi Journal of Anaesthesia* **11**(Suppl. 1): S8
DOI:https://doi.org/ 10.4103/sja.SJA_203_17

Walker B.R., Colledge N.R., Ralston S.H. & Penman I.D. (eds.) (2014). *Davidson's Principles and Practice of Medicine*, 22nd edition. Churchill Livinstone, Elsevier Health Sciences London, UK.

Ware Jr J.E. & Sherbourne C.D. (1992). A 36-item short-form health survey (SF-36): Results from the Medical Outcomes Study. *Med Care* 30: 473– 83.
DOI:https://doi.org/ 10.1080/17453674.2016.1181816.

Wong W.S. & Fielding R. (2010). Construct validity of the Chinese version of the Chalder Fatigue Scale in a Chinese community sample. *Journal of Psychosomatic Research* **68**(1): 89– 93.
DOI:https://doi.org/ 10.1016/j.jpsychores.2009.05.008.

World Health Organization. (2017). *Process of Translation and Adaptation of Instruments*. World Health Organization, Geneva, Switzerland.

J.Natn.Sci.Foundation Sri Lanka 2021 49 (3): 451 - 465
DOI: http://dx.doi.org/10.4038/jnsfsr.v49i3.10074

RESEARCH ARTICLE

Identification of potential TALEN and CRISPR/Cas9 targets of selected genes of some human pathogens which cause persistent infections

WGBP Dharmasena[1] and DHH Munasinghe[1,2*]
[1]*Department of Botany, Faculty of Applied Sciences, University of Sri Jayewardenepura, Gangodawila, Nugegoda.*
[2]*Centre for Plant Materials and Herbal Products Research, University of Sri Jayewardenepura, Gangodawila, Nugegoda.*

Submitted: Submitted: 09 July 2020; Revised: 15 April 2021; Accepted: 25 June 2021

Abstract: The human pathogens, Epstein Barr virus, human papilloma virus, herpes simplex virus-2, hepatitis B virus and *Leishmania* species can cause persistent infections, which cannot be cured with currently available treatments. The modern gene editing techniques, transcription activator like effector nuclease (TALEN) and clustered regularly interspaced short palindromic repeat / CRISPR associated protein 9 (CRISPR/Cas9), are potential candidates for their treatment. In this study, target sites for TALEN and CRISPR/Cas9 were identified *in silico* on selected essential and indispensable genes of the above pathogens, targeting the cease of the essential functions and curing the infection. The gene sequences of the pathogens were obtained from public databases and conserved sequences were identified. Then potential TALEN target sites were identified. For some selected targets, the off-target effects on the genomes of human, mouse, same pathogen and other organisms were tested and the putative functions of the mutated proteins were predicted. TALEN targets without having a potential off-target effect and not leading to mutated proteins with undesirable functions were selected for each gene. The potential CRISPR/Cas9 targets without off-target effect on human and murine genomes were identified and other off-target effects were evaluated. Results showed that potential TALEN and/or CRISPR/Cas9 targets with higher binding specificity and efficiency were available for the selected genes. It can be concluded that the selected targets can potentially be used to produce respective proteins, and *in vitro* and *in vivo* applications are potentially possible.

Keywords: Bioinformatics, CRISPR/Cas9, genome editing, persistent pathogens, TALEN

INTRODUCTION

Genome editing is a technique that is used to alter the genomes of organisms. There are mainly three types of genome editing techniques known as zinc finger nuclease (ZFN), transcription activator like effector nuclease (TALEN) and clustered regularly interspaced short palindromic repeat / CRISPR associated protein 9 (CRISPR/Cas9). All these are site - specific nucleases (Ochiai Yamamoto, 2015). Among them ZFN and TALEN are protein based chimeric nucleases while CRISPR/Cas is RNA based (Nemudryi *et al*., 2014). TALEN and CRISPR/Cas9 are of greater interest due to high specificity of binding (Yeadon, 2014). These techniques have recently been used in the gene editing of organisms such as zebra fish (Gonzales & Yeh, 2014), *Arabidopsis thaliana* (Feng *et al*., 2013), *Zea maize* (Char *et al*., 2015; Kelliher *et al*., 2017), *Oryza sativa* (Shen *et al*., 2017; Han *et al*., 2019; Usman *et al*., 2021) and of human cell lines (Yuen *et al*., 2015).

TALEN and CRISPR/Cas9 techniques have a high potential for application as alternative treatment strategies. The CRISPR/Cas9 method has been tested to cure some diseases such as β thalassemia (Xu *et al*., 2015; Frangoul *et al*., 2021), cystic fibrosis (Schwank *et al*., 2013; Fan *et al*., 2018), hemophilia A (Park *et al*., 2015), hepatitis B (Seeger Sohn, 2014) and human

* Corresponding author (helani@sci.sjp.ac.lk https://orcid.org/0000-0001-7229-4691)

immunodeficiency virus (HIV) infections (Liu *et al*., 2017). TALEN technology has been tested to cure Duchenne muscular dystrophy (Li *et al*., 2015), HIV (Hutter *et al*., 2015; Benjamin *et al*., 2016) and human papilloma viral infections (HPV) (Shankar *et al*., 2017).

Since these techniques are based on site - specific nucleases, specific sites to be bound to the nucleases can be previously determined. The essential or indispensable genes for infection, pathogenesis or persistence of the pathogens can be targeted and these genes could be mutated to cure the infection or to diminish the pathogens' activities. The techniques are highly effective to be used in persistent incurable infections or that have developed resistance against currently available drugs. Some examples of causative agents of these persistent infections are herpes simplex virus (HSV), hepatitis B virus (HBV), HPV, HIV, Epstein Barr virus (EBV), cytomegalovirus (CMV), some bacteria, fungi and protists (Boldogh *et al*., 1996).

The site where the TALEN or CRISPR/Cas9 binds in a specific gene can be predicted, according to that the TALEN or CRISPR/Cas9 proteins can be designed and synthesised. The main criterion in designing a TALEN is the positioning of a Thymine (T) nucleotide at the 5′ end of the target. In CRISPR/Cas9, the presence of Protospacer Adjacent Motif (PAM) is the essential criterion (Nemudryi *et al*., 2014).

The objective of this study was to identify TALEN and CRISPR/Cas9 potential targets and to predict their off-target effects to select the best target sites for some selected genes of EBV, HBV, HPV, HSV-2 and *Leishmania donovani* and *Leishmania infantum* pathogens, which are showing persistent infection and for which treatment is not currently available. The genes of these pathogens were selected with the aim to cease the pathogen persistency and replication. In EBV, the gene *LMP2A* is essential for the persistence of the virus (Longnecker, 2000), and *EBNA1* gene is essential for the replication and transcription (Sivachandran *et al*., 2012). It can be postulated that the introduction of a site-specific nuclease for *LMP2A* gene followed by that of *EBNA1* would diminish the viral content in the host leading to eradication of the virus with continual application. In the same way, the *UL21* and *UL30* of HSV-2 could be mutated which are essential in viral propagation (Le Sage *et al*., 2013) and replication (Liu *et al.,* 2006), respectively. The replication of the HPV can be inhibited by mutating the *E2* gene (Sanders & Stenlund, 2000; McBride, 2013) and also the oncogenic effect (Leykauf *et al*., 2008) could be diminished. *HBx* gene of HBV is indispensable for the development of the viremia and persistency (Tsuge *et al*., 2010) and its mutations might cease the viral infection in the host. *Leishmania* sp., a pathogen that causes persistent infection could be controlled by mutating the *tryR* gene that protects against oxidative stress (Paul *et al*., 2014).

METHODOLOGY

Selection of pathogen specific genes

The genes *LMP2A* and *EBNA1* of EBV, *E2* of HPV type 16, *UL21* and *UL30* of HSV-2, *HBx* of HBV and *tryR* of *L. donovani and L. infantum* were selected and subjected for TALEN and/or CRISPR/Cas9 target identification. Randomly selected entries were obtained from the databases 'GenBank' (Clark *et al*., 2016) and 'RefSeq' (O'Leary *et al*., 2015). These sequences were tested using the tool 'NCBI conserved domains' (Marchler-Bauer *et al.,* 2014) and from the results, the sequences that confirm the presence of the gene were selected. Then the open reading frame (ORF) responsible for the gene in each sequence was identified using the tool 'ORF Finder' (Rombel *et al*., 2002). The maximum length ORF of each sequence was obtained for the analysis.

Identification of conserved residues of the gene

Conserved residues of each selected gene were identified using the software 'Unipro UGENE' (Okonechnikov *et al*., 2012) by aligning the maximum length ORFs of selected sequences. Then some conserved residues above 60 nucleotides were selected for identification of potential TALEN and CRISPR/Cas9 targets.

Identification of TALEN target sites

The tool 'TALEN Targeter' was used to identify the TALEN target sites. The selected conserved sequences of a selected gene were used as the input data. The predesigned TALEN architecture by Miller *et al*. (2011) was used to design TALEN targets and 'NH' was selected as the G substitute repeat variable diresidue (RVD). Then the parameters were adjusted to hide redundant TALENs in output. Other than these, guidelines by Streubel *et al*. (2012) were applied in the analysis. From the output, several TALEN targets having highest percentage of 'HD or NH' RVDs in the respective TALENs and having at least one unique restriction site at the spacer region were selected for further analyses. This procedure was followed for all selected conserved sequences in selected genes of *LMP2A* and *EBNA1* of EBV, *E2* of HPV type 16, *HBx* of HBV and *tryR* of *L. donovani* and *L. infantum*.

Identification of the potential off-target effect of the respective TALENs of the selected TALEN targets

The potential off-target effect or target specificity of the selected TALENs were identified for human genome and murine genome using two bioinformatic tools, TAL Effector Nucleotide Target 2.0 (Doyle *et al*., 2012) and PROGNOS (Fine *et al*., 2013). The RVD sequences of the TALEN targets selected above were used as the input data for both tools.

Apart from these the probable unnecessary bindings on the genome of the selected pathogen was also determined using the tool Paired Target Finder. There, a genome sequence of the pathogen in 'RefSeq' database was used as the target sequence and the RVD sequences of the selected TALENs as the query sequence.

Identification of potential off-target effect of TALENs on other organisms

Basic local alignment search tool (BLAST) (Altschup *et al*., 1990) was used to identify the off-target effect of 'TALENs respective to the selected TALEN targets' in genomes of other organisms. Three methods were followed in the procedure. In the first method, TALEN target sequence was entered as the query sequence, where the spacer region was in lowercase letters, to the nucleotide BLAST tool. Filters were selected to mask lowercase letters and the search was carried out keeping all the other parameters default. In the second method instead of entering the whole TALEN target sequence as the query, only the TALE regions were entered as a continuous sequence by removing the spacer region. Next the same procedure as mentioned above was carried out except masking for lowercase letters. In the third method, all the nucleotides of the spacer region were replaced by the letter 'N' and used as the query sequence. The rest of this procedure is same as of the second method.

The suspected sequences for having off-target effect from the BLAST results were tested again using the tool 'Paired Target Finder'entering the NCBI accessions of the suspected sequences as the target and RVD sequence of the respective TALEN target as the query.

Identification of putative functions of the mutated protein

A sequence of a selected gene (Supplementary Table 1) was obtained and first, the nucleotides of the sequence were numbered from 5 to 3 end using the Group DNA option of the tool Sequence Manipulation suite (Stothard, 2000).

Then the probable cut site by the first TALEN target was marked in the sequence and one nucleotide adjacent to the cut site was deleted. The resulting sequence was filtered to remove unnecessary numbering and spaces using the option Filter DNA of the tool Sequence Manipulation Suite. Then the ORFs in the sequence were identified using the tool 'ORF Finder'. After that, the ORFs responsible for amino acid sequences greater than 75 amino acids and which passed through the cut site were obtained. Each of these sequences was BLAST searched in the Protein BLASTtool (Altschup *et al*., 1990). The same procedure was followed by deleting two nucleotides adjacent to the cut site. Then for all the other potential TALEN targets of the same gene and other selected genes, the same procedures were followed.

Identification of potential CRISPR/Cas9 target sites and their potential off-target effects

Potential CRISPR/Cas9 target sites were identified and off-target effect was predicted using the tool CCTop (Stemmer *et al*., 2015). First, a conserved residue of a selected gene of the selected pathogen was entered into the tool as a plain text as the input. Then the maximum mismatches that an off-target should possess were set as four and the human genome (*Homo sapiens* GRCH38/hg38) was selected to identify off-targets. Other categories were kept default and submitted for analysis. In the same way, off-targets in the murine genome were identified by selecting the mouse genome (*Mus musculus* GRCm38/mm10). Then the above procedures were carried out for all the selected conserved residues of the genes *LMP2A* and *EBNA1* of EBV, *E2* of HPV type 16, *UL21* and *UL30* of HSV-2, *HBx* of HBV and *tryR* of *L. donovani* and *L. infantum*.

The tool CCTop displays the CRISPR/Cas9 target sequences of the query sequence in the order of off-targets in the selected genome from targets with null off-target effect to the targets with the highest off-target effect. Among them, the targets with the null off-target effect on both human and the murine genome were selected and were overlaid to identify the targets common to both the human and murine genomes with null off-target effect. The binding efficacies of the designed CRISPR/Cas9 nucleases were determined using the tool CRISPRator (Labuhn *et al*., 2018).

Identification of the potential off-target effect of selected CRISPR/Cas9 targets on genomes of other organisms

The potential CRISPR/Cas9 target sites identified above were used as the query and BLASTN searches were carried out. From the results, the targets showing the off-target effect on other genomes were identified.

RESULTS AND DISCUSSION

Obtaining the ORF of confirmed gene sequence

The presence of the selected genes, *LMP2A* and *EBNA1* of EBV, *E2* of HPV type 16, *UL21* and *UL30* of HSV-2, *HBx* of HBV and *tryR* of *Leishmania* species in obtained sequences were confirmed from the results of the tool NCBI Conserved Domains. The ORF responsible for the gene in each sequence was identified from the results of the tool ORF Finder (Supplementary Table 2). The confirmation of a gene sequence is important because in some instances annotation errors are present in the sequences available in the NCBI database. The identification of ORF of the selected gene sequence is also of immense importance because a sequence obtained from the databases may contain areas that do not belong to the ORF of the gene. If such regions are present in the sequence, TALEN and/or CRISPR/Cas9 target sites may be identified for those regions too.

Selection of conserved residues to identify TALEN and CRISPR/Cas9 target sites

According to the selected TALEN architecture, the maximum length of a TALEN target site is 60 nucleotides. Therefore, from the gene sequence alignment results (Supplementary Table 3), the conserved sequences with minimum length of 60 nucleotides were selected for each gene (Supplementary Table 4). The same conserved sequences were used for the identification of CRISPR/Cas9 target sites. HPV types 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68 and 70 were initially selected because these types are the high-risk types for cancer (Burd, 2003). But any conserved domain greater than 60 nucleotides were not observed in them. Then HPV type 16 was considered because it is the type that is responsible for the highest percentage of cancers among the HPV types (National Cancer Institute of USA, 2017). For *Leshmania* sp. also any conserved sequence in enough length for a TALEN target was not identified. Therefore, the *Leishmanaia donovani* complex which includes the species *L. donovani*, *L. infantum* and *L. chagasi* was considered because it is the cause for visceral leishmaniasis (Sundar Rai, 2002). Visceral leishmaniasis is the most severe form of leishmaniasis among others (Das *et al*., 2016). But *tryR* gene sequence of *L. chagasi* was not available in the databases, GenBank, RefSeq, EMBL-EBI or DDJB and therefore the sequences of other two species were used. For HBV, a fully conserved residue longer than 60 nucleotides common to all genotypes was not identified but a partially conserved sequence was selected (Supplementary Table 4).

Designing of potential TALEN target sites

Potential TALEN target sites were identified in each of the above selected conserved residues of the *LMP2A* and *EBNA1* gene of EBV, *E2* gene of HPV type 16, and *tryR* gene of *L. donovani* and *L. infantum*. The tool TALEN Targeter' provides both TALEN targets and the RVD sequence for the targets, as the output. Apart from this, it shows the unique restriction sites at the spacer region. Supplementary Table 5 contains the TALEN target results obtained for each sequence. For the *HBx* gene of HBV, the targets were identified for each genotype separately using the whole *HBx* gene as the query, and the targets were identified for a selected, partially conserved sequence, considering all genotypes. Then for each gene, several TALEN targets with higher percentage of HD or NH in their RVDs and having at least one unique restriction site at the spacer region were selected. TALENs having high percentage of HD or NH were selected because the binding specificity and the efficiency are higher when the percentage of HD or NH is high (Streubel *et al*., 2012). The selection of target sites with unique restriction sites is beneficial in experimental identification of TALEN activity (Doyle *et al*., 2012).

Potential off-target effect of selected TALENs on human genome, murine genome and unnecessary areas of the pathogen genome

The off-target effect of the TALENs to their targets on human genome, murine genome and unnecessary loci of same pathogen genome were identified for *LMP2A* and *EBNA1* gene of EBV, *E2* gene of HPV type 16 and *tryR* gene of *L. donovani* and *L. infantum,* using the tools PROGNOS and Paired Target Finder. The off-target effect of the TALENs separately selected for the *HBx* gene of each genotype and the TALENs common to *HBx* gene of all the HBV genotypes were also identified.

The off-target effect on the human genome is essential to be identified because if any off-target is present for a 'TALEN respective to the identified target site', it may cause mutations in the human genome. The off-target effect on the mouse genome was also identified as preliminary toxicity tests were mostly carried out *in vivo* using mouse as the model organism, and as such, unnecessary mutations in the mouse genome was avoided. Two tools, Paired Target Finder and PROGNOS, were used in order to minimise the errors in identification of off-target effect. The off-target effects on human and murine genomes were identified with respect to the genomes already available in the tools. These genomes are consensus sequences and therefore, it cannot be concluded that the output given by the tools are valid for every human and mouse, but they would be valid for most. The use of the two tools minimises this effect because the genome entries are different in the two tools.

Other than these, unnecessary bindings on the same pathogen genome were identified in order to minimise the effect of mutations of other genes of the pathogen. If any undesired mutation occurs it might not be suitable for the host or might have a chance to elevate the pathogenic effect. Thus, the tool Paired Target Finder was used because other options were not available in the tool PROGNOS to check the off-targets in NCBI sequences other than the genomes already entered into the tool. Here a selected representative genome was used for each pathogen for convenience and it cannot be concluded about the off-target effect in every isolate and strain of the pathogen.

The score given by the tool, Paired Target Finder is the key by which the tool differentiates the off-targets from the on-targets. The score is given to the TALEN target based on the types of the RVDs present and the matching percentage of RVDs. The perfectly matching off-target gives the same score as the on-target of query TALEN sequence and the score increases when the mismatches increase. Doyle *et al*. (2012) suggested that the maximum score that an off-target would have is four times the score of the on-target. In the tool, only the off-targets below the maximum value are displayed. In the tool PROGNOS, the score has been adjusted to reduce when the off-target deviate from the on-target. Therefore, the perfect off-target is having the score same as that of the on-target. The results of the tool display the off-targets up to a selected number of mismatches, and the off-targets with higher potential of binding with the TALEN are mentioned. The maximum number of mismatches that off-target should contain is selected as five for all the TALEN target sites of the selected genes, and therefore, the off-target effect of those TALENs could be compared.

Identification of off-target effect of the TALENs on genomes of other organisms

The identification of the off-target effect of the TALENs on the genomes of other organisms is necessary. If the designed TALEN proteins are released to the environment, there is a chance of mutating the genomes of other organisms in the environment. The selection of target sites that are lacking off-targets in genomes of other organisms prevents this undesirable effect. Furthermore, human and mouse are inhabited with numerous species of commensals and mutation induction on them can also be predictively prevented with this step. Three methods of BLAST search gave desired results with comparative merits and demerits, and the results were in different formats (Figure 1). In this way the off-target effect of the selected TALENs of *LMP2A* and *EBNA1* of EBV, *E2* gene of HPV type 16, *tryR* gene of *L. donovanii* and *L. infantum*, *HBx* gene of all the genotypes of HBV and TALENs common for *HBx* gene of all the genotypes of HBV were identified. The 'nucleotide' search page of the tool BLAST was used because the tools Paired Target Finder and PROGNOS identify only the off-targets in selected genomes. But in BLAST tool it was a challenge to identify the targets/off-targets because the query (the TALEN target site) contained the spacer region which does not involve in the specific binding with a TALEN, and therefore three BLAST search methods were used. The method one of BLAST search displayed the highest number of probable targets/off-targets when compared with other two methods. But the identity and the query coverage the output have been calculated, including the spacer region, although it was masked in BLAST search. This interferes with the differentiation of off-targets from on-targets. In the result type 1 of method two, only the targets in the range one and range two that were lying in a distance not more than 30 nucleotides were selected because the maximum spacer length that a TALEN can be bound is thirty (30) nucleotides (Doyle *et al*., 2012). In the result type 2 of method two, the results are much effective because the spacer region has not been considered in calculating the percent identity and query coverage. In method three, the results of the on-targets/ off-targets are mostly similar to that of the 'result type 2 of method two. In these two results (result type 2 of method two and method three), the identification of the on-targets/off-targets are comparatively easier than other two, referring to the graphical alignment. The suspected sequences for having off-target effects were further tested with the tool Paired Target Finder, because BLAST tool is not specific for the purpose and a TALEN binding score is unavailable in the BLAST tool. Only the sequences

similar to the query sequence (TALEN target sequence) can be identified with BLAST search, and that does not reflect that the TALEN of that query sequence binds with the off-target sequence.

Putative functions of the mutated protein by TALEN

The putative functions of the mutated protein due to the double stranded break by each TALEN of the selected genes were identified. The TALENs that were shown to produce proteins with unnecessary function were avoided. The identification of the putative function of the mutated protein is important because the mutated proteins might have undesirable functions. The proteins that formed due to frame shift of one base pair and frame shift of two base pairs were considered, but frame shift of three base pairs was not considered because it will not change the reading frame and only lead to alterations of a few amino acids resulting slight deviations from the initial function of the protein. The ORFs of the mutated protein that were passing through the cut site of the TALEN were selected because other ORFs are not mutated. Other than that, only the ORFs leading to proteins with amino acid number greater than 75 were considered, because smaller proteins less than that might not possess specific functions. But there are small proteins with key functions in the cell (Reichman-Fried Raz, 2014), and at some point, mutations cause changes in the protein function and is a limitation of the present study. The putative functions of the proteins of the selected ORFs were identified with the protein BLAST tool, where the functions of similar proteins to the query were considered as the function of the query protein. Most of the results obtained for the ORFs were lacking BLAST queries and hence the putative function could not be determined. This might be a potential error because these queries also may have some undesirable functions. This can be identified up to some extent by *de novo* protein function analysis. The partial proteins of the original were not further considered because they might not have novel functions. But some partial proteins could have undesirable functions in the cell.

Selected TALEN target sites

The target sites were selected for each gene that did not show an off-target effect in the human genome and murine genome in both results of the tools Paired Target Finder and PROGNOS, which did not show an unnecessary binding effect in the same pathogen genome and did not produce mutated proteins with undesirable function. Table 1 shows the target sites selected for the *LMP2A* and *EBNA1* genes of EBV, E2 gene of HPV type 16, *tryR* gene of L. *donovani* and L. *infantum* and HBx gene of HBV.

Identification of potential CRISPR/Cas 9 targets

The results obtained for the *LMP2A* gene of EBV is shown in Figure 2. Among them T1, T2, T3, T4, T5, T6 and T7 were observed to be lacking off targets in the human genome (Supplementary Table 6). The same target sites were obtained and were numbered in the order of off-target effect on the murine genome as T65, T5, T21, T9, T30, T46 and T60 respectively. Among them T5 and T9 were not shown off target effect on the murine genome.

The selected targets were searched for the presence of off-target effect on the genomes of other organisms using 'nucleotide BLAST' tool. From the results it was identified that the target CRLMP2A01 possesses off-target effect on *Schistosoma rodhaini* genome. Therefore, target CRLMP2A02 was selected as the potential CRISPR/Cas9 target on *LMP2A* gene. Any genome with off-target effect with respect to the target CRLMP2A02 was not identified. In the same way CRISPR/Cas9 targets were selected and their potential of binding on other genomes was observed. But targets without off-target effect on human and murine genome were not observed for E2 gene of HPV type 16 and *HBx* gene of HBV. Table 2 shows the CRISPR/Cas9 targets selected for *EBNA1* gene of EBV, *UL21* and *UL30* gene of HSV-2 and *tryR* gene of *L. donovani* and *L. infantum* respectively. The efficacy of guideRNA of each selected target was analysed by the tool CRISPRater (Labuhn *et al.*, 2018). Score 0.56 or below shows low efficacy, score within 0.56 and 0.74 shows medium efficacy and high

These targets are same as T2 and T4 arrangement s of the off-targets in the human genome.

T2 target of human genome - 5'ACTTGGGATTGCAACACGACGGG3'
T5 target of murine genome - 5'ACTTGGGATTGCAACACGACGGG3'
} TARGET CRLMP2A01

T4 target of human genome - 5'TCACGTTCCTCATCGTTCGGTGG3'
T9 target of murine genome - 5'TCACGTTCCTCATCGTTCGGTGG3'
} TARGET CRLMP2A02

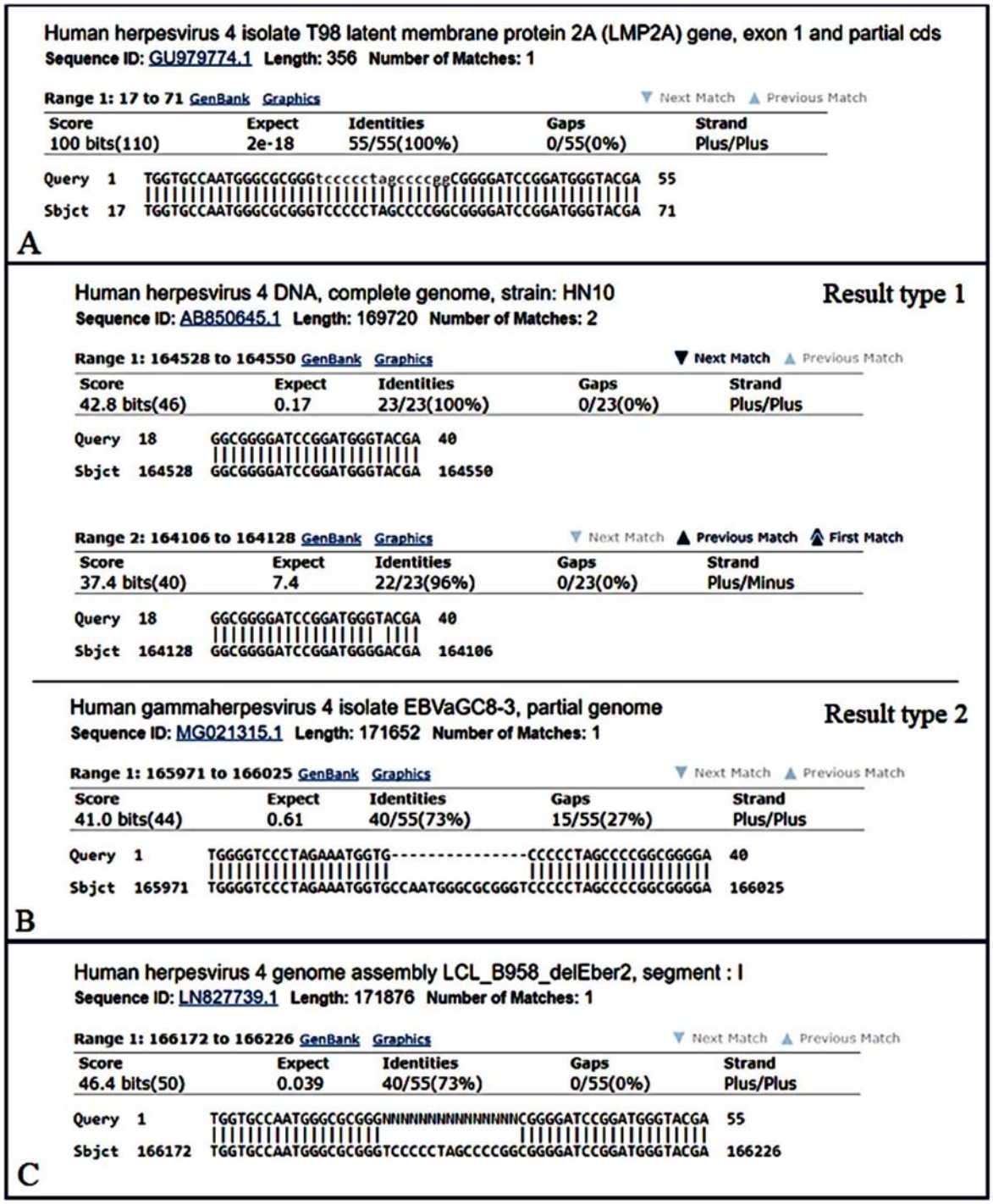


Figure 1: An example of the results obtained for the BLAST search for identification of off-targets in other genomes. A: 'BLAST' method one search result type of target site EBVLMP2A04. B: 'BLAST' method two search result type of target site EBVLMP2A05. C: 'BLAST' method three search result type of EBVLMP2A04 target site.

Table 1: Selected TALEN target sites for *LMP2A* and *EBNA1* genes of EBV, (No. 1) *E2* gene of HPV type 16, (No. 2) *tryR* gene *of Leishmania donovani* and *L. infantum* and *HBx* gene of HBV. 'Plus Strand Sequence' is the target sequence as denoted in the plus strand (5'–3'). (No. 3) denotes the TALEN targets separately in each genotype of HBV and (No. 5) denote the targets common for all the genotypes. Supplementary Figure 2 shows the ami no acid sequence of TAL01 of target EBVLMP2A01

No.	Pathogen	Gene	Sequence Name	Plus strand sequence
1	EBV	*LMP2A*	EBVLMP2A01	T GGGGTCCCTAGAAATGGTGC caatgggcgcgggtcc CCCTAGCCCCGGCGGGG A
			EBVLMP2A03	T GGGGTCCCTAGAAATGGTG ccaatgggcgcgggt CCCCCTAGCCCCGGCGGGG A
			EBVLMP2A05	T GGGGTCCCTAGAAATGGTGC caatgggcgcgggtc CCCCTAGCCCCGGCGGGG A
			EBVLMP2A06	T GCCAATGGGCGCGGGTCCCC ctagccccggcggggat CCGGATGGGTACGATGGCGG A
		EBNA1	EBVEBNA01	T GGAGGGGGCAGGAGTCTGCA ctccctgtattcactgagcgtcgg GGGCTGTTGGAGGGGGCAGG A
			EBVEBNA09	T GCACTCCCTGTATTCACTGA gcgtcgggggctgtt GGAGGGGGCAGGAGTCTGC A
2	HPV type 16	*E2*	HPVE201	T TACAAGGCCAGAGAAATGGG atttaaacatattaaccaccaggtg GTGCCAACACTGGCTGTATC A
			HPVE202	T ACAAGGCCAGAGAAATGGGA tttaaacatattaaccaccag GTGGTGCCAACACTGGCTGT A
			HPVE203	T AACTGCACCAACAGGATGTA taaaaaaacatggatata CAGTGGAAGTGCAGTTTG A
			HPVE204	T AACTGCACCAACAGGATGTA taaaaaaacatggatataca GTGGAAGTGCAGTTTGATGG A
			HPVE205	T TCATGCGGGTGGTCAG gtaatattatgtcct ACATCTGTGTTTAGC A
			HPVE207	T TAACAGCTCACACAAAGGAC ggattaactgtaata GTAACACTACACCCATAGT A
			HPVE208	T AACAGCTCACACAAAGGACG gattaactgtaatag TAACACTACACCCATAGTAC A
			HPVE209	T AACAGCTCACACAAAGGACG gattaactgtaatagt AACACTACACCCATAGTAC A
			HPVE210	T TGGACAGGACATAATGTAAA acataaaagtgcaat TGTTACACTTACATATGAT A
3	*Leshmania* spp.	*tryR*	LDTRYR02	T CCCGCGCGTACGACCTCGTG gtgcttggcgccggat CTGGAGGTCTGGAGGCGGG A
			LDTRYR03	T GCTTGGCGCCGGATCTGGAG gtctggaggcgggatg GAACGCGGCCGTCACGCACA A
			LDTRYR04	T GTCCCGCGCGTACGACCTC gtggtgcttggcgcc GGATCTGGAGGTCTGG A
			LDTRYR05	T CGGCGGCACGTGCGTGAACG tcggctgcgtgccaaa GAAACTCATGGTGAC A
			LDTRYR06	T CGGCGGCACGTGCGT gaacgtcggctgcgt GCCAAAGAAACTCATGGTG A
			LDTRYR07	T CCGTGAGTCTGGCGG cttcggatgggagatgg ACCGCGAATCGCTCTGCCCC A
			LDTRYR08	T TCGGATGGGAGATGGACCGC gaatcgctctgccccaact GGAAGACGCTCATCGCCGCG A
			LDTRYR09	T GGACCGCGAATCGCTCTGCC ccaactggaagacgctca TCGCCGCGAAGAACA A
4	HBV genotype A	*HBx*	HBVXGA01	T GGGACTCTCTCGTCCCCTTC tccgtctgccgttcca GCCGACCACGGGGCGC A
			HBVXGA02	T TGTCTACGTCCCGTCGGCGC tgaatcccgcggacgac CCCTCTCGGGGCCGCTTGGG A
	HBV genotype B	*HBx*	HBVXGB01	T GGGGCTCTACCGCCCGCTTC tccgcctgttgtacc GTCCGACCACGGGGCGC A
			HBVXGB02	T TGGGGCTCTACCGCCCGC ttctccgcctgttgt ACCGTCCGACCACGGGGCGC A
	HBV genotype C	*HBx*	HBVXGC01	T CTACGTCCCGTCGGCGCTGA atccagcggacgacc CGTCTCGGGGCCGCTTGGGG A
			HBVXGC02	T CTACGTCCCGTCGGCGCTGA atccagcggacgaccc GTCTCGGGGCCGCTTGGGG A
	HBV genotype D	*HBx*	HBVXGD02	T GCCAACTGGGTCCTGCGCGG gacgtcctttgttta CGTCCCGTCGGCGCTGA A
	HBV genotype E	*HBx*	HBVXGE01	T CCGGCCGACCACGGGGCGCA cctctctttacgcggtct CCCCGTCTGTGCCTTCTC A
	HBV genotype F	*HBx*	HBVXGF02	T GCCGGACCGTGTGCACTTCG cttcacctctgcacgt CGCATGGAGACCACCGTGA A
	HBV genotype G	*HBx*	HBVXGG01	T GGGGCTCTGTCGCCCCCTTC tccgtctgccgttcct GCCGACCACGGGGCGC A
	HBV genotype H	*HBx*	HBVXGH01	T CGCTTGGGGCTATGCCGCCC tcttctccgcctgccgtt CCGGCCGACGACGGGTCGC A
			HBVXGH02	T CTCGTGGTCGCTTGGGGCT atgccgccctcttct CCGCCTGCCGTTCCGGCCG A
5	HBV	*HBx*	HBVXCS01	T AGGAGGCTGTAGGCATAAA ttggtctgttcacca GCACCATGCAACTTTTTC A
			HBVXCS02	T AGGAGGCTGTAGGCATAA attggtctgttcacc AGCACCATGCAACTTTTTC A
			HBVXCS03	T AAATTGGTCTGTTCACCAGC accatgcaacttttt CACCTCTGCCTAATC A
			HBVXCS04	T AGGAGGCTGTAGGCATA aattggtctgttcac CAGCACCATGCAACTTTTTC A

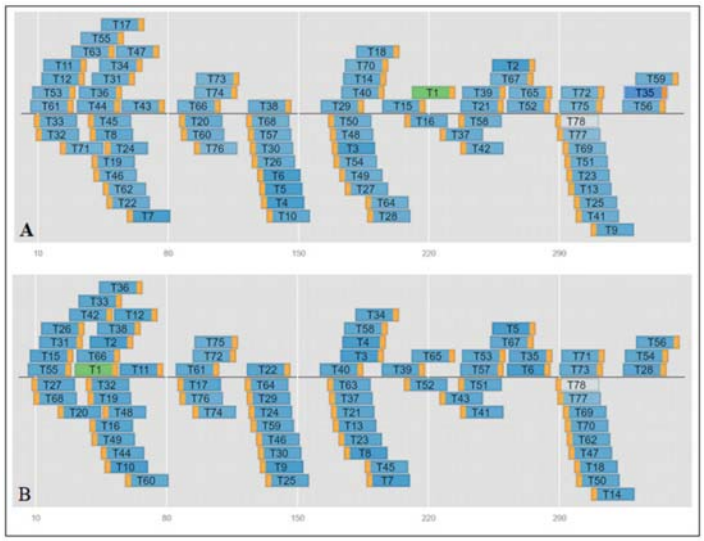


Figure 2: Graphical representation of the CRISPR/Cas9 target sites on *LMP2A* gene. A: target sites that arranged according to the increasing order of off-target effect on the human genome. B: target sites that arranged according to the increasing order of off-target effect on the murine genome.

efficacy is shown by scores equal or above 0.74. This scoring system is easily applicable for selecting efficient targets due to its ability to select efficient guideRNA. Supplementary Figure 1 represents the structure and action of CRISPR nuclease.

The protospacer adjacent motif (PAM) selected for the CRISPR/Cas9 targets in this study was NGG of *Streptococcus pyogenes*. But other PAM motifs can be substituted instead of NGG, which might change the off-target effect. The off-targets of the CRISPR/Cas9 nucleases might differ by two base pairs from an on-target (Cho *et al*., 2014). Therefore, in our study we considered the targets below three pair mismatches as off-targets. The targets above four base pair mismatches mostly prevented the double-strand break and therefore, they were omitted. The cleavage efficiency of CRISPR system greatly varies on different target sites or the cell type/organism. The efficacy of binding and cleavage of CRISPR system depends on several features; features of the guide RNA, genetic features including epigenetics and energetic properties that have been identified through various studies as factors involved in determining the efficacy of the guideRNA (Cong *et al*., 2013; Fu *et al*., 2013; Wang *et al*., 2014; Chari *et al*., 2015; Liu *et al*., 2020).The putative function of the mutated protein by CRISPR/Cas9 was not identified because the cut site cannot be exactly determined.

In this research, only the pathogens with double stranded DNA genomes were considered because the TALEN and CRISPR/Cas9 nucleases cannot function on single stranded genomes or in RNA genomes. But recently Abudayyeh *et al*. (2017) have identified the CRISPR/Cas13 system, which can be applied on RNA genomes.

Delivery of the CRISPR/cas9 nucleases or the TALENs to the required specific cell type is a question not solved yet. Specific strategies should be used for the successful transportation of mRNAs of the nucleases to the cytoplasm and the resulting nucleases to the nucleus of the cell. Viral vectors, microinjection, electroporation and chemical methods are a few currently used methods (Glass *et al*., 2018). The selection of a specific cell type by the nuclease is also important and Cheng *et al*. (2020) have described a tissue specific nanoparticle based method to deliver CRISPR mRNA.

The development of site specific nucleases is a concerns in terms of an ethical perspective (Rodriguez, 2016). The main ethical concern is the balance between risks and benefits. The loss of ecological equilibrium could occur. Apart from that the regulation of the product to the consumers is also an essential criterion to be evaluated. These ethical questions should be addressed in the development of TALEN and CRISPR/Cas nucleases as treatment strategies.

CONCLUSION

The identified potential TALEN and CRISPR/Cas9 targets may be applicable for specific mutagenic agents of EBV, HPV type 16, HSV-2, HBV and *L. donovani* and *L. infantum* and can be further developed as a treatment strategy. Furthermore, fully conserved residues of enough length for a TALEN target site are absent in the *HBx* gene considering all the genotypes of HBV, and in *E2* gene, considering all high-risk types of HPV.

Table 2: The CRISPR/Cas9 targets selected for the gene of *EBNA1* gene of EBV, *UL21* and *UL30* genes of HSV-2 and *tryR* gene of *Leishmania donovani* and *L. infantum*. (* - targets with high efficacy of mutagenesis)

Pathogen	Gene	Target name	Selected Target	Efficacy score of sgRNA
EBV	*EBNA1*	CREBNA101	5'GTCGCCGGTGTGTTCGTATATGG3'	0.66
HSV-2	*UL21*	CRUL2101	5'TTATTCCGTAGGGCGGCCTCGGG3'	0.76*
		CRUL2102	5'GAGCTCGCTTCGTACGTAGTTGG3'	0.72
		CRUL2103	5'GTACGTAGTTGGCGACCATGCGG3'	0.68
		CRUL2107	5'CGCGCCCGACGAACCGACGTTGG3'	0.71
		CRUL2108	5'TCGGCCAACGTCGGTTCGTCGGG3'	0.56
		CRUL2109	5'CAACGTCGGTTCGTCGGGCGCGG3'	0.81*
		CRUL2111	5'CGTCGGGCGCGGGAACGTACAGG3'	0.81*
		CRUL2113	5'ACTCGCAACGCCTGACCCCGGGG3'	0.71
		CRUL2114	5'CCCGAGCACGTCATGTACCTCGG3'	0.69
	UL30	CRUL3001	5'ATTTATCGCCCCGCGTTCGCTGG3'	0.67
		CRUL3002	5'CGCCCCTAAGGTGTACTGCGGGG3'	0.61
		CRUL3003	5'GCCGCGTCGCTTGCGCCTGTGGG3'	0.70
		CRUL3004	5'GCGTCCTCGTCCAGCGAACGCGG3'	0.57
		CRUL3005	5'CCAGCTCCACGAGCGATTTATGG3'	0.76*
		CRUL3006	5'CGCGTCGCCGTTCACGTCTACGG3'	0.75*
		CRUL3007	5'CCGCGTGCCGTAGACGTGAACGG3'	0.73
		CRUL3008	5'AGACGTGAACGGCGACGCGATGG3'	0.67
		CRUL3009	5'CCCCGCGATCAGGAAGTACGAGG3'	0.72
		CRUL3010	5'CCCGCGATCAGGAAGTACGAGGG3'	0.72
		CRUL3011	5'CGTCGAGTTTAACTGCACGGCGG3'	0.85*
		CRUL3012	5'GATCGGCGAGTATTGTGTGCAGG3'	0.83*
		CRUL3013	5'TTCCGCCGTCGCGCGCCTGGCGG3'	0.85*
		CRUL3014	5'GCGCATCATCTACGGGGACACGG3'	0.55
L. donovani and *L. infantum*	*tryR*	CRTRYR01	5'GTCCCGCGCGTACGACCTCGTGG3'	0.67
		CRTRYR03	5'CACCACGAGGTCGTACGCGCGGG3'	0.60

REFERENCES

Abudayyeh O.O. *et al.* (11 authors) (2017). RNA targeting with CRISPR–Cas13. *Nature* **550** (7675): 280–284. DOI: https://doi.org/10.1038/nature24049

Altschup S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**(3): 403–410. DOI: https://doi.org/10.1016/S0022-2836(05)80360-2

Benjamin R., Berges B.K., Solis-Leal A., Igbinedion O., Strong C.L. & Schiller M.R. (2016). TALEN gene editing takes aim on HIV. *Human Genetics* **135**(9): 1059–1070. DOI: https://doi.org/10.1007/s00439-016-1678-2

Boldogh I., Albrecht T. & Porter D.D. (1996). persistent viral infections. In: *Medical Microbiology* (eds. S. Baron). University of Texas Medical Branch at Galveston, Galveston, USA.

Burd E.M. (2003). Human papillomavirus and cervical cancer. *Clinical Microbiology Reviews* **16**(1): 1–17. DOI: https://doi.org/10.1128/CMR.16.1.1-17.2003

Char S.N., Unger-Wallace E., Frame B., Briggs S.A., Main M., Spalding M.H., Vollbrecht E., Wang K. & Yang B. (2015). Heritable site-specific mutagenesis using TALENs in maize. *Plant Biotechnology Journal* **13**(7): 1002–1010. DOI: https://doi.org/10.1111/pbi.12344

Chari R., Mali P., Moosburner M. & Church G.M. (2015). Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nature Methods* **12**(9): 823–826. DOI: https://doi.org/10.1038/nmeth.3473

Cheng Q., Wei T., Farbiak L., Johnson L.T., Dilliard S.A. & Siegwart D.J. (2020). Selective organ targeting (SORT) nanoparticles for tissue-specific mRNA delivery and CRISPR-Cas gene editing. *Nature Nanotechnology* **15**(4): 313–320. DOI: https://doi.org/10.1038/s41565-020-0669-6

Cho S.W., Kim S., Kim Y., Kweon J., Kim H.S., Bae S. & Kim J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Research* **24**(1): 132–141.

DOI: https://doi.org/10.1101/gr.162339.113

Clark K., Karsch-Mizrachi I., Lipman D.J., Ostell J. & Sayers E.W. (2016). GenBank. *Nucleic Acids Research* **44**(D1): D67–D72.
DOI: https://doi.org/10.1093/nar/gkv1276

Cong L. *et al.* (11 authors) (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**(6121): 819–823.
DOI: https://doi.org/10.1126/science.1231143

Das A., Karthick M., Dwivedi S., Banerjee I., Mahapatra T., Srikantiah S. & Chaudhuri I. (2016). Epidemiologic correlates of mortality among symptomatic visceral leishmaniasis cases: findings from situation assessment in high endemic foci in India. *PLoS Neglected Tropical Diseases* **10**(11): 0005150.
DOI: https://doi.org/10.1371/journal.pntd.0005150

Doyle E.L., Booher N.J., Standage D.S., Voytas D.F., Brendel V.P., VanDyk J.K. & Bogdanove A.J. (2012). TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Research* **40**(W1): W117–W122.
DOI: https://doi.org/10.1093/nar/gks608

Fan Z. *et al*. (11 authors) (2018). A sheep model of cystic fibrosis generated by CRISPR/Cas9 disruption of the CFTR gene. *JCI insight* **3**(19).
DOI: https://doi.org/10.1172/jci.insight.123529

Feng Z. *et al*. (11 authors) (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell Research* **23**(10): 1229.
DOI: https://doi.org/10.1038/cr.2013.114

Fine E.J., Cradick T.J., Zhao C.L., Lin Y. & Bao G. (2013). An online bioinformatics tool predicts zinc finger and TALE nuclease off-target cleavage. *Nucleic Acids Research* **42**(6): e42.
DOI: https://doi.org/10.1093/nar/gkt1326

Frangoul H. *et al*. (27 authors) (2021). CRISPR-Cas9 gene editing for sickle cell disease and β-thalassemia. *New England Journal of Medicine* **384**(3): 252–260.
DOI: https://doi.org/10.1056/NEJMoa2031054

Fu Y., Foden J.A., Khayter C., Maeder M.L., Reyon D., Joung J.K. & Sander J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology* **31**(9): 822–826.
DOI: https://doi.org/10.1038/nbt.2623

Glass Z., Lee M., Li Y. & Xu Q. (2018). Engineering the delivery system for CRISPR-based genome editing. *Trends in Biotechnology* **36**(2): 173–185.
DOI: https://doi.org/10.1016/j.tibtech.2017.11.006

Gonzales A.P. & Yeh J.R. (2014). Cas9-based genome editing in zebra fish. *Methods in Enzymology* **546**: 377–413.
DOI: https://doi.org/10.1016/B978-0-12-801185-0.00018-0

Han Y., Teng K., Nawaz G., Feng X., Usman B., Wang X., Luo L., Zhao N., Liu Y. & Li R. (2019). Generation of semi-dwarf rice (*Oryza sativa* L.) lines by CRISPR/Cas9-directed mutagenesis of OsGA20ox2 and proteomic analysis of unveiled changes caused by mutations. 3 *Biotech* **9**(11): 1–17.
DOI: https://doi.org/10.1007/s13205-019-1919-x

Hutter G., Bodor J., Ledger S., Boyd M., Millington M., Tsie M. & Symonds G. (2015). CCR5 targeted cell therapy for HIV and prevention of viral escape. *Viruses* **7**(8): 4186 – 4203.
DOI: https://doi.org/10.3390/v7082816

Hwang W.Y., Fu Y., Reyon D., Maeder M.L., Tsai S.Q., Sander J.D., Peterson R.T., Yeh J.J. & Joung J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology* **31**(3): 227–229.

Kelliher T. *et al.* (11 authors) (2017). MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. *Nature* **542**(7639): 105–109.
DOI: https://doi.org/10.1038/nature20827

Labuhn M., Adams F.F., Ng M., Knoess S., Schambach A., Charpentier E.M., Schwarzer A., Mateo J.L., Klusmann J.H. & Heckl D. (2018). Refined sgRNA efficacy prediction improves large-and small-scale CRISPR-Cas9 applications. *Nucleic Acids Research* **46**(3): 1375–1385.
DOI: https://doi.org/10.1093/nar/gkx1268

Le Sage V., Jung M., Alter J.D., Wills E.G., Johnston S.M., Kawaguchi Y., Baines J.D. & Banfield B.W. (2013). The herpes simplex virus 2 UL21 protein is essential for virus propagation. *Journal of Virology* **87**(10): 5904–5915.
DOI: https://doi.org/10.1128/JVI.03489-12

Leykauf K., Kabsch K., Gassler N., Gissmann L., Alonso A. & Schenkel J. (2008). Expression of the HPV11 E2 gene in transgenic mice does not result in alterations of the phenotypic pattern. *Transgenic Research* **17**(1): 1–8.
DOI: https://doi.org/10.1007/s11248-007-9130-y

Li H.L. *et al.* (11 authors) (2015). Precise correction of the dystrophin gene in duchenne muscular dystrophy patient induced pluripotent stem cells by TALEN and CRISPR-Cas9. *Stem Cell Reports* **4**(1): 143–154.
DOI: https://doi.org/10.1016/j.stemcr.2014.10.013

Liu S. *et al.* (11 authors) (2006). Crystal structure of the herpes simplex virus 1 DNA polymerase. *Journal of Biological Chemistry* **281**(26): 18193–18200.
DOI: https://doi.org/10.1074/jbc.M602414200

Liu Z. *et al.* (11 authors) (2017). Genome editing of the HIV co-receptors CCR5 and CXCR4 by CRISPR-Cas9 protects CD4+ T cells from HIV-1 infection. *Cell and Bioscience* **7**(1): 47.
DOI: https://doi.org/10.1186/s13578-017-0174-2

Liu G., Zhang Y. & Zhang T. (2020). Computational approaches for effective CRISPR guide RNA design and evaluation. *Computational and Structural Biotechnology Journal* **18**: 35–44.
DOI: https://doi.org/10.1016/j.csbj.2019.11.006

Longnecker R. (2000). Epstein-Barr virus latency: LMP2, a regulator or means for Epstein-Barr virus persistence?. *Advances in Cancer Research* **79**: 175 – 200.
DOI: https://doi.org/10.1016/S0065-230X(00)79006-3

Marchler-Bauer A., *et al* (11 authors) (2014). CDD: NCBI's conserved domain database. *Nucleic acids research* **43**(D1): D222 – D226.
DOI: DOI: https://doi.org/10.1093/nar/gku1221

McBride A.A. (2013). The papillomavirus E2 proteins. *Virology* **445**(1): 57–79.
DOI: https://doi.org/10.1016/j.virol.2013.06.006

Miller J.C., *et al* (11 authors) (2011). A TALE nuclease

architecture for efficient genome editing. *Nature Biotechnology* **29**(2): 143–148.
DOI: https://doi.org/10.1038/nbt.1755

National Cancer Institute of USA (2017). HPV and Cancer. Available at *https://www.cancer.gov/about-cancer/causes-prevention/risk/infectious-agents/hpv-fact-sheet*, Accessed 25 November 2017.

Nemudryi A.A., Valetdinova K.R., Medvedev S.P. & Zakian S.M. (2014). TALEN and CRISPR/Cas genome editing systems: tools of discovery. *Acta Natura* **6**(3): 19–40.
DOI: https://doi.org/10.32607/20758251-2014-6-3-19-40

Ochiai H. & Yamamoto T. (2015). Genome editing using zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs). In: *Targeted Genome Editing Using Site-Specific Nucleases* (ed. T. Yamamoto), pp3-24. Springer Japan, Hiroshima, Japan.
DOI: https://doi.org/10.1007/978-4-431-55227-7_1

Okonechnikov K., Golosova O., Fursov M. & UGENE team (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**(8): 1166–1167.
DOI: https://doi.org/10.1093/bioinformatics/bts091

O'Leary N.A., *et al* (11 authors) (2015). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* **44**(D1): D733–D745.
DOI: https://doi.org/10.1093/nar/gkv1189

Park C.Y., Kim D.H., Son J.S., Sung J.J., Lee J., Bae S., Kim J.H., Kim D.W. & Kim J.S. (2015). Functional correction of large factor VIII gene chromosomal inversions in hemophilia A patient-derived iPSCs using CRISPR-Cas9. *Cell stem cell* **17**(2): 213–220.
DOI: https://doi.org/10.1016/j.stem.2015.07.001

Paul M.S., Kaur A., Geete A. & Sobhia M.E. (2014). Essential gene identification and drug target prioritization in Leishmania species. *Molecular BioSystems* **10**(5): 1184–1195.
DOI: https://doi.org/10.1039/C3MB70440H

Reichman-Fried M. & Raz E. (2014). Small proteins, big roles: the signaling protein Apela extends the complexity of developmental pathways in the early zebrafish embryo. *BioEssays* **36**(8): 741–745.
DOI: https://doi.org/10.1002/bies.201400048

Rodriguez E. (2016). Ethical issues in genome editing using Crispr/Cas9 system. *Journal of Clinical Research and Bioethics* 7(266).
DOI: https://doi.org/10.4172/2155-9627.1000266

Rombel I.T., Sykes K.F., Rayner S. & Johnston S.A. (2002). ORF-FINDER: a vector for high-throughput gene identification. *Gene* **282**(1): 33–41.
DOI: https://doi.org/10.1016/S0378-1119(01)00819-8

Sanders C.M. & Stenlund A. (2000). Transcription factor-dependent loading of the E1 initiator reveals modular assembly of the papillomavirus origin melting complex. *Journal of Biological Chemistry* **275**(5): 3522–3534.
DOI: https://doi.org/10.1074/jbc.275.5.3522

Schwank G., *et al* (11 authors) (2013). Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* **13**(6): 653–658.
DOI: https://doi.org/10.1016/j.stem.2013.11.002

Seeger C. & Sohn J.A. (2014). Targeting hepatitis B virus with CRISPR/Cas9. *Molecular Therapy-Nucleic Acids* **3**(12): e216.
DOI: https://doi.org/10.1038/mtna.2014.68

Shankar S., Prasad D., Sanawar R., Das A.V. & Pillai M.R. (2017). TALEN based HPV-E7 editing triggers necrotic cell death in cervical cancer cells. *Scientific Reports* **7**(5500).
DOI: https://doi.org/10.1038/s41598-017-05696-0

Shen L., Hua Y., Fu Y., Li J., Liu Q., Jiao X., Xin G., Wang J., Wang X., Yan C. & Wang K. (2017). Rapid generation of genetic diversity by multiplex CRISPR/Cas9 genome editing in rice. *Science China Life Sciences* **60**(5): 506–515.
DOI: https://doi.org/10.1007/s11427-017-9008-8

Sivachandran N., Wang X. & Frappier L. (2012). Functions of the Epstein-Barr virus EBNA1 protein in viral reactivation and lytic infection. *Journal of Virology* **86**(11): 6146 – 6158.
DOI: https://doi.org/10.1128/JVI.00013-12

Soppe J.A. & Lebbink R.J. (2017). Antiviral goes viral: harnessing CRISPR/Cas9 to combat viruses in humans. *Trends in Microbiology* **25**(10): 833–850.
DOI: https://doi.org/10.1016/j.tim.2017.04.005

Stemmer M., Thumberger T., del Sol Keyer M., Wittbrodt J. & Mateo J.L. (2015). CCTop: an intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PloS One*, **10**(4): e0124633.
DOI: https://doi.org/10.1371/journal.pone.0124633

Stothard P.2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28**(6): 1102–1104.
DOI: https://doi.org/10.2144/00286ir01

Streubel J., Blücher C., Landgraf A. & Boch J. (2012). TAL effector RVD specificities and efficiencies. *Nature Biotechnology* **30**(7): 593–595.
DOI: https://doi.org/10.1038/nbt.2304

Sundar S. & Rai M. (2002). Laboratory diagnosis of visceral leishmaniasis. *Clinical and Diagnostic Laboratory Immunology* **9**(5): 951–958.
DOI: https://doi.org/10.1128/CDLI.9.5.951-958.2002

Tsuge M., *et al* (11 authors) (2010). HBx protein is indispensable for development of viraemia in human hepatocyte chimeric mice. *Journal of General Virology* **91**(7): 1854–1864.
DOI: https://doi.org/10.1099/vir.0.019224-0

Usman B., Nawaz G., Zhao N., Liao S., Qin B., Liu F., Liu Y. & Li R. (2021). Programmed editing of rice (*Oryza sativa* L.) OsSPL16 gene using CRISPR/Cas9 improves grain yield by modulating the expression of pyruvate enzymes and cell cycle proteins. *International Journal of Molecular Sciences* **22**(1): 249.
DOI: https://doi.org/10.3390/ijms22010249

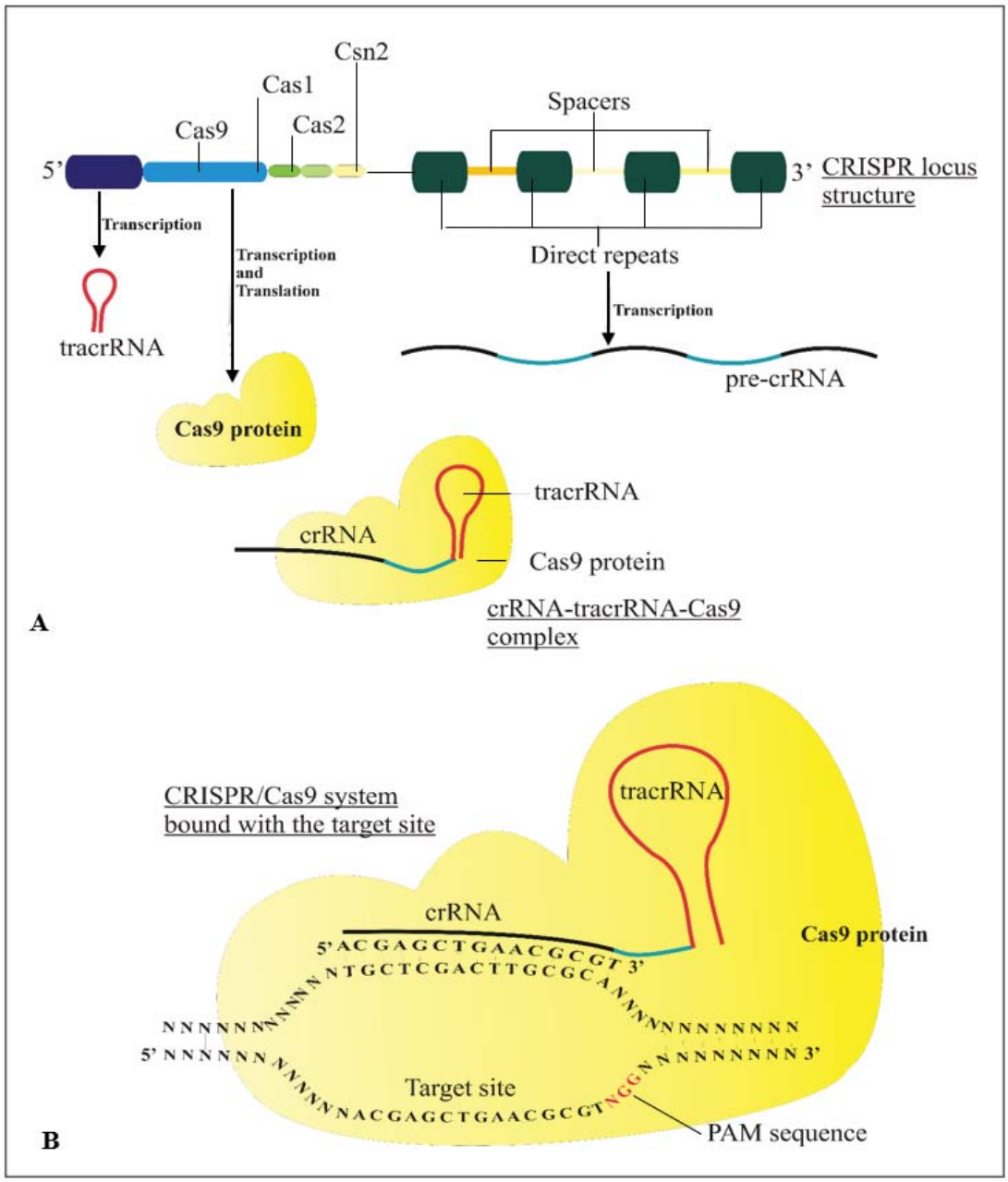
Wang T., Wei J.J., Sabatini D.M. & Lander E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**(6166): 80–84.
DOI: https://doi.org/10.1126/science.1246981

Xu P., Tong Y., Liu X.Z., Wang T.T., Cheng L., Wang B.Y., Lv X., Huang Y. & Liu D.P. (2015). Both TALENs and CRISPR/Cas9 directly target the HBB IVS2–654 (C>T) mutation in β-thalassemia-derived iPSCs. Scientific reports 5(12065).
DOI: https://doi.org/10.1038/srep12065

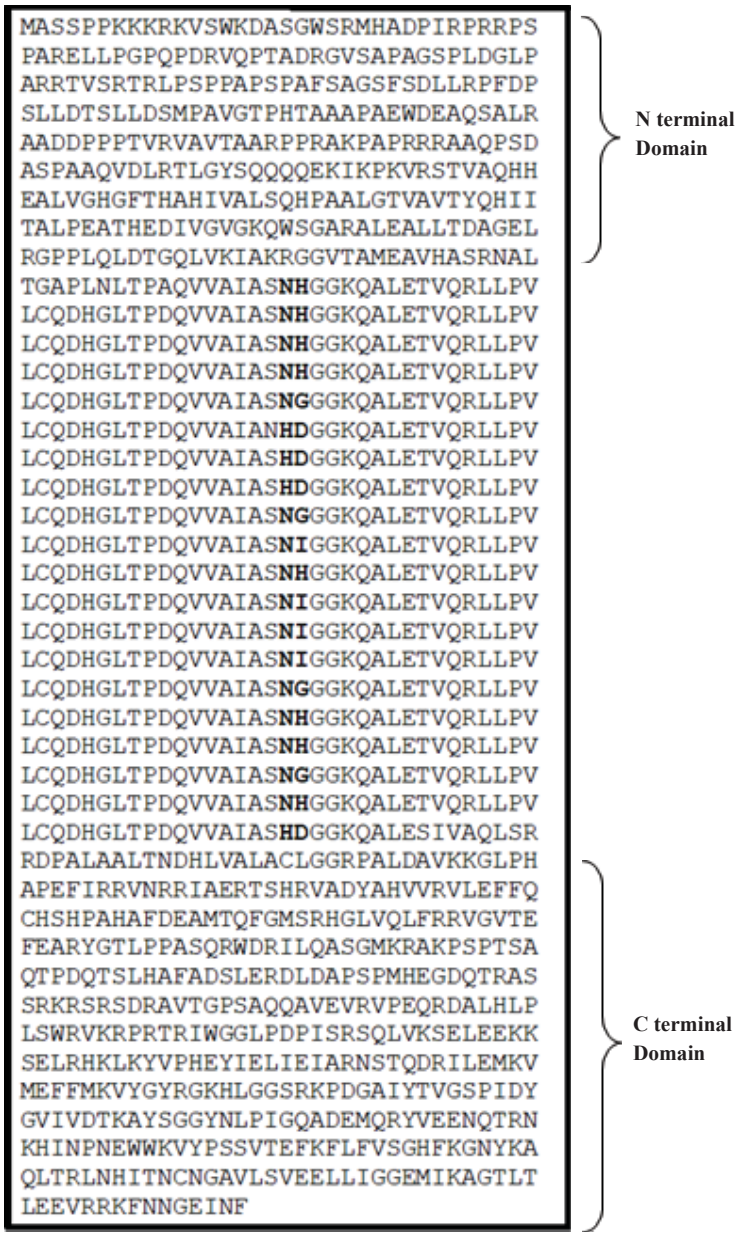
Yeadon J. (2014). Pros and cons of ZFNs, TALENs, and CRISPR/CAS. The Jackson Laboratory Blog Post. Available at https://www.jax.org/news-and-insights/jax-blog/2014/march/pros-and-cons-of-znfs-talens-and-crispr-cas, Accessed 4 November 2017.

Yuen K.S., et al (11 authors) (2015). CRISPR/Cas9-mediated genome editing of Epstein-Barr virus in human cells. Journal of General Virology 96(3): 626–636.
DOI: https://doi.org/10.1099/jgv.0.000012

Supplementary Figures



Supplementary Figure 1: CRISPR/Cas9 structure. A: CRISPR locus structure and its products. tracrRNA is formed from the transcription of tracrRNA locus, Cas9 region is responsible for the Cas9 protein and spacers and direct repeats are transcribed to pre-crRNA. Pre-crRNA is processed into crRNA with single spacer and repeat with the involvement of RNase III enzyme. These three molecules associate to form a complex known as crRNA-tracrRNA-Cas9. In artificial CRISPR system guide RNA is used instead of crRNA and tracrRNA. B: The recognition of target site by CRISPR/Cas9 system. The above formed complex bind complementarily with the target sequence and PAM is involved in the binding of Cas9 with the target site. (Designed based on Hwang *et al.* (2013), Nemydryi *et al*. (2014) and Soppe and Lebbink (2017) using the software CorelDRAW X6).



Supplementary Figure 2: The amino acid sequence of the TAL1 of TALEN EBVLMP2A01. The RVDs of the TALEN are shown in bold.

Supplementary Tables 1-6: Use below link to access Supplementary Tables 1-6
https://drive.google.com/drive/folders/1k19gtfOQsK6bcmR1EMH2qti2az-PGotI?usp=sharing

JOURNAL OF THE NATIONAL SCIENCE FOUNDATION OF SRI LANKA

GUIDANCE TO CONTRIBUTORS

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.

Categories of Submission

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 number of words is 10,000 including Figures, Tables and References. A Table/Figure occupying half a page will be considered as 400 words.

Reviews: Reviews are critical presentations on selected topics of Science or Technology. They should be well focused and organized and avoid general "textbook" style. A person/s wishing to submit a Review Article should have experience in the relevant field supported by an excellent publication record and 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). Maximum number of words is 12,000 including Figures, Tables and References. A Table/Figure occupying half a page will be considered as 400 words.

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. The article should include an Abstract, Keywords, Introduction, Methodology, Results & Discussion, Conclusion and References. Maximum number of words is 2,500 including Figures, Tables and References. A Table/Figure occupying half a page will be considered as 400 words.

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 number of words is 1,500 including Figures, Tables and References. A Table/Figure occupying half a page will be considered as 400 words.

SUBMISSION OF MANUSCRIPT

All manuscripts in MS Word format and PDF must be electronically submitted to the journal's online platform at https://jnsfsl.sljol.info/. Submissions *via* email are not encouraged.

Please make sure that no author information is mentioned in the article submitted.

A publication fee of US$ 150 will be levied for each manuscript other than where the corresponding author is affiliated to a Sri Lankan Institute.

Authors are required to provide their personal, validated ORCID ID (http://www.orcid.org). The ORCID ID will be displayed in the first page of the published article.

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). The JNSF does not accept manuscripts that have already been submitted to pre-print servers. The manuscripts submitted to the JNSF will initially be pre-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. If the revised manuscript is not received on time, the manuscript will not be processed further. The authors' response to the comments of referees should be tabulated with the comment, response and the line number/s for reference. The decision of the Editorial Board shall be final.

Submitted papers are subject to editing. The date of acceptance will be the day when the Editorial Board has decided it to be acceptable for publication.

The authors may suggest up to 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 supplied.

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.

A complimentary copy of the Journal will be supplied to each of the authors.

GUIDE TO PREPARATION OF MANUSCRIPTS

Manuscript

The manuscript should be free of errors and prepared in single column, using double-spaced text of Times New Roman 12 font throughout, leaving at least 2 cm margins on both sides, and liberal spacing at the top and bottom of each page. Sheets 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.

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. However, exceptions are allowed depending on the subject area. Pages should be arranged in the following order:

Page 1

Title page should include the title of manuscript, and no author information should be mentioned in the title page. The names and affiliations of all authors and the contact information of the corresponding author must be fed into the system during the online submission process. 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.

If a major part of the research has been published as an abstract in conference proceedings, it should be cited as a footnote on the title page.

Page 2 ff.

Title should accurately and concisely reflect the contents of the article.

Running title should be a shortened title that could be printed at the top of every other page of the article Journal and limited to a maximum of 50 characters.

Abstract **:** Should be between 200 - 250 words for full length articles and written as a single paragraph. 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. A Review Article should carry a summary of not more than 300 words.

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

Methodology **:** 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.
Collected data and relevant programming codes should be available upon the request of the reviewers.

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.

Conflict of interest statement **:** All authors should include a statement on conflict of interest disclosing any financial or other substantive conflicts of interest that may be construed to influence the results or interpretation of their research. All sources of financial support for the project should 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 in those indicated by the authors' affiliations given in the title page, this fact should be acknowledged. All those who have made substantial contribution to the research but do not qualify to be authors should be acknowledged.

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

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; b).
- Multiple citations should be made in chronological order and separated by a semi-colon, (i.e. Zimmerman *et al*., 1986; Able *et al*., 1997).
- Reference to unpublished work, work in preparation or work under review should be cited in italics as (*unpublished data*) or, with the author's initials and surname given; such works should not be included in the Reference section.
- Personal communications may be mentioned in the text with the date of communication as (*Personal communication, 2 June 2000*).

List of references:

- The list of References should be arranged in alphabetical order based on the last name of the first author.
- Names of all the authors should be given except when there are more than 10 authors. When there are more than 10 authors, only the name of the first author can be given followed by *et al*.
- 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 in bold type, 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.
- When books are listed, the order should be: author(s), year, book title, volume number, edition, pagination/ inclusive pages, publisher and place of publication. The book title should be in italics. When sections of a book are listed, the order should be: author(s) of chapter, year, title of the section, title of the book, edition, inclusive pages, publisher and place of publication.
- Digital object identifiers (DOIs) (https://doi.crossref.org/simpleTextQuery) should be included for all references where available.
- References should only be cited as 'in press' if the paper has been accepted for publication.

Examples of correct forms of references are given below.

Journal Articles

- Boutin C. & Harper J.L. (1991). A comparative study of the population dynamics of five species of *Veronica* in natural habitats. *Journal of Ecology* **79**(01): 199 – 221.
  DOI: https://doi.org/10.2307/2260793

Books

- Burnham K.P. & Anderson D.R. (2002). *Model Selection and Multimodal Inference*, 2nd edition, pp. 488. Springer Science and Business Media, Inc., New York, USA.

Book Chapters

- Hinrichsen R.A. & Holmes E.E. (2009). Using multivariate state-space models to study spatial structure and dynamics. In: *Spatial Ecology* (eds. R.S. Cantrell, C. Cosner & S. Ruan), pp. 145 – 166. CRC/ Chapman Hall, Boca Raton, Florida, USA.
  DOI: https://doi.org/10.1201/9781420059861.ch8

Edited Books

- KimatI H., Amorim L., Rezende J.A.M., Bergamin Filho A. & Camargo L.E.A. (eds.) (2005). *Manual de Fitopatologia*, volume 2. Doenças das Plantas Cultivadas, 4th edition. Ceres, São Paulo, Brazil.

Conference Papers

- Weaver D. (2002). Implementation of a learning management system using an integrated approach to professional development. In: Winds of change in the sea of learning. *Proceedings of the 19th Annual Conference of the Australasian Society for Computers in Learning and Tertiary Education (ASCILITE)* (eds. A. Williamson, C. Gunn, A. Young & T. Clear), volume 2, Auckland, New Zealand, 8-11 December. Unitec Institute of Technology, Auckland, New Zealand, pp. 711-720.

Agency Publications

- U.S. Census Bureau (2009). *World Population: 1950 – 2050*. U.S. Census Bureau, Washington DC, USA.
- Department of Health (2008). *Health Inequalities: Progress and Next Step* (pdf). Department of Health, London, UK. Available at *http:// www.dh.gov.uk/en/Publicationsandstatistics/ PublicationsPolicyAndGuidance/ DH _ 08 5307* , Accessed 9 June 2008.

Other

- Robinson L.J. (2003) Spatial scale and depletion models of farmland birds in a fragmented landscape. *PhD thesis*, University of Reading, Reading, UK.
- Efford M.G. (2008). Density 4.3: software for spatially explicit capture-recapture. Available at *http://www. otago.ac.nz/density* , Accessed 15 March 2009.

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. lf 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 : Should be clear and intelligible without reference to the text, and should not repeat data available elsewhere in the paper. 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 (^{a}, ^{b}, ^{c}, etc.).

Figures (Illustrations) : 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.

Note: Any reproduction of illustrations, tabulations, pictures etc. in the manuscript should be acknowledged. Tables and Figures should be kept to a minimum.

Units of measurement
Length: km, m, mm, µm, nm etc.
Area: ha, km^2, m^2, etc.
Capacity: kL, L, mL, µL etc.
Volume: km^3, m^3, cm^3 etc.
Mass: t, kg, g, mg, µg etc.
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

Tariffs for advertising in the JNSF (Local)

		TARIFF (Rs)			
		INSIDE PAGES		BACK INNER COVER	
NO. OF INSERTIONS	SIZE	B &W	COLOUR	B & W	COLOUR
1	Full page	10 000	20 000	20 000	30 000
	Half page	6 000	12 000	-	-
2	Full page	16 000	32 000	32 000	50 000
	Half page	10 000	20 000	-	-
4	Full page	30 000	60 000	60 000	100 000
	Half page	20 000	40 000	-	-

Tariffs for advertising in the JNSF (Foreign)

		TARIFF (US $)			
		INSIDE PAGES		BACK INNER COVER	
NO. OF INSERTIONS	SIZE	B &W	COLOUR	B & W	COLOUR
1	Full page	100	200	200	300
	Half page	60	120	-	-
2	Full page	160	320	320	500
	Half page	100	200	-	-
4	Full page	300	600	600	1000
	Half page	200	400	-	-

JOURNAL OF THE
NATIONAL SCIENCE FOUNDATION
OF SRI LANKA

Volume 49 Number 3 September 2021

CONTENTS

Printed by: NSF PRINTING UNIT