INNSPUB

Journal of Biodiversity and Environmental Sciences (JBES)
ISSN: 2220-6663 (Print) 2222-3045 (Online)
Vol. 19, No. 4, p. 36-46, 2021
http://www.innspub.net

RESEARCH PAPER OPEN ACCESS

Analysis of *rbcL* (Ribulose-1,5-Biphosphate carboxylase) gene sequences of identified noxious weed species (Poaceae)

Christian Joseph N. Ong[*1], Mariane Lou C. Obligar[1], Thelma DC. Arrieta[1], Oliver R. Alaijos[1,2]

[1]*Department of Biology, College of Science, Bulacan State University, City of Malolos, Bulacan, Philippines*

[2]*Graduate School, Bulacan State University, City of Malolos, Bulacan, Philippines*

Article published on October 26, 2021

Key words: BankIt, BLASTN, MEGA X, noxious, *rbcL* gene, SNPs.

Abstract

Noxious weed species usually ranks as a first or second problem in weed management, reduce farm productivity and increase weed control costs. Traditionally, identification of noxious grasses has generally relied on the morphological examination of grass floral material. Morphologically, about five species of grass weeds were identified and one species was identified only by genus level. These include *Eleusine indica* (L.) Gaertn., *Echinochloa crus-galli* (L.) P. Beauv., *Echinochloa stagnina* (Retz.) P. Beauv., *Ischaemum rugosum* Salisb. Var. *distachyum* (Cav.) Merr., *Leptochloa chinensis* (L.) Nees. and *Echinochloa* sp., respectively. DNA Barcoding may provide alternative means to identify noxious weed species. The utilization of genomic techniques depicted the *rbcL* gene in these plants, and have yielded a good amplification ranging from 500-850 base pairs (bp) in size. The BLASTN search results accumulated 95 - 98% maximum gene identity from the significant hits of specific species related from their *rbcL* gene sequences. The 9 *rbcL* gene sequences except Sample ID BS1 was input in MEGA X and were aligned and computed using Kimura- 2 Parameter (K2P) method. Single Nucleotide Polymorphisms (SNPs) variation analyses were used and identified 27 SNPs variant size among 9 *rbcL* gene sequences. Moreover, the 9 *rbcL* gene sequences were submitted and published to GenBank through BankIt, and have provided accession numbers. The molecular and phylogenetic relationships of *rbcL* gene in noxious weed species that were described and evaluated in this study may serve as baseline data for weed barcoding studies and enhancing weed management in rice cropping systems.

***Corresponding Author:** Christian Joseph N. Ong ✉ christiannavarro98.bsbio@gmail.com

Introduction

There are approximately 250,000 species of plants worldwide, about 3% or 8,000 species, behave as weeds. Of those 8,000, only 200 to 250 are considered to be a major problem in worldwide cropping systems (Lingenfelter and Hartwig, 2013). Weeds are regarded as unwanted and undesirable plants. These are the pests associated in any agriculture endeavor, which compete with rice and many other crops for sunlight, space, water and nutrients in soil (Hani *et al.*, 2017; Krishnan, 2018). Weeds possess adaptive characteristics which allow them to invade, survive and reproduce in different cropping system (Smith *et al.*, 2012).

The continuous invasion of weeds in field crops poses a serious problem in agriculture (Burlace, 2013). Weeds have a notorious nature, but managing its propagation can lessen its manifestation, which can help the crops to grow increasingly. In this context, the negative impacts of weeds indirectly affect all living beings (Lingenfelter and Hartwig, 2013). They are also considered as "noxious", because they are not actually native and considered as serious pest of rice cropping system (USDA-ARS, 2019); and they manifest around the world, specifically, in tropical and sub-tropical region of the continent (USDA-NRS, 2019). Based on the morphology, weeds are generally divided into three categories; the grasses (monocot), sedges and broad leaf (dicot) weeds. Grasses are under Poaceae family, which covers most of the land vegetation. The sedges belong to the Cyperaceae family with a modified stem, with or without tubers. On the other hand, broad leaf weeds are the major group of weeds; all the dicots have broader leaves (Walia *et al.*, 2011). Weeds can be classified based on how they complete their life cycle, namely the annuals, biennials and perennials (Dekker, 2011).

Among the categories of weeds, grasses (Poaceae) are the most difficult type of plant to identify by morphological features only. Traditionally, identifications were achieved after rigorous examination of morphological characteristics and subsequent consultation with the appropriate authoritative taxonomic literature.

However, in scenarios where the specimen is incomplete, traditional morphological methods can only generate reliable classifications at higher taxonomic levels. Considering molecular techniques are fast, more accessible and affordable, scientists globally are capitalizing on the discriminatory information contained in rapidly evolving regions of the genome to achieve species-level identifications (Meiklejohn *et al.*, 2019). Current researches suggest the process called DNA Barcoding as a tool for accessibly identifying up to the species level of grasses. DNA Barcoding makes a small typical sequence of genomic DNA (barcoding regions) of the grass, which can be a basis for a better floristic evaluation of grasses based on plant molecular analysis and its phylogenetic relationships (Saadullah *et al.*, 2016; Wang *et al.*, 2014). Utilization of *rbcL* gene sequencing for plant barcoding has been popular because of its power to discriminate or differentiate species with a marginal accuracy percentage.

The advantages of *rbcL* gene is it is easy to amplify and align sequence in most of the land plants at the family and genus level. This study was conducted to identify the species of grass weeds from selected rice fields in Malolos, Bulacan, Philippines based on their molecular and phylogenetic relationships. Results of this study would serve as baseline data about weed barcoding using *rbcL* barcoding genes as a universal marker used for plant molecular-based analysis and key for other potential studies in molecular taxonomy of these noxious weed species in the future.

Materials and Methods

Plant Material Collection and Preparation

The 10 samples of fresh leaf that were used from the whole grass weed in this study were collected in selected rice fields in Malolos, Bulacan, Philippines. The collected leaf samples were mounted in a clean resealable plastic container with label and were placed in a cooler with ice and salt to preserve their freshness and to protect from any contamination. An authentication was made by the Resident Taxonomist using the traditional method, such as the leaf arrangement, shape, inflorescences, seeds, etc. to

identify the grass weeds at the University of the Philippines Diliman-Institute of Biology (Jose Vera Santos Memorial Herbarium Section). Noxious weed species profiles with their scientific names and common names in each were given (Fig. 1).

DNA Extraction and Gel Electrophoresis

Genomic DNA extraction of the leaf samples were carried out by the University of the Philippines Diliman – Philippine Genome Center (PGC) using the modified cetyl trimethyl ammonium bromide (CTAB) method. After the gDNA were extracted, a 2µL of the genomic DNA sample was loaded to 1.5% agarose gel and separated for about 40 minutes at 4 V/cm. A 1 kilobase (kb) plus ladder (Invitogen®) was used.

PCR Amplification

rbcL gene amplification components include: gDNA, *rbcL* Forward (5'ATGTCACCACAAACAGAGACTA AAGC 3') and *rbcL* Reverse (5'GTAAAATCAAGTCC ACCRCG 3') universal primers (Costion *et al*., 2011; Maloukh *et al*., 2017), Titanium Taq buffer and Titanium DNA polymerase and Advantage Ultrapure dNTP mix. The PCR program was 96°C for 50 seconds for denaturation, 52°C for 50 seconds for annealing, and 72°C for 1 minute and 30 seconds for extension and 72°C for 10 minutes for final extension. All samples except sample ID BS2, DS3, and SB2 generated amplicons suitable for sequencing. The Sample ID BS2, DS3, and SB2 in the PCR program were subjected to re-amplification by following the same PCR condition above. However, samples BS2, DS3, and SB2 were then purified using 0.5x Agencourt AMPure XP beads and re-amplified again using the *rbcL* Forward and *rbcL* Reverse primers, with the resulting amplicons.

Sequencing and Data analysis

Sanger di-deoxy Sequencing was used in this portion. Cycle sequencing involves the incorporation of fluorescently labelled chain terminator ddNTPs. Components include: amplicons, primers, and ABI BigDye® Terminator v3.1 Cycle Sequencing Kit. The cycling parameters on Bio-Rad T100 Thermal Cycler were: pre-hold at 4°C; 96°C in 1 minute; 25 cycles of 96°C in 10 seconds, 50°C in 5 seconds, 62°C in 4 minutes; hold at 4°C. Ethanol precipitation was used to remove unincorporated ddNTPs, excess primers and primer dimers. Capillary electrophoresis was done on the ABI 3730xl DNA Analyzer using a 50cm 96-capillary array, POP7TM Polymer, and 3730xl Data Collection Software v3.1 base calling on the Sequencing Analysis Software v5.4.

The 10 samples of the collected grass weeds yielded good results of *rbcL* gene sequences. Cycle sequencing involves the incorporation of fluorescently labelled chain terminator dideoxynucleotide triphosphate (ddNTPs). The cycling parameters were pre-hold at 4°C. Several set of time and temperature were set at 4°C. The removal of unincorporated ddNTPs, excess primers and dimers was done using ethanol precipitation. ABI 3730xl DNA Analyzer using a 50cm – 96 capillary array, POP7TM Polymer and 3780xl data collection software was done using capillary electrophoresis. Base calling of the *rbcL* gene sequences on the sequencing analysis software v5.4 was used.

The quality control of the 10 *rbcL* gene sequences were carried out by trimming the contaminant peaks near base call near 3' and 5' ends to ensure the good quality of sequences after they were generated and processed. The *rbcL* gene sequences from both forward and reverse sequences were aligned and evaluated using BioEdit Sequence Alignment Editor v.7.0.5 to create consensus sequences of the *rbcL* gene sequences obtained from Capillary Sequencing. The consensus sequence of each sample ID was generated using Bio Edit Sequence Alignment Editor v7.0.5., and perform multiple sequence alignment through clustal W in the DNA Data Bank of Japan (http://clustalw.ddbj.nig.ac.jp/index.php?lang=en).

For species identification, BLASTN search was used in the National Center for Biotechnology Information (NCBI) website (Bethesda, 2008). Gen Bank submissions were made after the *rbcL* sequences were evaluated and processed using Bio Edit Sequence Alignment Editor. The 9 *rbcL* sequences were submitted in BankIt, and have used different annotations and features for the identification of sequences in the database search of NCBI.

Phylogenetic Analysis

The 9 *rbcL* gene sequences were aligned using BioEdit Sequence Alignment Editor version 7.0.5., then confirmed through MEGA X. Genetic distance was calculated using the Kimura-2-parameter model (K2P), and the phylogenetic tree was constructed by MEGA X (Muhammad *et al*., 2020).

Single Nucleotide Polymorphisms (SNPs) Variation Analysis

After the 9 *rbcL* sequences were aligned using Bio Edit Sequence Alignment editor software and in MEGA X, these sequences were subjected to SNP sequence variation analysis where specific SNPs variant positions can be found. SNPs positions were proceeded for the screening of *rbcL* sequences and to determine the unique SNPs variant in each of the sequences evaluated using Decision Tree SNPs Barcoding (DTSB) Approach (Yang *et al*., 2017).

Results and Discussion

rbcL Gene Amplification

The amplified regions of *rbcL* gene were determined and depicted its presence in 10 leaf samples of the collected grass weeds (Wattoo *et al*., 2016). The sample ID BS1, BS2, DS1, DS2, DS5, SB1, SB2, and SB3 have approximately ranges from ~500-600 base pairs (bp), while sample ID DS3 and BS2 are approximately ranges from ~500-850 base pairs (bp) in size. All samples yielded a good PCR amplification ranging from ~500-850 base pairs (bp) in size (Fig. 2). The PCR amplification success rate described in this study was 100% with *rbcL* universal primers.

BLASTN Species Identification of the Samples using rbcL Gene Sequences

The sample were assigned to their designated species using their consensus sequence except the BS1. However, the reverse complement of the BS1 was used to get the gene identity, the concensus and the reverse complement of BS1 were compared with the sequence from GenBank. The result were supported with a percent maximum identity that ranges from 95% to 98% (Bethesda, 2008). It is significant that the unidentified species of *Echinocloa* were assigned as *Echinochloa ugandensis*. Sample ID BS1 and SB3 were classified as *Dinebra panicea voucher*. In addition, Sample ID BS2, DS3 and SB2 were assigned as *Ischaemum aristatum*, respectively. Sample ID BS3 were assigned as *Echinochloa crus-galli* and the Sample ID DS5 was assigned as *Chloris virgata***.**

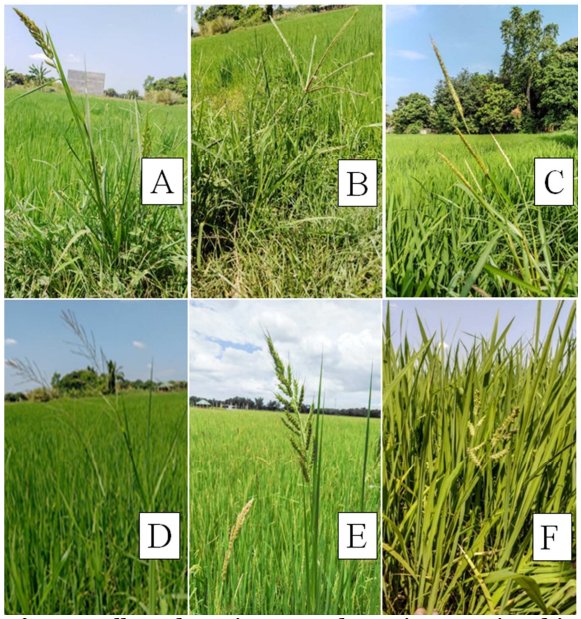

Fig. 1. Collected Noxious Weed Species associated in selected rice fields.
(a) *Echinochloa stagnina*, (b) *Eleusine indica,*
(c) *Ischaemum rugosum*, (d) *Leptochloa chinensis,*
(e) *Echinochloa crus-galli* and (f) *Echinochloa* sp.

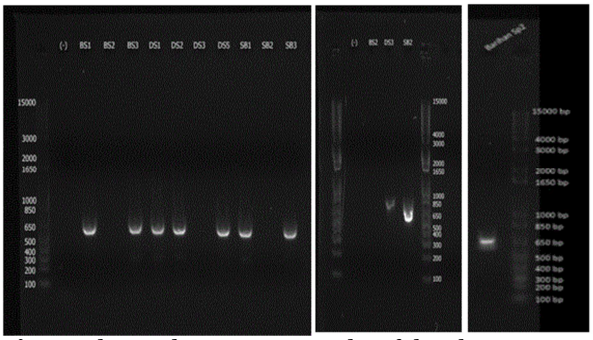

Fig. 2. Electrophoretogram results of the *rbcL* gene amplification of 10 weed species with ~500 to 850 bp size.

Comparison between Identified Weed Species and BLASTN Results

All samples accumulated percent gene identity of approximately 95% to 98% according to Nucleotide BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) search results. The discrepancies in their identity caused by differences in a single nucleotide sequence of *rbcL* gene sequences obtained.

Based on morphology, that there are six species of grass weed that were identified namely *Leptochloa chinensis* (L.) Neez., *Ischaemum rugosum* Salisb. var. distachyum (Cav.) Merr, *Echinochloa crus-galli* (L.) P. Beauv, *Echinochloa stagnina* (Retz.) P. Beauv, *Echinochloa* sp. and *Eleusine indica* (L.) Gaertn. (Naidu, 2012). On the other hand, the identification through molecular data have given significant search match data for the following results: In *Leptochloa chinensis* (L.) Neez with a sample ID of BS1 and SB3, search matches two significant hits of *Dinebra panicea voucher* have a different percent gene identity of 95.45% and 96.18%. In *Ischaemum rugosum* Salisb. var. *distachyum* (Cav.) Merr. with a sample ID of BS2, DS3 and SB2, search matches three significant hits of *Ischaemum aristatum* resulting to different percent gene identity of 97.68%, 98.60% and 98.76%. In *Echinochloa stagnina* (L.) P. Beauv. with a sample ID of DS1 and SB1, search matches two significant hits of *Echinochloa crus-galli var. praticola* with a percent gene identity of 98.81% and 97.10%.

In *Eleusine indica* (L.) Gaertn with a sample ID of DS5, search matches one significant hits of *Chloris virgata* with a percent gene identity of 98.02%. In *Echinochloa* sp. with sample ID of DS2, search matches one significant hit of *Echinochloa ugandensis* with a percent gene identity of 98.61%. On the other hand, *Echinochloa crus-galli* (L.) P. Beauv. with a sample ID of BS3 perfectly matches the result to both morphological and molecular data based from BLASTN, with a percent gene identity of 97.78% (Table 1). All the *Echinochloa* species have given significant data based on their search match from the GenBank database using *rbcL* gene sequences that were obtained (Ye *et al*., 2016).

Table 1. Search match results for Similarity Comparison Based on Identified weed species and Molecular Identification using *rbcL* gene barcodes.

Sample ID	Identified Noxious Weed Species	BLAST search match based on *rbcL* barcode	BLAST Similarity (%)
BS1	*Leptochloa chinensis* (L.) Nees	*Dinebra panicea voucher*	95.45%
BS2	*Ischaemum rugosum* Salisb. Var. *distachyum* (Cav.) Merr.	*Ischaemum aristatum*	97.68%
BS3	*Echinochloa crus-galli* (L.) P. Beauv.	*Echinochloa crus-galli*	97.78%
DS1	*Echinochloa stagnina* (L.) P. Beauv.	*Echinochloa crus-galli var. praticola*	98.81%
DS2	*Echinochloa* sp.	*Echinochloa ugandensis*	98.61%
DS3	*Ischaemum rugosum* Salisb. Var. *distachyum* (Cav.) Merr.	*Ischaemum aristatum*	98.60%
DS5	*Eleusine indica* (L.) Gaertn.	*Chloris virgata*	98.02%
SB1	*Echinochloa stagnina* (L.) P. Beauv.	*Echinochloa crus-galli var. praticola*	97.10%
SB2	*Ischaemum rugosum* Salisb. Var. *distachyum* (Cav.) Merr.	*Ischaemum aristatum*	98.76%
SB3	*Leptochloa chinensis* (L.) Nees	*Dinebra panicea voucher*	96.18%

GenBank Submission and Publication of the rbcL Gene Sequences

The *rbcL* gene sequences were submitted to GenBank, NCBI through BankIt for the publication of the nucleotide sequences online. There are six species of grass weeds that represents those 9 *rbcL* gene sequences that were submitted. The GenBank provides accession numbers for the following *rbcL* gene sequences. These are *I. rugosum* Isolate SB2 (with a GenBank Accession Number of MZ066627.1), *E. stagnina* Isolate SB1 (with a GenBank Accession Number of MZ066628.1), *E. indica* Isolate DS5 (with a GenBank Accession Number of MZ066629.1), *I. rugosum* Isolate DS3 (with a GenBank Accession Number of MZ066630.1), *Echinochloa* sp. Isolate DS2 (with a GenBank Accession Number of MZ066631.1), *E. stagnina* Isolate DS1 (with a GenBank Accession Number of MZ066632.1), *I. rugosum* Isolate BS2 (with a GenBank Accession Number of MZ066633.1), *E. crus galli* Isolate BS3

(with a GenBank Accession Number of MZ066634.1) and *D. chinensis* Isolate SB3 (with a GenBank Accession Number of MZ066626.1). These sequences have been recorded as the first manuscript of *rbcL* gene sequences from gDNA of the collected grass weeds. The NCBI Genbank published the 9 *rbcL* gene sequences and included in the Nucleotide BLAST Database of NCBI. In addition, these sequences are readily available to European Nucleotide Archive (ENA) and in DNA Data Bank of Japan (DDBJ) Database.

Phylogenetic Tree Analysis

The constructed Neighbor Joining (NJ) Tree of the 9 *rbcL* sequences and evolutionary relationships of taxa shows the relativity of the species given. Kimura-2 Parameter model was used to compute evolutionary distances in terms of numbers of substitution per site. There were total of 526 positions in the final dataset. The analysis involved 9 nucleotide sequences. Codon positions were included in the 1st+2nd+3rd+non-coding. All ambiguous positions were removed for each sequence pair (pairwise deletion option) (Kumar *et al.*, 2018). Neighbor-Joining (NJ) method in the analysis of phylogenetic can describe the clarity of weed species identification; the difference is limited by cluster and node.

The *rbcL* gene sequences of these noxious weed species can be in the same cluster even though they are from different areas. Phylogenetic relationships among the 6 identified noxious weed species using *rbcL* marker showed that *E. stagnina* Isolate SB1, *E. crus galli* Isolate BS3, *E. stagnina* Isolate DS1 and *Echinochola* sp. Isolate DS2 were clustered. The *I. rugosum* Isolate SB2 and DS3 were also clustered together, but the other *I. rugosum*, which is Isolate BS2 was clustered with *D. chinensis* Isolate SB3 and *E. indica* Isolate DS5 (Fig. 3). Species relationship based on genetic similarities is shown in the phylogenetic tree (Nurhasanah and Nurmaya, 2019). Bootstrapping method was applied to assure probability with 500 replicates in the NJ tree. In the constructed tree applied the boostrapping, the species were clustered accordingly but the *Echinochloa* sp. Isolate DS2 was the outgroup species in the tree (Fig. 4).

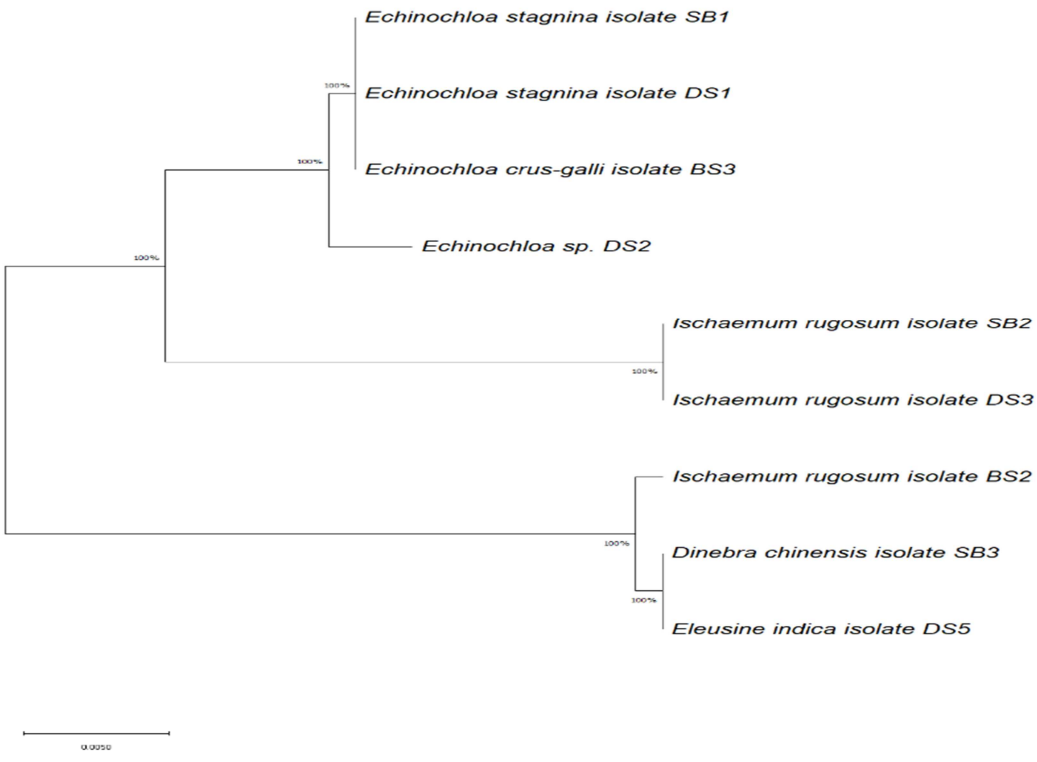

Fig. 3. Unrooted Neighbor-Joining (NJ) Tree using Kimura- 2 Parameter Model of 9 *rbcL* sequences from six species identified.

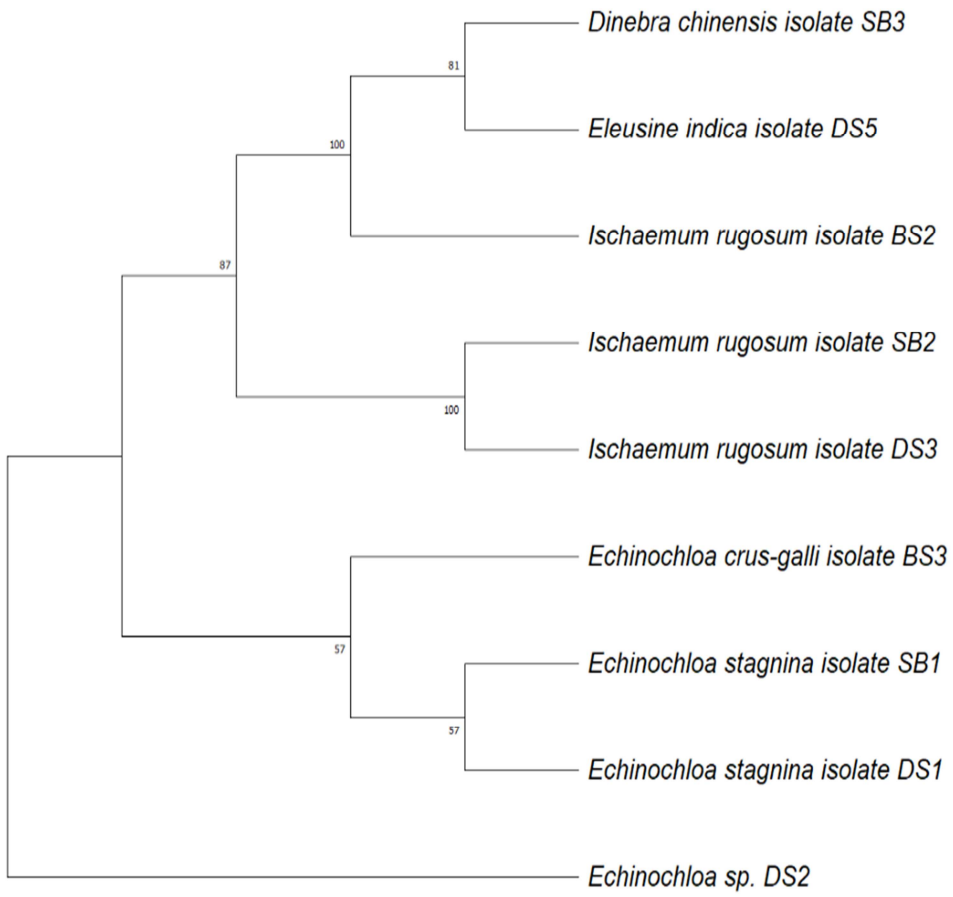

Fig. 4. Bootstrap Consensus Tree (500 replicates) using Kimura-2-Parameter Model of 9 *rbcL* sequences from six species identified.

The genetic distances based on the *rbcL* gene sequences ranged from 0.00% to 0.04% (Table 2). The smallest distance of 0.00% was observed between *E. crus galli* (MZ066634) and *E. stagnina* (MZ066628). However, *I. rugosum* (MZ066633), *E. indica* (MZ066629), and *I. rugosum* (MZ066627) were characterized by the maximum of divergence since the sequences exhibited the greatest genetic distance value of 0.04%. A similar study was observed in the genetic distances calculated using the program of MEGA 6 (Su *et al.*, 2017). The average AT nucleotide composition for *rbcL* of all collected noxious weed species was found higher to be 56.63% than GC content, 43.37% (Table 3). Similar findings have been reported by (Muhammad *et al.*, 2020), who found higher AT content (55.4%) than GC content (44.54%) in Acacia species using *rbcL* gene marker. Same study with (Muhammad *et al.*, 2020) that the possible explanation of higher AT contents than GC contents in *rbcL* marker could be due to high variability in nucleotide composition and higher nucleotide substitution rate in this marker. The polymorphism was further evaluated to estimate the interspecific and intraspecific divergence. Higher interspecific variation than intraspecific variation represents success of the study and may have wide scienific applications (Naeem *et al.*, 2014).

For the *rbcL* gene, the analysis involved 9 nucleotide sequences. There were a total of 525 positions in the final dataset. The T to G and C to G transversion rate was 3.36, that of A to T and G to T was 4.37, that of T to A and C to A was 4.06, and that of A to C and G to C was 3.11. Frequencies of the nucleotide substitution were A = 27.26%, T/U = 29.33%, C = 20.85%, and G = 22.56% (Table 4).

Table 2. Genetic distance matrix among the noxious weed species taxa based on *rbcL* gene sequence**.**

No.	Scientific Name (NCBI accession number)	1	2	3	4	5	6	7	8	9
1	*I. rugosum* (MZ066633)	-								
2	*E. crus-galli* (MZ066634)	0.03	-							
3	*E. stagnina* (MZ066628)	0.03	0.00	-						
4	*I. rugosum* (MZ066627)	0.04	0.02	0.02	-					
5	*D. chinensis* (MZ066626)	0.00	0.03	0.03	0.04	-				
6	*E. stagnina* (MZ066632)	0.03	0.00	0.00	0.02	0.03	-			
7	*E. indica* (MZ066629)	0.00	0.03	0.03	0.04	0.00	0.03	-		
8	*I. rugosum* (MZ066630)	0.04	0.02	0.02	0.00	0.04	0.02	0.04	-	
9	*Echinochloa* sp. (MZ066631)	0.03	0.00	0.00	0.02	0.03	0.00	0.03	0.02	-

Table 3. The average of AT% and GC% nucleotide composition of the Identified Noxious Weed Species based on *rbcL* marker.

Sample ID	AT	GC	Total	%AT	%GC
BS2	301	224	525	57.3	42.7
BS3	295	231	526	56.1	43.9
SB1	295	231	526	56.1	43.9
SB2	298	228	526	56.7	43.3
SB3	302	224	526	57.4	42.6
DS1	295	231	526	56.1	43.9
DS2	294	232	526	55.9	44.1
DS3	298	228	526	56.7	43.3
DS5	302	224	526	57.4	42.6
AVERAGE				56.63%	43.37%

Table 4. Transition and transversion rates of *rbcL* nucleotide sequences in Noxious weed species.

	A	T/U	G	C
A	-	4.37	3.11	18.72
T/U	4.06	-	11.99	3.36
C	4.06	16.87	-	3.36
G	22.62	4.37	3.11	-

Transition rates are shown in bold, and transversion rates are shown in *italics*. For simplicity, the sum of r values is 100.

SNPs Variation Analysis

All the sequences were edited manually using BioEdit Sequence Alignment Editor software, mismatch sites and all the gaps were removed for proper screening of SNPs. The 9 *rbcL* gene sequences were selected for the detection of SNPs. A total of 27 unique SNP based variant size were found in 9 *rbcL* sequences for differentiation through *rbcL* marker (Table 5). A similar study was obaserved and found the same output, these *rbcL* sequences found to have 4 specific SNPs (Yang *et al*., 2017; Yang *et al*., 2018).

Table 5. SNP Variation Analysis among the 9 *rbcL* gene sequences of noxious weed species.

Species Position index	BS2	BS3	SB1	SB2	SB3	DS1	DS2	DS3	DS5
3	**G**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**
13	**C**	**A**	**A**	**A**	**C**	**A**	**C**	**A**	**C**
21	**T**	**G**	**G**	**T**	**T**	**G**	**G**	**T**	**T**
27	A	G	G	G	A	G	G	G	A
48	C	C	C	T	C	C	C	T	C
57	A	G	G	G	A	G	G	G	A
110	C	C	C	T	C	C	C	T	C
111	C	C	C	C	C	C	G	C	C
135	**G**	**G**	**G**	**A**	**G**	**G**	**G**	**A**	**G**
237	A	G	G	G	A	G	G	G	A
252	A	G	G	C	A	G	G	C	A
253	G	C	C	C	G	C	C	C	G
254	A	C	C	C	A	C	C	C	A
255	C	A	A	A	C	A	A	A	C
256	A	G	G	G	A	G	G	G	A
257	G	A	A	A	G	A	A	A	G
274	A	A	A	G	A	A	A	G	A
291	T	C	C	C	T	C	C	C	T
336	G	G	G	A	G	G	G	A	G
361	C	C	C	T	C	C	C	T	C
366	T	C	C	C	T	C	C	C	T
396	T	C	C	C	T	C	C	C	T
400	G	A	A	G	G	A	A	G	G
406	T	T	T	G	T	T	T	G	T
420	A	G	G	A	A	G	G	A	A
432	T	C	C	C	T	C	C	C	T
468	T	T	T	C	T	T	T	C	T

In a related studies, the use of SNP variation analysis lead to the development of certain barcodes (in a form of QR barcode or strip barcode) for the collected species by using the *rbcL* gene sequences to easily access the identification of each species. However,

there are 38 species that were barcode (Yang *et al.*, 2018). When used, these barcode independently differentiated the species thus can be applied to identify the species through barcode scanning apps on smartphones like a tag price in a product or commodity (Muhammad *et al.*, 2020).

Conclusion

The present study identified five species of grass weeds (Poaceae) associated in rice field and one was identified up to genus only. These are *Echinochloa crus-galli, Echinochloa stagnina, Ischaemum rugosum, Eleusine indica, Leptochloa chinensis* and *Echinochloa* sp., respectively. The identified weed species and *rbcL* gene barcodes gave rise to the matching BLASTN results. The percent gene identity of the grass weeds ranges from 95 to 98%, which indicates a significant result. Most of the specific species obtained from BLASTN are related based on the genus level of grass weeds matches based on the identification through morphology. The clustering of species in the NJ Tree and through SNPs variation analysis proved that the utilization of *rbcL* markers have the slight power for species discrimination and identification of noxious weed species, effectively. This is more useful and a powerful tool that can countercheck the identity of an organism using barcodes that identify the identity of an unknown species. However, the genetic information obtained from the study may be used as a guide for future barcoding of noxious grass weeds.

Acknowledgements

This research was supported financially by the PMCP Foundation, Inc. of the College of Agriculture and Food Science, University of the Philippines Los Baños, Laguna. No conflicts of interest have been declared. Special thanks to Dr. Joeselle M. Serrana of Ehime University in Japan, for his invaluable help and guidance for the annotations of the *rbcL* gene sequences. Also, Thanh Dat Ta and Samantha Sevilleno of Samhyook University in South Korea for their expertise, guidance and help during the content review of the manuscript.

References

Bethesda. 2008. BLAST (Basic Local Alignment Search Tool) Command Line Applications User Manual. National Center for Biotechnology Information (US).

Burlace M. 2013. Understanding Weed Characteristics. https://farmref.com/understanding-weed-characteristics.html

Costion C, Ford A, Cross H, Crayn D, Harrington M., and Lowe A. 2011. Plant DNA Barcodes can Accurately Estimate Species Richness in Poorly Known Floras. PLoS ONE **6(11),** E26841. DOI: 10.1371/journal.pone.0026841

Dekker J. 2011. Evolutionary Ecology Models of Weed Life History. Weed Biology Laboratory, Agronomy Department, Iowa State University.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootsrap. Evolution **39,** 783-791.

Fisheries QD. 2011. Special edition of Environmental Weeds of Australia for Biosecurity Queensland., Australia:. Univ http://keyserver. lucidcentral.org/weeds/data/03030800-0b07-490a-8d04-0605030c0f01/media/Html/Index.htm

[FAO] Food Agriculture Organization. 2014. Grassland Species profiles. http;//www.fao.org/ag/AGP/AGPC/doc/Gbase/lDefault.html.

Hani M, Lebazda R, Fenni R. 2017. Studies of Morphological Characteristics and Production of Seeds Weeds of Species of Family Brassicaceae (Cruciferous) in Setifian High Plateau, Algeria. Annual Research & Review in Biology, DOI: 10.9734/ARRB/2017/33473.

Hartwig DD. 2013. Introduction to Weeds and Herbicides. The Pennsylvania State University.

Kumar SG, Lim KC. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution **35,** 1547-1549.

Krishnan N. 2018. The environmental impact of the world's most famous weed killer. https://envirobites.org/2018/01/19/the-environmental-impact-of-the-worlds-most-famous-weed-killer/

Maloukh L, Kumarappan A, Jarrar M, Salehi J, El-Wakil H, Lakshmi TVR. 2017. Discriminatory Power of *rbcL* Barcode locus for Authentication of some of United Arab Emirates (UAE) Native Plants. 3 Biotech (2017) **7,** 144. DOI: 10.1007/s13205-017-0746-1

Meiklejohn KA, Damaso N, Robertson JM. 2019. Assessment of BOLD and GenBank – Their accuracy and reliability for the identification of biological materials. PLoS ONE **14(6),** e0217084.

Muhammad I, Aftab A, Muhammad N, Muhammad AJ, Sulta HK, Iqra K, Aftab AS, Qari S, Alghanem S, Khalid AK, Muhammad FK, Samina Q. 2020. Development of DNA barcodes for selected *Acacia* species by using *rbcL* and *matK* DNA markers. Saudi Journal of Biological Sciences. DOI: 10.1016/j.sjbs.2020.08.020

Naeem A, Khan AA, Cheema HMN, Khan IA, Buerker A. 2014. DNA barcoding for species identification in the Palmae family. Genetics and Molecular Research **13(4),** 10341-10348 (2014). DOI: 10.4238/2014.December.4.29

Naidu VGSR. 2012. Hand Book on Weed Identification. Directorate of Weed Science Research, Jabalpur, India

Nurhasanah S, Nurmaya P. 2019. Amplification and Analysis of *rbcL* Gene (Ribulose-1,5-Biphosphate Carboxylase) of Clove in Ternate Island. IOP Conf. Series: Earth and Environmental Science 276 (2019) 012061. DOI:10.1088/1755-1315/276/1/012061

Randall RP. 2012. A global compendium of weeds. Dept. Agric. Food Western Australia, Saadullah Z. 2016. Identification of the grass family (Poaceae) by using the plant dna barcodes *rbcL* and matK . Journal of Biodiversity and Environmental Sciences (JBES) , Vol **8, No. 5,** p. 175-186.

Smith RG, Ryan MR, Menalled FD. 2011. Direct and Indirect Impacts of Weed Management Practices on Soil Quality. Pages 275-286 in Hatfield JL, Sauer TJ, eds. Soil Management: Building Stable Base for Agriculture. American Society of Agronomy and Soil Science Society of America.

Su JK, Kwang SC, Ki OY, Ki BL, Yoon JH, Dong CC, Ki SK. 2017. Sequence Analysis of the Internal Transcribed Spacer (ITS) Region of the Nuclear Ribosomal DNA (nrDNA) Chrysanthemum Species in Korea. Horticulture, Environment, and Biotechnology **56(1),** 44-53. 2015. DOI: 10.1007/ s13580-015-0085-2

[USDA-ARS] US Department of Agriculture-Agricultural Research Service. 2014. Germplasm Resource Information Network (GRIN) Online Database. Beltsville, Maryland, USA: National Germplasm Resources Laboratory. https://npgsweb.arsgrin.gov/gringlobal/taxon/taxonomysearch.aspx

[USDA-NRS] US Department of Agriculture-Natural Resource Service. 2019. Natural Resources Conservation Service. Online Database. http://plant.usda.gov/core/profile?symbol=LECH2

Walia US, Walia SS, Kler SS, Singh D. 2011. Science of Agronomy Scientific Publisher (India), pg. 61.

Wang A, David G, Wu H, Stanton R, Lepschi B. 2014. DNA Barcoding for Identification of Exotic grass species present in Eastern Australia. Pages 444-447 in Proceedings of the Nineteenth Australasian Weeds Conference.

Wattoo JI, Muhammad ZS, Muhammad SS, Amina A, Amir H, Mushtaq AS. 2016. DNA Barcoding: Amplification and sequence analysis of *rbcl* and matK genome regions in three divergent plant species. Advancements in Life Sciences. International Quarterly Journal of Biological Sciences **4(1),** 03-07.

Yang CH, Wu KC, Dahms HU, Chuang LY, Chang HW. 2017. Single nucleotide polymorphism barcoding of Cytochrome C oxidase I sequences for discriminating 17 species of Columbidae by decision tree algorithm. Ecology and Evolution. 2017; **7,** 4717-4725. DOI: 10.1002/ece3.3045

Yang CH, Wu KC, Chuang LY, Chang HW. 2018. Decision Tree Algorithm–Generated Single-Nucleotide Polymorphism Barcodes of *rbcL* Genes for 38 Brassicaceae Species Tagging. Evolutionary Bioinformatics Volume **14,** 1-9.

Ye CY, Lin Z, Li G, Wang YY, Qiu J, Fu F. 2014. Echinochloa Chloroplast Genomes: Insights into the Evolution and Taxonomic Identification of Two Weedy Species. PLoS ONE **9 (11),** e113657, DOI: 10.1371/journal.pone. 0113657.