A correction has been published for this article: https://peerj.com/articles/7965/correction-1/

Development of molecular markers for invasive alien plants in Korea: a case study of a toxic weed, *Cenchrus longispinus* L., based on next generation sequencing data

JongYoung Hyun*, Hoang Dang Khoa Do*, Joonhyung Jung and Joo-Hwan Kim

Department of Life Science, Gachon University, Seongnam, Gyeonggi, Korea
* These authors contributed equally to this work.

ABSTRACT

Background: Genomic data play an important role in plant research because of its implications in studying genomic evolution, phylogeny, and developing molecular markers. Although the information of invasive alien plants was collected, the genomic data of those species have not been intensively studied.
Methods: We employ the next generation sequencing and PCR methods to explore the genomic data as well as to develop and test the molecular markers.
Results: In this study, we characterize the chloroplast genomes (cpDNA) of *Cenchrus longispinus* and *C. echinatus*, of which the lengths are 137,144 and 137,131 bp, respectively. These two newly sequenced genomes include 78 protein-coding genes, 30 tRNA, and four rRNA. There are 56 simple single repeats and 17 forward repeats in the chloroplast genome of *C. longispinus*. Most of the repeats locate in non-coding regions. However, repeats can be found in *infA, ndhD, ndhH, ndhK, psbC, rpl22, rpoC2, rps14, trnA-UGC, trnC-GCA, trnF-GAA, trnQ-UUG, trnS-UGA, trnS-GCU*, and *ycf15*. The phylogenomic analysis revealed the monophyly of *Cenchrus* but not *Panicum* species in tribe Paniceae. The single nucleotide polymorphism sites in *atpB, matK*, and *ndhD* were successfully used for developing molecular markers to distinguish *C. longispinus* and related taxa. The simple PCR protocol for using the newly developed molecular markers was also provided.

Subjects Genomics, Molecular Biology, Plant Science
Keywords *Cenchrus longispinus*, *Cenchrus echinatus*, Invasive alien plants, Chloroplast genome, Molecular markers

Submitted 9 July 2019
Accepted 30 September 2019
Published 11 November 2019

Corresponding author
Joo-Hwan Kim,
kimjh2009@gachon.ac.kr

Academic editor
Vladimir Uversky

Additional Information and Declarations can be found on page 11

DOI **10.7717/peerj.7965**

INTRODUCTION

Cenchrus is a member of 780 genera of Poaceae and has widespread distributions in Asia, Africa, Australia, and America (*Govaerts, 2011*; *Christenhusz & Byng, 2016*). *Cenchrus longispinus*, also called long-spined sandbur or spiny burr grass, is native to North America. However, this species invaded into other continents and became a noxious weed (*Strat, Stoyanov & Holobiuc, 2017*; *Fagaras, 2018*). In Korea, *C. longispinus* has been recognized as one of 320 invasive alien plants, which locates in Incheon province (*Jung et al., 2017*). Among *Cenchrus* species, *C. longispinus* and *C. echinatus* share the

features of ovoid to globose burs with longer inner bristkes, sparse or numerous outer bristles, and the mostly fused or half fusion of the inner bristles (*Verloove & Sánchez Gullón, 2012*). However, *C. echinatus* is distinguished from *C. longispinus* by the characters of the burs which have numerous flexible and distinctly restrorsely barbellate outer bristles, and a single whorl inner bristles that form flattened spines. Previously, various studies on the management of *C. longispinus* have been conducted (*Anderson, 1997*; *Soltani et al., 2010*). Although the invasion and control of *C. longispinus* have been conducted, the study on its genomic data has not been approached. However, the genomic data of other *Cenchrus* (i.e., the chloroplast genome of *C. purpureus* and *C. ciliaris*) were reported (*Bhatt & Thaker, 2018*; *Wu & Zhou, 2018*). Chloroplast genome (cpDNA) plays an important role in plants because it contains essential genes for performing photosynthesis (*Sugiura, 1992*). Additionally, the highly conserved of cpDNA among plants resulted in the useful applications of cpDNA data in reconstructing phylogeny, exploring biogeography, surveying population genetics, and developing molecular markers (*Cui et al., 2017*; *Fischer et al., 2017*; *Leaché & Oaks, 2017*; *Pantoja et al., 2017*). In Poaceae, different studies on cpDNA have been reported (*Matsuoka et al., 2002*; *Hand et al., 2013*; *Zhang et al., 2016*; *Huang et al., 2017*). Specifically, there are different inversion event in cpDNA of the grass family (*Doyle et al., 1992*). Additionally, loss of gene (i.e., *accD*) was also found in the cpDNA of Poaceae (*Huang et al., 2017*). Also, the usage of cpDNA for phylogeny and development of molecular marker have been conducted among Poaceae species (*Zhang, 2000*; *Soreng et al., 2017*; *Sharma et al., 2018*).

The genomic data provide a profound understanding of the evolution and the potential solution for management of the invasive alien plants. Although the list of invasive alien plants in Korea was published, further studies on the management and genomic data have not been conducted. Therefore, in this study, we based on the next generation sequencing method to (1) complete and characterize the chloroplast genomes of *C. longispinus* and *C. echinatus*, (2) reconstruct the phylogeny of *Cenchrus* and related taxa in tribe Paniceae, and (3) develop molecular markers inferred from the single nucleotide polymorphisms (SNP) sites in *atpB, matK*, and *ndhD* to distinguish *C. longspinus* in Korea from related taxa.

MATERIALS AND METHODS

Taxon sampling, DNA extraction, chloroplast genome assembly, and comparison

The fresh leaves of *C. longispinus* and *C. echinatus* were collected in Daecheong Island (Incheon, South Korea) and Texas (USA), respectively. These samples were dried using silica gel powder before their total DNA was extracted based on the modified CTAB method (*Doyle & Doyle, 1987*). The high-quality DNA (>200 ng/1 µl) was used for constructing NGS data through Miseq platform (Illumina, Seoul, Korea). The complete chloroplast genome of *C. purpureus* (Accession number MF594682) was used as a reference genome for the assembly of cpDNA of *C. longispinus* and *C. echinatus*. After the NGS data were imported into Geneious program (*Kearse et al., 2012*), only reads that have over 95% similarity to reference cpDNA were isolated. Then, the isolated reads were assembled to complete the cpDNA sequences using *denovo* assemble function in Geneious.

Consequently, 95,386 out of 5,053,948 reads (1.9%) were assembled to complete the cpDNA of *C. longispinus* (Accession number MN078361) with the coverage of 208×. In the case of *C. echinatus*, only 20,267 out of 4,211,108 reads built up the complete cpDNA (Accession number MN078360) with the coverage of 44.3×. The complete cpDNA sequences of *C. longispinus* and *C. echinatus* were annotated in Geneious based on the previously published cpDNA of *C. purpureus*. The annotations that have over 95% similarity were kept and manually checked and adjusted the start and stop codons. The tRNA sequences were checked using tRNA Scan-SE (*Chan & Lowe, 2019*). The map of cpDNA was illustrated using OGDraw (*Greiner, Lehwark & Bock, 2019*). The complete cpDNA sequences of *Cenchrus* and related taxa were aligned using MAUVE embedded in Geneious (*Darling et al., 2004*). The REPuter program was used to locate the repeats in the chloroplast genomes with the minimum repeat length of 18 bp (*Kurtz et al., 2001*). To analyze the simple sequence repeat (SSR), the Phobos program was used with the setting of minimum repeated time of ten, five, four, three, and three for mono-, di-, tri-, tetra-, and pentanucleotide repeats, respectively (*Christoph, 2006–2010*).

Phylogenomic analysis

The complete cpDNA sequences of *Cenchrus* and related taxa were downloaded from NCBI (Table S1). The 75 protein-coding regions of cpDNA were extracted and aligned using MUSCLE embedded in Geneious (*Edgar, 2004*). Additionally, three partition data sets (including large single copy (LSC), small single copy (SSC), and inverted repeat (IR) regions) were used to test the phylogenetic utility of different regions in cpDNA of *Cenchrus* and related species. Three regions were also aligned using MUSCLE embedded in Geneious. Then, the aligned sequences were imported to jModeltest to find the best model for elucidating the phylogeny of *Cenchrus* (*Posada, 2008*). Consequently, the TVM+I+G model was selected as the best model for further analysis. The Maximum likelihood (ML) and Bayesian Inference (BI) analysis were conducted using IQtree (*Trifinopoulos et al., 2016*) and MrBayes (*Ronquist et al., 2012*). In the ML analysis, the bootstrapping test was repeated 1,000 times for calculating bootstrap values. Meanwhile, the BI analysis was run with one million generations. The phylogenetic tree was illustrated using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Development of molecular markers for *Cenchrus longispinus*

To develop the molecular marker for *C. longispinus*, the cpDNA sequences of *C. longispinus* and *C. echinatus* were aligned to locate the SNP sites. The primer pairs that cover the regions containing SNP sites were designed based on the conserved regions in two *Cenchrus*. Then, these primer pairs were used for *Pennisetum alopecuroides*, which has a close relationship to *Cenchrus* and has a wide distribution in Korea. The primer pairs that resulted in PCR products in *Pennisetum* sample were selected for sequencing. The new sequences from *Pennisetum* were aligned with those of *Cenchrus* to locate the specific SNP sites for *C. longispinus*. Consequently, the *atpB*, *matK*, and *ndhD* were selected to develop the molecular markers for *C. longispinus*. The primer pairs for three genes were designed using Primer3 (*Untergasser et al., 2012*; Fig. 1; Table S2). The specific

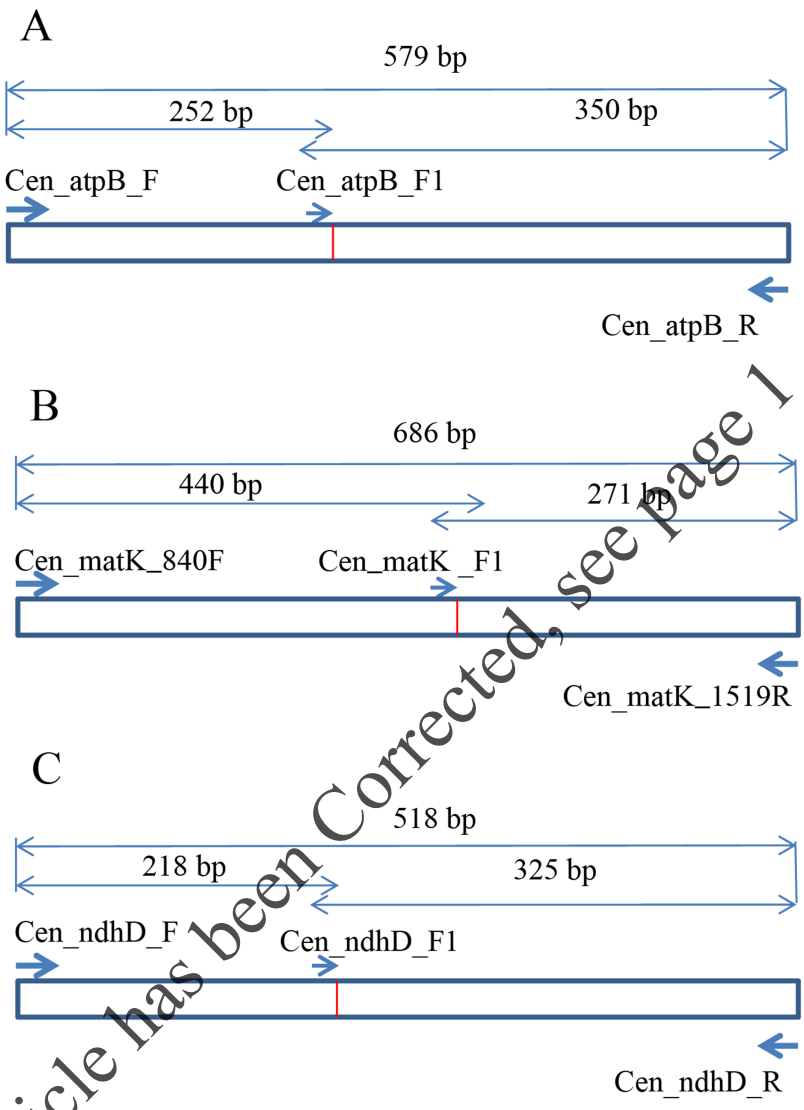

Figure 1 The design of primer pairs for SNP sites in *Cenchrus longispinus*. (A) The primer pairs in the *atpB* gene; (B) the primer pairs in the *matK* gene; (C) the primer pairs in the *ndhD* gene.
Full-size DOI: 10.7717/peerj.7965/fig-1

lengths of PCR products were designed for *C. longispinus* (Fig. 1). The PCR mixture and protocol were described in Table S3. For testing the efficiency of the primer pairs, further samples of three species were collected from different locations with the permission from Korea National Arboretum (KNA), Gachon University Herbarium (GCU); The University of Texas at Austin Herbarium (TEX), and The Queensland Herbarium (BRI) (Table S4). The PCR products were checked using agarose gel 1% and electrophoresis method.

RESULTS

Plastome features of *Cenchrus*

The complete cpDNA of *C. longispinus* has a quadripartite structure which consists of LSC region of 80,223 bp, a SSC region of 12,449 bp and two inverted repeat regions of

Table 1 Features of Chloroplast genomes and pairwise identity among *Cenchrus* species and related taxa.

Species	*Cenchrus longispinus*	*C. echinatus*	*C. purpureus*	*Eriochloa meyeriana*	*Panicum capillare*	*Dichanthelium acuminatum*	*Thyridolepsis xerophila*	*Lecomtella madagascariensis*
Accession number	MN078361	MN078360	MF594682	KU291498	KU291475	KU291496	KU291485	HF543599
Total length (bp)	137,144	137,131	138,199	139,890	134,520	140,122	140,644	139,073
GC content (%)	38.7	38.7	38.6	38.4	38.5	38.7	38.7	38.6
Lagre single copy (LSC-bp)	80,223	80,220	81,161	81,856	81,858	82,065	82,641	81,349
Small single copy (SSC-bp)	12,449	12,439	12,386	12,568	12,576	12,617	12,599	12,452
Inverted repeat (IR-bp)	22,236	22,236	22,236	22,733	20,043	22,720	22,702	22,636
Protein coding genes	78	78	78	78	76	78	78	78
tRNA	30	30	30	30	30	30	30	30
rRNA	4	4	4	4	4	4	4	4
LSC-IR junction	IGS (*rps19-rpl22*)	IGS (*rps19-rpl22*)	IGS (*rps19-rpl22*)	IGS (*rps19-rpl22*)	IGS (*rps19-rpl22*)	IGS (*rps19-rpl22*)	IGS (*rps19-rpl22*)	IGS (*rps19-rpl22*)
SSC-IR junction	*ndhF* (29 bp)	*ndhF* (29 bp)	*ndhF* (29 bp)	*ndhF* (29 bp)	*ndhF* (29 bp)	*ndhF* (29 bp)	*ndhF* (29 bp)	*ndhF* (17 bp)
Pairwise Identity (%)	100	99.9	97.3	94.6	92.3	94.4	94.9	94.4

22,236 bp (Table 1; Fig. 2). A similar structure was also found in cpDNA of *C. echinatus*; however, the lengths of LSC, SSC, and IR regions are 80,220, 12,439, and 22,236 bp, respectively. Although the lengths of LSC and SSC regions are different, the sizes of IR regions are identical in three examined *Cenchrus*. In comparison with related taxa, there is no identical length of three regions (Table 1). Albeit the lengths of three regions are various, the number of protein-coding genes (79), tRNA (30). and rRNA (4) are identical among *Cenchrus* and surveyed species (Table 1; Table S5), except *Panicum capillare* which do not have *ycf2* and *ycf15* regions. Among the three examined *Cenchrus*, *C. longispinus* is more similar to *C. echinatus* (99.9% pairwise identity) than *C. purpureus* (97.3% pairwise identity). Compared to other examined species, the pairwise similarity of *C. longispinus* cpDNA is less than 95% (Table 1). The IR-LSC junction locates in the intergenic space (IGS) between *rps19* and *rpl22*, whereas the SSR-IR border is in the *ndhF* coding region of *Cenchrus* and related taxa (Table 1). There are 56 SSRs and 17 forward repeats in the chloroplast genome of *C. longispinus* (Tables S6 and S7). Most of SSRs are composed of A and T nucleotides. Additionally, locations of these SSRs are in non-coding regions, except ten SSRs that were found in *infA, ndhD, ndhH, ndhK, rps14, rpl22, rpoC2, psbC*, and *ycf5* (Table S6). Similarly, 17 forward repeats

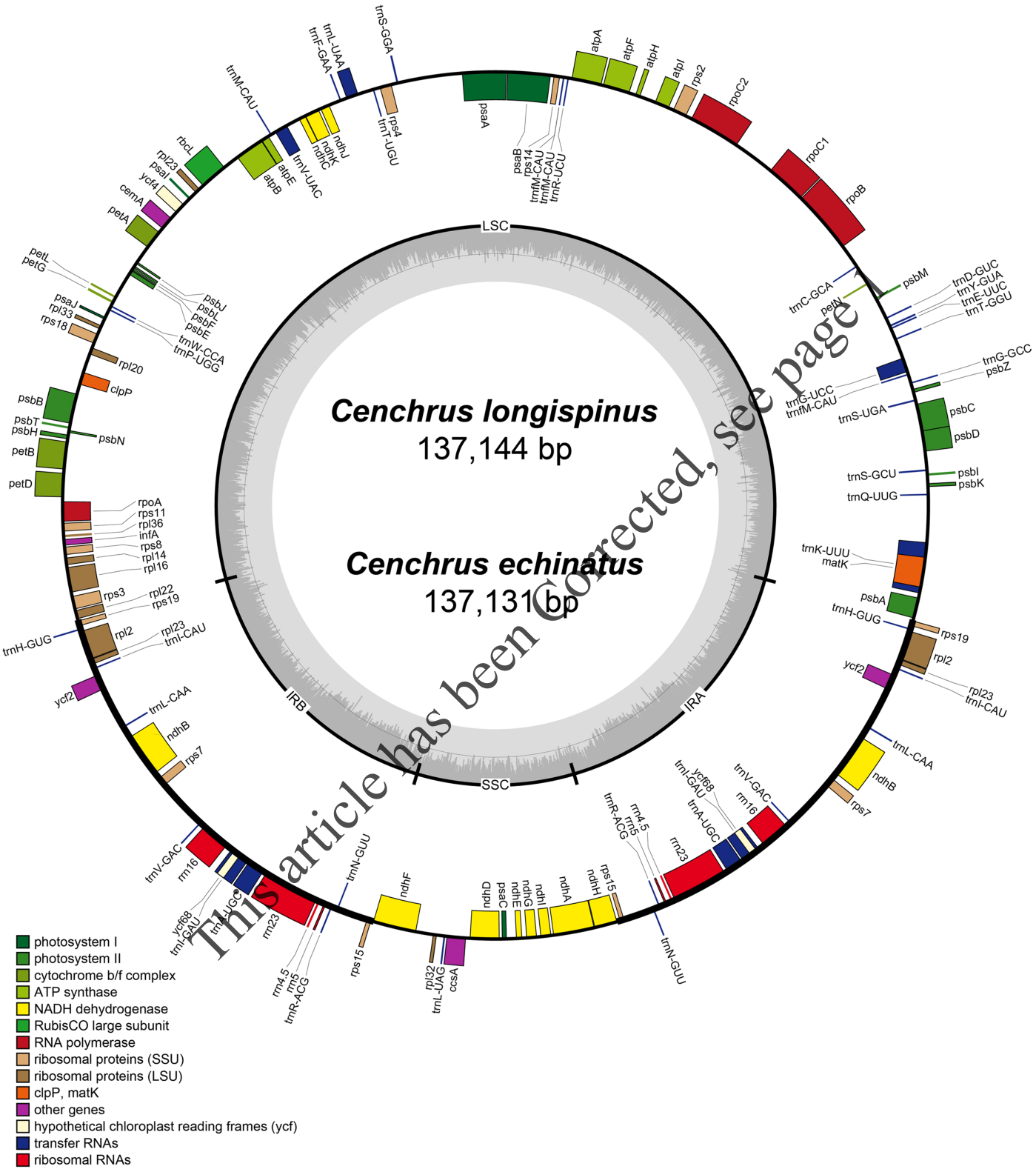

Figure 2 **The map of chloroplast genomes of *Cenchrus*.** The genes inside of the circle are transcribed clockwise, whereas the genes outside of the circle are transcribed counterclockwise. LSC, large single copy; SSC, small single copy; IRA-IRB, inverted repeat regions.

Full-size DOI: 10.7717/peerj.7965/fig-2

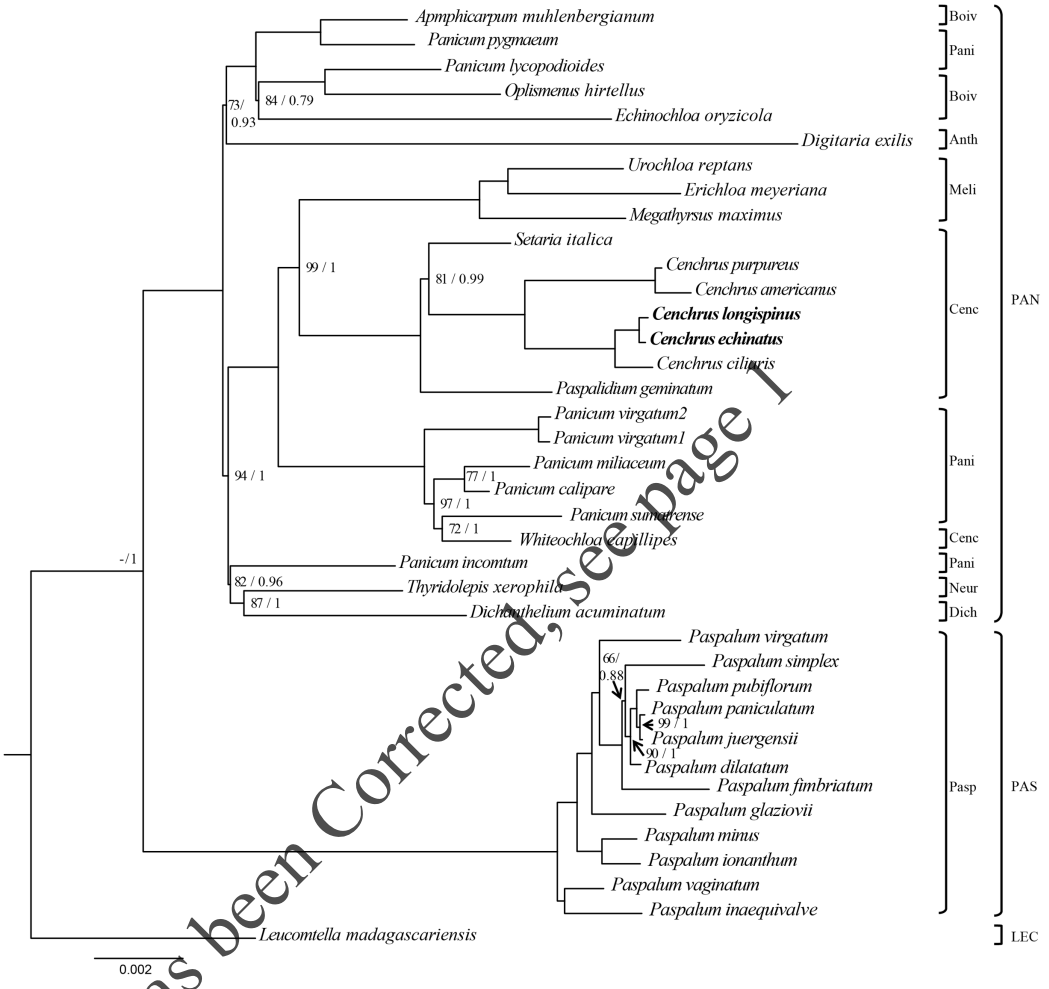

Figure 3 The Maximum likelihood tree of *Cenchrus* and related taxa inferred from 75 chloroplast genes. The numbers mean supporting values (Bootstrap (BP)/ Posterior probability (PP)). Only supporting values under (BP = 100/PP = 1) were shown. The dash means no value. LEC, Lecomtelleae; PAN, Paniceae; PAS, Paspaleae; Pasp, Paspalinae; Anth, Anthephorinae; Boiv, Boivinellinae; Neur, Neurachninae; Dich, Dichantheliinae; Pani, Panicinae; Meli, Melinidinae; Cenc, Cenchrinae.

Full-size DOI: 10.7717/peerj.7965/fig-3

locate in both non-coding (13 sites) and coding regions (five sites, Table S7). Although *C. longispinus* is highly similar to *C. echinatus*, the number and length of repeats are different between two species (Tables S6 and S7).

Phylogenomics analysis of *Cenchrus* and related taxa

The ML and BI analysis generated the same topology of phylogenetic trees, of which the monophyly of *Cenchrus* was revealed with high supporting value (Fig. 3). Although the tribe Paniceae is monophyletic, some subtribes are polyphyletic. Specifically, *Whiteochloa capillipes* located within subtribe Paniceae although it is a member of subtribe *Cenchrinae*. Another polyphyletic group is subtribe Panicinae of which *Panicum pygmaeum* and *Panicum lycopodioides* from a clade with the species of subtribe

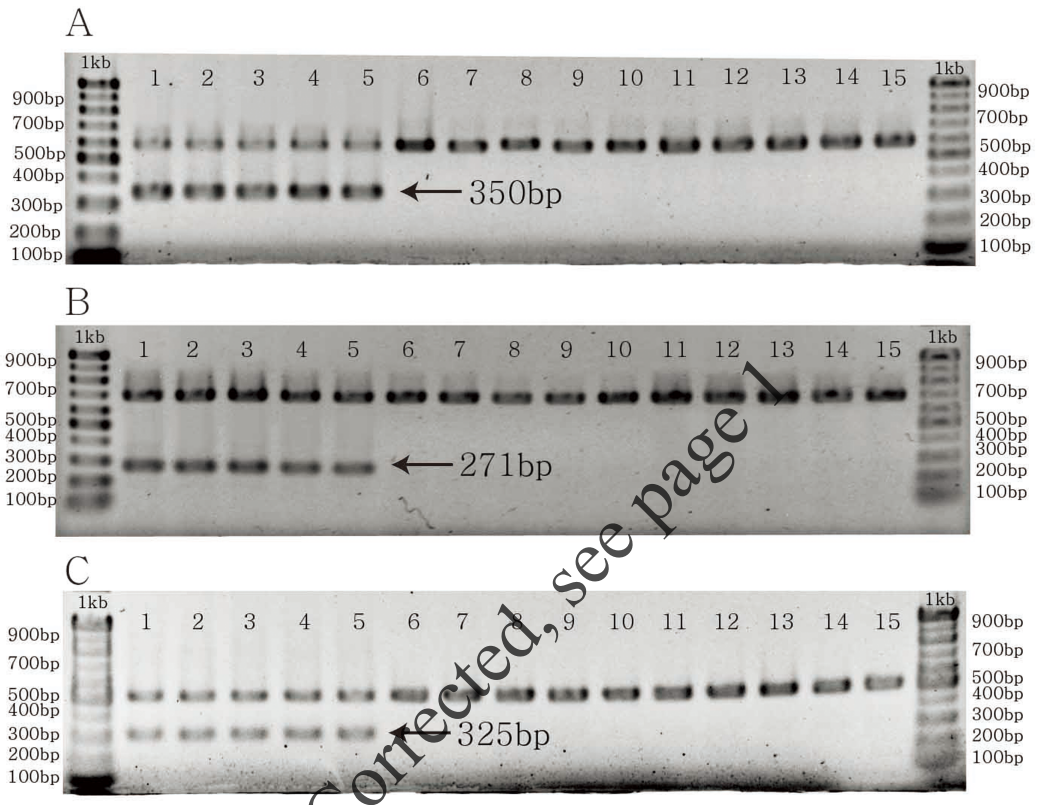

Figure 4 The PCR results of specific primer pairs for *Cenchrus longispinus*. (A) The specific primer pairs for *atpB* gene; (B) the specific primer pairs for *matK* gene; (C) the specific primer pairs for *ndhD* gene. The number from one to five: *Cenchrus longispinus*; from six to 10: *Cenchrus echinatus*; from 11 to 15: *Pennisetum alopecuroides*. Full-size DOI: 10.7717/peerj.7965/fig-4

Boivinellinae. Additionally, *Panicum incomtum* is sister to *Thyridopepsis xerophila* (subtribe Neurachninae) and *Dichanthelium acuminatum* (subtribe Dichantheliinae). The three partition data (LSC, SSC, and IR regions) resulted in different topologies of phylogenetic trees (Figs. S2–S4). However, the relationships of *Cenchrus* and related taxa inferred from LSC and SSC region data sets are quite similar to those of 75 protein-coding genes, except for *Paspalidium geminatum* and *Digitaria exilis*. In contrast, the IR region-based phylogenetic tree exhibits different relationships in tribe Paniceae compared to other data sets (Fig. S3). However, the monophyly of *Cenchrus* was found in all data sets with strong supporting values.

Specific SNP-based molecular markers for *Cenchrus longispinus*

The specific primer pairs for *C. longispinus* inferred from *atpB, matK*, and *ndhD* resulted in bands of 350, 271, and 325 bp, respectively (Fig. 4). There are no bands for *C. echinatus* and *Pennisetum alopecuroides*, except the control bands of 579, 686, 518 bp from *atpB*, *matK*, and *ndhD*, respectively. The primer pairs are also effective in other examined species from Korea, the USA, and other countries (Table S4; Fig. S1).

DISCUSSION

The chloroplast genomes of *Cenchrus* and phylogenomic implication

The chloroplast genomes of *Cenchrus* are highly conserved regards structure, gene content and order in comparison with other species in tribe Paniceae (Table 1). Previously, the inversion event of 28 kb-region was reported in the grass family (*Doyle et al., 1992*). Additionally, *Huang et al. (2017)* suggested a model for rearrangement of LSC in the Poaceae. The process includes two inversion events of which the first inversion of 28 kb-region was followed by a second inversion between *trnS-GCU* and *trnT-GGU*. Besides inversions, the loss of intron was also recorded in Poaceae (*Huang et al., 2017*). In *Cenchrus* and other members of tribe Paniceae, the cpDNA also have the inversions and loss of intron in *rpoC1* (Table 1).

In the chloroplast genome, the IR regions can be contracted or expanded, which resulted in various boundaries among LSC, SSC, and IR regions of different green plants (*Wang et al., 2008*; *Downie & Jansen, 2015*; *Wang et al., 2015*; *Raman & Park, 2016*; *Cauz-Santos et al., 2017*). Notably, the loss of IR region was found to be reversed in legumes (*Choi, Jansen & Ruhlman, 2019*). In the previously published cpDNA of Poaceae, the loss of IR regions was not found and there are different LSC-IR-SSC junctions. However, the IR-LSC junction is quite stable, whereas there is more variable in SSC-IR boundary (*Huang et al., 2017*). Both IR-LSC junction and SSC-IR border are stable at the tribal and genus level in Poaceae. For example, in tribe Oryzeae, the IR-LSC junction is in the IGS between *rpl22* and *rps19* whereas the SSC-IR border is in the coding region of *ndhH* (*Zhang et al., 2016*). A similar trend was also found in *Festuca* and *Lolium* (tribe Poeae; *Qiu et al., 2019*). In this study, the junctions among *Cenchrus* and related taxa in tribe Paniceae are similar (the IR-LSC and IR-SSC junctions in the IGS of *rps19- rpl22*, and the *ndhF*, respectively), which provides more evidence to support the structural stability of junctions at tribal and genus level in Poaceae (Table 1). However, the available complete cpDNA have not covered all tribes and genera in Poaceae. Also, the study in evolution of chloroplast genomes, which includes the representatives of all subfamilies, tribes, and genera in Poaceae, has not been conducted. Therefore, further studies that cover more samples in Poaceae should be conducted to trace the structural evolution of chloroplast genomes in Poaceae, an ecologically and economically important family among green plants.

The phylogenomic analysis inferred from 75 protein-coding regions highly support the monophyly of *Cenchrus* (Fig. 3). Previously, different studies revealed the close relationship between *Cenchrus* and *Pennisetum* and the merge of *Pennisetum* to *Cenchrus* was also suggested (*Chemisquy et al., 2010*; *Veldkamp, 2014*). In this study, the relationship between *Pennisetum* and *Cenchrus* was not checked due to the lack of complete cpDNA sequences of *Pennisetum*. Therefore, further studies on the comparative genomics of these two genera should be conducted to give a better understanding of their evolutionary histories in Poaceae.

In contrast to the monophyly of *Cenchrus*, *Panicum* species is polyphyletic (Fig. 3). In previous studies, *Panicum* is sister to *Whiteochloa capillipes* (*Burke et al., 2016*;

Saarela et al., 2018). As a result, *Saarela et al. (2018)* suggested transferring *Whiteochloa* from subtribe Cenchrinae to subtribe Panicinae. In this study, *Panicum* species forms a clade with not only *Whiteochloa* but also other species in subtribes Boivinellinae, Neurachninae, and Dichantheliinae (Fig. 3). *Zuloaga, Salariato & Scataglini (2018)* provided a deeper understanding of the phylogeny of *Panicum* and suggested a new circumscription of *Panicum*, including seven sections, based on *ndhF* data. However, the sample of *Panicum incomtum* was not included in their data. In the present study, the ML and BI analysis revealed a close relationship between *Panicum incomtum, Thyridopepsis xerophila* (subtribe Neurachninae), and *Dichanthelium acuminatum* (subtribe Dichantheliinae, Fig. 3), which was not shown in the study of *Zuloaga, Salariato & Scataglini (2018)*. Therefore, more samples and molecular data of *Panicum* should be covered to clarify its relationship to other members of tribe Panicinae in further studies.

Previously, partition data inferred from LSC, SSC, and IR regions were used to test the phylogenetic utility of different regions (*Li et al., 2013*; *Evans, Joshi & Wang, 2019*). It is clear that the incongruent relationships resulted from different partition data sets. In the present study, three partition data revealed different topologies compared to 75 protein-coding genes (Fig. 3; Figs. S2–S4). However, these data support the monophyly of *Cenchrus*, suggesting the sufficiency of the phylogenetic utility of different chloroplast genome regions at genus level in Poaceae.

The efficiency of SNP-based molecular markers for *Cenchrus longispinus*

A molecular marker is a useful tool for exploring genetic variations, identifying traits, and distinguishing species (*Hayward et al., 2015*; *Garrido-Cardenas, Mesa-Valle & Manzano-Agugliaro, 2018*). In Poaceae, various molecular markers have been developed (*Liu et al., 2012*; *Hammami et al., 2014*; *Gao, Jia & Kong, 2016*). For example, traits in important species of Poaceae (i.e., bamboo, and wheat) can be traced using the SNP markers (*Jiang, Zhang & Ding, 2013*; *Hu et al., 2015*; *Przewieslik-Allen et al., 2019*). Besides that, plastid data were used for species discrimination of *Stipa* (*Krawczyk et al., 2018*). Previously, although there are reports on the complete cpDNA sequences of *Cenchrus*, there have no studies on developing markers for *Cenchrus* (*Bhatt & Thaker, 2018*; *Wu & Zhou, 2018*). Being classified as an invasive alien plant, more data of *C. longispinus* should be collected. In this study, we employ the SNP sites in *C. longispinus* to develop molecular markers which can be used under the same PCR protocol (Table S3). These newly developed markers will be useful for identifying *C. longispinus* during its invasion in Korea. The information on invasive patterns of *C. longispinus* will be useful to make profound strategies for control its invasion in Korea. Previously, the hybrids between *Cenchrus* and related species were identified and characterized (*Read & Bashaw, 1974*; *Marchais & Tostain, 1997*; *Quiroga et al., 2013*). Therefore, the present genomic data will be essential for further identifying the hybrid between *C. longispinus* and related species. Addition to the SNP sites, the SSR data were used to develop molecular markers for different species in Poaceae (*Hodkinson, De Cesare & Barth, 2013*; *Jiang, Zhang & Ding, 2013*; *Hammami et al., 2014*). In this study, although the SSR markers were not developed,

the availability of SSR in the cpDNA of *C. longispinus* will provide fundamental information for further studies on both population genetics and SSR-based molecular markers in Poaceae.

CONCLUSIONS

This study provides the first complete chloroplast genome sequences of *C. longispinus* and *C. echinatus* which can be used for further studies on population genetics, phylogeny, and comparative genomics of *Cenchrus* in particular and the family Poaceae in general. Also, the newly developed molecular markers based on SNP sites and simple PCR protocol can be useful for the identification and management of the invasion of *C. longispinus* in Korea.

ACKNOWLEDGEMENTS

The authors would like to thank Korea National Arboretum (KNA), Gachon University Herbarium (GCU); The University of Texas at Austin Herbarium (TEX), The Queensland Herbarium (BRI) for supporting materials for this study. Also, we appreciate anonymous reviewers for helpful comments.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This research was supported by the funds ('Risk assessment and taxonomic characteristics analysis of putative invasive alien plants in Korea, 2017' and 'Development of molecular markers for putative invasive alien plants that threaten biodiversity based on NGS technology, 2018') by Research of Animal and Plant Quarantine Agency, South Korea. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Scientific research of Animal and Plant Quarantine Agency (2017, 2018), South Korea.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- JongYoung Hyun performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Hoang Dang Khoa Do performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Joonhyung Jung performed the experiments, analyzed the data, prepared figures and/or tables, approved the final draft.

- Joo-Hwan Kim conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

Field Study Permissions

The following information was supplied relating to field study approvals (i.e., approving body and any reference numbers):

The destruction of samples from specimens was approved by Korea National Arboretum (KNA), Gachon University Herbarium (GCU), The University of Texas at Austin Herbarium (TEX), and The Queensland Herbarium (BRI).

Data Availability

The following information was supplied regarding data availability:

Cenchrus longispinus and Cenchrus echinarus data are available at GenBank: MN078361 and MN078360.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.7965#supplemental-information.

REFERENCES

Anderson RL. 1997. Longspine Sanbur (*Cenchrus longispinus*) ecology and interference in irrigated Corn (*Zea mays*). *Weed Technology* **11(4)**:667–671 DOI 10.1017/S0890037X00043220.

Bhatt PP, Thaker VS. 2018. The complete chloroplast genome of *Cenchrus ciliaris* (Poaceae). *Mitochondrial DNA Part B* **3(2)**:674–675 DOI 10.1080/23802359.2018.1481795.

Burke SV, Wysocki WP, Zuloaga FO, Craine JM, Pires JC, Edger PP, Mayfield-Jones D, Clark LG, Kelchner SA, Duvall MR. 2016. Evolutionary relationships in Panicoid grasses based on plastome phylogenomics (Panicoideae; Poaceae). *BMC Plant Biology* **16(1)**:140 DOI 10.1186/s12870-016-0823-3.

Cauz-Santos LA, Munhoz CF, Rodde N, Cauet S, Santos AA, Penha HA, Dornelas MC, Varani AM, Oliveira GCX, Bergès H, Vieira MLC. 2017. The chloroplast genome of *Passiflora edulis* (Passifloraceae) assembled from long sequence reads: structural organization and phylogenomic studies in Malpighiales. *Frontiers in Plant Science* **8**:334 DOI 10.3389/fpls.2017.00334.

Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods in Molecular Biology* **1962**:1–14 DOI 10.1007/978-1-4939-9173-0_1.

Chemisquy MA, Giussani LM, Scataglini MA, Kellogg EA, Morrone O. 2010. Phylogenetic studies favour the unification of *Pennisetum*, Cenchrus and *Odontelytrum* (Poaceae): a combined nuclear, plastid and morphological analysis, and nomenclatural combinations in *Cenchrus*. *Annals of Botany* **106(1)**:107–130 DOI 10.1093/aob/mcq090.

Choi I-S, Jansen R, Ruhlman T. 2019. Lost and found: return of the inverted repeat in the legume clade defined by its absence. *Genome Biology and Evolution* **11(4)**:1321–1333 DOI 10.1093/gbe/evz076.

Christenhusz MJM, Byng JW. 2016. The number of known plants species in the world and its annual increase. *Phytotaxa* **261(3)**:201–217 DOI 10.11646/phytotaxa.261.3.1.

Christoph M. 2006–2010. *Phobos 3.3.11. Available at http://www.rub.de/ecoevo/cm/cm_phobos.htm*.

Cui C, Mei H, Liu Y, Zhang H, Zheng Y. 2017. Genetic diversity, population structure, and linkage disequilibrium of an association-mapping panel revealed by genome-wide SNP markers in sesame. *Frontiers in Plant Science* **8**:1189 DOI 10.3389/fpls.2017.01189.

Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* **14(7)**:1394–1403 DOI 10.1101/gr.2289704.

Downie SR, Jansen RK. 2015. A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Systematic Botany* **40(1)**:336–351 DOI 10.1600/036364415X686620.

Doyle JJ, Davis JI, Soreng RJ, Garvin D, Anderson MJ. 1992. Chloroplast DNA inversions and the origin of the grass family (Poaceae). *Proceedings of the National Academy of Sciences of the United States of America* **89(16)**:7722–7726 DOI 10.1073/pnas.89.16.7722.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**:11–15.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32(5)**:1792–1797 DOI 10.1093/nar/gkh340.

Evans DL, Joshi SV, Wang J. 2019. Whole chloroplast genome and gene locus phylogenies reveal the taxonomic placement and relationship of *Tripidium* (Panicoideae: Andropogoneae) to sugarcane. *BMC Evolutionary Biology* **19(1)**:33 DOI 10.1186/s12862-019-1356-9.

Fagaras M. 2018. *Cenchrus longispinus* (Hack) Fernald, one of the most aggressive alien plants on the romanian black sea coast. In: Finkl CW, Makowski C, eds. *Diversity in Coastal Marine Sciences*. Cham: Springer International Publishing, 383–395 DOI 10.1007/978-3-319-57577-3_22.

Fischer MC, Rellstab C, Leuzinger M, Roumet M, Gugerli F, Shimizu KK, Widmer A. 2017. Estimating genomic diversity and population differentiation – an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. *BMC Genomics* **18(1)**:69 DOI 10.1186/s12864-016-3459-7.

Gao L, Jia J, Kong X. 2016. A SNP-based molecular barcode for characterization of common wheat. *PLOS ONE* **11(3)**:e0150947 DOI 10.1371/journal.pone.0150947.

Garrido-Cardenas JA, Mesa-Valle C, Manzano-Agugliaro F. 2018. Trends in plant research using molecular markers. *Planta* **247(3)**:543–557 DOI 10.1007/s00425-017-2829-y.

Govaerts RHA. 2011. World checklist of selected plant families published update. Facilitated by the Trustees of the Royal Botanic Gardens, Kew.

Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* **47(W1)**:W59–W64 DOI 10.1093/nar/gkz238.

Hammami R, Jouve N, Soler C, Frieiro E, González JM. 2014. Genetic diversity of SSR and ISSR markers in wild populations of *Brachypodium distachyon* and its close relatives *B. stacei* and *B. hybridum* (Poaceae). *Plant Systematics and Evolution* **300(9)**:2029–2040 DOI 10.1007/s00606-014-1021-0.

Hand ML, Spangenberg GC, Forster JW, Cogan NOI. 2013. Plastome sequence determination and comparative analysis for members of the *Lolium-Festuca* grass species complex. *G3: Genes, Genomes, Genetics* **3(4)**:607–616 DOI 10.1534/g3.112.005264.

Hayward AC, Tollenaere R, Dalton-Morgan J, Batley J. 2015. Molecular marker applications in plants. In: Batley J, ed. *Plant Genotyping. Methods in Molecular Biology (Methods and Protocols).* Vol. 1245. New York: Humana Press.

Hodkinson TR, De Cesare M, Barth S. 2013. Nuclear SSR markers for *Miscanthus, Saccharum*, and related grasses (Saccharinae, Poaceae). *Applications in Plant Sciences* **1(11)**:1300042 DOI 10.3732/apps.1300042.

Hu X, Ren J, Ren X, Huang S, Sabiel SAI, Luo M, Nevo E, Fu C, Peng J, Sun D. 2015. Association of agronomic traits with SNP markers in Durum Wheat (*Triticum turgidum* L. durum (Desf.)). *PLOS ONE* **10(6)**:e0130854 DOI 10.1371/journal.pone.0130854.

Huang Y-Y, Cho S-T, Haryono M, Kuo C-H. 2017. Complete chloroplast genome sequence of common bermudagrass (*Cynodon dactylon* (L.) Pers.) and comparative analysis within the family Poaceae. *PLOS ONE* **12(6)**:e0179055 DOI 10.1371/journal.pone.0179055.

Jiang W-X, Zhang W-J, Ding Y-L. 2013. Development of polymorphic microsatellite markers for *Phyllostachys edulis* (Poaceae), an important bamboo species in China. *Applications in Plant Sciences* **1(7)**:1200012 DOI 10.3732/apps.1200012.

Jung SY, Lee JW, Shin HT, Kim SJ, An JB, Heo TI, Chung JM, Cho YC. 2017. *Invasive Alien plants in South Korea*. Pocheon: Korea National Arboretum.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Drummond A. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28(12)**:1647–1649 DOI 10.1093/bioinformatics/bts199.

Krawczyk K, Nobis M, Myszczyński K, Klichowska E, Sawicki J. 2018. Plastid super-barcodes as a tool for species discrimination in feather grasses (Poaceae: *Stipa*). *Scientific Reports* **8(1)**:1924 DOI 10.1038/s41598-018-20399-w.

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research* **29(22)**:4633–4642 DOI 10.1093/nar/29.22.4633.

Leaché AD, Oaks JR. 2017. The utility of single nucleotide polymorphism (SNP) data in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* **48(1)**:69–84 DOI 10.1146/annurev-ecolsys-110316-022645.

Li R, Ma P-F, Wen J, Yi T-S. 2013. Complete sequencing of five araliaceae chloroplast genomes and the phylogenetic implications. *PLOS ONE* **8(10)**:e78568 DOI 10.1371/journal.pone.0078568.

Liu H, Guo X, Wu J, Chen G-B, Ying Y. 2012. Development of universal genetic markers based on single-copy orthologous (COSII) genes in Poaceae. *Plant Cell Reports* **32(3)**:379–388 DOI 10.1007/s00299-012-1371-4.

Marchais L, Tostain S. 1997. Analysis of reproductive isolation between pearl millet (*Pennisetum glaucum* (L.) R.Br.) and *P. ramosum*, *P. schweinfurthii*, *P. squamulatum*, *Cenchrus ciliaris*. *Euphytica* **93**:97 DOI 10.1023/A:1002991721159.

Matsuoka Y, Yamazaki Y, Ogihara Y, Tsunewaki K. 2002. Whole chloroplast genome comparison of rice, maize, and wheat: implications for chloroplast gene diversification and phylogeny of cereals. *Molecular Biology and Evolution* **19(12)**:2084–2091 DOI 10.1093/oxfordjournals.molbev.a004033.

Pantoja PO, Simón-Porcar VI, Puzey JR, Vallejo-Marín M. 2017. Genetic variation and clonal diversity in introduced populations of *Mimulus guttatus* assessed by genotyping at 62 single nucleotide polymorphism loci. *Plant Ecology & Diversity* **10(1)**:5–15 DOI 10.1080/17550874.2017.1287785.

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

Przewieslik-Allen AM, Amanda JB, Wilkinson PA, Winfield MO, Shaw DS, McAusland L, King J, King IP, Edwards KJ, Barker GLA. 2019. Developing a high-throughput SNP-based marker system to facilitate the introgression of traits from Aegilops species into bread wheat (*Triticum aestivum*). *Frontiers in Plant Science* **9**:1993 DOI 10.3389/fpls.2018.01993.

Qiu Y, Hirsch CD, Yang Y, Watkins E. 2019. Towards improved molecular identification tools in fine fescue (*Festuca* L., Poaceae) turfgrasses: nuclear genome size, ploidy, and chloroplast genome sequencing. *bioRxiv* 708149 DOI 10.1101/708149.

Quiroga M, Grunberg K, Ribotta A, López Colomba E, Carloni E, Tommasino E, Luna C, Griffa S. 2013. Obtaining sexual genotypes for breeding in buffel grass. *South African Journal of Botany* **88**:118–123 DOI 10.1016/j.sajb.2013.04.016.

Raman G, Park SJ. 2016. The complete chloroplast genome sequence of *Ampelopsis*: gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. *Frontiers in Plant Science* **7**:341 DOI 10.3389/fpls.2016.00341.

Read JC, Bashaw EC. 1974. Intergeneric hybrid between pearl millet and buffelgrass. *Crop Science* **14(3)**:401–403 DOI 10.2135/cropsci1974.0011183X001400030018x.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61(3)**:539–542 DOI 10.1093/sysbio/sys029.

Saarela JM, Burke SV, Wysocki WP, Barrett MD, Clark LG, Craine JM, Peterson PM, Soreng RJ, Vorontsova MS, Duvall MR. 2018. A 250 plastome phylogeny of the grass family (Poaceae): topological support under different data partitions. *PeerJ* **6(10)**:e4299 DOI 10.7717/peerj.4299.

Soltani N, Kumagai M, Brown L, Sikkema PH. 2010. Long-spine sandbur [*Cenchrus longispinus* (Hack. in Kneuck.) Fernald] control in corn. *Canadian Journal of Plant Science* **90(2)**:241–245 DOI 10.4141/CJPS09132.

Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Teisher JK, Clark LG, Barberá P, Gillespie LJ, Zuloaga FO. 2017. A worldwide phylogenetic classification of the Poaceae (Gramineae) II: an update and a comparison of two 2015 classifications. *Journal of Systematics and Evolution* **55(4)**:259–290 DOI 10.1111/jse.12262.

Strat D, Stoyanov S, Holobiuc I. 2017. The occurrence of the alien plants *species Cenchrus longispinus* on the Danube Delta Shore (North West Black Sea Coast)-threats and possible impacts on the local biodiversity. *Acta Horti Botanici Bucurestiensis* **44**:17–31.

Sugiura M. 1992. The chloroplast genome. *Plant Molecular Biology* **19(1)**:149–168 DOI 10.1007/BF00015612.

Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research* **44(W1)**:W232–W235 DOI 10.1093/nar/gkw256.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40(15)**:e115 DOI 10.1093/nar/gks596.

Veldkamp JF. 2014. A revision of *Cenchrus* incl. *Pennisetum* (Gramineae) in Malesia with some general nomenclatural notes. *Blumea - Biodiversity, Evolution and Biogeography of Plants* **59(1)**:59–75 DOI 10.3767/000651914X684376.

Verloove F, Sánchez Gullón E. 2012. A taxonomic revision of non-native *Cenchrus* s.str. (Paniceae, Poaceae) in the Mediterranean area. *Willdenowia* **42(1)**:67–75 DOI 10.3372/wi.42.42107.

Wang R-J, Cheng C-L, Chang C-C, Wu C-L, Su T-M, Chaw S-M. 2008. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evolutionary Biology* **8(1)**:36 DOI 10.1186/1471-2148-8-36.

Wang M, Cui L, Feng K, Deng P, Du X, Wan F, Weining S, Nie X. 2015. Comparative analysis of Asteraceae chloroplast genomes: structural organization, RNA editing and evolution. *Plant Molecular Biology Reporter* **33(5)**:1526–1538 DOI 10.1007/s11105-015-0853-2.

Wu Y, Zhou H. 2018. The complete chloroplast genome sequence of *Cenchrus purpureus*. *Mitochondrial DNA Part B* **4(1)**:51–52 DOI 10.1080/23802359.2018.1536451.

Zhang W. 2000. Phylogeny of the grass family (Poaceae) from *rpl16* intron sequence data. *Molecular Phylogenetics and Evolution* **15(1)**:135–146 DOI 10.1006/mpev.1999.0729.

Zhang D, Li K, Gao J, Liu Y, Gao L-Z. 2016. The complete plastid genome sequence of the wild rice *Zizania latifolia* and comparative chloroplast genomics of the rice tribe Oryzeae, Poaceae. *Frontiers in Ecology and Evolution* **4**:88 DOI 10.3389/fevo.2016.00088.

Zuloaga FO, Salariato DL, Scataglini A. 2018. Molecular phylogeny of *Panicum* s. str. (Poaceae, Panicoideae, Paniceae) and insights into its biogeography and evolution. *PLOS ONE* **13(2)**:e0191529 DOI 10.1371/journal.pone.0191529.