Biochemical Systematics and Ecology 69 (2016) 91–100

Contents lists available at ScienceDirect

Biochemical Systematics and Ecology

journal homepage: www.elsevier.com/locate/biochemsyseco

Studies on genetic diversity and phylogenetic relationships of limpograss (*Hemarthria altissima*) and related species based on combined chloroplast DNA intergenic spacer data

Jian Zhang [a, **, 1], Xiu Huang [b, 1], De-jun Huang [a], Yu Zhang [c], Lin-kai Huang [b, *], Lu Lu [d], Han-dong Yan [b]

[a] Chongqing Academy of Animal Sciences, Herbivorous Livestock Research Institute, Rongchang, Chongqing, 404100, China
[b] Department of Grassland Science, Animal Science and Technology College, Sichuan Agricultural University, Chengdu, Sichuan, 611130, China
[c] Institute of Agrifood Research and Technology (IRTA), Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Barcelona, 08193, Spain
[d] Department of Biochemistry and Molecular Biology, Biosciences Faculty, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Barcelona, 08193, Spain

ARTICLE INFO

Article history:
Received 25 April 2016
Received in revised form 27 July 2016
Accepted 30 July 2016

Keywords:
Chloroplast DNA
Hemarthria altissima (Poir.) Stapf et C. E. Hubb.
Hemarthria compressa (L. f.) R. Br.
Genetic diversity
Genetic structure
trnL-F
trnC-ycf6
psbC-trnS

ABSTRACT

Hemarthria R. Br. is a genus which includes important forage grasses. However, there is currently a lack of data analysis on the chloroplast DNA (cpDNA) of *Hemarthria* species. This study is to use three cpDNA intergenic spacers (*trnL-F*, *trnC-ycf6* and *psbC-trnS*) to obtain phylogenetic information in 36 *Hemarthria* samples including four *Hemarthria* species: *Hemarthria altissima* (Poir.) Stapf et C. E. Hubb., *Hemarthria compressa* (L. f.) R. Br., *Hemarthria uncinata* R. Br., and *Hemarthria japonica* (Hack.) Roshev. Data analysis revealed that non-significant genetic diversity existed in our samples, which was implied by nucleotide sequences information and the results of haplotypic and nucleotide diversity. The results of phylogenetic trees based on maximum likelihood (ML) and Bayesian inference (BI) revealed that *H. altissima* and *H. compressa* samples were not entirely distinct, suggesting that the two species share an intimate genetic relationship. A haplotype median-joining (MJ) network revealed broadly similar results to those derived from the ML and BI trees and implied that haplotype H3 may represent an ancient haplotype. Analysis of the population statistic F_{ST} revealed little genetic differentiation among the seven populations of *H. altissima* in Africa.

1. Introduction

The genus *Hemarthria* R. Br. in the Poaceae family consists of about 20 species that are geographically widely distributed. Among these species, the two that are the most important agriculturally and widely studied are *Hemarthria altissima* (Poir.)

* Corresponding author.
** Corresponding author.
E-mail addresses: zjturf@163.com (J. Zhang), huanglinkai@sicau.edu.cn (L.-k. Huang).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.bse.2016.07.024
0305-1978/

Stapf et C. E. Hubb. and *Hemarthria compressa* (L. f.) R. Br. *H. altissima*, also known as limpograss, is a warm-season perennial grass that is native to Africa (Yang et al., 2004; Newman et al., 2011). After being introduced to Florida, limpograss has been widely used by beef cattle producers as a summer forage plant in the southeastern part of the United States due to its high forage yield, high quality, and high tolerance of poorly drained soils (Newman et al., 2011). *H. compressa* (whipgrass), a relative of limpograss, is another particularly important species of *Hemarthria* and is mainly found in China. These two species have been extensively used and commercially grown in subtropical and tropical areas, and made important contributions to agro-animal husbandry ecosystem development and maintenance (Yang et al., 2004).

Despite their economic and ecological importance, there is a serious lack of genomic information available for *H. altissima* and *H. compressa*. Before the 21th century, studies on *H. altissima* and *H. compressa* were mainly focused on naturally occurring germplasm morphology (Schank et al., 1973), cytology (Quesenberry et al., 1982), stress resistance (Hudson, 1986), and breeding (Yang et al., 2004). In recent years, molecular studies, including the estimation of genetic diversity, construction of DNA fingerprints, and association of molecular markers with agronomically important traits, have been conducted on *H. altissima* and *H. compressa* using a variety of molecular markers. However, all of these markers have so far been based on nuclear DNA (Huang et al., 2008, 2014, 2014a; uang et al., 2014; Huang et al., 2014a, 2014b; Chen et al., 2011). To date, there have been no molecular research studies based on analysis of chloroplast DNA (cpDNA) in *Hemarthria*.

Outside of the cell nucleus, the cpDNA of higher plants represents an additional source of genetic information. The chloroplast genomes of most angiosperms exhibit a simple, circular structure with a low substitution rates and high conservation, and they are inherited in a non-Mendelian, uniparental fashion (Clegg et al., 1994). Currently, cpDNA sequence data is widely used to study the origin, evolution, and genetic diversity of natural populations of the majority of important plant species (Juszczak et al., 2012). Although cpDNA is conserved in plants, non-coding and coding regions differ greatly in the evolution rate (Yan et al., 2015). The non-coding regions are more hypervariable and are richer in variations such as substitutions, translocations, inversions, insertions, and deletions (Katayama et al., 2012). These represent ideal segments for interspecific phylogenetic studies at low taxonomic levels, such as hybrid cultivar identification, reconstruction of genetic relationships of plants species, and intraspecific phylogeographic studies (Bakker et al., 1999; Kimura et al., 2003).

Previous molecular research on *Hemarthria* based on DNAmarkers revealed distinctive genetic differences and extensive genetic diversity in wild plant resources. These studies also suggested a close genetic relationship between *H. altissima* and *H. compressa*. However, in-depth analysis of the relationship between these two significant *Hemarthria* species is lacking. Therefore, in this study, we sought to take advantage of the virtues of cpDNA non-coding regions by assessing three cpDNA intergenic spacers (*trnL-F*, *trnC-ycf6*, and *psbC-trnS*) in 36 *Hemarthria* germplasm resources (including 26 *H. altissima*, 8 *H. compressa*, 1 *Hemarthria uncinata* R. Br., and 1 *Hemarthria japonica* (Hack.) Roshev.). Our aims were mainly to (1) estimate the genetic diversity of the 36 *Hemarthria* samples based on cpDNA sequence variation, (2) explore phylogenetic relationships between *H. altissima* and *H. compressa*, and (3) reveal the genetic structure of African *H. altissima*. Our results serve to enrich the existing knowledge on the genetic diversity of *H. altissima* and *H. compressa*, providing useful information for the development of conservation strategies for *Hemarthria* and for future phylogenetic and phylogeographic studies of *Hemarthria* species.

2. Material and methods

2.1. Plant materials

In this study, 26 *H. altissima* samples, one *H. compressa*, and one *H. uncinata* included, were collected from the USDA Germplasm Resources Information Network (GRIN) program in October 2012 (Huang et al., 2014). Another seven *H. compressa* samples and one H. japonica sample were obtained from the Sichuan Agricultural University (Yaan), Sichuan, China. Detailed information on these 36 *Hemarthria* samples is summarized in Table 1. Unfortunately, records of two samples were ambiguous. One sample was described as deriving from 'South Africa', which could refer to either the Limpopo or KwaZulu-Nata region. Another source was recorded as simply 'Japan', which may refer to the city of Yokohama. All plant samples were maintained as rhizomes. Fresh and young leaf tissues from each sampled individual clone were harvested and dried at room temperature in collection bags with color silica gel.

2.2. DNA extraction

Dry and young leaf tissues were first ground by a Tissue Lyser (Grinder, Beijing, China) and were later used to be extracted genomic DNA using a Plant Genomic DNA Kit (Tiangen, Beijing, China) in accordance with the manufacturer's directions. DNA quality was analyzed by 0.8% (w/v) agarose gel electrophoresis, and DNA concentration was quantified by NanoDrop 2000 spectrophotometer (Thermo, USA). The DNA stock was diluted to a working concentration of 20 ng/μL and stored at −20 °C.

2.3. PCR amplification and sequence determination

A set of universal cpDNA primers were used to amplify the target regions of *trnL-F* (5′-CGAAATCGGTAGACGCTACG-3′, 5′-ATTTGAACTGGTGACACGAG-3′), *trnC-ycf6* (5′-CCAGTTCAAATCTGGGTGTC-3′,5′-CATTAAAGCAGCCCAAGC-3′) (Demesure et al., 1995) and *psbC-trnS* (5′-GGTCGTGACCAAGAAACCAC-3′, 5′-GGTTCGAATCCCTCTCTCTC-3′) (Murakami et al., 2006). PCR was

Table 1
Sources of the 36 *Hemarthria* samples.

No.	Code	Accession	Plantid	Materials origin	Species	Types
1	H201	PI 299039	H.A.1	Africa, Zimbabwe	*Hemarthria altissima*	Wild materials
2	H228	PI 379613	–	Africa, Zimbabwe	*H. altissima*	Wild materials
3	H202	PI 299993	Redalta	South Africa, Transvaal	*H. altissima*	Cultivar
4	H204	PI 299995	Bigalta	South Africa, Transvaal	*H. altissima*	Cultivar
5	H208	PI 349754	–	South Africa, Transvaal	*H. altissima*	Wild materials
6	H230	PI 409751	3009	South Africa, Transvaal	*H. altissima*	Wild materials
7	H203	PI 299994	Greenalta	South Africa, Limpopo	*H. altissima*	Cultivar
8	H206	PI 349751	–	South Africa, Limpopo	*H. altissima*	Wild materials
9	H207	PI 349752	–	South Africa, Limpopo	*H. altissima*	Wild materials
10	H225	PI 364890	1517	South Africa, Limpopo	*H. altissima*	Wild materials
11	H231	PI 410128	1799	South Africa, Limpopo	*H. altissima*	Wild materials
12	H232	PI 410129	1806	South Africa, Limpopo	*H. altissima*	Wild materials
13	H233	PI 410133	1810	South Africa, Limpopo	*H. altissima*	Wild materials
14	H205	PI 349750	–	South Africa, KwaZulu-Nata	*H. altissima*	Wild materials
15	H211	PI 364863	1036	South Africa, KwaZulu-Nata	*H. altissima*	Wild materials
16	H218	PI 364876	1200	South Africa, KwaZulu-Nata	*H. altissima*	Wild materials
17	H220	PI 364878	1265	South Africa, KwaZulu-Nata	*H. altissima*	Wild materials
18	H221	PI 364881	1284	South Africa, KwaZulu-Nata	*H. altissima*	Wild materials
19	H234	PI 410134	1887	South Africa, KwaZulu-Nata	*H. altissima*	Wild materials
20	H235	PI 410137	2067	South Africa, KwaZulu-Nata	*H. altissima*	Wild materials
21	H236	PI 410138	2068	South Africa, KwaZulu-Nata	*H. altissima*	Wild materials
22	H213	PI 364868	1077	South Africa	*H. altissima*	Wild materials
23	H227	PI 365145	1697	South Africa, Swaziland	*H. altissima*	Wild materials
24	H237	PI 410139	2460	South Africa, Cape Province	*H. altissima*	Wild materials
25	H239	PI 413186	–	East Africa, Mauritius	*H. altissima*	Wild materials
26	H240	PI 508606	385–79	South America, Argentina	*H. altissima*	Wild materials
27	H241	PI 404118	Line607	Japan	*Hemarthria compressa*	Wild materials
28	H037	–	–	China, Sichuan, Daxian	*H. compressa*	Wild materials
29	H021	–	–	China, Sichuan, Mianyang	*H. compressa*	Wild materials
30	Guangyi	–	Guangyi	China, Sichuan, Guangyi	*H. compressa*	Cultivar
31	Yaan	–	Yaan	China, Sichuan, Ya'an	*H. compressa*	Cultivar
32	Chonggao	–	Chonggao	China, Chongqing	*H. compressa*	Cultivar
33	H046	–	–	China, Guizhou, Libo	*H. compressa*	Wild materials
34	H051	–	–	China, Yunnan, Qiaojia	*H. compressa*	Wild materials
35	H242	PI 400272	–	Australia, New South Wales	*Hemarthria uncinata*	Wild materials
36	H244	–	–	China, Heilongjang, meadow of the Songnen Plain	*Hemarthria japonica*	Wild materials

carried out in a total volume of 50 μL, containing 6 μL DNA template (20 ng/μL), 2 μL each forward and reverse primers (10 pmol/μL), 25 μL Premix *Taq* (TakaRa *Taq* Version 2.0 plus dye; TakaRa Bio Inc., China), and 15 μL distilled water. PCR cycling conditions observed the following settings: 5 min pre-denaturation at 94 °C; 35 cycles of 1 min denaturation at 94 °C, 1 min annealing at 62–65 °C, and 1 min extension at 72 °C; and a final 10 min extension at 72 °C. PCR products were detected on 1.5% agarose gels at 130 V for 30 min and visualized under UV light. Appropriately sized PCR products were then purified by gel extraction and sent to the Beijing Genomics Institute for bi-directional gene sequencing.

2.4. Data analysis

Obtained sequences of the 25 African *H. altissima* samples or all samples pertaining to the three cpDNA intergenic spacers were respectively spliced using DNASTAR SeqMan and then manually adjusted (Swindell and Plasterer, 1997). The multiple splice sites of each cpDNA intergenic spacer sequence were aligned using the ClustalW algorithm in MEGA 5.0 and improved by visual inspection (Yan et al., 2015). A BLASTN search (http://blast.ncbi.nlm.nih.gov) was then performed using default settings in order to identify homologous sequences in other plants. To assess levels of genetic variation, we estimated the number of haplotypes (h), haplotype diversity (H_d), and nucleotide diversity (π) of each cpDNA intergenic spacer in the 36 *Hemarthria* samples using DnaSP 5.1 software (Librado and Rozas, 2009).

To determine whether the aligned sequences of each cpDNA intergenic spacer evolved neutrally in the 25 African *H. altissima*, we performed neutrality tests, calculating Tajima's D, Fu and Li's F^*, and Fu and Li's D^* using DnaSP 5.1 following the method published by Batnini et al. (2014). Next, mismatch distribution was assessed to identify historical processes such as population equilibrium and recent demographic expansion, which are characterized by a multimodal distribution and unimodal distribution, respectively (Hu et al., 2011). These were implemented in DnaSP 5.1 and Arlequin 3.1, respectively (D'Aloia et al., 2015). These two programs were used to draw plots of mismatch distribution based on different theories.

Since a single cpDNA intergenic spacer sequence does not provide sufficient phylogenetic information, we decided to combine the three cpDNA sequence data sets using SequenceMatrix 1.7 software for subsequent phylogenetic analysis (Vaidya et al., 2011). Before doing this, we investigated whether the three cpDNA sequence data sets were appropriate for pooling based on congruence between sequence datasets (Dolphin et al., 2000). The incongruence length difference (ILD) test

was implemented by PAUP 4.0 (Swofford, 1998). The likelihood heterogeneity values in this study were not significant at $P = 0.10$, indicating no conflicting phylogenetic signals among the three cpDNA intergenic spacers. Hence, we combined the data from the *trnL-F*, *trnC-ycf6,* and *psbC-trnS* regions.

Combined data were then tested for substitutional saturation in order to detect whether the sequence data were suited for phylogenetic reconstruction. We applied Xia's test by calculating the saturation parameters I_{ss} and $I_{ss.c}$ using DAMBE 4.5 (Xia and Lemey, 2009). In order to provide a more intuitive assessment, substitution saturation was also investigated by plotting the observed divergence against the corrected pairwise divergence calculated by applying the evolutionary model (GTR) using DAMBE 4.5 (Xia and Lemey, 2009).

Phylogenetic analyses on the combined data of all materials were performed according to the Maximum likelihood (ML) approach using PhyML 3.1 (Guindon et al., 2009) and the Bayesian inference (BI) approach with MrBayes 3.2 (Huelsenbeck and Ronquist, 2001). The best-fitting nucleotide substitution models of sequence evolution with the lowest Akaike information criterion (AIC) scores were determined by jModelTest 2.1 (Posada, 2008). Clade credibility of the ML phylogram was assessed with 100 bootstrap replicates. For BI analysis, default priors were applied to each analysis using four Markov chain Monte Carlo (MCMC) chains that were run for 2,000,000 generations with sampling frequency assigned at every 100 trees. Node support was evaluated with posterior probabilities of the tree topology, and parameter values were summarized from 20,000 trees (burn-in of 5000 trees).

To evaluate which ML and BI tree topologies were optimal, several testing procedures were performed, including observed log-likelihood difference (Obs), approximately unbiased (AU), bootstrap probability (BP), approximate Bayesian posterior probability (PP), Shimodaira-Hasegawa (SH) test, Kishino-Hasegawa (KH) test, weighted Kishino-Hasegawa (WKH), and weighted Shimodaira-Hasegawa (WSH) test using CONSEL 0.20 with default settings (Shimodaira and Hasegawa, 2001). CONSEL calculated *p*-values from the output (the site-wise log-likelihoods for the trees) of the software package TREE-PUZZLE (Schmidt et al., 2002).

To infer relationships among haplotypes of the 36 *Hemarthria* samples, the phylogenetic median-joining (MJ) network algorithm based on parsimony criteria was performed by Network 4.6 software (Forster et al., 2004). The contribution to genetic variation and parameter of genetic differentiation (F_{ST}) of populations of the 25 African *H. altissima* samples based on chloroplast sequences were computed in Arlequin 3.5 (Excoffier and Lischer, 2010).

3. Results

3.1. Chloroplast DNA sequence analysis of 36 Hemarthria samples

3.1.1. cpDNA sequence polymorphisms

Three cpDNA intergenic spacers, *trnL-F*, *trnC-ycf6,* and *psbC-trnS,* were amplified and sequenced independently in 36 *Hemarthria* samples. After visual inspection and adjustment, the total lengths of the aligned *trnL-F*, *trnC-ycf6,* and *psbC-trnS* sequences were 803 bp, 862 bp, and 539 bp, respectively. Though the *psbC-trnS* sequences were incomplete, they contain some parsimony-informative sites, so we decided to add the sequence data in subsequent analyses. We conducted a BLASTN search with above sequences, afterwards findingthat the *trnL-F* sequence exhibited strong homology (approximately 99% identity) to known sequences of other *Hemarthria* species (*Hemarthria pratensis* and *Hemarthria longiflora*). The *trnC-ycf6* sequence showed highhomology (97% identity) to sequences of the awn species *Miscanthus sinensis*, while the incomplete *psbC-trnS* sequence similarly showed high homology (99% identity) to the complete *psbC-trnS* sequences of *M. sinensis* and *Miscanthus* × *giganteus*. These results suggest that the sequences obtained in our study indeed belong to *Hemarthria*, and that among the data in NCBI's GenBank, other *Hemarthria* sequences (those of *H. pratensis* and *H. longiflora*) or those belonging to *Miscanthus* are the most similar to our *Hemarthria* samples. This highlights the fact that the three cpDNA intergenic spacers are conserved between *Hemarthria* and related species of the genus *Miscanthus*.

As summarized in Table 2, we obtained 803 bp of *trnL-F* and 862 bp *trnC-ycf6* sequence, which included 789 and 845 invariable sites, 14 and 17 variable sites, and four and three parsimony-informative sites, respectively. The incomplete, 539 bp *psbC-trnS* intergenic spacer consisted of 532 invariable sites, seven variable sites, and two parsimony-informative sites. The GC contents of the amplified sequences were 0.312 for *trnL-F*, 0.359 for *trnC-ycf6*, and 0.441 for *psbC-trnS*. The *trnC-ycf6* sequence exhibited that highest rate of polymorphism (1.97%) with the most haplotypes (11). Although *trnL-F* exhibited the second-highest rate of polymorphism (1.74%) with the second most haplotypes (9), it demonstrated the greatest haplotype

Table 2
Chloroplast DNA sequence polymorphisms in the 36 *Hemarthria* samples.

cpDNA intergenic spacer	*NS*	*IS*	*VS*	*PIS*	*GCc*	*h*	H_d	π ($\times 10^{-2}$)
trnL-F	803	789	14	4	0.312	9	0.575	0.151
trnC-ycf6	862	845	17	3	0.359	11	0.522	0.131
psbC-trnS	539	532	7	2	0.441	7	0.354	0.092
trnL-F/trnC-ycf6/psbC-trnS	2204	2166	38	9	0.362	21	0.851	0.129

Note: *NS* = number of sites; *IS* = invariable sites; *VS* = variable sites (no insertions/deletions); *PIS* = parsimony informative sites; *GCc* = GC content; *h* = number of haplotypes; H_d = haplotype diversity; π = nucleotide diversity.

diversity (0.579) and nucleotide diversity (0.00151). The incomplete *psbC-trnS* sequence exhibited the lowest rate pf polymorphism (1.30%) with the fewest haplotypes (7), the lowest haplotype diversity (0.354), and the lowest nucleotide diversity (0.00092). All haplotype sequences had been submitted to NCBI's GenBank, the accession numbers were from KX599423 to KX599454 (Supplementary Table 1).

According to the ILD test, no conflicting phylogenetic signal was discovered among the three cpDNA intergenic spacers, so we combined the three sequence datasets for follow-up analyses. The total sequence length of the aligned, combined data was 2204 bp, including 2166 invariable sites and 38 variable sites. Among these variable sites, we explore nine parsimony-informative sites, including eight sites with two variants and one site with five variants. The overall GC content was 0.362, and a total of 21 haplotypes were presented. The average values of haplotype diversity (H_d) and nucleotide diversity (π) were 0.851 and 0.00129, respectively (Table 2).

3.1.2. Test for substitution saturation

In order to detect whether the sequence data were suited for phylogenetic reconstruction, the combined sequence data were tested for substitutional saturation applying Xia's test by calculating the saturation parameters I_{ss} and $I_{ss.c}$ using DAMBE 4.5. Once the parameter $I_{ss.c}$ for a set of sequences was obtained, the I_{ss} value would be calculated and compared. With a I_{ss} vule less than $I_{ss.c}$ value, we deduced that the sequences experiencing little substitution saturation could be used for phylogenetic reconstruction. In this study, we obtained the average I_{ss} for subsets of 4, 8, 16, and 32 OTUs ($I_{ss} = 0.016$, 0.019, 0.023, and 0.027, respectively) and found that they were significantly less than the corresponding $I_{ss.c}$ values ($I_{ss.c} = 0.819$, 0.733, 0.638, and 0.526, respectively; $P = 0.000$) assuming a symmetrical topology. This means the combined sequence dataset of 36 *Hemarthria* samples is useful for phylogenetic analyses. A saturation plot (Supplementary Fig. 1a) showed a basic linear regression, further demonstrating that there was little substitution saturation.

3.1.3. Phylogenetic tree inference and topology comparison

To clarify the genetic relationships among the 36 *Hemarthria* samples, the combined sequences of the three cpDNA intergenic spacers were used to construct ML and BI phylogenetic trees. The log likelihood scores of 88 substitution models varied from 3318.4581 to 3417.8324, and the jModelTest program suggested that the best-fit model under AIC was TIM1+I (-lnL = 3318.8688). The obtained ML and BI trees were presented in Figs. 1 and 2, respectively. Both trees revealed the existence of three main groups of *Hemarthria* samples. Each of the two groups contained only one single sample each: *H. uncinata* sample H242 (Australia, New South Wales) and *H. japonica* sample H244 (China, Heilongjang, Songnen Plain meadow). The third group covered the remaining 34 samples belonging to *H. altissima* and *H. compressa*. From these results, we submitted an inference that hierarchical classification structure was mainly independent of species and geographic origin of the samples.

The ML and BI trees exhibited similar patterns, though they converged on different topologies. To further evaluate the optimal tree topology, *p*-values for the two trees were calculated using CONSEL. As seen in Supplementary Table 2, most *p*-values from the ML tree (AU, BP, PP, KH, SH, WKH) were larger than those of the BI tree, indicating that for our experimental data, the ML method was more appropriate than the BI method.

3.1.4. Median-joining (MJ) network analysis

The relationships among haplotypes in the 36 *Hemarthria* samples were further assessed by constructing a phylogenetic median-joining (MJ) network with the combined sequences of the three cpDNA intergenic spacers (*trnL-F*, *trnC-ycf6*, and *psbC-trnS*) (Fig. 3). According to the network tree, 21 haplotypes were analyzed with a distribution to the four species. Haplotype 3 (H3), containing 10 *H. altissima* (H201, H211, H213, H218, H220, H221, H228, H233, H234, and H236) and 4 *H. compressa* ('Chonggao', H046, H051, and H241) samples, was found to be the most ancestral haplotype. This revealed the close genetic relationship between *H. altissima* and *H. compressa*. The remaining 20 haplotypes were derived directly from the H3 haplotype. The two independent lineages on the ML and BI trees – *H. uncinata* sample H242 (haplotype H20) and *H. japonica* sample H244 (H21) – were segregated in the network from *H. altissima* and *H. compressa* by numerous mutations and were connected to H3 by median vector 2. Haplotypes H10 (*H. altissima* sample H206) and H12 (*H. altissima* sample H208) were also segregated and connected to H3 by MV1. In addition, haplotypes H6 (H202), H9 (H205), H11 (H207), and H16 (H235) were jointly connected to haplotype H14 (H227, H230, and H232). H13 (H225) was connected to H18 (H239), which was subsequently linked to H19 (H240). These results illustrated a broadly similar distribution of samples to those suggested by the ML and BI trees.

3.2. Genetic structure of African H. altissima

3.2.1. Genetic diversity and neutrality test

Twenty-five *H. altissima* coming from seven populations were collected from Africa: five populations from South Africa, one from Zimbabwe, and one from East Africa. We performed saturation analyses of the combined sequence dataset in the 25 *H. altissima* samples, obtaining an I_{ss} value of 0.0241, which was much less than the $I_{ss.c}$ (0.7951 assuming a symmetrical topology and 0.5796 assuming an asymmetrical topology). This indicated that the sequences experienced little substitution saturation, which was further confirmed by a linear regression in the saturation plot (Supplementary Fig. 1b). Next, we obtained the genetic diversity and performed neutrality tests for the cpDNA sequences of the 25 African samples

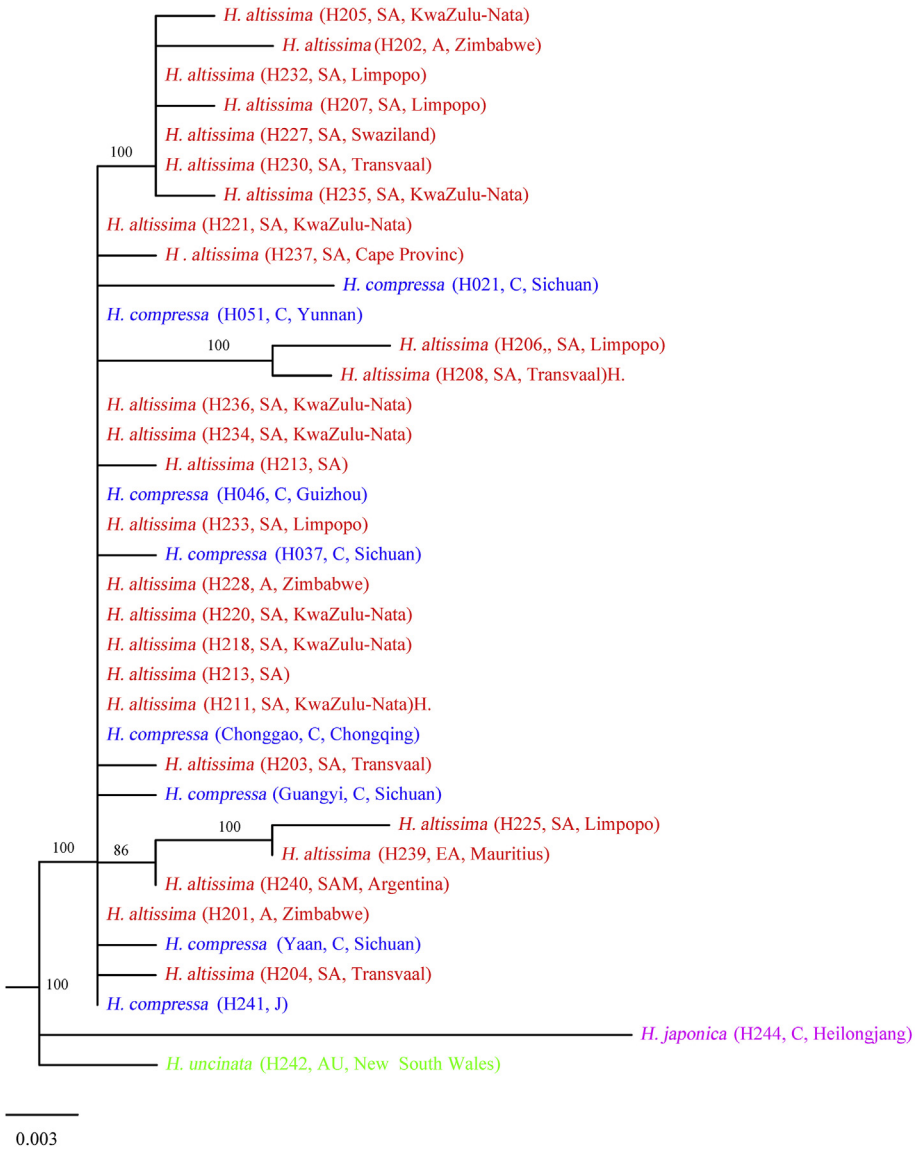

Fig. 1. Maximum likelihood (ML) tree based on the combined sequences of three cpDNA intergenic spacers (*trnL-F*, *trnC-ycf6*, and *psbC-trnS*) in 36 *Hemarthria* samples. Numbers above branches are bootstrap values, which were computed based upon 100 replicates. A: Africa; Au: Australia; SA: South Africa; EA: East Africa; C: China; SAM: South America; J: Japan.

(Supplementary Table 3). Statistical neutrality tests (Tajima's *D*, Fu and Li's F^*, and Fu and Li's D^*) of *trnC-ycf6* and the combined sequence data yielded highly significant ($P < 0.05$) negative values, but only the observed mismatch distribution of the combined sequence data was unimodal (Supplementary Fig. 2), closely fitting the expected distribution under a model of sudden expansion. For *trnL-F* and *psbC-trnS*, no distinct signal of population equilibrium or expansion was observed, as the neutrality tests were not significant and the mismatch distribution did not reflect multimodal or unimodal distributions (Supplementary Table 3, Supplementary Fig. 2).

3.2.2. Population genetic structure analysis

We calculated the F_{ST} value of the seven African populations of *H. altissima* using Arlequin. As described in Table 3, the F_{ST} was extremely low (0.0298, $P = 0.343$), indicating little genetic differentiation between these populations. In other words, the genetic structure between subpopulations was weak (Balloux and Lugon-Moulin, 2002). Moreover, a high percentage of total variance (97.02%) rooted in within-population differences, while only 2.98% of the variance resulted from differences between populations. This might be due to the strong capacity of *Hemarthria* to propagate asexually and its ecological adaptation to

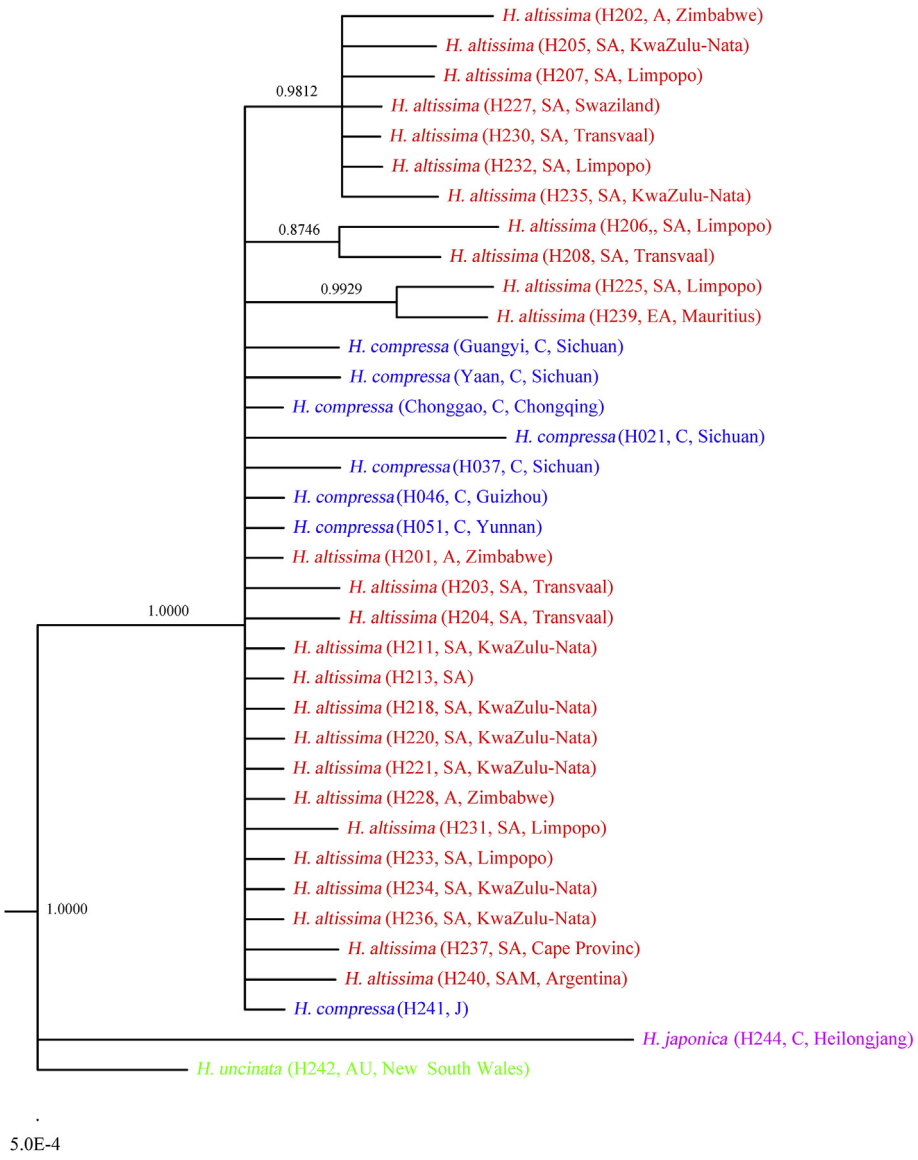

Fig. 2. Phylogenetic tree constructed from the combined data sequences of three cpDNA intergenic spacers (*trnL-F*, *trnC-ycf6*, and *psbC-trnS*) in 36 *Hemarthria* samples using Bayesian inference (BI). Numbers above branches are Bayesian posterior probability values. A: Africa; Au: Australia; SA: South Africa; EA: East Africa; C: China; SAM: South America; J: Japan.

disperse, which could lead to high gene flow between populations, generating genetic relatedness between samples with further geographic distant.

4. Discussion

This is the first report analyzing the genetic diversity of *Hemarthria* species using cpDNA molecular markers. Although cpDNA non-coding regions are generally hypervariable and rich in variation, the 36 *Hemarthria* samples exhibited relatively low genetic diversity of the cpDNA intergenic spacers *trnL-F*, *trnC-ycf6*, and *psbC-trnS* (H_d = 0.575, 0.522, and 0.354, respectively; π = 0.00151, 0.00131, and 0.00092, respectively). The combined sequence data also exhibited low haplotype diversity ($H_d = 0.851$) and nucleotide diversity ($\pi = 0.129$). This low level of genetic diversity resulted in low efficiency of the cpDNA markers for resolving *Hemarthria* genotypes or, in other words, poor detection of variable sites (the highest proportion of variable sites was only 2.0%). The proportion of variable sites is a major index that reflects the level of genetic diversity. Similar findings were reported previously for *trnC-ycf6* (proportion of variable sites = 0.23%) in *Anemoclema glaucifolium* (Guan et al., 2013) and for *psbC-trnS* (0.13%) in pear germplasm resources (Chang et al., 2014) but stood in contrast to results

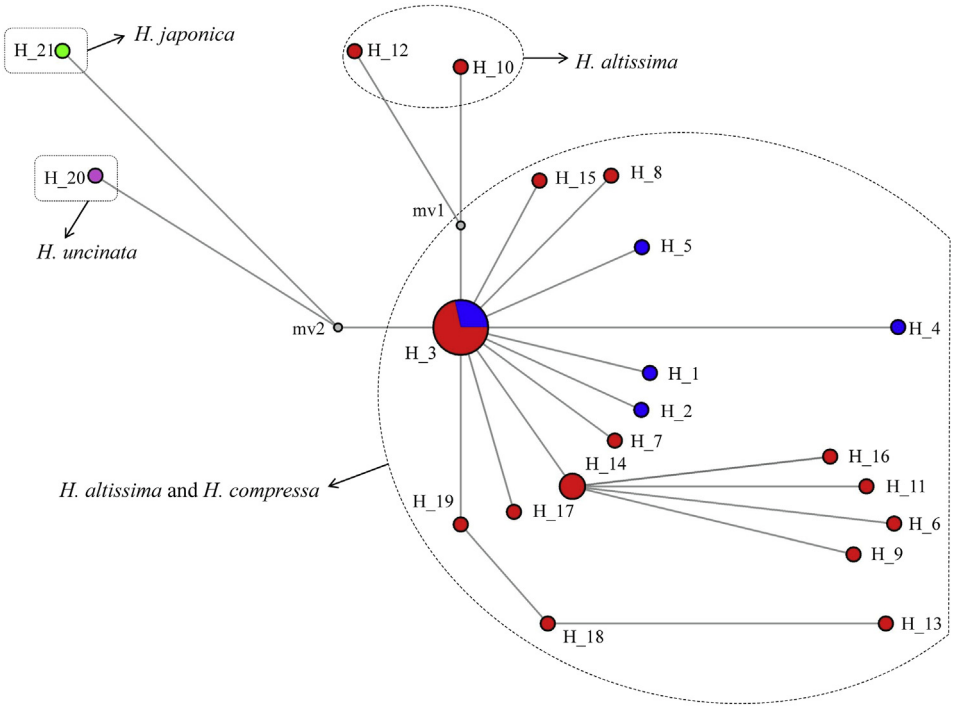

Fig. 3. Median-joining haplotype network based on the combined sequences of three cpDNA intergenic spacers (*trnL-F*, *trnC-ycf6*, and *psbC-trnS*) in 36 *Hemarthria* samples. H and mv denote haplotype and median vector, respectively.

Table 3
Molecular variance and F_{ST} in populations of *H. altissima* based on combined sequence data.

Source of population	Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	F_{ST}
All population	Among populations	7	12.357	0.04994	2.98%	$F_{ST} = 0.02980$
	Within populations	17	27.643	1.62605	97.02%	($P = 0.34311$)
	Total	24	40	1.67599		

for *trnL-F* (44%) in field vittarioid gametophytes (Chen et al., 2013). Moreover, such low genetic diversity in our cpDNA analysis contradicts previous studies based on nuclear DNA molecular markers (SCoT and EST-SSR markers) with higher levels of genetic diversity in *Hemarthria* samples (Huang et al., 2014). A similar situation – low diversity in cpDNA and higher diversity in nuclear markers – was also found in the *Miscanthus sinensis* (Yan et al., 2015). Lower genetic diversity in cpDNA non-coding spacers compared to nuclear genes may be due to the influences of random genetic drift and genetic bottlenecks (Gong et al., 2011).

Two methods of phylogenetic inference (ML and BI) were applied to group samples to elucidate relationships between samples. Although the ML and BI trees displayed similar patterns, the hierarchical classification structures were not directly associated with either the species labels or the geographic distributions of the samples. These results largely conflict with results from nuclear EST-SSR and SCoT markers (Huang et al., 2014). Samples of the two species *H. altissima* and *H. compressa* could not be clearly distinguished, indicating that they are closely related. This could be explained by the fact that the selected markers did not cover the entire chloroplast genome, or it could reflect the strong conservation of the *trnL-F*, *trnC-ycf6*, and *psbC-trnS* regions across plant species. The phylogenetic inferences were found to be consistent with a median-joining network analysis, which showed that 10 *H. altissima* samples and 4 *H. compressa* samples were included together in Haplotype 3. In addition, the results of the ML and BI analyses showed that some *Hemarthria* samples that were genetically divergent were geographically close in proximity, which agrees with the mismatch distribution analysis of the combined data, the population genetic structure analysis, and previous studies using cpDNA sequences (Batnini et al., 2014). This could be explained by 1) the strong capacity of *Hemarthria* individuals to undergo asexual propagation and ecological adaptation, which may be susceptible to natural factors (e.g., river flow), livestock activity (e.g., foraging), and human activities (e.g., domestication and cultivation) resulting in widespread dispersal of *Hemarthria* clones (Yang et al., 2004); 2) the likely occurrence of natural hybridization between *Hemarthria* clones, resulting in the production of new genetic combinations that may increase the complexity of the genetic and geographic distribution (Wu and Du, 2000); and 3) the probable occurrence of gene mutations during plant growth.

There is a common belief that branch-support values from each partition of a tree can be interpreted as the probability that the tree is correct or as an indicator of phylogenetic accuracy (Huelsenbeck et al., 2002). However, some believe that this is misleading and branch-support values should not be interpreted as probabilities that clades are correctly resolved, as the validity of ML or BI methods depend on the validity of the likelihood model and subsets of data (Simmons et al., 2004). Although the bootstrap support values of the ML tree and the Bayesian posterior probability values of the BI tree were quite weak for most of the clades in this study, we used normative processes to simulate the phylogenetic trees. Our low branch-support values may be due to the restricted genetic background of the tested samples, as branch-support values are poor when taxon sampling increases without an increase in genetic diversity (Simmons et al., 2004).

The ML and BI methods have been shown to be accurate in numerous simulation studies (Gill and Fast, 2006) and are now broadly recognized as better approaches than the neighbor-joining (NJ) or maximum parsimony (MP) methods (Guindon et al., 2010). In the case that an explicit evolutionary model was fitted for ML or BI, Huang (2012) held the ML method was more consistent with evolutionary fact. By testing a variety of conditions, Hall (2005) reported that the BI method was the more accurate. Negrisolo et al. (2004) believed that the ML method was very flexible due to its plasticity, and was theoretically sound and statistically consistent. For this study, we employed the CONSEL program, which calculates probability values (i.e., *p*-values), to assess the tree topologies of the ML and BI trees in order to reveal which method was better suited to our data. The results demonstrated that the ML tree exhibited a better topological structure, which may indicate that the ML method is better suited at illustrating genetic relationships between samples with a narrow genetic background than the BI method.

F_{ST} is the most commonly reported statistic for the assessment of genetic differentiation (Balloux and Lugon-Moulin, 2002). The F_{ST} value reflects the amount of genetic differentiation between populations. An F_{ST} value between 0 and 0.05 indicates little genetic differentiation, from 0.05 to 0.15 suggests moderate differentiation, from 0.15 to 0.25 indicates strong differentiation, and over 0.25 reflects very strong genetic differentiation (Curnw and Wright, 1979). Our F_{ST} of 0.0298 calculated for the seven African populations of *H. altissima* is generally considered to imply low genetic differentiation; however, such a low F_{ST} value may actually indicate important genetic differentiation that is by no means negligible (Curnw and Wright, 1979).

Author contributions

Lin-kai Huang and Jian Zhang conceived and designed the experiments; Xiu Huang and Lu Lu performed the experiments; Xiu Huang, De-jun Huang, Lu Lu, and Han-dong Yan analyzed the data; Jian Zhang, Xiu Huang, De-jun Huang, Yu Zhang, Lin-kai Huang, Lu Lu, and Han-dong Yan contributed reagents/materials/analysis tools; Jian Zhang and Xiu Huang wrote the paper.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

The research was supported by Strategic Planning and Management Innovation of Science and Technology Support Demonstration Project of Chongqing Modern Cattle and Sheep Industry (cstc2014zktjccxbx0021), and Modern Agro-industry Technology Research System (CARS-35-05). We thank Dr. Fang-luan Gao (Fujian Agriculture and Forestry University) for comments and suggestions that improved the manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.bse.2016.07.024.

References

Bakker, F.T., Culham, A., Daugherty, L.C., Gibby, M., 1999. A *trnL-F* based phylogeny for species of *Pelargonium* (Geraniaceae) with small chromosomes. Plant Syst. Evol. 216, 309–324.
Balloux, F., Lugon-Moulin, N., 2002. The estimation of population differentiation with microsatellite markers. Mol. Ecol. 11, 155–165.
Batnini, M.A., Bourguiba, H., Trifi-Farah, N., Krichen, L., 2014. Chloroplast DNA sequence data provides new insights into genetic diversity and phylogenetic relationships of Tunisian apricot germplasm. Sci. Hortic. 178, 241–247.
Chang, Y.J., Cao, Y.F., Zhang, J.M., Tian, L.M., Dong, X.G., Zhang, Y., Qi, D., Zheng, Y.C., 2014. Studies on genetic diversity of pear germplasm resources in Liaoning province of China based on chloroplast DNA analysis. Acta Hortic. Sin. 41, 1307–1316.
Chen, C.W., Huang, Y.M., Kuo, L.Y., Nguyen, Q.D., Luu, H.T., Callado, J.R., Farrar, D.R., Chiou, W.L., 2013. trn*L-F* is a powerful marker for DNA identification of field vittarioid gametophytes (Pteridaceae). Ann. Bot. 111, 663–673.
Chen, Y.X., Zhang, X.Q., Ma, X., Xie, W.G., 2011. Association analysis between SSR molecular markers and agronomic characters of Hemarthria compressa. Hubei Agric. Sci. 7, 059.
Clegg, M.T., Gaut, B.S., Learn, G.H., Morton, B.R., 1994. Rates and patterns of chloroplast DNA evolution. Proc. Natl. Acad. Sci. 91, 6795–6801.
Curnw, R.N., Wright, S., 1979. Evolution and the Genetics of Populations, Volume 4: Variability within and among Natural Populations. Biometrics 35, 359.
Demesure, B., Sodzi, N., Petit, R.J., 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol. Ecol. 4, 129–134.

Dolphin, K., Belshaw, R., Orme, C.D.L., Quicke, D.L.J., 2000. Noise and incongruence: interpreting results of the incongruence length difference test. Mol. Phylogenet. Evol. 17, 401–406.
D'Aloia, C.C., Azodi, C.B., Sheldon, S.P., Trombulak, S.C., Ardren, W.R., 2015. Genetic models reveal historical patterns of sea lamprey population fluctuations within Lake Champlain. PeerJ 3, e1369.
Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10 (3), 564–567.
Forster, P., Bandelt, H.J., Röhl, A., 2004. Network 4.2. 0.1. Fluxus Technology Ltda. Software available free at www. fluxus-engineering. com.
Gill, E.E., Fast, N.M., 2006. Assessing the microsporidia-fungi relationship: combined phylogenetic analysis of eight genes. Gene 375, 103–109.
Gong, X., Luan, S.S., Hung, K.H., Hwang, C.C., Lin, C.J., Chiang, Y.C., Chiang, T.Y., 2011. Population structure of *Nouelia insignis* (Asteraceae), an endangered species in southwestern China, based on chloroplast DNA sequences: recent demographic shrinking. J. Plant Res. 124, 221–230.
Guan, M.M., Ma, R., Gong, X., 2013. Conservation genetics of and endemic plant, *Anemoclema glaucifolium*, in the Jinsha river valley. Plant Divers. Resour. 35, 555–562.
Guindon, S., Delsuc, F., Dufayard, J.F., Gascuel, O., 2009. Estimating Maximum Likelihood Phylogenies with PhyML. Bioinformatics for DNA Sequence Analysis. Springer, pp. 113–137.
Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.
Hall, B.G., 2005. Comparison of the accuracies of several phylogenetic methods using protein and DNA sequences. Mol. Biol. Evol. 22, 792–802.
Hu, Z.M., Uwai, S., Yu, S.H., Komatsu, T., Ajisaka, T., Duan, D.L., 2011. Phylogeographic heterogeneity of the brown macroalga *Sargassum horneri* (Fucaceae) in the northwestern Pacific in relation to late Pleistocene glaciation and tectonic configurations. Mol. Ecol. 20, 3894–3909.
Huang, L.K., Huang, X., Yan, H.D., Yin, G.H., Zhang, X.Q., Tian, Y., Zhang, Y., Jiang, X.M., Yan, Y.H., Ma, X., 2014a. Constructing DNA fingerprinting of *Hemarthria* cultivars using EST-SSR and SCoT markers. Genet. Resour. crop Evol. 61, 1047–1055.
Huang, L.K., Zhang, X.Q., Ma, X., Liu, W., Li, F., Zeng, B., 2008. Genetic differentiation among *Hemarthria compressa* populations in south China and its genetic relationship with *H. japonica*. Hereditas 145, 84–91.
Huang, L.K., Zhang, Y., Zhang, J., Zhang, X.Q., Xie, W.G., Jiang, X.M., Peng, F., Yan, Y.H., Ma, X., Liu, W., et al., 2014b. Genetic stability and DNA fingerprinting of the *Hemarthria compressa* cultivar "Guangyi". Biochem. Syst. Ecol. 55, 310–316.
Huang, X., Zhang, X.Q., Huang, L.K., Ma, Y.M., Yin, G.H., Lee, S., Zeng, J., Liu, H., 2014. Genetic diversity of *Hemarthria altissima* and its related species by EST-SSR and SCoT markers. Biochem. Syst. Ecol. 57, 338–344.
Huang, Y., 2012. Molecular Phylogenetics. Science Press, Beijing.
Hudson, W.G., 1986. Mole cricket (Orthoptera: Gryllotalpidae) damage to *Hemarthria altissima*: resistance or nonpreference? J. Econ. Entomol. 79, 961–963.
Huelsenbeck, J.P., Larget, B., Miller, R.E., Ronquist, F., 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. Syst. Biol. 51, 673–688.
Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.
Juszczak, I., Rudnik, R., Pietzenuk, B., Baier, M., 2012. Natural genetic variation in the expression regulation of the chloroplast antioxidant system among *Arabidopsis thaliana* accessions. Physiol. Plant. 146, 53–70.
Katayama, H., Tachibana, M., Iketani, H., Zhang, S.L., Uematsu, C., 2012. Phylogenetic utility of structural alterations found in the chloroplast genome of pear: hypervariable regions in a highly conserved genome. Tree Genet. Genomes 8, 313–326.
Kimura, T., Iketani, H., Kotobuki, K., Matsuta, N., Ban, Y., Hayashi, T., Yamamoto, T., 2003. Genetic characterization of pear varieties revealed by chloroplast DNA sequences. J. Hortic. Sci. Biotechnol. 78, 241–247.
Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.
Murakami, A., Darby, P., Javornik, B., Pais, M.S.S., Seigner, E., Lutz, A., Svoboda, P., 2006. Molecular phylogeny of wild hops, *Humulus lupulus* L. Heredity 97, 66–74.
Negrisolo, E., Minelli, A., Valle, G., 2004. The mitochondrial genome of the house centipede Scutigera and the monophyly versus paraphyly of myriapods. Mol. Biol. Evol. 21, 770–780.
Newman, Y., Vendramini, J., Sollenberger, L.E., Quesenberry, K., 2011. Limpograss (Hemarthria Altissima): Overview and Management. EDIS Publication. SS-AGR-320. http://edis.ifas.ufl.edu.
Posada, D., 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.
Quesenberry, K.H., Oakes, A.J., Jessop, D.S., 1982. Cytological and geographical characterizations of *Hemarthria*. Euphytica 31, 409–416.
Schank, S.C., Klock, M.A., Moore, J.E., 1973. Laboratory evaluation of quality in subtropical grasses: II. Genetic variation among *Hemarthrias* in in vitro digestion and stem morphology. Agron. J. 65, 256–258.
Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18, 502–504.
Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247.
Simmons, M.P., Pickett, K.M., Miya, M., 2004. How meaningful are Bayesian support values? Mol. Biol. Evol. 21, 188–199.
Swindell, S.R., Plasterer, T.N., 1997. Seqman. Methods in Molecular Biology, pp. 75–89.
Swofford, D., 1998. PAUP 4.0: Phylogenetic Analysis Using Parsimony. Encyclopedia of Genetics, Genomics, Proteomics and Informatics, p. 1455.
Vaidya, G., Lohman, D.J., Meier, R., 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27, 171–180.
Wu, Y.Q., Du, Y., 2000. Some reproductive characteristics of *Hemarthria compressa*. J. Sichuan Agric. Univ. 18, 262–264.
Xia, X.H., Lemey, P., 2009. Assessing Substitution Saturation with DAMBE. The Phylogenetic Handbook: a Practical Approach to DNA and Protein Phylogeny, 2, pp. 615–630.
Yan, H.D., Zhang, X.Q., Fu, C., Huang, L.K., Yin, G.H., Nie, G., Xu, W.Z., Liu, X.C., Chen, T., Lee, S., et al., 2015. Chloroplast DNA variation and genetic structure of *Miscanthus sinensis* in southwest China. Biochem. Syst. Ecol. 58, 132–138.
Yang, C.H., Zhang, X.Q., Li, X.L., Du, Y., Wu, Y.Q., 2004. *Hemarthria* germplasm resources and breeding. Acta Prataculturae Sin. 13, 7–12.