Published October 19, 2015

RESEARCH

Novel Polymorphic Expressed-Sequence Tag–Simple-Sequence Repeat Markers in *Campeiostachys nutans* for Genetic Diversity Analyses

Dong Luo, Qiang Zhou, Lichao Ma, Wengang Xie, Yanrong Wang, Xiaowen Hu, and Zhipeng Liu*

ABSTRACT

Campeiostachys nutans (Griseb.) J. L. Yang, B. R. Baum et C. Yen is a widely cultivated grass that is used for forage. Considering its limited genomic resources, there is a strong need to increase the available data on this species, which would allow for the identification of genes useful for molecular breeding and lead to increased crop quality and productivity. Complementary DNA (cDNA) samples from 11 different tissues were sequenced using Illumina paired-end sequencing. A total of 106,220 unigenes were obtained via de novo assembly, and 82,863 (78.01%) unigenes were annotated based on similarity searches using known proteins. In addition, 8727 expressed sequence tag–simple-sequence repeat (EST-SSR) loci were identified. A subset of 480 EST-SSRs was randomly synthesized and screened for validation across a panel of 16 *C. nutans* accessions using 48 individual plants, with 113 polymorphic EST-SSRs generating 439 alleles. The number of alleles, observed heterozygosity, expected heterozygosity, and polymorphism information content ranged from three to nine, 0.00 to 1.00, 0.51 to 0.84, and 0.39 to 0.82, respectively. Overall, the *C. nutans* transcriptomic sequences generated in this study provide both an important platform that will facilitate gene discovery for functional genomic studies and EST-SSR markers for use in breeding.

State Key Lab. of Grassland Agro-ecosystems, School of Pastoral Agricultural Science and Technology, Lanzhou Univ., Lanzhou 730020, China. Financial support was provided by the National Basic Research Program of China (2014CB138704) and the National Natural Science Foundation of China (31272492). Received 8 Jan. 2015. Accepted 16 Apr. 2015. *Corresponding author (lzp@lzu.edu.cn).

Abbreviations: cDNA, complementary DNA; COG, clusters of orthologous groups; dNTP, deoxynucleoside triphosphate; ds-cDNA, double-stranded cDNA; EST-SSR, expressed sequence tag–simple-sequence repeat; GC, guanine–cytosine; GO, gene ontology; H_E, expected heterozygosity; H_O, observed heterozygosity; KEGG, Kyoto Encyclopedia of Genes and Genomes; mRNA, messenger RNA; N_A, number of alleles; nr, nonredundant protein; nt, nonredundant nucleotide; PCR, polymerase chain reaction; PIC, polymorphism information content; QTP, Qinghai–Tibetan Plateau; SSR, simple-sequence repeat.

Campeiostachys nutans is a perennial, cespitose, self-pollinating, allohexaploid ($2n = 6x = 42$) species that contains St, H, and Y genomes (Lu, 1993; Dou et al., 2009; Fan et al., 2013). *Campeiostachys nutans* was previously part of the genus *Elymus* but was recently reclassified into a resurrected, legitimate genus, *Campeiostachys* Drobov, in accordance with the International Code of Botanical Nomenclature, which is based on monophyly, discrimination, and practicality (Baum et al., 2011). *Campeiostachys nutans* is widely distributed in eastern Mongolia, the Asian part of Russia, and the Himalayan areas of northern India (Chen et al., 2009a). In China, this species is commonly distributed in north, northwest, and southwest regions, particularly on the Qinghai–Tibetan Plateau (QTP) at elevations between 3000 and 5000 m (Chen et al., 2013). An important livestock grass, this species is well adapted to alpine, cold, drought environments, and biotic stress, which may be a valuable genetic reservoir for the improvement of four taxonomically related genera: *Elymus*, *Anthosachne*, *Kengyilia*, and *Roegneria* (Yan et al., 2009; Wang and Lu, 2014). Additionally,

Published in Crop Sci. 55:2712–2718 (2015).
doi: 10.2135/cropsci2015.01.0012
© Crop Science Society of America | 5585 Guilford Rd., Madison, WI 53711 USA
All rights reserved.

it may be useful in environmental sustainability, such as in the construction of artificial grasslands and in ecological restoration (Chen et al., 2013). These characteristics have stimulated interest in the development of molecular resources for *C. nutans*.

To date, several studies have reported on the genetic diversity of *C. nutans* with regard to gliadins (Miao et al., 2011), sequence-related amplified polymorphism markers (Chen et al., 2009b), random amplified polymorphic DNA markers (Miao et al., 2011), amplified fragment-length polymorphism markers (Yan et al., 2009, 2010; Guo et al., 2014), intersimple sequence repeat markers (Chen et al., 2009a), and homologous simple-sequence repeats (SSRs) from wheat (*Triticum aestivum* L.) and *Elymus* species (Chen et al., 2013). However, there is little relevant genomic information or species-specific molecular markers available for *C. nutans*, which limits the opportunities for genetic diversity and molecular breeding to improve the yield and nutritional quality of this species.

Recent studies have shown that next-generation transcriptome sequencing technology (e.g., Illumina Genome Analyzer, Applied Biosystems SOLiD, and Roche/454 Genome Sequencer FLX platforms) is an effective approach for generating many genomic resources in a short time with reasonable cost and effort, even for nonmodel plant species (Wang et al., 2010; Kaur et al., 2011; Wei et al., 2011; Parra-González et al., 2012; Rowland et al., 2012; Shahin et al., 2012; Liu et al., 2013; Shirasawa et al., 2013; Chung et al., 2014). Because of the ability to extensively investigate alternative splicing, discover large-scale novel transcripts, and identify gene boundaries at single-nucleotide resolution, massively parallel transcriptome sequencing has great potential to revolutionize the field of species transcriptomics (Chen et al., 2011). Indeed, based on transcriptomic data, EST-SSR markers can be developed using bioinformatic data mining. Compared with genomic SSR markers, these EST-SSR markers are easily accessible, are present in gene-rich regions, are associated with transcription, can identify candidate functional genes, and are more easily transferrable between closely related species, attributes that can help accelerate marker-assisted selection for crop improvement programs (Thiel et al., 2003; Chung et al., 2014; Liu et al., 2014; Zhou et al., 2014b).

To characterize novel polymorphic EST-SSR markers in *C. nutans*, we sequenced 11 different tissue samples and assembled the first global transcriptome of this grass species using Illumina paired-end sequencing. By exploiting sequence databases and developing EST-SSR markers, we aim to promote studies on the genetic diversity of this species and provide tools for *C. nutans* molecular breeding programs.

MATERIALS AND METHODS

Plant Material and RNA Isolation

Campeiostachys nutans accession 'Xiahe No. 2' was grown in a field at the Yuzhong campus of Lanzhou University, Lanzhou, China (Supplemental Table S1). Between 10 April and 1 June 2013, a total of 11 tissue samples were collected at different stages of development. The samples were as follows: callus cells, plumules (7 d after seed germination), whole seedlings (3 wk after germination), tufted leaves in the tillering stage, flag leaves in the heading stage, partially lignified stems, moderately lignified stems, highly lignified stems, young inflorescences (10 d before fertilization), inflorescences (5 d before fertilization), and old inflorescences (5 d after fertilization). The callus cells were induced from young spikes of Xiahe No. 2 on solid MS medium containing 2,4-dichlorophenoxyacetic acid (3.0 mg L^{-1}) at 25°C for 30 d under 16 h light–8 h dark cycles. All tissues were immediately frozen in liquid nitrogen and stored at −80°C until RNA extraction. For transcriptome sequencing, total RNA from each sample was isolated using RNeasy Plant Mini Kit (Qiagen). The integrity of the RNA was determined using both a NanoDrop ND1000 (Thermo Scientific) and an Agilent 2100 Bioanalyzer (Agilent Technologies). For preparation of the cDNA library, RNA from the 11 tissues was separately extracted and pooled by equal amounts to approximately 20 μg.

Complementary DNA Library Construction and Sequencing

The cDNA library was constructed using an mRNA-Seq sample preparation kit (Illumina Inc.) according to the manufacturer's instructions. In brief, total RNA from all 11 *C. nutans* samples was purified to enrich poly(A) messenger RNA (mRNA) using magnetic oligo(dT) beads. To avoid priming bias, the purified mRNA was cut into short fragments using an RNA fragmentation kit (Ambion); these short fragments were reverse transcribed into first-strand cDNA using random hexamer primers and reverse transcriptase (Invitrogen). Second-strand cDNA was then synthesized using buffer, deoxynucleoside triphosphates (dNTPs), RNase H (Invitrogen), and DNA polymerase I (New England BioLabs). The paired-end library was synthesized using a genomic sample preparation kit (Illumina) following the manufacturer's instructions. The double-stranded cDNA (ds-cDNA) was purified with a MinElute polymerase chain reaction (PCR) purification kit (Qiagen). After eluting the ds-cDNA with EB buffer (Qiagen) in preparation for end reparation and poly(A) addition, the short fragments were connected with sequencing adapters, and suitable fragments (200 ± 25 bp) were selected as templates for PCR amplification (15 cycles) by agarose gel electrophoresis. Finally, the sequencing library was constructed and sequenced using the Illumina Genome Analyzer IIx sequencing platform. All reads were deposited into the National Center for Biotechnology Information (NCBI) Single-Read Archive database (SRS627415).

Sequence Analyses, Assembly, and Annotation

Clean reads were separated from the raw data by filtering out the dirty reads, which contained adapters, low-quality reads (reads with ambiguous N bases), or reads with more than 10% of the

bases having Q < 20. These clean reads were assembled using the Trinity method, which recovers more full-length transcripts across a wider range of expression levels than other available de novo transcriptome assembly tools (Grabherr et al., 2011). In this step, the transcriptome reads were first broken into *k*-mers, which were continuous DNA sequences; these *k*-mers were then used to construct contigs based on overlaps, and a de Bruijn graph was constructed for each cluster of related contigs. Finally, the paths were analyzed within the context of the corresponding de Bruijn graph, and all plausible transcript sequences were reported. Unigenes were obtained via these steps.

For further analysis, all of the unigenes were initially aligned using BLASTx (E-value < 10^{-5}) to protein databases, including NCBI nonredundant protein (nr), Swiss-Prot, clusters of orthologous groups (COG), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases, and then aligned by BLASTn (E-value < 10^{-5}) to the NCBI nonredundant nucleotide (nt) nucleic acid database, retrieving proteins with the highest sequence similarity with the given unigenes along with their protein functional annotations (Wang et al., 2014). Blast2GO software was used to obtain gene ontology (GO) annotations of the unigenes based on BLASTx (E-value < 10^{-5}) hits against the NCBI nr database. WEGO software (Ye et al., 2006) was used to perform GO functional classification of all of the unigenes and to plot the macro-level distribution of *C. nutans* gene functions. The unigene sequences were also aligned with the COG database to predict and classify their functions. Pathway assignments were performed using the KEGG database.

Detection of Expressed Sequence Tag–Simple-Sequence Repeat Markers and Primer Design

Expressed sequence tag–simple-sequence repeat mining of the 106,220 *C. nutans* unigenes was performed using the MIcroSAtellite program (http://pgrc.ipkgatersleben.de/misa/) with the following parameters: at least 12 repeats for mononucleotide SSRs, six repeats for dinucleotide SSRs, five repeats for tri- and tetranucleotide SSRs, and four repeats for penta- and hexanucleotide SSRs (Thiel et al., 2003). Both perfect and compound EST-SSRs were identified. Batchprimer3 software was used for primer design (http://probes.pw.usda.gov/cgi-bin/batchprimer3/batchprimer3.cgi).

Expressed Sequence Tag–Simple-Sequence Repeat Amplification and Diversity Analysis

A total of 16 *C. nutans* accessions (Supplemental Table S1) were selected for polymorphism validation of the EST-SSRs. These accessions originate from six different counties in the QTP region and represent a relatively broad genetic diversity. Using the hexadecyltrimethyl ammonium bromide method, total genomic DNA was extracted from the young leaves of three individual plants per accession (Liu et al., 2013). Polymerase chain reactions were performed in a final volume of 10 μL containing 40 ng of template DNA, 1× PCR buffer, 2.0 mM $MgCl_2$, 2.5 mM dNTPs, primers (4 μM each), and 0.8 U Taq polymerase (TaKaRa). The PCR cycling parameters included 3 min at 94°C, 35 cycles of 30 s at 94°C, 30 s at the annealing temperature (Supplemental Table S2), 20 s at 72°C, and a final extension step of 7 min at 72°C. The PCR products were resolved via electrophoresis on 8.0% nondenaturing polyacrylamide gels and stained with ethidium bromide (Liu et al., 2013). The indices of the number of alleles (N_A), the observed heterozygosity (H_O), the expected heterozygosity (H_E), and the polymorphism information content (PIC) were calculated as previously described (Botstein et al., 1980; Anderson et al., 1993; Liu et al., 2007). Cluster analysis was performed to generate a dendrogram using the unweighted pair-group method with arithmetic mean and Nei's unbiased genetic distance using the NTSYSPC 2.0 software (Nei, 1978).

RESULTS AND DISCUSSION

Illumina Sequencing, De Novo Assembly, and Functional Annotation

Gene expression in tissues or organs varies during different developmental stages in *C. nutans* (Liu et al., 2013). To achieve greater global and comprehensive coverage of the *C. nutans* transcriptome, total RNA was extracted from 11 different *C. nutans* tissues; the morphology of each tissue is shown in Supplemental Fig. S1. Equal amounts of total RNA from each *C. nutans* tissue were pooled and the cDNA was subjected to Illumina HiSeq 2000 sequencing. After stringent quality assessment and data filtering, we obtained 77,057,690 clean reads with a total of 6,935,192,100 nucleotides; the Q20, the proportion of uncertain nucleotides between each pair of contigs after date filtering (*N*), and guanine–cytosine (GC) percentages were 98.02, 0.00, and 54.32%, respectively. These reads were considered high quality and were used for further analysis.

Using the Trinity program (Grabherr et al., 2011), all high-quality reads were de novo assembled into 285,832 contigs, with a total of 73,438,351 nt and an N50 length of 324 bp (Table 1); 106,220 unigenes were obtained from 67,398,704 nt, with an N50 length of 926 bp (Table 1). The average unigene length was 635 bp, which was longer than those assembled in studies of sesame (*Sesamum indicum* L.; 629 bp) (Wei et al., 2011), sweet potato [*Ipomoea batatas* (L.) Lam.; 581 bp] (Wang et al., 2010), *Epimedium sagittatum* (Sieb. et Zucc.) Maxim (246 bp) (Zeng et al., 2010), safflower (*Carthamus tinctorius* L.; 446 bp) (Huang et al., 2012), and watermelon [*Citrullus lanatus* (Thunb.) Matsum. et Nakai; 302 bp] (Guo et al., 2011). For validation and annotation of the assembled unigenes, unique sequences were compared with the following databases using BLAST: NCBI nr, NCBI nt, Swiss-Prot, KEGG, and COG. The best alignment was selected from the matches with E-values less than 10^{-5}. Annotations were successfully attained for 82,863 (78.01%) unigenes in the above databases (Table 2), suggesting that the Illumina paired-end sequencing used in this study recovered a substantial fraction of *C. nutans* genes. In the present study, 62.12% (66,514 of 106,220) of the *C. nutans* unigenes had homologs in the NCBI nr (Table 2), while the homolog percentages for the five species mentioned above were 53.19, 48.54, 38.54, 53.92, and

Table 1. Summary of the de novo assembly and expressed sequence tag (EST)–simple-sequence repeat (SSR) analysis for *Campeiostachys nutans*.

Items	Number
Total unigenes	106,220
Total annotated unigenes	82,863
Total raw reads	86,742,888
Total clean reads	77,057,690
Total clean nucleotides (nt)	6,935,192,100
Average unigene length (nt)	635
Q20%[†]	98.02%
N percentage	0.00%
GC[‡] percentage	54.32%
Total number of identified SSRs	8727
Number of SSR-containing sequences	7671
Number of sequences containing more than 1 SSR	932
Number of compound SSRs	419

[†] Q20% is the proportion of nucleotides with a quality value greater than 20.

[‡] GC, guanine–cytosine.

Table 2. Functional annotation of *Campeiostachys nutans* unigenes.

Database[†]	No. of unigenes
nr	66,514
nt	76,340
Swiss-Prot	40,848
KEGG	37,701
COG	22,636
GO	43,882
Total	82,863

[†] nr, nonredundant protein (National Center for Biotechnology Information [NCBI]) ; nt, nonredundant nucleotide (NCBI); COG, clusters of orthologous groups; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, gene ontology.

54.94%, respectively. The higher percentage of hits found in our research was partially a result of the increased average unigene length in our database.

Based on 66,514 nr annotations, 43,882 (41.31%) unigenes were assigned to one or more GO terms using the Blast2GO program. GO, which can be used for data from any organism, is a useful tool for annotating and analyzing the functions of a large number of genes and their products. The GO-annotated unigenes were found to have three ontologies—biological process, cellular component, and molecular function—distributed across 55 subfunctional groups, with 42.61% in biological process: 40.46% in molecular function and 16.93% in cellular component (Supplemental Fig. S2). In the biological process group, metabolic process, cellular process, and single-organism process were the three most frequent terms, with 23,353; 23,273; and 13,887 unigenes, respectively, indicating the high metabolic activity of *C. nutans*. For the cellular component category, cell (31,329 unigenes), cell part (31,328 unigenes), and organelle (26,976 unigenes) were highly represented, whereas only a few unigenes were assigned to virion (21 unigenes), virion part (21 unigenes), and extracellular region part (26 unigenes). With regard to molecular function, the majority of the unigenes were assigned to binding (21,975 unigenes), followed by catalytic activity (21,684 unigenes) and transporter activity (2567 unigenes), indicating the dominance of gene regulation, signal transduction, and enzymatically active processes. The categories of metabolic process, cell, and binding activity were the most abundant in the biological process, cellular component, and molecular function groups, respectively. These results are consistent with previous studies in sweet potato (Wang et al., 2010), sesame (Wei et al., 2011), rubber tree [*Hevea brasiliensis* (Willd. ex A. Juss.) Müll. Arg.] (Li et al., 2012), alfalfa (*Medicago sativa* L.) (Liu et al., 2013), and pumpkin (*Cucurbita moschata* Duchesne) (Wu et al., 2014).

To predict and classify the possible functions of the *C. nutans* unigenes, 22,636 of the 66,514 sequences showing nr hits were assigned COG classifications (Supplemental Fig. S3). In this database, every protein is assumed to have evolved from the same ancestral protein. Among the 25 COG categories, general function prediction only was the largest group (8831 unigenes), followed by function unknown (7222 unigenes), transcription (6437 unigenes), replication, recombination and repair (5586 unigenes), and translation, ribosomal structure, and biogenesis (5507 unigenes). Conversely, only 16 and 25 unigenes were assigned to nuclear structures and extracellular structures, respectively. As the *C. nutans* unigenes were assigned to a wide range of COG classifications, our paired-end sequencing data appear to reflect an extensive diversity of transcripts.

To further analyze the *C. nutans* transcriptome, all 106,220 unigenes were analyzed using KEGG pathway tools, which can improve our understanding of the biological functions of genes. This process predicted a total of 128 pathways that were represented by 37,701 (35.49%) unigenes (Supplemental Table S3).

Simple-Sequence Repeat Discovery and Polymorphism Analysis

Of the 106,220 unigenes, 7671 (7.2%) were identified as containing EST-SSRs (Table 1), a result that is consistent with the EST frequency of between 2.65 and 10.62% reported for 49 dicot species (Kumpatla and Mukhopadhyay, 2005). Of the 7671 EST-SSR-containing unigenes, 932 (12.1%) were found to contain two or more SSRs, with 419 (5.5%) containing compound EST-SSRs. A total of 8727 EST-SSR markers were identified from the 7671 SSR-containing ESTs (Table 1).

The frequency, type, and distribution of the 8727 potential SSRs were analyzed. As shown in Table 3, we found that the most highly represented repeat unit of the potential SSRs was five repeats, accounting for 4133 (47.36%) SSRs, followed by six (2310; 26.47%), seven (881; 10.10%), and eight repeats (330; 3.78%). Furthermore, trinucleotide

Table 3. Length distribution of expressed sequence tag (EST)–simple-sequence repeats (SSR) based on the number of nucleotide repeat units.

No. of repeats	Di-	Tri-	Tetra-	Penta-	Hexa-	Total	%
4	0	0	0	382	176	558	6.39
5	0	3840	256	37	0	4133	47.36
6	961	1299	50	0	0	2310	26.47
7	446	435	0	0	0	881	10.10
8	295	35	0	0	0	330	3.78
9	212	0	0	0	0	212	2.43
10	172	0	0	0	0	172	1.97
11	123	0	0	0	0	123	1.41
12	8	0	0	0	0	8	0.09
≥13	0	0	0	0	0	0	0
Total	2217	5609	306	419	176	8727	
%	25.40	64.27	3.51	4.80	2.02		

repeats were the most abundant type (5609; 64.27%), followed by di- (2217; 25.40%), penta- (419; 4.80%), tetra- (306; 3.51%), and hexanucleotide (176; 2.02%) repeats. There was a large proportion of both di- and trinucleotide repeats (89.67%), whereas the other types of repeats amounted to less than 10.33%. This result is consistent with the EST-SSR distributions reported for alfalfa (Liu et al., 2013; Wang et al., 2014), common vetch (*Vicia sativa* subsp. *sativa* L.) (Liu et al., 2014), lentil (*Lens culinaris* Medik.) (Kaur et al., 2011), narrow-leaved vetch (*V. sativa* subsp. *nigra* L.) (Chung et al., 2014), sweet potato (Wang et al., 2010), and yellow lupine (*Lupinus luteus* L.) (Parra-González et al., 2012). It has been proposed that the deleterious effects of frame-shift mutations in translated regions could suppress the expansion or contraction of dinucleotide repeats in exons (Xin et al., 2012), which might explain the dominance of trinucleotide repeat motifs observed in this study. In total, 195 types of sequence motifs were identified; of these, the CCG/CGG repeat was the most abundant (1991; 22.81%), followed by AG/CT (1321; 15.14%), AGG/CCT (1025; 11.75%), and AGC/CTG (990; 11.34%) (Supplemental Fig. S4). This finding was also consistent with a previous report showing that the CCG/CGG motif is very common in monocots but rare in dicotyledons (Wang et al., 2011).

Of the 8727 SSR-containing unigenes in *C. nutans*, 2647, 2940, and 3067 unigenes have GO annotations of biological process, molecular function, and cellular component, respectively. Using the annotations of the assembled 106,220 unigenes as a reference, GO enrichment was executed by agriGO (Du et al., 2010) (http://bioinfo.cau.edu.cn/agriGO/), and the results showed a statistically significant increase in the proportion of transcription (GO: 006350)-related unigenes (Supplemental Fig. S5). Additionally, we performed similar GO analyses of the SSR-containing unigenes for our published data for common vetch (Liu et al., 2014) and alfalfa (Liu et al., 2013); interestingly, transcription-related unigenes were also significantly enriched (Supplemental Fig. S6), which was consistent with the results for *C. nutans*. The similar results found in two dicots (common vetch and alfalfa) and one monocot (*C. nutans*) indicate that unigenes related to the category transcription might be more likely to contain SSRs than other unigenes.

A total of 480 primer pairs were randomly designed and synthesized from 8727 identified SSRs. The 480 EST-SSR primer pairs were assayed in five *C. nutans* individuals (accessions No. 1, No. 2, No. 3, No. 13, and No. 14) using 8.0% nondenaturing polyacrylamide gel electrophoresis. Among the 480 primer pairs, 405 showed successful PCR amplification. The remaining 75 primers failed to generate PCR products, which may have been due to the location of the primers across splice sites, large introns, or low-quality sequences. The 405 primer pairs were then assayed in 16 *C. nutans* accessions and 48 individuals. Among the 405 working primer pairs, amplified products of the expected sizes were found for 379 pairs: 22 pairs generated products larger than expected, and the remaining four generated products that were smaller than expected. Of the 379 primer pairs, only one band or several bands without polymorphisms were found for 266 PCR products. The other 113 PCR amplifications resulted in more than one band or bands containing polymorphisms, which might be due to the high heterozygosity of *C. nutans* germplasm at these loci (Supplemental Fig. S7).

The polymorphic ratio of EST-SSR markers in *C. nutans* (23.54%) was lower than that in alfalfa (27%) (Liu et al., 2013), blueberry (*Vaccinium corymbosum* L.) (43.0%) (Rowland et al., 2012), *Eucalyptus grandis* Hill ex Maiden (90.8%) (Zhou et al., 2014a), lentil (47.5%) (Kaur et al., 2011), narrow-leaved vetch (49.0%) (Chung et al., 2014), sweet potato (41.9%) (Wang et al., 2011), and yellow lupine (59.2%) (Parra-González et al., 2012) but higher than that in peanut (*Arachis hypogaea* L.) (23.3%) (Koilkonda et al., 2012) and common vetch (21.1%) (Liu et al., 2014). The differences among these reports might be related to the pollination characteristics of the species, the ploidy of the species, the number and genetic relationship of the accessions used, the use of bulk DNA or individual DNA across accessions, the sequencing technology employed, and the primer design (Kuleung et al., 2004; Wang et al., 2011; Li et al., 2012).

The 113 polymorphic EST-SSRs generated a total of 439 alleles. The N_A values ranged from three to nine (mean 3.88), and the ranges for H_O, $H_{E,}$ and PIC were 0.00 to 1.00 (mean 0.50), 0.51 to 0.84 (mean 0.59), and 0.39 to 0.82 (mean 0.50), respectively (Supplemental Table S2; Supplemental Fig. S7). These results indicated a high level of polymorphism in *C. nutans* populations, even though it is a self-pollinated species (Yan et al., 2006). These characteristics might exist because this species is widely distributed in the QTP region, which has complicated and diverse environmental conditions such as a

diverse climate, different elevations, intricate landforms, and various soil types (Chen et al., 2009a).

To determine the possible functions of the 113 EST-SSR markers, a BLAST analysis was performed. The results showed that 110 (97.3%) have homology to a large number of known proteins; the remaining three EST-SSRs resulted in no hits (2.7%) and may represent *C. nutans*-specific transcripts (Supplemental Table S2). Additionally, among the 8727 identified EST-SSRs, only 1264 primer pairs for EST-SSRs were successfully designed with high quality. In addition to the 480 pairs of primers synthesized in the present study, another 784 pairs of candidate primers may be useful in future studies (Supplemental Table S4).

A phylogenetic tree was constructed based on the observed polymorphisms, with the results indicating that the 16 *C. nutans* accessions evaluated can be grouped into three clusters (Supplemental Fig. S8). As expected, the cultivated accession Pop 10 clustered separately from the other 15 wild accessions and grouped into Cluster 3, indicating that the artificial domestication of this cultivar might have affected its genetic diversity. However, Cluster 1 (containing 12 accessions: Pop 1, Pop 2, Pop 3, Pop 4, Pop 5, Pop 6, Pop 7, Pop 8, Pop 9, Pop 11, Pop 12, and Pop 13) and Cluster 2 (containing three accessions: Pop 14, Pop15, and Pop 16) did not show a clear relationship with regard to the clustering pattern and geographical distance, suggesting that future studies will require the use of a greater number of accessions from close geographical locations, as well as more individual plants per accession, to verify the genetic diversity of *C. nutans*.

CONCLUSIONS

In conclusion, we herein describe the de novo transcriptome sequencing analysis of mixed RNAs from 11 different tissues in *C. nutans*. A total of 6.94 Gb of data were generated and assembled into 106,220 unigenes. Based on these sequences, 8727 EST-SSRs were predicted and characterized as potential molecular markers. To detect polymorphisms among 16 *C. nutans* accessions, 480 EST-SSRs primer pairs were randomly selected, 113 of which successfully amplified fragments, revealing abundant polymorphisms. To our knowledge, this is the first attempt to investigate the global transcriptome of the allohexaploid species *C. nutans* and to use Illumina paired-end sequencing technology to assemble reads without a reference genome for the development of EST-SSR markers. This valuable dataset significantly augments the available genomic resources for the allohexaploid grass *C. nutans*, and it will provide a very effective platform for future molecular breeding studies.

References

Anderson, J.A., G.A. Churchill, J.E. Autrique, S.D. Tanksley, and M.E. Sorrells. 1993. Optimizing parental selection for genetic linkage maps. Genome 36:181–186. doi:10.1139/g93-024

Baum, B.R., J.L. Yang, C. Yen, and A.V. Agafonov. 2011. A taxonomic synopsis of the genus *Campeiostachys* Drobov. J. Syst. Evol. 49:146–159. doi:10.1111/j.1759-6831.2010.00106.x

Botstein, D., R.L. White, M. Skolnick, and R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32:314–331.

Chen, C.Y., H.S. Ai, J. Ren, W.B. Li, P.H. Li, R.M. Qiao, J. Ouyang, M. Yang, J.W. Ma, and L.S. Huang. 2011. A global view of porcine transcriptome in three tissues from a full-sib pair with extreme phenotypes in growth and fat deposition by paired-end RNA sequencing. BMC Genomics 12:448. doi:10.1186/1471-2164-12-448

Chen, S.Y., X. Ma, X.Q. Zhang, and Z.H. Chen. 2009a. Genetic variation and geographical divergence in *Elymus nutans* Griseb. (Poaceae: Triticeae) from west China. Biochem. Syst. Ecol. 37:716–722. doi:10.1016/j.bse.2009.12.005

Chen, S.Y., X.Q. Zhang, X. Ma, and L.K. Huang. 2013. Assessment of genetic diversity and differentiation of *Elymus nutans* indigenous to Qinghai-Tibet Plateau using simple sequence repeats markers. Can. J. Plant Sci. 93:1089–1096. doi:10.4141/cjps2013-062

Chen, Z.H., J.M. Miao, J.C. Zhong, X. Ma, S.Y. Chen, and X.Q. Zhang. 2009b. Genetic diversity of wild *Elymus nutans* germplasm detected by SRAP markers. (In Chinese with English abstract.) Acta. Pratac. Sin. 18:192–200 .

Chung, J.W., T.S. Kim, S. Sundan, G.A. Lee, J.H. Park, G.T. Cho, H.S. Lee, J.Y. Lee, M.C. Lee, H.J. Baek, and S.Y. Lee. 2014. New EST-SSR markers in the narrow-leaved vetch (*Vicia sativa* subsp *nigra*) using 454 pyrosequencing. Mol. Breed. 33:749–754. doi:10.1007/s11032-013-9980-3

Dou, Q.W., Z.G. Chen, Y.A. Liu, and H. Tsujimoto. 2009. High frequency of karyotype variation revealed by sequential FISH and GISH in plateau perennial grass forage *Elymus nutans*. Breed. Sci. 59:651–656. doi:10.1270/jsbbs.59.651

Du, Z., X. Zhou, Y. Ling, Z.H. Zhang, and Z. Su. 2010. agriGO: A GO analysis toolkit for the agricultural community. Nucleic Acids Res. 38:W64–W70. doi:10.1093/nar/gkq310

Fan, X., L.N. Sha, Z.Z. Dong, H.Q. Zhang, H.Y. Kang, Y. Wang, X.L. Wang, L. Zhang, C.B. Ding, R.W. Yang, Y.L. Zheng, and Y.H. Zhou. 2013. Phylogenetic relationships and Y genome origin in *Elymus* L. sensu lato (Triticeae; Poaceae) based on single-copy nuclear *Acc1* and *Pgk1* gene sequences. Mol. Phylogenet. Evol. 69:919–928. doi:10.1016/j.ympev.2013.06.012

Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644–652. doi:10.1038/nbt.1883

Guo, S.G., J.G. Liu, Y. Zheng, M.Y. Huang, H.Y. Zhang, G.Y. Gong, H.J. He, Y. Ren, S.L. Zhong, Z.J. Fei, and Y. Xu. 2011. Characterization of transcriptome dynamics during watermelon fruit development: Sequencing, assembly, annotation and gene expression profiles. BMC Genomics 12:454. doi:10.1186/1471-2164-12-454

Guo, Y.X., C.Z. Wang, X.B. Yan, R.Y. Wang, S.S. Liu, and M. Tian. 2014. Markers associated with altitude and ecological factors in hexaploid *Campeiostachys nutans* (Griseb.) J. L. Yang, B. R. Baum et C. Yen and tetraploid *Roegeneria nutans* (Keng) Keng. Genet. Resour. Crop Ev. 61:511–521. doi:10.1007/s10722-013-0055-4

Huang, L.L., X. Yang, P. Sun, W. Tong, and S.Q. Hu. 2012. The first Illumina-based *de novo* transcriptome sequencing and analysis of safflower flowers. PLoS ONE 7:e38653. doi:10.1371/journal.pone.0038653

Kaur, S., N.O. Cogan, L.W. Pembleton, M. Shinozuka, K.W. Savin, M. Materne, and J.W. Forster. 2011. Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene

assembly and SSR marker discovery. BMC Genomics 12:265. doi:10.1186/1471-2164-12-265

Koilkonda, P., S. Sato, S. Tabata, K. Shirasawa, H. Hirakawa, H. Sakai, S. Sasamoto, A. Watanabe, T. Wada, Y. Kishida, H. Tsuruoka, T. Fujishiro, M. Yamada, M. Kohara, S. Suzuki, M. Hasegawa, H. Kiyoshima, and S. Isobe. 2012. Large-scale development of expressed sequence tag-derived simple sequence repeat markers and diversity analysis in *Arachis* spp. Mol. Breed. 30:125–138. doi:10.1007/s11032-011-9604-8

Kuleung, C., P.S. Baenziger, and I. Dweikat. 2004. Transferability of SSR markers among wheat, rye, and triticale. Theor. Appl. Genet. 108:1147–1150. doi:10.1007/s00122-003-1532-5

Kumpatla, S.P., and S. Mukhopadhyay. 2005. Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. Genome 48:985–998. doi:10.1139/g05-060

Li, D.J., Z. Deng, B. Qin, X.H. Liu, and Z.H. Men. 2012. *De novo* assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). BMC Genomics 13:192. doi:10.1186/1471-2164-13-192

Liu, Z.P., T.L. Chen, L.C. Ma, Z.G. Zhao, P.X. Zhao, Z.B. Nan, and Y.R. Wang. 2013. Global transcriptome sequencing using the Illumina platform and the development of EST-SSR markers in autotetraploid alfalfa. PLoS ONE 8:e83549. doi:10.1371/journal.pone.0083549

Liu, Z.P., P. Liu, D. Luo, W.X. Liu, and Y.R. Wang. 2014. Exploiting Illumina sequencing for the development of 95 novel polymorphic EST-SSR markers in common vetch (*Vicia sativa* subsp. *sativa*). Molecules 19:5777–5789. doi:10.3390/molecules19055777

Liu, Z.P., G.S. Liu, and Q.C. Yang. 2007. A novel statistical method for assessing SSR variation in autotetraploid alfalfa (*Medicago sativa* L.). Genet. Mol. Biol. 30:385–391.

Lu, B.R. 1993. Meiotic studies of *Elymus nutans* and *Elymus jacquemontii* (Poaceae, Triticeae) and their hybrids with *Pseudoroegneria spicata* and seventeen *Elymus* species. Plant Syst. Evol. 186:193–211. doi:10.1007/BF00940798

Miao, J.M., X.Q. Zhang, S.Y. Chen, X. Ma, Z.H. Chen, J.C. Zhong, and S.Q. Bai. 2011. Gliadin analysis of *Elymus nutans* Griseb. from the Qinghai–Tibetan Plateau and Xinjiang, China. Grassl. Sci. 57:127–134. doi:10.1111/j.1744-697X.2011.00219.x

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

Parra-González, L.B., G.A. Aravena-Abarzua, C.S. Navarro-Navarro, J. Udall, J. Maughan, L.M. Peterson, H.E. Salvo-Garrido, and I.J. Maureira-Butler. 2012. Yellow lupin (*Lupinus luteus* L.) transcriptome sequencing: Molecular marker development and comparative studies. BMC Genomics 13:425. doi:10.1186/1471-2164-13-425

Rowland, L.J., N. Alkharouf, O. Darwish, E.L. Ogden, J.J. Polashock, N.V. Bassil, and D. Main. 2012. Generation and analysis of blueberry transcriptome sequences from leaves, developing fruit, and flower buds from cold acclimation through deacclimation. BMC Plant Biol. 12:46. doi:10.1186/1471-2229-12-46

Shahin, A., M. van Kaauwen, D. Esselink, J.W. Bargsten, J.M. van Tuyl, R.G.F. Visser, and P. Arens. 2012. Generation and analysis of expressed sequence tags in the extreme large genomes *Lilium* and *Tulipa*. BMC Genomics 13:640. doi:10.1186/1471-2164-13-640

Shirasawa, K., K. Ishii, C. Kim, T. Ban, M. Suzuki, T. Ito, T. Muranaka, M. Kobayashi, N. Nagata, S. Isobe, and S. Tabata. 2013. Development of *Capsicum* EST-SSR markers for species identification and *in silico* mapping onto the tomato genome sequence. Mol. Breed. 31:101–110. doi:10.1007/s11032-012-9774-z

Thiel, T., W. Michalek, R.K. Varshney, and A. Graner. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. 106:411–422.

Wang, R.R.C., and B.R. Lu. 2014. Biosystematics and evolutionary relationships of perennial Triticeae species revealed by genomic analyses. J. Syst. Evol. 52:697–705. doi:10.1111/jse.12084

Wang, Z., G.H. Yu, B.B. Shi, X.M. Wang, H.P. Qiang, and H.W. Gao. 2014. Development and characterization of simple sequence repeat (SSR) markers based on RNA-sequencing of *Medicago sativa* and *in silico* mapping onto the *M. truncatula* genome. PLoS ONE 9:e92029. doi:10.1371/journal.pone.0092029

Wang, Z.Y., B.P. Fang, J.Y. Chen, X.J. Zhang, Z.X. Luo, L.F. Huang, X.L. Chen, and Y.J. Li. 2010. *De novo* assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*). BMC Genomics 11:726. doi:10.1186/1471-2164-11-726

Wang, Z.Y., J. Li, Z.X. Luo, L.F. Huang, X.L. Chen, B.P. Fang, Y.J. Li, J.Y. Chen, and X.J. Zhang. 2011. Characterization and development of EST-derived SSR markers in cultivated sweet potato (*Ipomoea batatas*). BMC Plant Biol. 11:139. doi:10.1186/1471-2229-11-139

Wei, W.L., X.Q. Qi, L.H. Wang, Y.X. Zhang, W. Hua, D.H. Li, H.X. Lv, and X.R. Zhang. 2011. Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. BMC Genomics 12:451. doi:10.1186/1471-2164-12-451

Wu, T.Q., S.B. Luo, R. Wang, Y.J. Zhong, X.M. Xu, Y.E. Lin, X.M. He, B.J. Sun, and H.X. Huang. 2014. The first Illumina-based *de novo* transcriptome sequencing and analysis of pumpkin (*Cucurbita moschata* Duch.) and SSR marker development. Mol. Breed. 34:1437–1447. doi:10.1007/s11032-014-0128-x

Xin, D.W., J.Y. Sun, J.L. Wang, H.W. Jiang, G.H. Hu, C.Y. Liu, and Q.S. Chen. 2012. Identification and characterization of SSRs from soybean (*Glycine max*) ESTs. Mol. Biol. Rep. 39:9047–9057. doi:10.1007/s11033-012-1776-8

Yan, X.B., Y.X. Guo, F.Y. Liu, C. Zhao, Q.L. Liu, and B.R. Lu. 2010. Population structure affected by excess gene flow in self-pollinating *Elymus nutans* and *E. burchan-buddae* (Triticeae: Poaceae). Popul. Ecol. 52:233–241. doi:10.1007/s10144-009-0169-x

Yan, X.B., Y.X. Guo, C. Zhao, F.Y. Liu, and B.R. Lu. 2009. Intra-population genetic diversity of two wheatgrass species along altitude gradients on the Qinghai-Tibetan Plateau: Its implication for conservation and utilization. Conserv. Genet. 10:359–367. doi:10.1007/s10592-008-9596-z

Yan, X.B., Y.X. Guo, H. Zhou, B.R. Lu, and K. Wang. 2006. Genetic patterns of ten *Elymus* species from the Tibetan and Inner Mongolian plateaus of China. Grass Forage Sci. 61:398–404. doi:10.1111/j.1365-2494.2006.00547.x

Ye, J., L. Fang, H. Zheng, Y. Zhang, J. Chen, Z. Zhang, J. Wang, S. Li, R. Li, L. Bolund, and J. Wang. 2006. WEGO: A web tool for plotting GO annotations. Nucleic Acids Res. 34:W293–W297. doi:10.1093/nar/gkl031

Zeng, S.H., G. Xiao, J. Guo, Z.J. Fei, Y.Q. Xu, B.A. Roe, and Y. Wang. 2010. Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. et Zucc.) Maxim. BMC Genomics 11:94. doi:10.1186/1471-2164-11-94

Zhou, C.P., X.D. He, F.G. Li, Q.J. Weng, X.L. Yu, Y. Wang, M. Li, J.S. Shi, and S.M. Gan. 2014a. Development of 240 novel EST-SSRs in *Eucalyptus* L'Hérit. Mol. Breed. 33:221–225. doi:10.1007/s11032-013-9923-z

Zhou, Q., T.L. Chen, Y.R. Wang, and Z.P. Liu. 2014b. The development of 204 novel EST-SSRs and their use for genetic diversity analyses in cultivated alfalfa. Biochem. Syst. Ecol. 57:227–230. doi:10.1016/j.bse.2014.08.023