

RESEARCH

Open Access



Development of unigene-derived SSR markers from RNA-seq data of *Uraria lagopodioides* (Fabaceae) and their application in the genus *Uraria* Desv. (Fabaceae)

Chaoyu Liu¹, Maomao Zhang¹ and Xueli Zhao^{1,2*}

Abstract

Background *Uraria* Desv. belongs to the tribe Desmodieae (Fabaceae), a group of legume plants, some of which have medicinal properties. However, due to a lack of genomic information, the interspecific relationships, genetic diversity, population genetics, and identification of functional genes within *Uraria* species are still unclear.

Results Using RNA-Seq, a total of 66,026 *Uraria lagopodioides* unigenes with a total sequence content of 52,171,904 bp were obtained via de novo assembly and annotated using GO, KEGG, and KOG databases. 17,740 SSRs were identified from a set of 66,026 unigenes. Cross-species amplification showed that 54 out of 150 potential unigene-derived SSRs were transferable in *Uraria*, of which 19 polymorphic SSRs were developed. Cluster analysis based on polymorphisms successfully distinguished seven *Uraria* species and revealed their interspecific relationships. Seventeen samples of seven *Uraria* species were clustered into two monophyletic clades, and phylogenetic relationships of *Uraria* species based on unigene-derived SSRs were consistent with classifications based on morphological characteristics.

Conclusions Unigenes annotated in the present study will provide new insights into the functional genomics of *Uraria* species. Meanwhile, the unigene-derived SSR markers developed here will be invaluable for assessing the genetic diversity and evolutionary history of *Uraria* and relatives.

Keywords *Uraria*, Transcriptome sequencing, Unigene-derived SSR, Interspecific transferability, Genetic diversity

Background

Uraria is a genus of legumes that contains ca. 20 species that are mainly distributed in tropical and subtropical Asia, Australia, and Africa [1–3]. Several species of

Uraria (e.g., *U. lagopodioides*, *U. crinita* and *U. picta*) are used for medicinal purposes. The flavonoids, triterpenes, megastigmanes, nucleoside compounds, and 3-hydroxy-7,4'-dimethoxyflavone they produce display a wide range of medicinal properties used in the treatment of asthma, dysentery, ulcers, and malaria-induced fever [4–7].

Previous studies of *Uraria* have mainly focused on morphology, geographical distribution, and palynology, with limited phylogenetic analyses of *Uraria* within the tribe Desmodieae. [1–3, 8–17]. Jabbour et al. conducted molecular phylogenetic and historical biogeographic analyses of Desmodieae genera endemic to New Caledonia using both nuclear (ITS1) and chloroplast (*rbcL*,

*Correspondence:

Xueli Zhao

zhaoxueli@swfu.edu.cn

¹ College of Forestry, Southwest Forestry University, Kunming 650224, China

² Key Laboratory of National Forestry and Grassland Administration on Biodiversity Conservation in Southwest China, Southwest Forestry University, Kunming 650224, China



psbA-trnH) fragments, though only three *Uraria* species were included [18]. Ohashi et al. provided a new classification of Desmodieae using a single nuclear (ITS) and nine chloroplast (5'*trnK* intron, *ndhJ-trnL-trnF*, *trnT-trnL*, *trnG-trnS*, *trnQ-rps16*, *trnL-rpl32*, *rpl16* intron, *trnC-rpoB* and *ndhA* intron) fragments, including sequences from six *Uraria* species. New taxonomic treatments were proposed based on phylogenetic analyses with morphological and palynological characters. *Desmodium oblongum* Wall. ex Benth. was transferred to *Uraria*, as a synonym of *Uraria oblonga* (Wall. ex Benth.) H. Ohashi & K. Ohashi [19]. However, the phylogenetic relationships and evolutionary history of *Uraria*, especially for phylogenetically related species, are still largely uncharacterized.

DNA-based molecular markers such as restriction length fragment polymorphisms (RFLPs), random amplified polymorphic DNAs (RAPDs), and simple sequence repeats (SSRs) have been employed effectively in numerous studies of genetic diversity [20–22], among which SSRs are popular for the differentiation of heterozygotes and homozygotes, reliable reproducibility, and cost-effectiveness. SSRs consist of tandem units of short nucleotide motifs of 1–6 bp in length [23–26]. SSRs can be developed from both non-expressed regions and expressed regions, referred to as genomic SSRs and genic SSRs, respectively [27–30]. Unigenes from the expressed regions are the longest transcripts in genes and have been widely used for SSR marker development. Compared to genomic SSRs, unigene-derived SSRs are more likely to be transferable and orthologous and have been widely used in phylogenetic and population genetic studies, especially for analyses of genetic diversity among phylogenetically related species [31, 32]. Next-generation sequencing, especially RNA-Seq using an Illumina platform, has been used as a rapid and cost-effective solution for identifying and developing SSR markers in non-model plants [33–35].

The objectives of this study were: (1) to enrich *Uraria* transcriptome data and better understand the functional significance of expressed genes, (2) to develop unigene-derived SSRs and examine both their cross-species transferability and levels of polymorphism, and (3) to reconstruct the genetic relationships of *Uraria* species.

Results

Illumina sequencing and de novo transcriptome assembly

A total of 8.23 Gb of clean data were obtained, and the Q20, Q30, and GC contents were 97.34%, 92.37%, and 43.98%, respectively. A total of 66,026 unigenes were assembled, of which there were 337,837 unigenes with a length of 200–500 bp, 12,769 with a length of 500–1000 bp, 9,297 with a length of 1–2 kb, and 6,123 with

a length of more than 2 kb. The N50 of the unigenes was 1,850, indicating a high-quality assembly.

Functional annotation of unigenes

To annotate *U. lagopodioides* unigenes, sequences from 66,026 unigenes were searched against different universal databases. 31,065 (47.04%) unigenes were aligned to sequences in the Nr database, 35,722 (54.10%) in the Nt database, 23,160 (35.07%) in the Swiss-Prot database, and 21,178 (32.07%) in the Pfam database. The annotation of 39,915 (60.45%) unigenes was achieved in at least one database.

According to gene ontology (GO) analyses, 21,178 (32.08%) annotated unigenes could be assigned to three functional categories: biological processes, molecular functions, and cellular components (Fig. 1a). In “biological process”, the largest classes were “cellular process” (11,806, 17.88%), “metabolic process” (11,153, 16.89%), and “single organization process” (8,708, 13.19%). The cellular component category mainly consists of genes assigned to “cell” (5,993, 9.08%) and “cell part” (5,993, 9.08%) categories. The largest class identified in the molecular function category was “binding” (11,790, 17.86%). According to the KOG database, 6,614 unigenes (10.01%) were categorized into 25 functional groups (Fig. 1b), of which 879 were annotated as “general functional” genes, followed by “post-translational modification, protein turnover, chaperones” (852), and “translation, ribosomal structure, and biogenesis” (630). “Cell motility” (7) and “extracellular structures” (8) were the least frequently observed KOG classifications. According to the KEGG database, 10,956 unigenes were categorized into 19 biological pathways in five large groups (cellular processes, environmental information processing, genetic information processing, metabolism, and organismal systems) (Fig. 1c). Among them, the three most frequently observed functional pathways were “carbohydrate metabolism” (1,007), “translation” (783), and “overview” (676).

Identification and characteristics of unigene-derived SSRs

A total of 17,740 potential unigene-derived SSRs were identified from the set of 66,026 unigenes (52,171,904 bp), with 2,952 unigenes containing more than one SSR locus. Of the 17,740 SSRs, 1,156 were presented in compound formation. These SSRs were further divided into six different types based on unit size, of which the mono-nucleotide repeats exhibited the highest frequency of occurrence (10,769, 60.70%), followed by tri-nucleotides (3,337, 18.81%), di-nucleotides (3,288, 18.53%), tetra-nucleotides (288, 1.62%), penta-nucleotides (36, 0.20%), and hexa-nucleotides (22, 0.12%) (Table 1). The most frequent mono-nucleotide

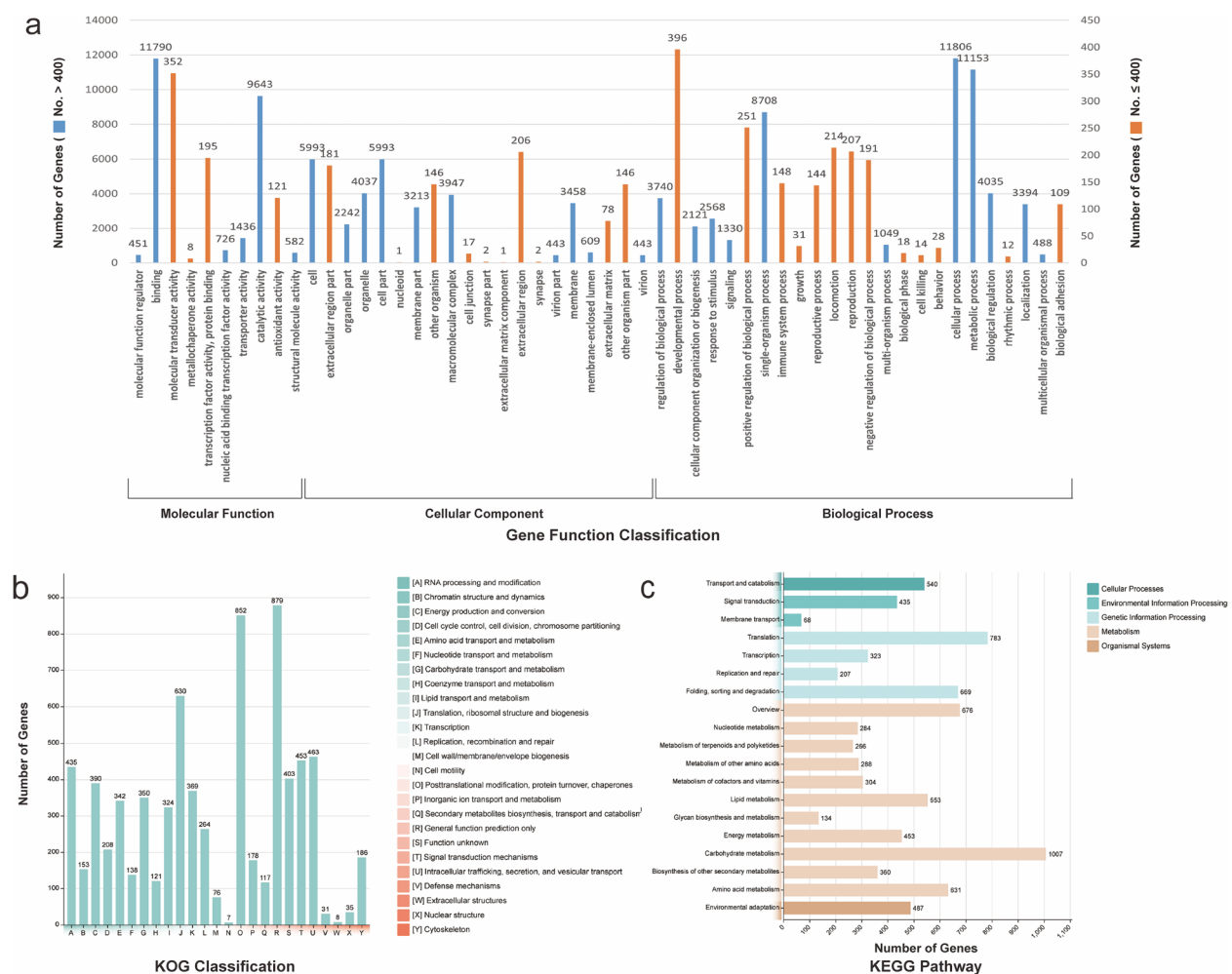


Fig. 1 Gene annotations of *U. lagopodioides* unigenes based on GO, KOG, and KEGG databases. **a.** GO annotations of *U. lagopodioides* unigenes. **b.** KOG classifications of *U. lagopodioides* unigenes. **c.** KEGG classifications of *U. lagopodioides* unigenes

repeats were A/T (10,635), accounting for 59.95% of the total SSRs. Of the tri-nucleotide repeats, AAG/CTT (799, 4.50%) was the most abundant motif, followed by AAT/ATT (662, 3.73%) and AAC/GTT (558, 3.15%). The most abundant di-nucleotide, tetra-nucleotide, and penta-nucleotide repeats were AG/CT (1,731, 9.76%), AAAT/ATTT (68, 0.38%), and AACAC/GTGT (3, 0.02%), respectively. The number of repeats ranged from 5 to 36, with 10, 5, and 6 being the most frequent (Additional file 1: Table S1).

We analyzed the distribution of SSRs in the 3' UTR, 5' UTR and CDS regions of the genome (Fig. 2). There were 2,977, 2,357 and 942 SSRs distributed in 5'UTR, 3'UTR and CDS, respectively. The trinucleotide repeat sequence is the most abundant in the CDS region. The number of SSRs in the UTR region was significantly higher than that

in the CDS region, and most SSRs were distributed in 5' UTR.

Development of polymorphic unigene-derived SSR markers

To validate primers designed to detect unigene-derived SSRs, 150 potential unigene-derived SSRs were randomly selected and tested in *Uraria*. Fifty-four of these were successfully amplified and produced amplicons of expected size using genomic DNA as a template, while the remaining 96 failed to amplify despite trying a range of annealing temperatures (Additional file 2: Table S2). Using 17 individuals from seven *Uraria* species, 19 of 54 unigene-derived SSRs showed high levels of polymorphism and good transferability among different *Uraria* species (Table 2).

Table 1 Length distributions of the unigene-derived SSRs of *U. lagopodioides* based on the number of nucleotide repeat units

No. of repeats	Mono-	Di-	Tri-	Tetra-	Penta-	Hexa-	Total	Percentage (%)
5			1789	251	34	7	2081	11.73
6		1082	919	31	1	9	2042	11.51
7		595	588	5	1	4	1193	6.72
8		462	35			2	499	2.81
9		504	2	1			507	2.86
10	3234	480	4				3718	20.96
11	1776	158					1934	10.9
12	1146	5					1151	6.49
13	815	1					816	4.6
14	659						659	3.71
15	582						582	3.28
16	534						534	3.01
17	489						489	2.76
18	511						511	2.88
19	465						465	2.62
20	324						324	1.83
21	163						163	0.92
22	54						54	0.3
23	15						15	0.08
24	2						2	0.01
36		1					1	0.01
Total	10,769	3288	3337	288	36	22	17,740	
Percentage (%)	60.7	18.53	18.81	1.62	0.2	0.12		

Cluster analysis of *Uraria* based on unigene-derived SSRs

The r -value of matrix correlation was 0.847, and the value of the approximate mantel t -test was 9.869. The topology of the unweighted pair-group method analysis (UPMGA) tree based on genetic distance was used to show the relationships of *Uraria* species (Fig. 3). UPMGA cluster

analysis revealed that 17 samples from seven *Uraria* species were clustered into two monophyletic clades. *Uraria oblonga*, *U. lacei*, and *U. sinensis* were clustered into Clade I. *U. lagopodioides*, *U. rufescens*, *U. crinita*, and *U. picta* were clustered into Clade II, indicating close genetic relationships.

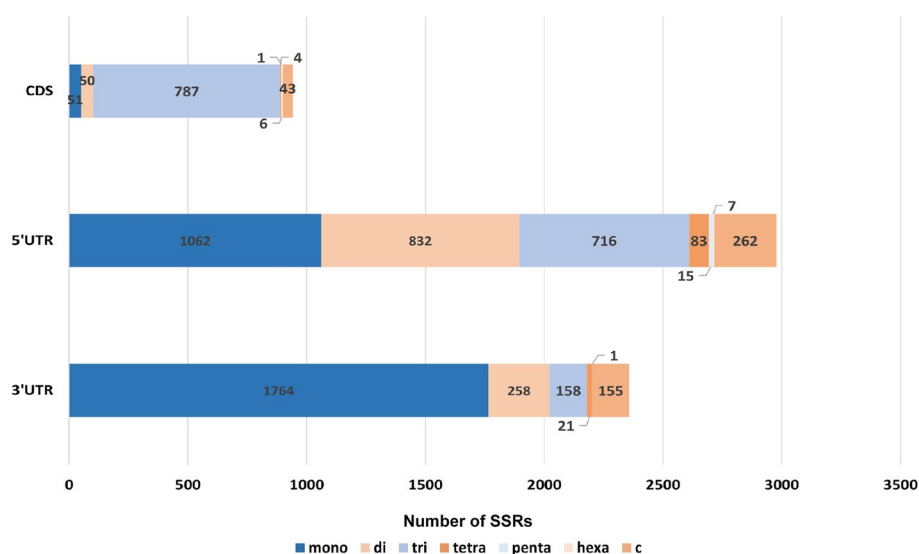
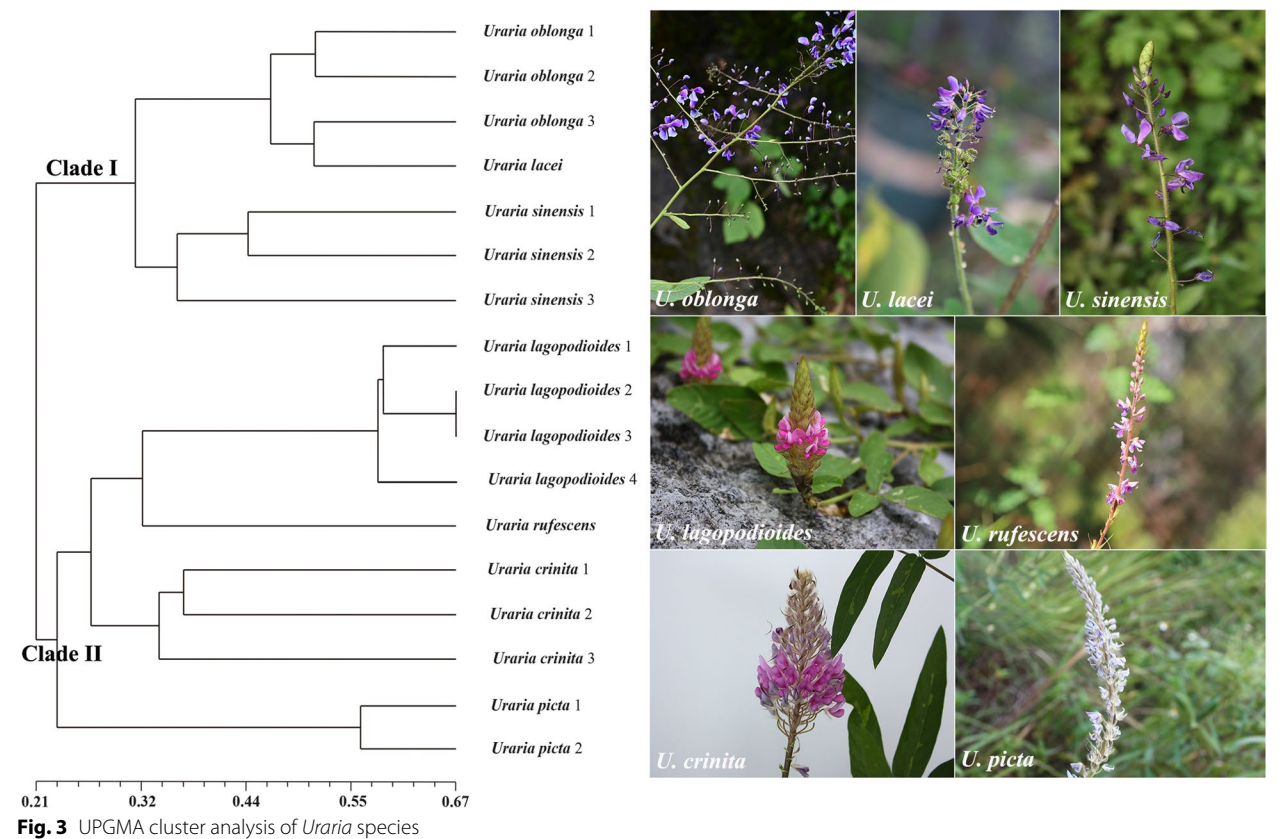
**Fig. 2** Frequency and distribution of SSRs in coding sequence (CDS) and untranslated region (UTRs) of *U. lagopodioides*

Table 2 Characteristics of 19 polymorphic unigene-derived SSRs

SSR locus	Forward primer (5'-3')	Reverse primer (5'-3')	Repeat motif	Ta (°C)	Observed product size (bp)	Na
P6	GACACGCTCATGTCGCATTC	GAGCATTGAAGTCGGTGCG	(TG)8	55	194–227	8
P8	ACGGTCGATGAGTTTGTCTG	ACTGTGTGCTGTCTGTAGAA	(AG)8	55	231–297	2
P25	TGATGTCCTCCACATGTGCC	GCCACTCACCTAGAAGAGCC	(GTGA)7	55	268–281	5
P30	AGATTGCAGCAAGGGTCTCA	GGCACCTCCAATTCCTAGTCA	(TTCT)5	54	177–195	3
P32	CCAGCAGGTACATGTCTGGG	AGGTATTGTAGTCAAAGAACCAGT	(GT)9	53	168–177	3
P38	AGAGTTGCGACGCATAACCA	CCGTAAAGTCGCGTGAAAC	(TA)8	55	225–265	3
P39	ACGCATATCGGGTCAGTAAACA	GTATTACAAATATTACATGCACACTGC	(GT)6(T)11	53	205–247	5
P40	TGGTTGACCGCAGCATAGTT	GAAGGTGCTTGTCCATGGGA	(TGG)6	55	458–498	7
P43	TGCTGCACCTTCACATTCA	GGGTGAGTTTGAGGAGTCG	(AC)8	55	224–255	4
P46	GGAGCAATGTCACCAAGGGA	CCATCTTAGAGCTGGCCACA	(GATG)5	55	355–370	1
P55	GCAGGTTTACGTCCAATGCT	ACAGTGTGCCCTCACGAAAT	(TATC)5	54	217–247	3
P59	ACCCTTCACAACTCCCTCTC	AGGCTGAAAGAATGCCTCC	(TTG)6	54	260–274	2
P71	CCACACCACAGCAGAACAGA	GTTGCGGCAAGGGAAGAGAA	(AC)6*(CT)9	55	154–225	10
P127	TCAACCACCGTAGCTGAACC	CGGATCTGAGAGCGGAAGAC	(GGA)6	55	231–277	5
P131	TGACAAGCCTTCCCGAACTC	CTCTAGGCACCTGCAAGAGA	(TTTA)5	54	243–251	2
P132	CGTTGAGACCCCAAGTCAGT	AGCATCACAGCCATCAACGA	(ATTA)5	55	237–286	2
P138	CAGCGTCCTCTGTGTTCTGT	AGTTCCGATTTCGGAAGCGA	(AT)9	55	217–285	7
P141	GAGGAGAAGGCGGTGATGAC	TAGAAGCAGCACCACCACTG	(TGG)7	55	275–296	3
P149	ACCGCAACTATTACCTCCT	AGAATTCAGCACACGGAGCA	(TA)10	54	120–176	5

*GCACCTTACTCACTCTCTGTACTCTTCACACA



Discussion

RNA-Seq is a cost-efficient and powerful technology for generating high-coverage transcriptome data, and it has been increasingly used for detecting functional genes and identifying molecular markers in non-model plants such as *Panax vietnamensis* [36], *Acrocomia aculeata* [37], *Luculia yunnanensis* [38], *Paris polyphylla* [39], and *Bromus catharticus* [31]. However, no transcriptome sequencing of *Uraria* species has been reported thus far. In the present study, we reported the first transcriptome sequence data of *U. lagopodioides* using Illumina RNA-Seq technology. A total of 55,933,282 paired-end raw reads were generated, of which 54,843,810 were high-quality clean reads. 97.34% of reads had minimum quality scores of Q20, indicative of high-quality sequencing [28, 40, 41].

Previous studies have shown that unigenes longer than 500 bp are more amenable to annotation efforts, while reads with shorter lengths are more difficult to annotate and categorize [42–45]. In the present study, a total of 66,026 unigenes were assembled from the *U. lagopodioides* transcriptome with an average length of 1,041 bp and N50 length of 1,850 bp, which was longer than those reported in the studies of *Panax vietnamensis* (598.32 bp and 1,268 bp, respectively) [36], *Brassica napus* (834 bp and 1,245 bp) [46], *Parrotia subaequalis* (890 bp and 1,591 bp) [47], and *Vigna aconitifolia* (937.78 bp and 1,227 bp) [48], but shorter than *Phoebe bournei* (1,019 bp and 2,016 bp) [49] and *Lathyrus sativus* (1,250 bp and 1,781 bp) [50]. The unigenes generated in this study will be valuable for characterizing molecular mechanisms and exploring novel functional genes in *Uraria* and related taxa. To obtain comprehensive gene function categories of *U. lagopodioides*, we performed gene function annotations using the public databases KOG [51], GO [52], and KEGG [53]. In sum, 4,095 of 66,026 unigenes were functionally annotated in all three databases, and 39,915 were functionally annotated in at least one database. The low percentage of annotated unigenes may be a consequence of the relative dearth of related species in these databases or a relatively large proportion of non-coding regions in the *U. lagopodioides* transcriptome sequence [38, 40, 54, 55].

As a result of gene function annotation, 21,178 unigenes (32.07%) were classified into GO categories. The largest GO category was “cellular process”, followed by “binding”. A total of 10,956 unigenes (16.59%) were annotated using the KEGG database, with the largest group of genes categorized as “carbohydrate metabolism”, followed by “translation”. According to the KEGG database, many unigenes were classified in metabolism or genetic information processing categories, which will be useful for

future characterization of the physiology, biochemistry, and functional genomics of *Uraria*.

Unigene-derived SSR markers have been widely used in studies of genetic diversity and population genetics, especially for phylogenetically related species [56–58]. In this study, polymorphic SSR markers of *U. lagopodioides* were developed using NGS technology. A total of 17,740 potential SSRs were identified from the set of 66,026 unigenes, with 26.9% of unigenes containing an SSR and an average distribution density of one SSR per 2.94 kb. The number and distribution density of SSRs in *U. lagopodioides* were significantly higher than those in *Argyranthemum broussonetii* (2.3% and 27 kb, respectively) [59], *Opisthopappus* (7.78% and 10.30 kb) [60], and *Arachis hypogaea* (17.7% and 3.30 kb) [61]. The differences in SSR abundance and frequency among different species may be partially attributed to the size of the unigene assembly dataset, SSR search criteria, sequence redundancy, database mining tools, and actual differences between species [62–64].

Among the identified SSRs, mono-nucleotide repeats are the most frequently observed, followed by tri-nucleotide and di-nucleotide repeats. For mono-nucleotide motifs, the proportion of the A/T motif (59.95%) was significantly higher than that of G/C (0.76%), which was consistent with most previous studies of other plants [25, 65–67]. The most abundant di-nucleotide motif was AG/TC (5.93%), followed by AT/TA (5.08%). The number of AT-containing repeats was significantly higher than that of GC-containing repeats, which suggests that these sequences are relatively unstable and prone to base substitution and gene mutation [68, 69]. The results of this study showed that the number of simple repeats of *U. lagopodioides* was negatively correlated with the size of SSR bases. Mono-nucleotide, di-nucleotide and tri-nucleotide repeats accounted for the majority of SSR loci (98.04%), while tetra-nucleotide, penta-nucleotide and hexa-nucleotide repeat unit combinations accounted for only 1.96%. The existence of a large number of short-repeat SSR loci may be due to the high mutation frequency and high rate of evolution of the genome itself. There are significant differences in the distribution of SSRs in different functional regions of the genome. SSRs located in the CDS region can affect gene activation and protein expression, while those located in the non-coding region and UTR region may affect gene regulation and translation.

Most previous studies of *Uraria* focused on their medicinal value [4, 6, 7], while studies on the taxonomy and evolution of *Uraria* were limited. DNA fragments involved in the previous phylogenetic studies were relatively conserved, limiting their value for analyses within

Uraria [18, 19]. Therefore, unigene-derived SSR markers developed in the present study will be invaluable for further population genetic studies of *Uraria* species.

Using cluster analysis, 17 samples of seven *Uraria* species were clustered into two monophyletic clades, with samples from each species forming monophyletic clusters. *Uraria oblonga*, *U. lacei*, and *U. sinensis* were clustered in clade I. *U. lagopodioides*, *U. rufescens*, *U. crinita*, and *U. picta* were clustered in clade II. Interspecific relationships revealed by the cluster analysis based on the 19 unigene-derived SSRs were consistent with the inflorescence type of *Uraria* species. The inflorescence type of species in clade I is panicles, while that of species nested within clade II is racemes. The results of this study demonstrated that phylogenetic analysis based on unigene-derived SSRs can provide valuable evidence for the taxonomy and evolution of *Uraria*.

Conclusions

In this study, we assembled and annotated a large number of unigenes of *U. lagopodioides* using RNA-Seq technology and also characterized and evaluated a number of unigene-derived SSR markers derived from the transcriptome of *U. lagopodioides*. A total of 54 unigene-derived SSRs were verified to be of cross-species transferability in *Uraria*, 19 of which displayed polymorphisms useful for phylogenetic studies. These results will provide a theoretical basis for further functional genomics, population genetics, and phylogenetic analyses of *Uraria* and relatives.

Methods

Plant materials and RNA / DNA extraction

The *U. lagopodioides* plant materials for RNA isolation and transcriptome sequencing were collected from Yuanjiang County, Yunnan Province in June 2018. Fresh leaf tissues were cleaned and immediately preserved in liquid nitrogen until RNA extraction. Total RNA was isolated using TRIzol Reagent (Invitrogen, CA, USA), and RNA purity was checked using the NanoPhotometer spectrophotometer (IMPLEN, CA, USA). RNA integrity was assessed using the RNA Nano 6000 Assay Kit on the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA) by Novogene (Beijing, China). The transcriptome of *Uraria lagopodioides* was sequenced using the Illumina HiSeq 2500 platform by Novogene (Beijing, China).

For identifying polymorphisms and testing the cross-species transferability of the developed unigene-derived SSR markers, 17 individuals representing seven *Uraria* species were sampled. Voucher information is provided in Table 3. The materials were identified by Dr. Xueli Zhao according to *Flora of China* [1] and deposited at the Herbarium of Southwest Forestry University (SWFU).

Table 3 Voucher information of the materials used in this study

Species	Voucher	Locality
<i>U. lagopodioides</i> 1	WL (SWFU)	Qingyuan, Guangdong
<i>U. lagopodioides</i> 2	ZXL592 (SWFU)	Baise, Guangxi
<i>U. lagopodioides</i> 3	ZXL441 (SWFU)	Chongzuo, Guangxi
<i>U. lagopodioides</i> 4	ZXL381-2 (SWFU)	Yvxi, Yunnan
<i>U. oblonga</i> 1	CB2019121901 (SWFU)	Changjiang, Hainan
<i>U. oblonga</i> 2	ZHL2019 (SWFU)	Dali, Yunnan
<i>U. oblonga</i> 3	ZXX20211003 (SWFU)	Maguan, Yunnan
<i>U. lacei</i>	WL201809 (SWFU)	Qingyuan, Guangdong
<i>U. picta</i> 1	ZXL663-3 (SWFU)	Chuxiong, Yunnan
<i>U. picta</i> 2	0911-1 (SWFU)	Chongzuo, Guangxi
<i>U. rufescens</i>	YLZB4792 (SWFU)	Mengla, Yunnan
<i>U. sinensis</i> 1	ZZM1143 (SWFU)	Leibo, Sichuan
<i>U. sinensis</i> 2	ZZM1045-2 (SWFU)	Gongshan, Yunnan
<i>U. sinensis</i> 3	ZZM1177-11 (SWFU)	Luding, Sichuan
<i>U. crinita</i> 1	ZXL558-1 (SWFU)	Huizhou, Guangdong
<i>U. crinita</i> 2	YLZB4783 (SWFU)	Mengla, Yunnan
<i>U. crinita</i> 3	DTT201804 (SWFU)	Wenchang, Hainan

Total genomic DNA was extracted from silica-gel-dried leaves with the TIANGEN plant genomic DNA extraction kit (TIANGEN Biotech, Beijing, China) following the manufacturer's protocol.

RNA-Seq library construction, sequencing, and transcriptome assembly

A total amount of 3 µg RNA per sample was used as input material for the RNA-Seq sample preparations. Sequencing libraries were generated using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, USA). mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation buffer was added to mRNA samples, and these were then randomly sheared into 150–200 bp fragments. The library preparations were sequenced on an Illumina HiSeq 2500 platform by Novogene (Beijing, China). Transcriptome assembly was performed using Trinity [70]. The RNA-seq data have been submitted to the NCBI Sequence Read Archive (SRR21474487, <https://www.ncbi.nlm.nih.gov/sra/SRR21474487>). Gene function annotations using multiple databases were performed to obtain comprehensive gene function information. Diamond v0.8.22 (<http://www.ncbi.nlm.nih.gov/COG/>) was used to annotate gene functions via the KOG database with the parameter e-value = 1e-3. The KEGG (<http://www.genome.jp/kegg/>) Automatic Annotation Server was used for functional annotation of metabolic pathways and gene products, with the parameter set to e-value = 1e-10 [71]. Protein annotation analysis for GO was performed using Blast2GO v2.5 (<http://www.geneontology.org/>) with the parameter e-value = 1e-6.

Unigene-derived SSR detection, primer design, and marker validation

Potential unigene-derived SSRs were screened using the program MISA 1.0 [72]. The mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides were designed with minimum repeat numbers of 10, 6, 5, 5, 5, and 5 for the SSRs, respectively. 150 SSR primers with no more than 4 consecutive repeat units and a length greater than 18 nucleotides were randomly selected. These SSR primers were synthesized by Sangon Biotech (Shanghai, China).

These 150 unigene-derived SSRs were then tested for proper PCR amplification. PCR reactions were carried out with a 25 µL reaction volume containing 0.5 ng genomic DNA template, 1 µL of each primer (100 µM), 12.5 µL of 2 × SanTaq PCR Mix (Sangon Biotech, Shanghai, China), and 10 µL of ddH₂O. PCR amplification conditions were as follows: initial denaturation at 94°C for 5 min, followed by 35 cycles of 94°C for 30 s, 54°C for 35 s, and 72°C for 60 s, and a final extension of 10 min at 72°C. PCR products were visualized via electrophoresis in 1% agarose gels and 8% polyacrylamide gels (Additional file 3: Fig. S1), and SSRs that could be successfully amplified were selected for polymorphism assessment.

The capillary electrophoresis (CE) PCR amplification was performed in a 25 µL solution containing 20–50 ng DNA, 0.5 µL of each forward primer (10 µM) labeled with a fluorescent dye (FAM, HEX, and TAMRA), 0.5 µL of unlabeled reverse primer (10 µM), 0.5 µL of 5 µM dNTP (mix), 2.5 µL 10 × Taq Buffer (with MgCl₂), and finally ddH₂O to 25 µL. Amplification was performed with initial denaturation of 95°C for 5 min, followed by 10 cycles of 94°C for 30 s, 60°C (-0.5°C/cycle) for 30 s, and 72°C for 30 s. This was followed by a further 30 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s, and a final extension of 10 min at 72°C. The amplification results of the SSR primers were analyzed with GeneMapper software (Applied Biosystems).

SSR primer data and population genetic analyses

For SSR data analysis, CE products were manually scored based on allele size. Data were scored as “0” if no band was present and “1” if it was present. UPMGA cluster analysis was conducted using the NTSYSpc program [73].

Abbreviations

GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
KOG	Eukaryotic Ortholog Groups
Pfam	Protein family
UPGMA	Unweighted Pair-Groups Method with Arithmetic Averages
Ta	Annealing temperature

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-023-04086-1>.

Additional file 1: Table S1. Summary of SSRs identified from the transcriptome of *U. lagopodioides*.

Additional file 2: Table S2. Details of 54 transferable SSR markers.

Additional file 3: Fig. S1. Representative polyacrylamide gel profiles.

Acknowledgements

The authors are grateful to Zhangming Zhu and Xinxin Zhou for providing some of the plant materials and photos. We also thank to Tingting Duan, Li Wen, Bin Chen, Hailei Zheng for providing some of the plant materials. Many thanks to Bo Xu, Jia Miao, Zuping Xu and Chao Yuan for their help in the field work.

Authors' contributions

CYL and XLZ conceived the study and designed the experiments. CYL and MMZ performed the experiments. The manuscript was written and revised by CYL and XLZ. All authors read and approved the final version of the manuscript.

Funding

This research was funded by the National Natural Science Foundation of China (Grant No. 31800170); and Scientific Research Fund Project of Yunnan Provincial Department of Education (Grant No. 2021Y241).

Availability of data and materials

The Illumina NGS reads generated in this study have been submitted to the BioProject database of the National Center for Biotechnology Information (SRR21474487).

Declarations

Ethics approval and consent to participate

The authors complied with all relevant institutional, national and international guidelines.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 11 September 2022 Accepted: 25 January 2023

Published online: 10 February 2023

References

- Huang PH, Ohashi H, Oikawa Y. *Uria* Desvaux. In Flora of China, Wu ZY, Raven PH, Eds. Science Press: Beijing, China & Missouri Botanical Garden Press: St. Louis, United State of America. 2010;10:286–8.
- Ohashi H, Iokawa Y, Phon PD. The genus *Uria* (Leguminosae) in China. J Jpn Bot. 2006;81(6):332.
- Yang YC, Huang PH. A revision of the genus *Uria* Desv. (Leguminosae) in China. Bull Bot Res. 1981;1(3):1–20.
- Bhusare BP, Ahire ML, John CK, Nikam TD. *Uria picta*: a comprehensive review on evidences of utilization and strategies of conservation. J Phytol. 2021;13:41–7.
- Oyesiku OO, Okusanya OT, Olowokudejo JD. Morphological and anatomical investigations into the mechanism of leaf pair unrolling in *Uria picta* (Jacq.) Desv. Ex DC. (Papilionaceae), a medicinal plant in Nigeria. Afr J Tradit Complement Altern Med. 2013;10(4):144–50.

6. Thien DD, Tai BH, Dai TD, Sa NH, Thuy TT, Hoang Anh NT, Tam NT. New phenolics from *Uraria crinita* (L.) DC. Nat Prod Res. 2022;36(13):3381–8.
7. Hamid H, Abdullah S, Ali A, Alam M, Ansari SH. Anti-inflammatory and analgesic activity of *Uraria lagopoides*. Pharm Biol. 2008;42(2):114–6.
8. Schindler AK. *Desmodii* generumque affinium species et combinationes novae. II Repert Spec Nov Regni Veg. 1926;22(13–21):250–88.
9. Kumar S, Sane PV. Legumes of South Asia. London, UK: Royal Botanic Gardens. Kew; 2003.
10. De Haas A, Bosman MT, Geesink R. *Urariopsis* reduced to *Uraria* (Leguminosae-Papilionoideae). Blumea. 1980;26(2):439–44.
11. Van Meeuwen MS, Nooteboom H, Steenis C. Preliminary revisions of some genera of Malaysian Papilionaceae I. Reinwardtia. 1961;5(4):426.
12. Gagnepain F, Humbert H. Supplément à la Flora Générale de l'Indochine. 1st ed. Paris: Muséum National d'Histoire naturelle; 1938.
13. Ohashi H. A taxonomic study of the tribe Coronilleae (Leguminosae), with a special reference to pollen morphology. J Fac Sci Univ Tokyo. 1971;11:25–92.
14. Zhu MJ, Miu J, Zhao XL. Simulation of potential distribution of *Uraria* in China based on maximum entropy model. Plant Sci J. 2020;38(04):476–82.
15. Azani N, Babineau M, Bailey CD, Banks H, Barbosa A, Pinto RB, Boatwright J, Borges L, Brown G, Bruneau A, et al. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny—The Legume Phylogeny Working Group (LPWG). Taxon. 2017;66(1):44–77.
16. Ohashi K, Ohashi H, Nemoto T, Ikeda T, Izumi H, Kobayashi H, Muragaki H, Nata K, Sato N, Suzuki M. Phylogenetic analyses for a new classification of the *Desmodium* group of Leguminosae tribe Desmodieae. J Jpn Bot. 2018;93(3):165–89.
17. Zhao XL, Zhu ZM. Comparative genomics and phylogenetic analyses of *Christia vespertilionis* and *Urariopsis brevissima* in the tribe Desmodieae (Fabaceae: Papilionoideae) based on complete chloroplast genomes. Plants. 2020;9(9):1116.
18. Jabbour F, Gaudeul M, Lambourdiere J, Ramstein G, Hassanin A, Labat JN, Sarthou C. Phylogeny, biogeography and character evolution in the tribe Desmodieae (Fabaceae: Papilionoideae), with special emphasis on the New Caledonian endemic genera. Mol Phylogenet Evol. 2018;118:108–21.
19. Ohashi H, Ohashi K. *Grona*, a genus separated from *Desmodium* (Leguminosae tribe Desmodieae). J Jpn Bot. 2018;93(2):104–20.
20. Raizada A, Soufaramanien J. Transcriptome sequencing, de novo assembly, characterisation of wild accession of blackgram (*Vigna mungo* var. *silvestris*) as a rich resource for development of molecular markers and validation of SNPs by high resolution melting (HRM) analysis. BMC Plant Biol. 2019;19(1):358.
21. Peng X, Khayatnezhad M, Ghezjelzheidan L. RAPD profiling in detecting genetic variation in *Stellaria* L. (Caryophyllaceae). Genetika. 2021;53(1):349–62.
22. Chen J, Dong S, Zhang X, Wu Y, Zhang H, Sun Y, Zhang J. Genetic diversity of *Prunus sibirica* L. superior accessions based on the SSR markers developed using restriction-site associated DNA sequencing. Genet Resour Crop Evol. 2020;68(2):615–28.
23. Powell W, Machray GC, Provan J. Polymorphism revealed by simple sequence repeats. Trends in Plant Sci. 1996;1(7):215–22.
24. Tautz D, Renz M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucl Acid Res. 1984;12(10):4127–38.
25. Park S, Son S, Shin M, Fujii N, Hoshino T, Park S. Transcriptome-wide mining, characterization, and development of microsatellite markers in *Lychnis kiusiana* (Caryophyllaceae). BMC Plant Biol. 2019;19(1):14.
26. Preethi P, Rahman S, Naganeeswaran S, Sabana AA, Gangaraj KP, Jerard BA, Niraj V, Rajesh MK. Development of EST-SSR markers for genetic diversity analysis in coconut (*Cocos nucifera* L.). Mol Biol Rep. 2020;47(12):9385–97.
27. Lachheb M, Merzougui SE, Boudadi I, Caid MBE, Mousadik AE, Serghini MA. Assessing genetic diversity using the first polymorphic set of EST-SSRs markers and barcoding of Moroccan saffron. J App Res Med Aromat Plant. 2022;29:100376.
28. Zhang C, Wu Z, Jiang X, Li W, Lu Y, Wang K. De novo transcriptomic analysis and identification of EST-SSR markers in *Stephanandra incisa*. Sci Rep. 2021;11(1):1059.
29. Gulyaeva EN, Tarelkina TV, Galibina NA. Functional characteristics of EST-SSR markers available for Scots pine. Math Biol Bioinform. 2022;17(1):82–155.
30. Debbabi OS, Mnasri SR, Amar FB, Naceur MB, Montemurro C, Miazzi MM. Applications of microsatellite markers for the characterization of olive genetic resources of Tunisia. Genes. 2021;12(2):286.
31. Sun M, Dong Z, Yang J, Wu W, Zhang C, Zhang J, Zhao J, Xiong Y, Jia S, Ma X. Transcriptomic resources for prairie grass (*Bromus catharticus*): expressed transcripts, tissue-specific genes, and identification and validation of EST-SSR markers. BMC Plant Biol. 2021;21(1):264.
32. Chen W, Yang H, Zhong S, Zhu J, Zhang Q, Li Z, Ren T, Tan F, Shen J, Li Q, et al. Expression profiles of microsatellites in fruit tissues of *Akebia trifoliata* and development of efficient EST-SSR markers. Genes. 2022;13(8):1451.
33. Tong Y, Gao LZ. Development and characterization of EST-SSR markers for *Camellia reticulata*. Appl Plant Sci. 2020;8(5):e11348.
34. Li X, Liu X, Wei J, Li Y, Tigabu M, Zhao X. Development and transferability of EST-SSR markers for *Pinus koraiensis* from cold-stressed transcriptome through Illumina sequencing. Genes. 2020;11(5):500.
35. Biswas MK, Bagchi M, Nath UK, Biswas D, Natarajan S, Jesse DMI, Park JI, Nou IS. Transcriptome wide SSR discovery cross-taxa transferability and development of marker database for studying genetic diversity population structure of *Lilium* species. Sci Rep. 2020;10(1):18621.
36. Vu DD, Shah SNM, Pham MP, Bui VT, Nguyen MT, Nguyen TPT. De novo assembly and transcriptome characterization of an endemic species of Vietnam, *Panax vietnamensis* Ha et Grushv., including the development of EST-SSR markers for population genetics. BMC Plant Biol. 2020;20(1):358.
37. Bazzo BR, de Carvalho LM, Carazzolle MF, Pereira GAG, Colombo CA. Development of novel EST-SSR markers in the macaúba palm (*Acrocomia aculeata*) using transcriptome sequencing and cross-species transferability in Arecaceae species. BMC Plant Biol. 2018;18(1):276.
38. Zhang Y, Liu X, Li Y, Liu X, Ma H, Qu S, Li Z. Basic characteristics of flower transcriptome data and derived novel EST-SSR markers of *Luculia yunnanensis*, an endangered species endemic to Yunnan, Southwestern China. Plants. 2022;11(9):1204.
39. Gao X, Su Q, Yao B, Yang W, Ma W, Yang B, Liu C. Development of EST-SSR markers related to polyphenol biosynthesis reveals genetic diversity and population structure in *Paris polyphylla*. Diversity. 2022;14(8):589.
40. Zhang Y, Zhang X, Wang YH, Shen SK. De novo assembly of transcriptome and development of novel EST-SSR markers in *Rhododendron rex* Lévl. through Illumina sequencing. Front Plant Sci. 2017;8:1664.
41. Wei W, Qi X, Wang L, Zhang Y, Hua W, Li D, Lv H, Zhang X. Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. BMC Genomics. 2011;12(1):1–13.
42. Novaes E, Drost DR, Farmerie WG, Pappas GJ, Grattapaglia D, Sederoff RR, Kirst M. High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. BMC Genomics. 2008;9:312.
43. Parchman TL, Geist KS, Grahnen JA, Benkman CW, Buerkle CA. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. BMC Genomics. 2010;11:180.
44. Cardoso-Silva CB, Costa EA, Mancini MC, Balsalobre TW, Canesin LE, Pinto LR, Carneiro MS, Garcia AA, de Souza AP, Vicentini R. De novo assembly and transcriptome analysis of contrasting sugarcane varieties. PLoS ONE. 2014;9(2):e88462.
45. Feng C, Chen M, Xu CJ, Bai L, Yin XR, Li X, Allan AC, Ferguson IB, Chen KS. Transcriptomic analysis of Chinese bayberry (*Myrica rubra*) fruit development and ripening using RNA-Seq. BMC Genomics. 2012;13(1):1–15.
46. Wang D, Yang C, Dong L, Zhu J, Wang J, Zhang S. Comparative transcriptome analyses of drought-resistant and -susceptible *Brassica napus* L. and development of EST-SSR markers by RNA-Seq. J Plant Biol. 2015;58(4):259–69.
47. Zhang Y, Zhang M, Hu Y, Zhuang X, Xu W, Li P, Wang Z. Mining and characterization of novel EST-SSR markers of *Parrotia subaequalis* (Hamamelidaceae) from the first Illumina-based transcriptome datasets. PLoS ONE. 2019;14(5):e0215874.
48. Suranjika S, Pradhan S, Nayak SS, Parida A. De novo transcriptome assembly and analysis of gene expression in different tissues of moth bean (*Vigna aconitifolia*) (Jacq.) Marechal. BMC Plant Biol. 2022;22(1):198.

49. Zhou Q, Zhou PY, Zou WT, Li YG. EST-SSR marker development based on transcriptome sequencing and genetic analyses of *Phoebe bournei* (Lauraceae). *Mol Biol Rep*. 2021;48(3):2201–8.
50. Hao X, Yang T, Liu R, Hu J, Yao Y, Burlyaeva M, Wang Y, Ren G, Zhang H, Wang D, et al. An RNA sequencing transcriptome analysis of grasspea (*Lathyrus sativus* L.) and development of SSR and KASP markers. *Front Plant Sci*. 2017;8:1873.
51. Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol*. 2004;5(2):1–28.
52. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005;21(18):3674–6.
53. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res*. 2008;36:D480–484.
54. Liu L, Fan X, Tan P, Wu J, Zhang H, Han C, Chen C, Xun L, Guo W, Chang Z, et al. The development of SSR markers based on RNA-sequencing and its validation between and within *Carex* L. species. *BMC Plant Biol*. 2021;21(1):17.
55. Xing W, Liao J, Cai M, Xia Q, Liu Y, Zeng W, Jin X. De novo assembly of transcriptome from *Rhododendron latoucheae* Franch. using Illumina sequencing and development of new EST-SSR markers for genetic diversity analysis in *Rhododendron*. *Tree Genet Genomes*. 2017;13(3):1–4.
56. Liu Y, Fang X, Tang T, Wang Y, Wu Y, Luo J, Wu H, Wang Y, Zhang J, Ruan R, et al. Inflorescence transcriptome sequencing and development of new EST-SSR markers in common buckwheat (*Fagopyrum esculentum*). *Plants*. 2022;11(6):742.
57. Yang W, Bai Z, Wang F, Zou M, Wang X, Xie J, Zhang F. Analysis of the genetic diversity and population structure of *Monochasma savatieri* Franch. ex Maxim using novel EST-SSR markers. *BMC Genomics*. 2022;23(1):597.
58. Sahoo A, Behura S, Singh S, Jena S, Ray A, Dash B, Kar B, Panda PC, Nayak S. EST-SSR marker-based genetic diversity and population structure analysis of Indian *Curcuma* species: significance for conservation. *Braz J Bot*. 2021;44(2):411–28.
59. White OW, Doo B, Carine MA, Chapman MA. Transcriptome sequencing and simple sequence repeat marker development for three Macaronesian endemic plant species. *Appl Plant Sci*. 2016;4(8):1600050.
60. Chai M, Ye H, Wang Z, Zhou Y, Wu J, Gao Y, Han W, Zang E, Zhang H, Ru W, et al. Genetic divergence and relationship among *Opisthopappus* species identified by development of EST-SSR markers. *Front Genet*. 2020;11:177.
61. Wang H, Lei Y, Yan L, Wan L, Cai Y, Yang Z, Lv J, Zhang X, Xu C, Liao B. Development and validation of simple sequence repeat markers from *Arachis hypogaea* transcript sequences. *Crop J*. 2018;6(2):172–80.
62. Chen H, Wang L, Wang S, Liu C, Blair MW, Cheng X. Transcriptome sequencing of mung bean (*Vigna radiata* L.) genes and the identification of EST-SSR markers. *PLoS ONE*. 2015;10(4):e0120273.
63. Taheri S, Lee Abdullah T, Yusop MR, Hanafi MM, Sahebi M, Azizi P, Shamshiri RR. Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants. *Molecules*. 2018;23(2):399.
64. Kumpatla SP, Mukhopadhyay S. Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome*. 2005;48(6):985–98.
65. Gao Z, Wu J, Liu ZA, Wang L, Ren H, Shu Q. Rapid microsatellite development for tree peony and its implications. *BMC Genomics*. 2013;14(1):1–11.
66. Taheri S, Abdullah TL, Rafii MY, Harikrishna JA, Werbrouck SPO, Teo CH, Sahebi M, Azizi P. De novo assembly of transcriptomes, mining, and development of novel EST-SSR markers in *Curcuma alismatifolia* (Zingiberaceae family) through Illumina sequencing. *Sci Rep*. 2019;9(1):3047.
67. Sun M, Zhao Y, Shao X, Ge J, Tang X, Zhu P, Wang J, Zhao T. EST-SSR marker development and full-length transcriptome sequence analysis of tiger lily (*Lilium lancifolium* Thunb). *Appl Bionics Biomech*. 2022;2022:7641048.
68. Fryxell KJ, Zuckerkandl E. Cytosine deamination plays a primary role in the evolution of mammalian isochores. *Mol Biol Evol*. 2000;17(9):1371–83.
69. Yakovchuk P, Protozanova E, Frank-Kamenetskii MD. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res*. 2006;34(2):564–74.
70. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29(7):644–52.
71. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30.
72. Beier S, Thiel T, Munch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics*. 2017;33(16):2583–5.
73. Rohlf FJ. NTSYS: numerical taxonomy and multivariate analysis system, V. 2.0. Applied Biostatistics Inc., New York. 1998.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

