

De novo transcriptome analysis and development of novel EST-SSR markers in *Bergenia ciliata* (Haw.) Sternb. (Saxifragaceae) through Illumina sequencing

Harish Chandra Singh

CSIR-National Botanical Research Institute, Rana Pratap Marg, Lucknow

Vandana Tiwari

CSIR-National Botanical Research Institute, Rana Pratap Marg, Lucknow

Avinash Tiwari

Jiwaji University, Gwalior

Tikam Singh Rana (✉ ranats@nbri.res.in)

CSIR-National Botanical Research Institute, Rana Pratap Marg, Lucknow

Research Article

Keywords: *Bergenia ciliata*, de-novo transcriptome, EST-SSRs, RNA-Seq

Posted Date: October 12th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-2145830/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Bergenia ciliata (Haw.) Sternb. is an important herb predominantly found in Indian Himalayan Region (IHR). It is widely used in medicines and healthcare system, cosmetics, food, fodder, and ornamental purposes. Illumina sequencing and *de novo* transcriptome assembly were carried out in *B. ciliata* to develop and identify simple sequence repeat markers for genetic diversity and conservation studies. The assembled data generated a total of 65,010 unigenes that showed significant similarities when compared with seven functional databases including 53,577 (Non-Redundant Protein Sequence Database: 82.41%), 44,297 (Nucleotide Sequence Database: 68.14%), 42,287 (Swiss Prot: 65.05%), 15,027 (Eukaryotic Orthologous Groups: 23.11%), 22,540 (KEGG Orthology: 34.67%), 29,477 (Gene Ontology: 45.34%) and 20,609 (Pfam: 31.7%) unigenes. In this study, a total of 18,226 SSRs and 14,497 SSR containing sequences were identified. Dinucleotides were found to be abundant (47.88%) in *B. ciliata* followed by mononucleotides (35.04%), and trinucleotides repeat (15.90%). AG/CT was the most common di-nucleotide repeat (40.33%). A total of 11,839 EST-SSR primers were designed, of which 96 primer pairs were synthesized randomly. Finally, 18 primer pairs were selected that revealed clear, distinct polymorphic bands when examined in eight diverse *B. ciliata* accessions. Furthermore, the transcriptome data and the EST-SSR markers will be an important resource for investigating genetic diversity in *B. ciliata* and other species of the family Saxifragaceae.

Introduction

The genus *Bergenia* Moench (Saxifragaceae) consists of about 10 species of perennial rhizomatous herbs distributed mainly in the temperate and subtropical regions in Central and East Asia (Pandey et al. 2017). In India, the four species of *Bergenia* viz., *Bergenia ciliata* (Haw.) Sternb., *Bergenia ligulata* (Wall.) Engl., *Bergenia stracheyi* (Hook. & Thoms.) Engl., and *Bergenia purpurascens* (Hook. & Thoms.) Engl., are mainly found in the Himalayan regions (Tiwari et al. 2015). *Bergenia ciliata* (Haw.) Sternb. (Saxifragaceae), commonly known as '*Pashanbheda*' is a perennial, rhizomatous herb occurs in temperate and sub-temperate Himalayan regions at an elevation of 350–4000 m from Afghanistan to eastward in Pakistan, Nepal, Bhutan and India (Ahmad et al. 2018; Tiwari et al. 2020). In India, it is found mostly in humid and moist habitats, crevices, on rock surfaces, hill slopes and near to water channels (Tiwari et al. 2020). *B. ciliata* is medicinally important and traditionally being used to treat several diseases like kidney stones, urinary problems, piles, boils, burns, fever, liver disorders, thirst, cuts and wounds, asthma, ophthalmia, and diarrhoea in cattle (Kirtikar and Basu 1935; Rana and Samant 2011). It is used as an important botanical source for preparation of many herbal medicines due to the presence of numerous bioactive metabolites. It is one of the significant constituents in indigenous drugs such as Calculi and Cystone used in clinically to cure urinary bladder and kidney stones (Asolkar et al. 1992). To estimate the genetic diversity and population genetic structure with high reproducibility within a large number of species, microsatellites or SSRs (Simple sequence repeats) are recognized as significant molecular markers (Taheri et al. 2019). SSRs are tandem repeats of 2–6 nucleotides widely distributed in both eukaryotic and prokaryotic genomes (Zhai et al. 2014). Microsatellites can be developed either from EST database or from coding sequences by RNA sequence technology and found to be less expensive as compared to genomic SSRs which are costly, time-consuming, and labor-intensive (Taramino et al. 1997; Rota et al. 2005). EST-SSRs are generated from the highly conserved regions of the genome, therefore, show high rates of transferability than gSSRs and are therefore considered as an efficient tool for exploring the population dynamics, evolutionary studies, comparative genetic mapping, and determination of differentially expressed genes (Hina et al. 2020). A large number of transcribed sequences with EST-SSR markers have been generated for studying cross transferability in medicinal plant species such as *Saxifraga sinomontana* (Li et al. 2019), *Melilotus* germplasm (Yan et al. 2017), *Hibiscus cannabinus* (Kim et al. 2019), and *Phoebe bournei* (Zhou et al. 2021). Earlier genetic diversity studies on *B. ciliata* were restricted to universal, multi-allelic, dominant, single primer amplification reaction (SPAR) methods like Inter Simple Sequence Repeat (ISSR) and Directed Amplification of Minisatellites region-DNA (DAMD). Development of SSR markers in *B. ciliata* will be valuable tools for studies of genetic diversity, population structure, cross-transferability, conservation biology, and plant breeding. As such there is no report available on the development of microsatellites and transcriptome analysis (RNA-seq) of *B. ciliata*, therefore, to the best of our knowledge, the present study seems to be the first report on *de novo* transcriptome analysis, and development of SSR markers based on Illumina HiSeq 4000 sequencing for exploring the population genetics of *B. ciliata* in India and establishing its relationship with allied species of the family Saxifragaceae.

Materials And Methods

Plants material and RNA extraction

Rhizomes of *B. ciliata* were collected from the Galla Dhar, Rajgarh, Jammu and Kashmir (N33°11'39.41" E 75°22'44.74") and planted in a glasshouse at CSIR-National Botanical Research Institute, Lucknow, India. The Fresh leaf samples were collected, cleaned, and stored immediately in liquid nitrogen and kept at -20°C until RNA extraction. Total RNA was isolated using a commercially available QuickRNA Plant Miniprep Plus kit (ZYMO Research) as per the manufacturer's instruction. Qualitative analysis of the isolated RNA samples, RNA degradation and contamination were monitored on 1% denaturing RNA agarose gel. RNA purity was checked using the Nano Photometer® spectrophotometer (IMPLEN, CA, USA) and RNA integrity and quantitation were assessed using RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

Cdna Library Construction And Rna-sequencing

A total amount of 1 µg RNA per sample was used as input material for preparation of RNA samples. More than 30 ng of RNA was equally pooled from the three individuals for preparing a complementary DNA (cDNA) library. Sequencing libraries were generated using NEBNext® Ultra™ RNA Library Preparation Kit for Illumina® (NEB, USA) as per the manufacturer's protocol and index codes were added to attribute sequences to each sample. Purification of mRNA was done from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). The first strand of cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H), and the second strand was synthesized subsequently using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, The NEBNext Adaptor with hairpin loop structure was ligated to prepare for hybridization.

To select cDNA fragments of preferentially 250 ~ 300 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then 3 µl of USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95°C before PCR. PCR was performed with Phusion high-fidelity DNA polymerase, universal primers, and index (X) primer. At last, PCR products were purified (AMPure XP system) and library quality was assessed (Agilent Bioanalyzer 2100 system). The clustering of the index-coded samples was performed on a cBot Cluster Generation System using PE Cluster Kit, cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced, and paired-end reads were generated.

The sequenced raw data were stored in FASTAQ format. RNA isolation, Illumina library preparation, and sequencing were outsourced to Eurofins Scientific, Bengaluru, India. The raw reads of the transcriptome sequence were submitted to the sequence read archive (SRA) under accession number PRJNA775824.

De novo transcriptome assembly and analysis of sequence data

De novo assembly of the transcriptome was carried out in Trinity (Grabherr et al. 2011). Gene Ontology (GO) annotations as well as KOG functional analysis of the unigenes and pathway were performed based on four public databases (Cai et al. 2018; Yue et al. 2017). The GO annotation and functional classification of the unigenes were carried out in automatic and high throughput software, Blast2GO (Conesa et al. 2005).

Identification And Development Of Simple Sequence Repeat Markers

The MicroSatellite program (MISA; <http://pgrc.ipk-gatersleben.de/misa/>) was used to detect SSRs in the unigenes. Simple sequence repeats consist of a repeating unit of 1–6 bp, and can be of three types according to their distribution and composition (i.e., compound, pure, and interrupted). In this study, SSRs comprised at least four uninterrupted repeats of 2–6 bp. Minimum number of repeats of each unit size obtained were: 1–10; 2–6; 3–5; 4–5; 5–5 and 6–5. SSRs occurred at a frequency of 1/kb of cDNA. PCR primers (20 bp) with a melting temperature (T_m) of 55–65 °C, were designed to produce an amplification product of 100–200 bp, using primer designing tool, Primer3 v 2.3.5 (Untergasser et al. 2012).

Dna Extraction, Est-ssr Amplification And Ssr Marker Validation

The extraction of genomic DNA from the fresh leaves of eight *B. ciliata* accessions collected from eight different geographical locations (Table 1.) was carried using CTAB method (Doyle and Doyle 1990). The SSR-PCR reactions were carried in 10 µL volume, containing 2 µL water, 6 µL PCR Master Mix 2X (Thermo Scientific Pvt. Ltd.), 0.5 µL of each primer at a concentration of 10 µM, and 1 µL of genomic DNA (30–50 ng). The reaction conditions for SSR primers were optimized to an initial denaturation at 95 °C for 4 min, followed by 35 cycles of 45 sec at 94 °C, annealing temperature for 45 sec, and 72 °C for 1 min, and a final extension at 72 °C for 5 min. The amplified products were separated on 2.5% agarose gels and stained with ethidium bromide (10 mg/mL) for visualization. Step Up (50 bp) DNA Ladder (GeNei Laboratories Pvt. Ltd.) was used as a reference for comparison of the band size of the amplified products. The gel was documented using UV Tech Gel Documentation System (UK) and the gel patterns were documented as digital images for further processing.

Table 1
Accessions of *B. ciliata* used in validating of EST-SSRs

SN	Accessions number	Locality	Latitude	Longitude	Altitude (m)
1	255845	J&K, Rajouri, Dera Gali Forest, RJR	33.58432°	74.35805°	2036
2	253516	HP, Kothi to Rahla, Kullu, KLU	32.34808°	77.22172°	2277
3	335347	UK, Dehradun, Chakarata, CHK	30.681388°	77.880828°	2097
4	335316	UK, Dehradun, Mussoorie, MSR	30.447476°	78.091282°	1902
5	335301	UK, Bageshwar, Gansi, GSI	30.04853°	79.97393°	1615
6	253459	UK, Almora, Binsar Wildlife Sanctuary, BWS	29.68313°	79.72982°	1923
7	255752	SK, East Sikkim, Penlong, PLG	27.37408°	88.62213°	1646
8.	255715	WB, Darjeeling, DRG	27.05383°	88.25267°	2087

Results

Illumina sequencing and de novo assembly

In Illumina sequencing platform, the cDNA library was constructed using a total of 21,490,725 paired-end raw reads, which included null reads, low-quality sequences, and adapter-primer sequences. Total of 21,277,286 high-quality clean reads with 98.02% (Q20) and 93.98% (Q30) bases were obtained after a stringent quality check and data filtering. The GC percentage for the clean reads had 4.71 and clean reads were total nucleotide number of 133,174,507 (66,597,750 transcripts + 66,576,757 unigenes) (Table 2). Total unigenes identified with paired-end reads were 65,010 bp, with total length of 665,767,57 bp encompassing an average length of 1024.10 bp. From a given set of contigs, the values of N50 and N90 were generated as 1349 and 494 bp, respectively. In the 65,010 unigenes, 17,527 unigenes (26.96%) had a length of 200-500bp; 21,591 unigenes (33.21%) ranged from 500-1 kbp; 19,547 bp unigenes (30.06%) ranged from 1kbp-2kbp, and the length of 6,345 unigenes (9.76%) ranged > 2kbp (Fig. 1).

Table 2
Statistics of the transcriptome data generated after Illumina sequencing in *B. ciliata*

Category	Items	Number
Raw reads	Total raw read	21,490,725
Clean reads	Total clean reads	21,277,286
	Total clean nucleotides (nt)	133,174,507
	Q20 percentage	98.02%
	Q30 percentage	93.98%
	GC percentage	44.71%
Unigenes	Total sequence number	65,010
	Total sequence base	665,767,57
	Largest	6,531
	Smallest	301
	Average	1,024
	N50(bp)	1,349
	N90 (bp)	494
EST-SSR	Total number of examined sequences	65,010
	Total size of examined sequences (bp)	665,767,57
	Total number of identified SSRs	18,226
	Number of SSR-containing sequences	14,497
	Number of sequences containing more than one SSR	2,913
	Number of SSRs present in compound formation	1,468

Frequency And Distribution Of Est-ssrs In The Unigenes

Out of 65,010 unigenes, 18,226 potential EST-SSRs were identified, and 1,468 compound microsatellites obtained from the 18,226 EST-SSRs (Table 3). The SSR frequency of unigenes in *B. ciliata* was 28.03%. A total of 18,226 EST-SSRs were identified, wherein the most prominent type repeat was dinucleotides (8,728, 47.89%), followed by mono (6,327, 35.04%) and trinucleotide repeats (2,899, 15.91%) (Table 4; Fig. 2).

Table 3
EST-SSR markers identified from *de novo* transcriptome sequencing
in *B. ciliata*

Searching Items	Numbers
Total number of sequences examined	65,010
Total size of examined sequences (bp)	66,576,757
Total number of identified SSRs	18,226
Number of SSR containing sequences	14,497
Number of sequences containing more than 1 SSR	2,913
Number of SSRs present in compound formation	1,468
Mono-nucleotide	6,387
Di-nucleotide	8,728
Tri-nucleotide	2,899
Tetra-nucleotide	101
Penta-nucleotide	41
Hexa-nucleotide	70

Table 4
Length distributions of microsatellites in *B. ciliata* based on the number of nucleotide repeat units

Number of repeats	Mono-	di-	Tri-	tetra	Penta-	Hexa-	Total	Percentage (%)
5			1,737	72	30	54	1,893	10.38
6		2,468	593	24	8	9	3,102	17.01
7		1,709	297	2		7	2,015	11.05
8		1,292	147	1	1		1,441	7.90
9		901	79	2	1		983	5.39
10	3,194	741	29				3,964	21.74
11	1,127	448	4				1,579	8.66
12	683	340	8		1		1,032	5.66
13	364	217	5				586	3.21
14	275	234					509	2.79
15	198	190					388	2.12
16	123	24					147	0.80
17	108	50					158	0.86
18	56	25					81	0.44
19	50	19					69	0.37
20	43	17					60	0.32
21	13	23					36	0.19
22	20	10					30	0.16
23	18	5					23	0.12
24	20	4					24	0.13
25	12	5					17	0.09
26	15	4					19	0.10
27	8	1					9	0.04
28	6	1					7	0.03
29	25						25	0.13
30								0
< 30	29						29	0.15
Total	6,387	8,728	2,899	101	41	70	18,226	
Percentage (%)	35.04	47.89	15.91	0.55	0.22	0.38		

The ten tandem repeats of EST-SSR (3,964, 21.74%) were found to be the most common (Table 5), followed by six (3,102, 17.01%), seven tandem repeats (2,015, 11.05%), and five, eleven, and eight tandem repeats were 1,893 (10.38%), 1,579 (8.66%) and 1,441 (7.90%), respectively, while the remaining of tandem repeat for individual contributed < 10% of EST-SSR. The AG/CT was most dominant motif (7,351; 40.33%) followed by AT/AT (1050; 5.76%), AC/GT (301; 1.65%), and CG/CG (26; 0.14%) in the di-nucleotide repeats. The most abundant repeat motif was AAG/CTT (706; 3.87%) followed by ACC/GGT (519; 2.84%), ATC/ATG (387; 2.12%), AGC/CTG (309; 1.69%), and AGG/CCT (281; 1.54%) in the tri-nucleotide repeats (Table 5; Fig. 2).

Table 5
Frequency and distribution of microsatellites in *B. ciliata* based on SSRs repeat motifs

Repeat motifs	Number of repeats							Total	Frequency (%)
	5	6	7	8	9	10	> 10		
Mono-nucleotide									
A/T	—	—	—	—	—	3164	3081	6245	34.26
C/G	—	—	—	—	—	30	112	142	0.77
								6387	35.04
Di-nucleotide									
AG/CT	—	2033	1471	1088	749	634	1376	7351	40.33
AT/AT	—	293	165	163	125	89	215	1050	5.76
AC/GT	—	121	70	41	25	18	26	301	1.65
CG/CG	—	21	3	—	2	—	—	26	0.14
								8,728	47.88
Tri-nucleotide									
AAG/CTT	377	159	73	47	31	11	8	706	3.87
ACC/GGT	308	108	62	32	6	3	—	519	2.84
ATC/ATG	254	67	32	12	13	6	3	387	2.12
AGC/CTG	202	49	33	20	1	4	—	309	1.69
AGG/CCT	189	52	26	5	7	2	—	281	1.54
Others	407	158	71	31	21	3	6	697	3.82
								2,899	15.90
Quad-nucleotide									
AAAT/ATTT	24	—	—	—	—	—	—	24	0.13
AAAC/GTTT	11	2	—	—	—	—	—	13	0.07
AATC/ATTG	11	2	—	—	—	—	—	13	0.07
AGAT/ATCT	7	3	1	—	2	—	—	13	0.07
Others	19	17	1	1	—	—	—	38	0.21
								101	0.55
Penta-nucleotide	30	8	—	1	1	—	1	41	0.22
Hexa-nucleotide	54	9	7	—	—	—	—	70	0.38
								111	0.60
Total	1,893	3,102	2,015	1,441	983	3,964	4,829	18227	100
Frequency (%)	10.38569	17.01871	11.05503	7.905854	5.393098	21.74796	26.49366	100	

Functional annotation of *Bergenia* transcriptome

De novo assembled unigenes of *B. ciliata* were annotated against the functional public databases; Nt (Non-redundant nucleotide sequence), Nr (Non-redundant protein sequence), KO (KEGG Orthology), Swiss-Prot, Pfam, GO (Gene Ontology), and KOG (Eukaryotic Orthologous Group) databases (Table 6; Fig. 3). All the sequences were generated by Blast and splicing algorithm, which was applied for the comparison and to obtain a relevant sequence and associating annotation.

Table 6
Summary of functional annotation of unigenes of *B. ciliata* with seven databases

	Number of Unigenes	Percentage (%)
Annotated in NR	53577	82.41
Annotated in NT	44297	68.14
Annotated in KO	22540	34.67
Annotated in Swissport	42287	65.05
Annotated in PFAM	20609	31.7
Annotated in GO	29477	45.34
Annotated in KOG	15027	23.11
Annotated in all Databases	4954	7.62
Annotated in at least one Database (overall*)	54732	84.19
Total Unigenes	65010	100
* the number of unigenes which can be annotated with at least one functional database		

Annotation Of Non-redundant And Nucleotide Database

Out of the 65,010 unigenes, 54,732 unigenes were successfully annotated. However, 53,577 (82.41%) unigenes showed efficient homology with the proteins in the Nr database, while 44,297 (68.14%) of the control sequences relate with the Nt database entries (Fig. 3). As for the species distribution annotations of *B. ciliata*, *Vitis vinifera* (Vitaceae) has the highest similarity score (25.7%, followed by *Quercus suber* (Fagaceae) 7.6%, *Juglans regia* (Juglandaceae) 5.1%, *Nelumbo nucifera* (Nelumbonaceae) 3.3% and *Hevea brasiliensis* (Euphorbiaceae) 3.2% (Fig. 4).

Kog Classification, Go Annotation, Kegg Pathway And Swiss-prot Annotation

Functional annotation in the KOG database was based on 25 functional groups, including metabolic functions, cellular structure, and signal transduction. The post-translational modification, protein turnover and chaperones (2266 genes, 15.07%) represents the largest group, followed by general function prediction (1861 genes, 12.38%), translation, ribosomal structure, and biogenesis (1638 genes, 10.90%), and nuclear structures and cell motility as the smallest groups (Fig. 5).

Annotation in GO database grouped 29,477 (45.34%) unigenes into three major categories such as cellular component (51232, 37.8%), biological process (50744, 37.5%), and molecular functions (33274, 24.6%), with 51 subcategories (Fig. 6). Most of the unigenes in the molecular function are specified for binding (15346) and catalytic activity (14136), while metabolic process (14927) and cellular process (14264) are the major subcategories in the biological process. In total, 22,540 (34.67%) of unigenes were identified in the database, which were significantly assigned to 125 metabolic pathways.

Annotation in KEGG database, categorized metabolic pathways into 11 main divisions. Metabolic information processing (with 21717 genes) found to be the largest division, followed by genetic information processing (5,167), organismal systems (4,282),

cellular processing (2,482), and environmental information processing (2,286) Further, in Swiss-Prot database 42,287 (65.05%) unigenes were found matched (Table 6; Fig. 7).

Development And Validation Of Novel Est-ssr Markers

Total 96 primer pairs were synthesized and checked for their amplification and polymorphism. Out of 96 primer pairs, 37 were successfully amplified, while the remaining 59 primers did not show any amplification even at different annealing temperatures. Among the 37 primer pairs, 32 successfully produced the desired amplified products, while the remaining 5 PCR products were larger or smaller than the expected size (ESM Fig. 1). Total of eight individuals from eight different populations of *B. ciliata* were used as PCR templates, and from 37 primers, 18 primer pairs were found polymorphic (ESM Fig. 2; ESM Fig. 3), whereas 14 pairs were identified as monomorphic (Table 7).

Table 7
Characterization of the 18 novel EST-SSR polymorphic primer pairs synthesized from the transcriptome of *B. ciliata*

SSR	Repeat Type	Repeat motif	Forward Primer	Tm (°C)	Reverse Primer	Tm (°C)	Product Size (bp)
BC2	Di	(AG)10	CTGAGGCCAAAGAAAGTGCG	59.7	ACAAAGTCACACGGGCATCT	59.8	190–250
BC7	Di	(AG)6	ACAATCAACAAGGCATCATGC	57.7	TCCAACCTTACTGGGCAGGAA	58.5	180–250
BC8	Di	(AG)7	TGGTCTGACAGTGAGTTCGC	59.9	TCGCCATCACAGAAGCCTTT	59.9	140–160
BC17	Di	(AT)6	TACAAATACACCGGTGCAGG	57.8	AAATCTGGAGGGTTGCCAGG	59.9	125–150
BC23	Di	(CT)10	TCACTCGTAAAGTCGACCCT	58.0	GGACGTCGAGCGAACAAATG	59.9	140–180
BC26	Di	(CT)11	CAGCCAGTACTCTGCCCAA	59.9	ACTCTCCACCTCCTGACCTC	59.9	130–150
BC29	Di	(CT)6	ACGCCATTCTCACTGTACCT	58.7	TCAGCGGAGAAACAACCTCC	59.9	180–210
BC33	Di	(CT)6	CATTGTTTCCTCCGTTGCC	59.7	CTCCGTTTGGTTCTCGGGAA	59.9	200–250
BC38	Di	(CT)8	TCGCAAACCTCTCTCACTCTCC	59.4	AAACTTCAACCGCGGGATCT	59.9	140–160
BC50	Di	(GA)8	TCCTCGAGTATTTGTGCGAG	57.4	GCGTTGAGAATCATTGCCCC	59.9	140–160
BC53	Di	(GA)9	ACCGCCAAGAGCTTGATGTA	59.3	TGTTGAGTCGTTCTGCTTCC	57.8	110–150
BC58	Di	(TA)8	ACACATGTTTACACGCGCAT	59.1	GAAGTGCACCCAAAGCATGA	59.0	175–210
BC67	Tri	(AGA)5	ACCAATGTGAGGGTTCCTTCT	58.9	CCAACACACAGCAAGACAGC	59.9	160–180
BC71	Tri	(CAC)6	AGAGGCACAATGTGGAAGAGA	59.0	TTCATGTAGTCCGGCAGCTC	59.8	190–210
BC73	Tri	(GAA)5	AGTGTGGTACTCCTCGCTCT	59.9	ATCACGTCGTCGGAGAATCG	59.9	220–250
BC74	Tri	(GAC)5	GGCAAACCTCCTCCAAGAA	59.8	TTCCCTTGCCAGTTCCTCAC	59.8	230–250
BC84	Tri	(TTC)5	GCTTGCAGTTTACACCCACA	58.9	CGCCTCCACGTCTATGTCTC	59.9	160–190
BC87	Tri	(TTTA)5	GGAAAGGTTGGATTGCTCCC	58.8	GATCTGCTGCAGAACTGGGT	60.0	100–130

Discussion

Characterization of *B. ciliata* transcriptome

Next-generation sequencing has been gradually used to analyze the transcriptome sequencing and assembly in various plants due to the accuracy, high efficiency, high speed, and low cost (Zhou et al. 2018; Dhiman et al. 2020; Shah et al. 2020; Xie et al. 2021). EST-SSR markers have been used widely in assessing genetic diversity, DNA fingerprinting, marker-assisted breeding, and gene mapping due to their high polymorphism, codominant inheritance, and repeatability (Ercisli et al. 2011; Palumbo et al. 2018; Wang

et al. 2020; Gaurav et al. 2021). An enormous data generated by transcriptome sequencing delivers extensive details and a significant source for the SSRs and gene development in several life forms (Li et al. 2019; Chabikwa et al. 2020; Saina et al. 2021). Therefore, transcriptome sequences have been successfully employed in both model and non-model plants for detection of marker-based functional variation and gene-associated genetic analysis (Klepikova et al. 2016; Zhang et al. 2020; Raizada et al. 2021). The lacks information on transcriptome data and EST sequences prompted us to sequence the transcriptome, and eventually develop the EST-SSR markers through NGS technology in *B. ciliata*. A total of 21,490,725 paired-end raw reads were constructed, and 21,277,286 best-quality clean reads were generated with 98.02% Q20 level (base quality > 20) during the present study (Table 2), which assures the sequencing quality. This is in congruence with the earlier studies on *Neolitsea sericea* (Chen et al. 2015), and *Stephanandra incisa* (Zhang et al. 2021). The mean N50 sizes (1,349 bp) of unigenes constructed in this study were found smaller than *Stephanandra incisa* (7,212 bp; Zhang et al. 2021) and *Paeonia* cultivars (1,780bp; He et al. 2020). However, total of 65,010 unigenes were assembled to transcriptome of *B. ciliata* with the average length of 1,024 bp, which was significantly larger than earlier reported transcriptome studies of *Cynanchum komarovii* (604 bp; Ma et al. 2015), *Raphanus sativus* (576 bp; Wu et al. 2015), *Rhododendron rex* (526.74 bp; Zhang et al. 2017), and *Magnolia wufengensis* (695 bp; Wang et al. 2019), but shorter than *Cyamopsis tetragonoloba* (1583.43 bp; Rawal et al. 2017), *Phragmites karka* (1354.16 bp; Nayak et al. 2020), and *Lathyrus sativus* (1250 bp; Hao et al. 2017). This could be due to a mismatch between the parameters and the assembler, as well as differences in the nature of the species (Xing et al. 2017). These unigenes were enriched for proteins that maintain the essential functions of *B. ciliata*. Ultimately, the massive amount of transcriptome sequences generated in this study will be further significant for investigating gene function and molecular mechanisms.

The assembled unigenes of *B. ciliata* were annotated to known public databases successfully; Nt, Nr, KOG, KEGG, GO, Pfam, and Swiss-Prot (Table 6). These annotations may provide useful facts for the future molecular studies in the genus *Bergeria*. The limited annotation or short sequence length of the genus *Bergeria* and related species in the present database could be the reason for the genes which did not match any functional annotations (Xing et al. 2017). In the KOG classification, the first and second groups found in this study were the post-translational modification, protein turnover chaperones, and general function prediction (Fig. 5), which were similar to the studies performed by Du et al. (2019) but distinct from the work of Cao et al. (2018). Overall, 4,954 (7.62%) unigenes in the database had significant matches and were classified into seven main categories, including 271 KEGG pathways. In this study the biggest group in three GO categories were metabolic process, and cellular process under the biological process and the cell and cell part in cellular component (Fig. 6), which were similar to *Hevea brasiliensis* (Li et al. 2012). These outcomes revealed that *B. ciliata* has active metabolic processes and can synthesize plethora of metabolites. GO functional annotation helped to describe the macro level of gene functions and predict the physiological role of each unigene (Kumar et al. 2014). The results revealed various molecular functions of assembled unigenes, suggesting their involvement in diverse metabolic pathways. These findings indicate that *B. ciliata* makes a huge investment in gene transcription control and capacity, as well as cell maintenance and defence. Briefly, functional analysis revealed that RNA-seq-based *de novo* transcriptome analysis for *B. ciliata* a non-model organism with a complex genome, will facilitate further research on the physiology, molecular genetics, and biochemistry of *B. ciliata* or related species.

Frequency And Distribution Of Est-ssrs

EST-SSR markers play a significant role in the genetic diversity and population structure analysis, development of genetic maps, marker-assisted selection, genomics comparison, breeding, and species conservation (Chen et al. 2005; Marconi et al. 2011; Ahmad et al. 2018; Sahoo et al. 2021). Total of 18,226 potential EST-SSRs were recognized to 65,010 unigenes in this study, which has significantly enhanced the SSR resources availability for marker development in the *B. ciliata* and other closely related taxa of the family Saxifragaceae. The di-nucleotide (47.88%) repeats were the most abundant type, which are corresponding with other species of the plants studied earlier (Chen et al., 2015; Jia et al., 2016; Biswas et al., 2020), followed by mono- nucleotide (35.04%) and tri-nucleotide (15.90%) repeats (Fig. 2; Table 5). In dicotyledonous plants, di-nucleotide repeats have been found to be the most abundant SSR repeat type (Yang et al. 2020). Overexpression of UTRs as compared to open reading frames may lead to di-nucleotide repeat sequence (Qiu et al. 2010). Additionally, AG/CT was the most common di-nucleotide repeat (40.33%).

In an mRNA population, the AG/CT motif can reveal UCU and CUC codons, which translate to the Ala and Leu amino acids, are found in proteins at a higher frequency than other amino acids and can be identified in an mRNA population (Chen L. Y. et al.

2015). As a result, AG/CT motifs can be found in abundance in plant EST libraries (Morgante et al. 2002; Chen L. Y. et al. 2015). In our study, the CG/CG (0.14%) was the lowest prominent motif. It could be due to cytosine methylation, which inhibits the transcription in some plants (Chen L. Y. et al. 2015; Xing et al. 2017). The AAG/CTT repeat motif was the most dominant in tri-nucleotide repeat (3.87%). The earlier studies on *Ricinus communis*, *Neolitsea sericea*, *Sesamum indicum*, and *Cucumis sativus*, also revealed the tri-nucleotide AAG motif is highly dominant and valuable in dicotyledonous plants (Qiu et al. 2010; Cavagnaro et al. 2010; Wei et al. 2011; Chen L. Y. et al. 2015; Zhang et al. 2021).

EST-SSRs frequency of unigenes in *B. ciliata* was (Table 2) is significantly higher than *Citrus reticulata* (Long 2014), *Quercus austrocochinchinensis* (An et al. 2016), *Rosa roxburghii* (Yan et al. 2015), *Rhododendron fortune* (Yang et al. 2018), but lower than *Tagetes erecta* (Zhang et al. 2018), and *Lonicera caerulea* (Zhang et al. 2016). The frequency of SSRs, diversified from species to species depends upon the SSR database size, database-mining tools, and the search criteria, in different studies (Gao et al. 2003; Varshney et al. 2005). A significant number of high-quality EST-SSRs generated during present investigation will be immensely helpful in better understanding of the genetic diversity, population structure, and in breeding programs.

Est-ssr Makers Validation

A total 96 primer pairs were randomly selected for PCR validation, out of which 37 (38.54%) were successfully amplified by genomic DNA of *B. ciliata*, 18 (18.75%) primer pairs were polymorphic among the eight geographically isolated populations of *B. ciliata* tested, and the remaining 59 (61.45%) primer pairs failed to amplify the PCR products at different annealing temperatures or else that were formed bigger than expected size of PCR products (Fig. S1).

The PCR success rate was higher in *B. ciliata* than the rates reported in other species like *Curcuma alismatifolia* (11.33%; Taheri et al. 2019), *Magnolia sinostellata* (15.33%; Wang et al. 2019), *Salix psammophila* (16.07%; Jia et al. 2016). However, the success rate was found lower than *Sesamum indicum* (80%; Wei et al. 2011) and *Hevea brasiliensis* (55.45%; Li et al. 2012). Thus, in this study the polymorphic ratio of EST-SSR markers was found significantly high. This indicates that newly developed EST-SSR markers will be significant and useful for the population genetic structure and evolutionary studies among the *Bergenia* species.

The transferability of EST-SSRs markers among related species is higher because it contains conserved sequences among homologous genes and their origin from the transcribed regions in genomes (Wu et al. 2014; Guo et al. 2014). In the current study, the 18 newly developed polymorphic markers will be applied to estimate genetic diversity in the genus *Bergenia* and cross transferability in allied species of the family Saxifragaceae. Thus, these markers will be helpful in providing valuable sequence resources for the development of molecular markers in *Bergenia* species.

Conclusion

Complete transcriptomic information obtained from *B. ciliata*, was useful in mining SSR markers, thus providing a large and well-assembled transcript dataset. A total 65,010 unigenes with the mean length of 1,024 bp were generated, 54,732 unigenes were efficiently annotated into the Nr (53,577: 82.41%), Nt (44,297: 68.14%), Pfam (20,609: 31.7%), Swiss-Prot (42,287: 65.05%), KEGG (22,540: 34.67%), KOG (15,027: 23.11%), and GO (29,477: 45.34%) databases. Based on these unigene sequences, 18,226 potential EST-SSRs were identified and developed. The ninety-six primer pairs were selected randomly, synthesized, and utilized for the evaluation of polymorphism. Eighteen primer pairs (18.75%) were found polymorphic and showed polymorphism among the eight individuals of eight populations. The present study has provided a rich dataset of *B. ciliata* transcripts, that will enrich the available genomic resources and serve as an important resource for future genomic and genetic studies. These outcomes would also be helpful in exploring the biosynthetic pathways and regulatory mechanisms of active phytoconstituents of therapeutic value present in *B. ciliata*.

Declarations

Acknowledgements

The authors are thankful to the Director, CSIR-National Botanical Research Institute, Lucknow for facilities. The Science and Engineering Research Board (SERB), Government of India, New Delhi (SERB No: EEQ/2016/000096) is thankfully acknowledged for financial support, and we also thank the Forest Departments of Jammu and Kashmir, Himachal Pradesh, and Uttarakhand for permission to collect the representative samples of *B. ciliata* from different forest locales. This manuscript has the institutional communication number: CSIR-NBRI_MS/2022/05/10.

Author contributions

Tikam Singh Rana conceived and designed the study. Harish Chandra Singh and Vandana Tiwari conducted the experiments and wrote the first draft of the manuscript. Tikam Singh Rana and Avinash Tiwari finalized the manuscript. All authors read and approved the manuscript.

Availability of data and materials

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive of NCBI under accession number PRJNA775824.

Funding

This research was partially funded by the Science and Engineering Research Board (SERB), Government of India, New Delhi (SERB No: EEQ/2016/000096).

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Ahmad A, Wang J-D, Pan Y-B, et al (2018) Development and Use of Simple Sequence Repeats (SSRs) Markers for Sugarcane Breeding and Genetic Studies. *Agronomy* 8:260. <https://doi.org/10.3390/agronomy8110260>
2. An M, Deng M, Zheng S-S, et al (2016) De novo transcriptome assembly and development of SSR markers of oaks *Quercus austrocochinchinensis* and *Q. kerrii* (Fagaceae). *Tree Genet Genomes* 12:103. <https://doi.org/10.1007/s11295-016-1060-5>
3. Asolkar LV, Kakkar KK, Chakre OJ (1992) Second supplement to glossary of Indian medicinal plants with active principles: part-1 (A-K), (1965-1981). Publications and information directorate, New-Dehli
4. Biswas MK, Bagchi M, Nath UK, et al (2020) Transcriptome wide SSR discovery cross-taxa transferability and development of marker database for studying genetic diversity population structure of *Lilium* species. *Sci Rep* 10:18621. <https://doi.org/10.1038/s41598-020-75553-0>
5. Cao D, Liu Y, Ma L, et al (2018) Transcriptome analysis of differentially expressed genes involved in selenium accumulation in tea plant (*Camellia sinensis*). *PLoS One* 13:e0197506. <https://doi.org/10.1371/journal.pone.0197506>
6. Cavagnaro PF, Senalik DA, Yang L, et al (2010) Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genom* 11:569. <https://doi.org/10.1186/1471-2164-11-569>
7. Chabikwa TG, Barbier FF, Tanurdzic M, et al (2020) De novo transcriptome assembly and annotation for gene discovery in avocado, macadamia and mango. *Sci Data* 7:9. <https://doi.org/10.1038/s41597-019-0350-9>
8. Chen H (2005) Development, chromosome location and genetic mapping of EST-SSR markers in wheat. *Chinese Sci Bull* 50:2328. <https://doi.org/10.1360/982005-379>
9. Chen L-Y, Cao Y-N, Yuan N, et al (2015) Characterization of transcriptome and development of novel EST-SSR makers based on next-generation sequencing technology in *Neolitsea sericea* (Lauraceae) endemic to East Asian land-bridge islands. *Mol Breeding* 35:187. <https://doi.org/10.1007/s11032-015-0379-1>
10. Conesa A, Götz S, García-Gómez JM, et al (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>

11. Dhiman N, Kumar A, Kumar D, et al (2020) De novo transcriptome analysis of the critically endangered alpine Himalayan herb *Nardostachys jatamansi* reveals the biosynthesis pathway genes of tissue-specific secondary metabolites. Sci Rep 10:17186. <https://doi.org/10.1038/s41598-020-74049-1>
12. Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13-15. doi: 10.1007/978-3-642-83962-7_18.
13. Du X, Zhu X, Yang Y, et al (2019) De novo transcriptome analysis of *Viola xwittrockiana* exposed to high temperature stress. PLoS One 14:e0222344. <https://doi.org/10.1371/journal.pone.0222344>
14. Ercisli S, Ipek A, Barut E (2011) SSR Marker-Based DNA Fingerprinting and Cultivar Identification of Olives (*Olea europaea*). Biochem Genet 49:555–561. <https://doi.org/10.1007/s10528-011-9430-z>
15. Gao L, Tang J, Li H, et al (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol Breeding 12:245–261. <https://doi.org/10.1023/A:1026346121217>
16. Gaurav AK, Namita, Raju DVS, et al (2022) Genetic diversity analysis of wild and cultivated *Rosa* species of India using microsatellite markers and their comparison with morphology based diversity. J Plant Biochem Biotechnol 31:61–70. <https://doi.org/10.1007/s13562-021-00655-3>
17. Grabherr MG, Haas BJ, Yassour M, et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652. <https://doi.org/10.1038/nbt.1883>
18. Guo R, Mao Y-R, Cai J-R, et al (2014) Characterization and cross-species transferability of EST–SSR markers developed from the transcriptome of *Dysosma versipellis* (Berberidaceae) and their application to population genetic studies. Mol Breeding 34:1733–1746. <https://doi.org/10.1007/s11032-014-0134-z>
19. Hao X, Yang T, Liu R, et al (2017) An RNA Sequencing Transcriptome Analysis of Grasspea (*Lathyrus sativus* L.) and Development of SSR and KASP Markers. Front Plant Sci 8:1873. doi: 10.3389/fpls.2017.01873
20. He D, Zhang J, Zhang X, et al (2020) Development of SSR markers in Paeonia based on De Novo transcriptomic assemblies. PLoS One 15:e0227794. <https://doi.org/10.1371/journal.pone.0227794>
21. Hina F, Yisilam G, Wang S, et al (2020) De novo Transcriptome Assembly, Gene Annotation and SSR Marker Development in the Moon Seed Genus *Menispermum* (Menispermaceae). Front Genet 11:380. <https://doi.org/10.3389/fgene.2020.00380>
22. Jia H, Yang H, Sun P, et al (2016) De novo transcriptome assembly, development of EST-SSR markers and population genetic analyses for the desert biomass willow, *Salix psammophila*. Sci Rep 6:39591. <https://doi.org/10.1038/srep39591>
23. Kim JM, Lyu JI, Lee M-K, et al (2019) Cross-species transferability of EST-SSR markers derived from the transcriptome of kenaf (*Hibiscus cannabinus* L.) and their application to genus Hibiscus. Genet Resour Crop Evol 66:1543–1556. <https://doi.org/10.1007/s10722-019-00817-2>
24. Kirtikar KR, Basu BD (1935) Indian Medicinal Plants. Lalit Mohan Basu Publication. Allahabad.
25. Klepikova AV, Kasianov AS, Gerasimov ES, et al (2016) A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. Plant J 88:1058–1070. <https://doi.org/10.1111/tpj.13312>
26. Kumar S, Shah N, Garg V, Bhatia S (2014) Large scale in-silico identification and characterization of simple sequence repeats (SSRs) from de novo assembled transcriptome of *Catharanthus roseus* (L.) G. Don. Plant Cell Rep 33:905–918. <https://doi.org/10.1007/s00299-014-1569-8>
27. Kumpatla SP, Mukhopadhyay S (2005) Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. Genome 48:985–998. <https://doi.org/10.1139/g05-060>
28. La Rota M, Kantety RV, Yu J-K, et al (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. BMC Genom 6:23. <https://doi.org/10.1186/1471-2164-6-23>
29. Li D, Deng Z, Qin B, et al (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 Genom 13:192. <https://doi.org/10.1186/1471-2164-13-192>
30. Li Y, Jia L-K, Zhang F-Q, et al (2019) Development of EST-SSR markers in *Saxifraga sinomontana* (Saxifragaceae) and cross-amplification in three related species. Appl Plant Sci 7:e11269. <https://doi.org/10.1002/aps3.11269>
31. Long D (2014) Large scale development of SSR markers based on transcriptome sequencing of precocious trifoliate orange (M.Sc. dissertation). Wuhan (China): Huazhong Agriculture University.

32. Ma X, Wang P, Zhou S, et al (2015) De novo transcriptome sequencing and comprehensive analysis of the drought-responsive genes in the desert plant *Cynanchum komarovii*. BMC Genom 16:753. <https://doi.org/10.1186/s12864-015-1873-x>
33. Marconi TG, Costa EA, Miranda HR, et al (2011) Functional markers for gene mapping and genetic diversity studies in sugarcane. BMC Res Notes 4:264. <https://doi.org/10.1186/1756-0500-4-264>
34. Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30:194–200. <https://doi.org/10.1038/ng822>
35. Nayak SS, Pradhan S, Sahoo D, et al (2020) De novo transcriptome assembly and analysis of *Phragmites karka*, an invasive halophyte, to study the mechanism of salinity stress tolerance. Sci Rep 10:5192. <https://doi.org/10.1038/s41598-020-61857-8>
36. Palumbo F, Galla G, Vitulo N, et al (2018) First draft genome sequencing of fennel (*Foeniculum vulgare* Mill.): identification of simple sequence repeats and their application in marker-assisted breeding. Mol Breeding. <https://doi.org/10.1007/s11032-018-0884-0>
37. Pandey R, Kumar B, Meena B, et al (2017) Major bioactive phenolics in *Bergenia* species from the Indian Himalayan region: Method development, validation and quantitative estimation using UHPLC-QqQLIT-MS/MS. PLoS One 12:e0180950. <https://doi.org/10.1371/journal.pone.0180950>
38. Qiu L, Yang C, Tian B, et al (2010) Exploiting EST databases for the development and characterization of EST-SSR markers in castor bean (*Ricinus communis* L.). BMC Plant Biol 10:278. <https://doi.org/10.1186/1471-2229-10-278>
39. Raizada A, Souframanien J (2021) SNP genotyping and diversity analysis based on genic-SNPs through high resolution melting (HRM) analysis in blackgram [*Vigna mungo* (L.) Hepper]. Genet Resour Crop Evol 68:1331–1343. <https://doi.org/10.1007/s10722-020-01064-6>
40. Rana MS, Samant SS (2011) Diversity, indigenous uses and conservation status of medicinal plants in Manali wildlife sanctuary, North Western Himalaya. IJTK Vol10(3) [July 2011]
41. Rawal HC, Kumar S, Mithra SVA, et al (2017) High Quality Unigenes and Microsatellite Markers from Tissue Specific Transcriptome and Development of a Database in Clusterbean (*Cyamopsis tetragonoloba*, L. Taub). Genes 8:313. <https://doi.org/10.3390/genes8110313>
42. Sahoo A, Behura S, Singh S, et al (2021) EST-SSR marker-based genetic diversity and population structure analysis of Indian *Curcuma* species: significance for conservation. Braz J Bot 44:411–428. <https://doi.org/10.1007/s40415-021-00711-1>
43. Shah M, Alharby HF, Hakeem KR, et al (2020) De novo transcriptome analysis of *Lantana camara* L. revealed candidate genes involved in phenylpropanoid biosynthesis pathway. Sci Rep 10:13726. <https://doi.org/10.1038/s41598-020-70635-5>
44. Taheri S, Abdullah TL, Rafii MY, et al (2019) De novo assembly of transcriptomes, mining, and development of novel EST-SSR markers in *Curcuma alismatifolia* (Zingiberaceae family) through Illumina sequencing. Sci Rep 9:3047. <https://doi.org/10.1038/s41598-019-39944-2>
45. Taramino G, Tarchini R, Ferrario S, et al (1997) Characterization and mapping of simple sequence repeats (SSRs) in Sorghum bicolor: Theor Appl Genet 95:66–72. <https://doi.org/10.1007/s001220050533>
46. Tiwari V, Mahar KS, Singh N, et al (2015) Genetic variability and population structure of *Bergenia ciliata* (Saxifragaceae) in the Western Himalaya inferred from DAMD and ISSR markers. Biochem Sys Ecol 60:165–170. <https://doi.org/10.1016/j.bse.2015.04.018>
47. Tiwari V, Meena B, Nair NK, et al (2020) Molecular analyses of genetic variability in the populations of *Bergenia ciliata* in Indian Himalayan Region (IHR). Physiol Mol Biol Plants 26:975–984. <https://doi.org/10.1007/s12298-020-00797-z>
48. Untergasser A, Cutcutache I, Koressaar T, et al (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40:e115–e115. <https://doi.org/10.1093/nar/gks596>
49. Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55. <https://doi.org/10.1016/j.tibtech.2004.11.005>
50. Wang L, Gong X, Jin L, et al (2019a) Development and validation of EST-SSR markers of *Magnolia wufengensis* using de novo transcriptome sequencing. Trees 33:1213–1223. <https://doi.org/10.1007/s00468-019-01853-2>

51. Wang X, Chen W, Luo J, et al (2019b) Development of EST-SSR markers and their application in an analysis of the genetic diversity of the endangered species *Magnolia sinostellata*. Mol Genet Genomics 294:135–147. <https://doi.org/10.1007/s00438-018-1493-7>
52. Wang Y, Jia H-M, Shen Y-T, et al (2020) Construction of an anchoring SSR marker genetic linkage map and detection of a sex-linked region in two dioecious populations of red bayberry. Hortic Res 7:53. <https://doi.org/10.1038/s41438-020-0276-6>
53. Wei W, Qi X, Wang L, et al (2011) Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. BMC Genom 12:451. <https://doi.org/10.1186/1471-2164-12-451>
54. Wu G, Zhang L, Yin Y, et al (2015) Sequencing, de novo assembly and comparative analysis of *Raphanus sativus* transcriptome. Front Plant Sci 6: 198. doi: 10.3389/fpls.2015.00198
55. Xie X, Jiang J, Chen M, et al (2021) De novo Transcriptome Assembly of *Mylocerinus aurolineatus* Voss in Tea Plants. Front Sustain Food Syst 5: 631990. doi: 10.3389/fsufs.2021.631990
56. Xing W, Liao J, Cai M, et al (2017) 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 13:53. <https://doi.org/10.1007/s11295-017-1135-y>
57. Yan XQ, Lu M, An HM (2015) Analysis on SSR information in transcriptome and development of molecular markers in *Rosa roxburghii*. Acta Hort Sinica 42:341–349. doi: 10.1016/j.gene.2015.02.054.
58. Yan Z, Wu F, Luo K, et al (2017) Cross-species transferability of EST-SSR markers developed from the transcriptome of *Melilotus* and their application to population genetics research. Sci Rep 7:17959. <https://doi.org/10.1038/s41598-017-18049-8>
59. Yang B, Xu QW, Niu MY, et al (2018) SSR analysis and molecular marker development of the transcriptome of *Rhododendron fortunei* Lindl.. Acta Agri Nucl Sin 32:53–63.
60. Yang Y, He R, Zheng J, et al (2020) Development of EST-SSR markers and association mapping with floral traits in *Syringa oblata*. BMC Plant Biol 20:436. <https://doi.org/10.1186/s12870-020-02652-5>
61. Yue L, Twell D, Kuang Y, et al (2017) Transcriptome Analysis of *Hamelia patens* (Rubiaceae) Anthers Reveals Candidate Genes for Tapetum and Pollen Wall Development. Front Plant Sci 7: 1991. doi: 10.3389/fpls.2016.01991
62. Zhai L, Xu L, Wang Y, et al (2014) Novel and useful genic-SSR markers from de novo transcriptome sequencing of radish (*Raphanus sativus* L.). Mol Breeding 33:611–624. <https://doi.org/10.1007/s11032-013-9978-x>
63. Zhang C, Wu Z, Jiang X, et al (2021) De novo transcriptomic analysis and identification of EST-SSR markers in *Stephanandra incisa*. Sci Rep 11:1059. <https://doi.org/10.1038/s41598-020-80329-7>
64. Zhang H, Cong R, Wang M, et al (2018) Development of SSR molecular markers based on transcriptome sequencing of *Tagetes erecta*. Acta Hortic Sin 45:159-167. doi: 10.16420/j.issn.0513-353x.2017-0166.
65. Zhang K, Li Y, Zhu W, et al (2020) Fine Mapping and Transcriptome Analysis of Virescent Leaf Gene v-2 in Cucumber (*Cucumis sativus* L.). Front Plant Sci 11:570817. doi: 10.3389/fpls.2020.570817
66. Zhang QT, Li XY, Yang YM, et al (2016) Analysis on SSR information in transcriptome and development of molecular markers in *Lonicera caerulea*. Acta Hortic Sin 43:557–563. doi: 10.16420/j.issn.0513-353x.2017-0283.
67. Zhang Y, Zhang X, Wang Y-H, et al (2017) De Novo Assembly of Transcriptome and Development of Novel EST-SSR Markers in *Rhododendron rex* Lévl. through Illumina Sequencing. Front Plant Sci 8:1664. <https://doi.org/10.3389/fpls.2017.01664>
68. Zhou Q, Zhou P-Y, Zou W-T, et al (2021) EST-SSR marker development based on transcriptome sequencing and genetic analyses of *Phoebe bournei* (Lauraceae). Mol Biol Rep 48:2201–2208. <https://doi.org/10.1007/s11033-021-06228-w>
69. Zhou S, Wang C, Frazier TP, et al (2018) The first Illumina-based de novo transcriptome analysis and molecular marker development in Napier grass (*Pennisetum purpureum*). Mol Breeding 38:95. <https://doi.org/10.1007/s11032-018-0852-8>

Figures

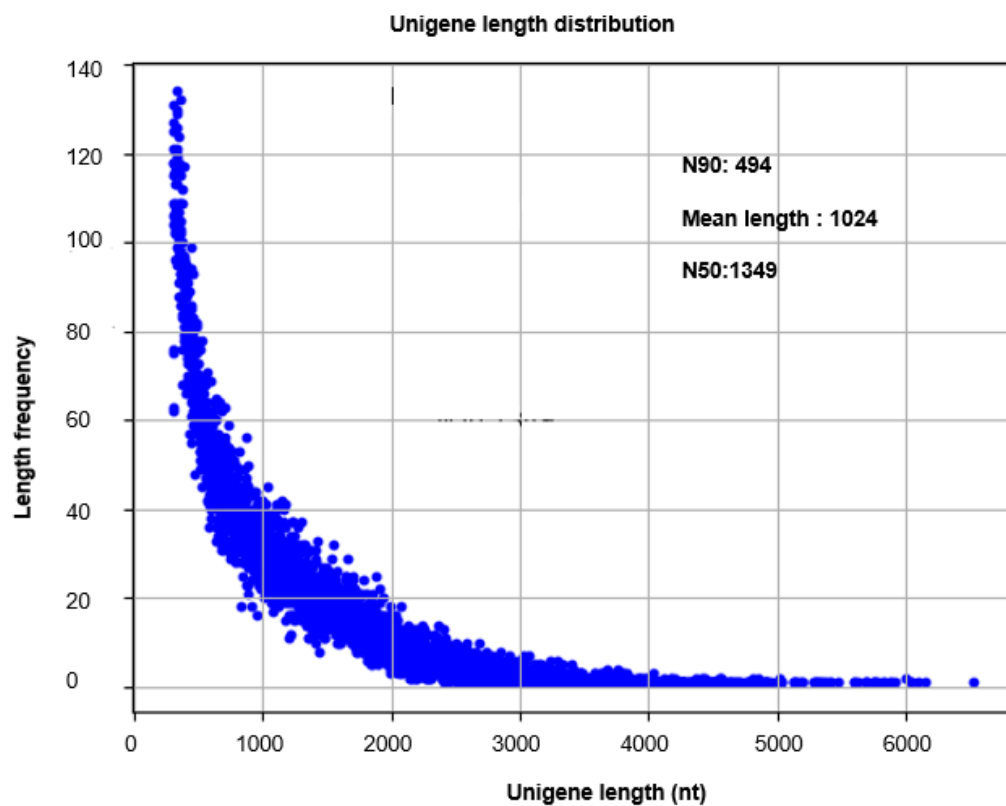


Fig. 1

Figure 1

Unigene length distribution in *B. ciliata*. The X and Y-axis represent the unigene length and length frequency, respectively

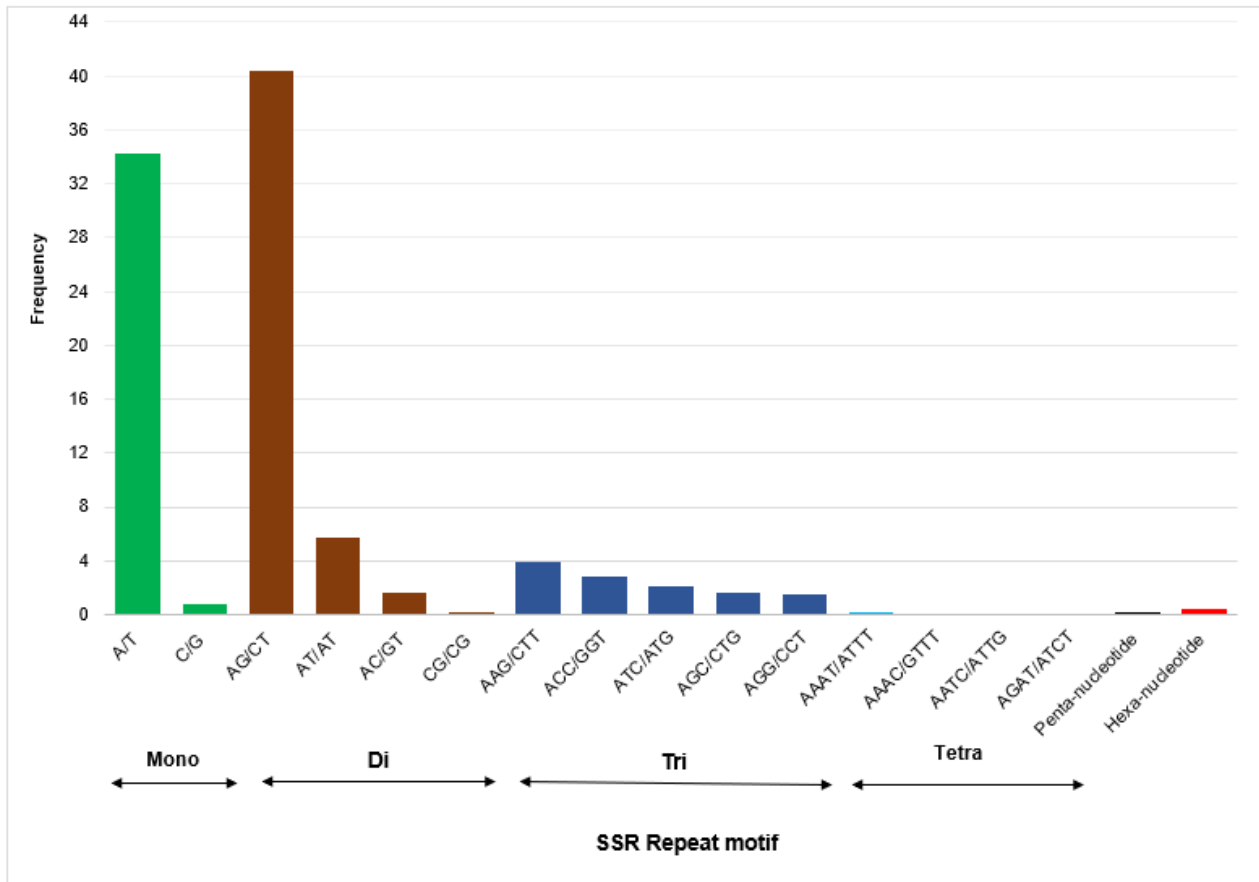


Fig. 2

Figure 2

Frequency and types of SSR repeat motif in *B. ciliata*

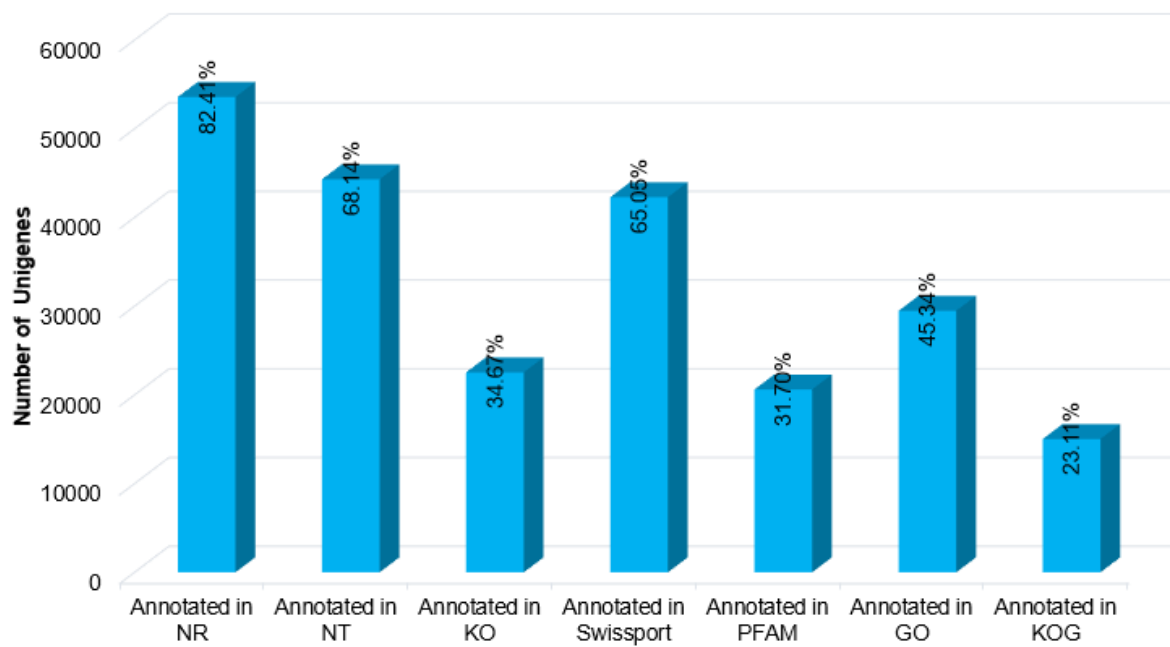


Fig. 3

Figure 3

Functional annotation of unigene in *B. ciliata*. The x-axis indicates different databases and y-axis indicates number of unigenes

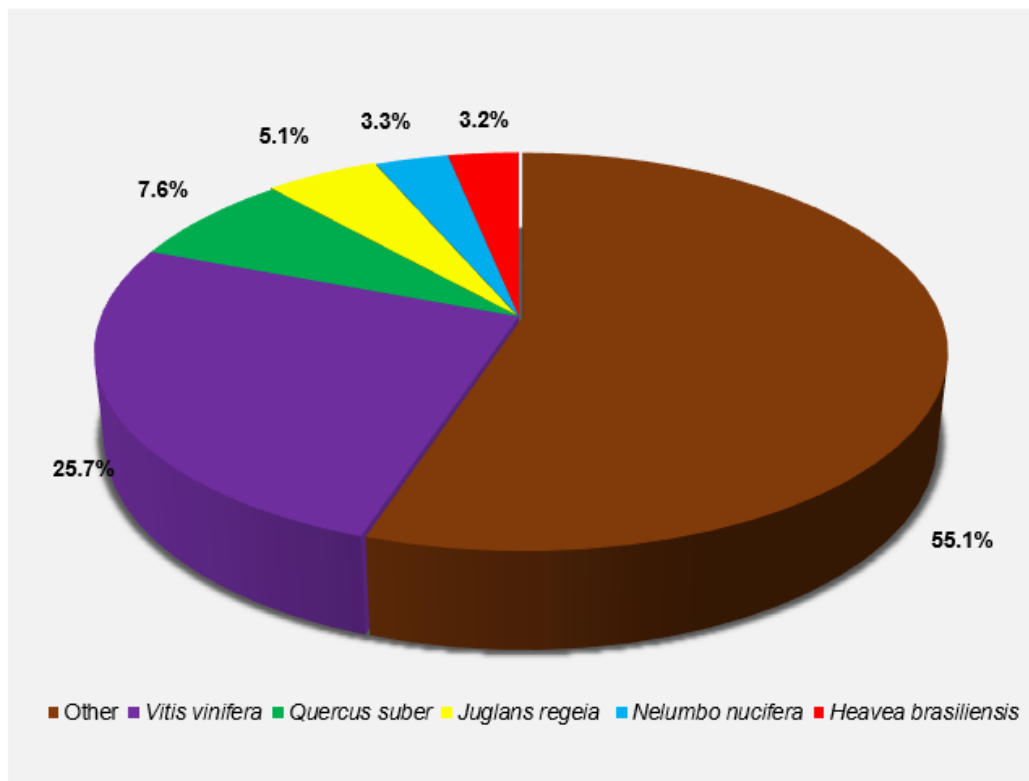


Fig. 4

Figure 4

Pie chart showing distribution of species and their similarity score as annotated from *B. ciliata* transcriptome

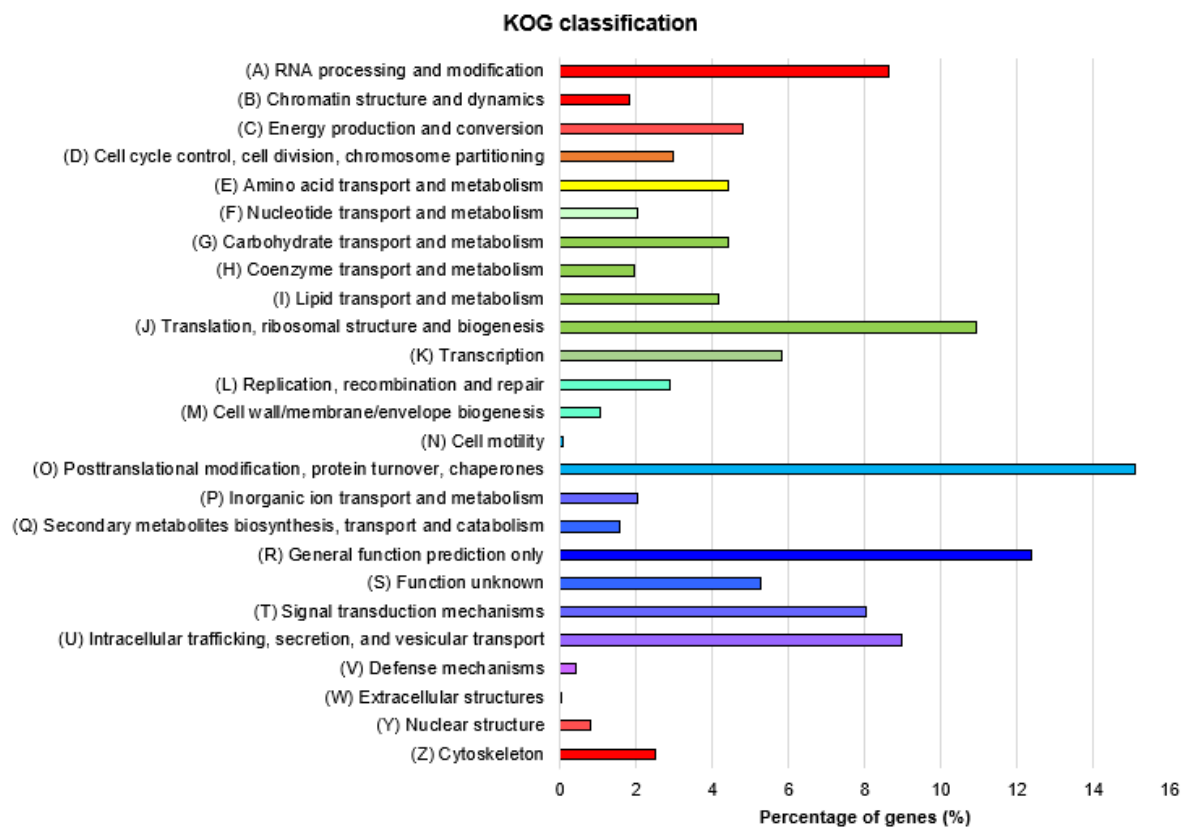


Fig.5

Figure 5

KOG analysis of the unigene sequences of *B. ciliata*. The x-axis indicates percentage of genes, and the y-axis. indicates the function class

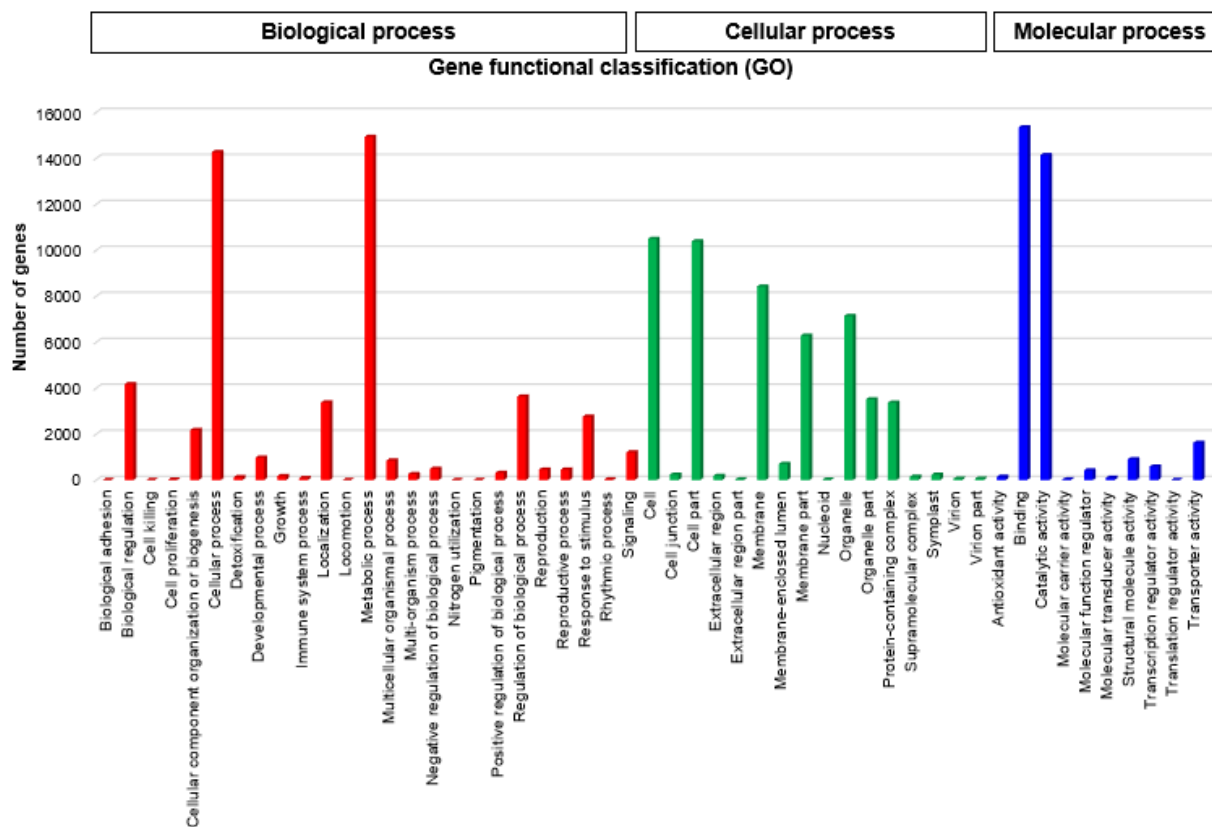


Fig.6

Figure 6

GO analysis of the unigene sequences of *B. ciliata*. The y-axis indicates the number of genes in a category and y-axis indicates the number of genes in that main category

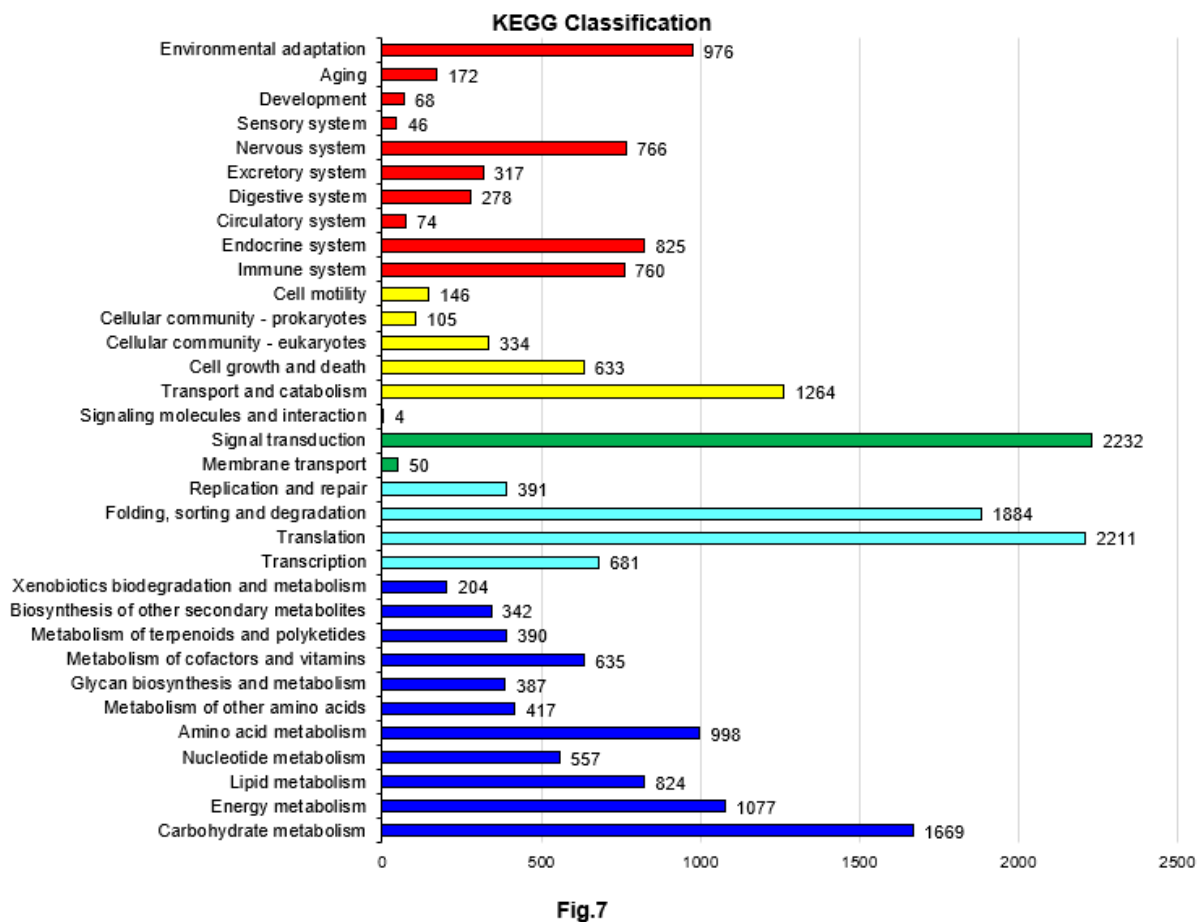


Figure 7

KEGG metabolic pathway of *B. ciliata*. the x- indicates number of genes and y-axis indicates name of the KEGG metabolic pathway

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ESMFig.1.pptx](#)
- [ESMFig.2.pptx](#)
- [ESMFig.3.pptx](#)